

## THE INVOLVEMENT LOAD HYPOTHESIS: ITS EFFECT ON VOCABULARY LEARNING IN PRIMARY EDUCATION

GEMA ALCARAZ MÁRMOL\*

ÁNGELA ALMELA SÁNCHEZ-LAFUENTE

*Universidad Católica San Antonio de Murcia*

**ABSTRACT.** *The present study explores the effect of four types of tasks on English as a Foreign Language (EFL) vocabulary acquisition. Four groups of students in their fifth year of Primary Education worked with a list of English words by means of a different task: reading comprehension with marginal glosses; reading comprehension and gap-filling; writing with marginal glosses; writing and dictionary use. Each task is characterised by a different involvement load. After each group worked with the words in a different way, all students took a receptive and a productive vocabulary test in order to know the degree of acquisition of the target words. The group doing the task with the highest degree of involvement load is expected to obtain the best results in the vocabulary tests.*

**KEY WORDS.** *Elementary Education, Foreign Language, involvement, vocabulary acquisition.*

**RESUMEN.** *El presente estudio pretende explorar el efecto de cuatro tipos de tareas sobre la adquisición léxica en inglés como lengua extranjera (EFL). Un total de cuatro grupos de estudiantes de EFL cursando quinto curso de Educación Primaria trabajan una serie de palabras en inglés. Cada grupo trabaja dicho vocabulario a través de una tarea distinta: comprensión lectora con glosas; comprensión lectora y relleno de huecos; producción escrita con glosas; producción escrita y uso del diccionario. Las tareas se diferencian entre ellas por el grado de procesamiento léxico (involvement) que cada una conlleva. Tras los cuatro tratamientos léxicos, todos los participantes se someterán a dos tests distintos, uno productivo y otro receptivo. Se espera que el grupo que trabaje las palabras en la actividad con mayor grado de procesamiento léxico obtenga mayores resultados que el resto.*

**PALABRAS CLAVE.** *Educación Primaria, Lengua Extranjera, involucramiento, adquisición léxica.*

### 1. INTRODUCTION

Nowadays the importance of vocabulary in language learning is beyond question. Many studies suggest that those students with high vocabulary level normally have better performance in Foreign Language (FL) written and oral tasks. That is why teachers and

learners would like to optimize the time they spend on vocabulary teaching and learning in a context where the contact with the Second Language (L2) is considerably limited. In other words, the FL teaching process develops within a framework of formal instruction, inside the classroom, taking from two to three hours a week. Consequently, the way vocabulary activities are approached has a main role in this teaching-learning context, since they are the open door to vocabulary acquisition. The present study aims to investigate the effect of the Involvement Load Hypothesis on L2 vocabulary acquisition in Primary Education students. This first section offers a review of the most relevant literature about the Involvement Load Hypothesis including empirical evidence on this issue. In the second section, the focus of the study is specified. Next, in the methodology we offer information about participants, materials, procedure and data analysis. In the fourth section, results are displayed and discussed. Finally, some final remarks are presented.

### 1.1. *The construct of involvement and the Involvement Load Hypothesis*

Craik and Lockhart (1972) stated that the fact that new linguistic data are stored in long-term memory depends on how deeply information has been initially processed, and not on the amount of time those data have been stored in short-term memory. Alongside the same lines, three years later Craik and Tulving (1975) suggested that it is the richness with which the material is encoded and not just the presence or absence of semantic encoding what should be considered critical to retention. This is one of the reasons why activities are so important in FL contexts, because they are one of the main ways students have to access Foreign Language.

According to cognitive psychology, deep elaboration of lexical information leads to better retention (Broeder and Plunkett 1994). Elaborated processing of lexical information involves both paying attention –whether intentionally or unintentionally– and somehow noticing. In fact, attention should not be understood as intentional learning. As Hulstijn and Laufer stated, “careful attention can be paid to a certain word during intentional learning (e.g. preparation for a vocabulary test) just as well as during incidental learning (e.g. when a word occurs in a text and successful completion of the reading task requires such attention)” (Hulstijn and Laufer 2001: 542).

Previous studies show that activities with deeper and more elaborated lexical processing are more effective. Ellis (1994) and Newton (1995) suggested that negotiation in communicative tasks promotes better vocabulary acquisition than those tasks with no negotiation. Alongside the same lines, Ellis and He (1999) and Joe (1998) confirmed that activities which require production were more efficient for vocabulary acquisition than those activities where production does not occur.

Traditionally, the concepts of Elaboration and Processing have been understood in a general way from the viewpoint of cognition (Anderson 1995; Baddeley 1997). However, Laufer and Hulstijn (2001) have more recently felt the need to operationalize and adapt these notions to the field of L2 vocabulary acquisition. They suggested what they called the construct of *involvement*. In their view, *involvement* can be conceived as able to explain

and predict learners' success in the retention of unfamiliar words. This construct includes both cognitive and motivational aspects, adopting the term cognitive as referring exclusively to information processing, excluding affective aspects of cognition.

The three basic components of this construct are *need*, *search* and *evaluation*. The first one corresponds to the motivational non-cognitive dimension of the construct. It refers to the need for achievement. It should be stated here that the notion of achievement should not be understood in the sense of fear of failure, but in a more positive way, as a way to comply with the task requirements, whether these requirements be externally imposed or self-imposed. Search and evaluation are the second and third dimensions of involvement, and they have a cognitive origin. They are related to noticing and paying attention to the form meaning relationship (Schmidt 2000). Search concerns the attempt to find the L2 form or meaning by using a dictionary or any other authority. Evaluation focuses on deciding whether an L2 word can be used in a specific linguistic context or not (see Table 1). It can entail a comparison of a given word with other words, or a specific meaning with other meanings of the same words. In fact, evaluation implies some kind of decision based on specific criteria.

The components of *need* and *evaluation* comprehend three levels of involvement which are called level. Level 0 occurs when a component is not part of the activity or task. Level 1 is described as 'moderate'. *Need* is moderate when it is triggered by an external agent such as the teacher. In the case of *evaluation*, it is moderate when there is the possibility of choosing an L2 word among different options. Level 2 is called the 'strong' level. It is given when the learning need comes from the learner himself. The strong level in evaluation appears when the learner uses an L2 word in a sentence or text, without having the possibility to choose among several options. Contrary to the components of *need* and *evaluation*, the component of *search* only comprises two levels: level 0 and level 1, indicating no gradation but just whether this component is included or not in the activity.

Components	Degrees of Involvement	Definition
Need	Index 0 (none)	The learner does not feel the need to learn the word
	Index 1 (moderate)	The learner is required to learn the word
	Index 2 (strong)	The learner decides the learn the word
Search	Index 0 (absence)	The learner does not look for the meaning or form of the word with a lexical instrument
	Index 1 (existence)	The meaning and form of the word are found by the learner
Evaluation	Index 0 (none)	The word is not compared with any other word
	Index 1 (moderate)	The word is compared with other words in the provided context
	Index 2 (strong)	The word is compared with other words in self-provided context (the learner's mental lexicon)

Table 1. Components and levels of Involvement.

The construct of involvement can result in different combinations and levels of the three components. They are represented by different degrees of involvement load. Technically, we can find 15 different combinations of components and levels (see Table 2).

Need	Search	Evaluation
0	0	0
0	1	0
0	1	1
0	1	2
1	0	0
1	0	1
1	0	2
1	1	1
1	1	2
2	0	0
2	0	1
2	0	2
2	1	0
2	1	1
2	1	2

Table 2. Potential combinations of components and levels.

Some of them are theoretically possible, but they would not make sense or would not be very practical. For instance, an activity with involvement load 0 means, among other things, that it has no *need*, something which is theoretically possible but not useful. Given its particular nature, an activity is to be done; otherwise it is obsolete. If an activity has 0 level of *need*, it loses its essence, that is to say, to be done.

All these combinations differ in the involvement load they generate. The so called involvement load was defined by Laufer and Hulstijn (2001: 15) as “the combination of the presence or absence of the involvement factors need, search and evaluation” in different activities. For instance, looking up unknown words in a dictionary while reading implies search, while the same reading with words glossed in the margin does not. In the first type of task the learner pays attention to form in a way that is much more overt than in the second one, where search is not given. In a similar way, a writing task that asks learners to use previously known words to write a composition implies stronger evaluation than choosing among different words offered in the task.

Taking the previous information about involvement into consideration, Laufer and Hulstijn proposed what they called the Involvement Load Hypothesis (2001), which suggests that the possibility of learning an L2 word is directly proportional to the involvement load during the word processing. That is to say, the higher the involvement load, the better the acquisition. The authors stated that retention of words is conditional upon the components of need, search and evaluation explained above. In fact, they state 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” (Laufer and Hulstijn 2001: 15).

### 1.2. *Some empirical evidence on the Involvement Load Hypothesis*

To our knowledge, three degrees of involvement load have been explored up to now in empirical studies. They correspond to moderate *need*, no *search* and no *evaluation* (level 1); moderate *need*, no *search* and moderate *evaluation* (level 2); moderate *need*, no *search* and strong *evaluation* (level 3). We are unaware of cases where the effect of activities with *search* have been explored.

What is more specifically, up to now, research about the involvement construct has almost exclusively focused on learners with intermediate or advanced L2 level, and their results are somehow conflicting. For instance, Folse (2006) studied the effects of different writing tasks on the learning of L2 words by university students whose proficiency levels ranged from lower intermediate to advanced level. Curiously enough, his results were against the Involvement Load Hypothesis, since the tasks involving strong and moderate evaluation were equally effective. On the contrary, Keating (2008) found that beginning Spanish learners completing a sentence-writing exercise after reading obtained better retention of meaning and form than by means of a blank-filling exercise. Nevertheless, the first condition scored comparatively lower on the delayed form-retention posttest than the second condition. More recently, Lu (2013) showed that the blank-filling task was more beneficial to vocabulary learning than the composition task for lower-intermediate learners, probably because of time constraints in FL classrooms.

Accordingly, the present study aims to take one step further in two senses: (1) by including *search* in the task, and (2) by working with students with an elementary L2 level.

## 2. AIM

According to the Involvement Load hypothesis, the higher the involvement load in activities, the better the vocabulary acquisition. With this premise in mind, the aim of the present study is twofold:

- a. To confirm the Involvement Load hypothesis as regards elementary students of English as a Foreign Language.

- b. To introduce a new degree of involvement load by adding the component of search, which has not been explored up to now.

### 3. METHODOLOGY

#### 3.1. *Participants*

A group of 28 ten-year-old learners studying their fifth year of Primary Education took part in the study. According to Piaget's theory of cognitive development (1936), participants belong to the concrete operational stage, being at the doors of pre-adolescence. This stage is believed to affect children between seven and twelve. During this stage the thought process becomes more rational, mature and adult-like. In other words, the child has the ability to develop logical thought about and object. However, they still cannot tackle a problem with several variables in a systematic way. They study in a state school in Murcia. All of them speak Spanish as their mother tongue and have English as a Foreign Language. At the moment the study was carried out, students had been exposed to the English language for about 648 hours. The teaching method we followed was reportedly Communicative, although lessons were not completely developed in English.

Participants were divided into four groups of seven ten-year-old students. Each group was assigned a different task, so that group 1 (G1) did task 1, group 2 (G2) performed task 2, group 3 (G3) fulfilled task 3 and group 4 (G4) carried out task 4.

#### 3.2. *Materials and procedure*

3.2.1 Words. The words which were selected for the study belonged to a short graded reader entitled *A pretty face* by John Escott (2008), published by Dominoes (Oxford University Press). It targets Elementary students of English as a Foreign Language, as it is the case of the participants in this study. The story is about a girl who works as a bookshop assistant, but her dream is to become a star. Up to 18 words were randomly selected from the reading: six nouns, six adjectives and six verbs (see Table 3).

Nouns	Adjectives	Verbs
Actress	Amazed	Believe
Costume	Difficult	Follow
Desk	Free	Hate
Interview	Funny	Hide
Lie	Kind	Pay
Movie	Pretty	Refuse

Table 3. Target words.

3.2.2. Tasks. Four tasks with different degrees of involvement load were designed. Each task was done by a different group of students, namely G1, G2, G3, and G4 (as specified in section 3.1). Groups consisted of seven students who had been randomly selected. Table 4 shows the four tasks involved in the study with a brief description of each one and the involvement load in each of them.

- a. Task 1: Reading comprehension with marginal glosses. This task entailed a reading activity on the basis of a text with short paragraphs and subsequent reading comprehension question for each paragraph. The paragraphs were extracted from the graded reader and contained between 20 and 30 words. They had to be slightly edited and adapted to the level of participants. Questions were designed ad hoc. Afterwards, G1 had to show that they had understood the paragraphs by answering those questions correctly. Target words were highlighted in bold and their meaning was found at the end of each paragraph. This first task involved moderate need (1), no search (0) and evaluation (0).
- b. Task 2: Reading comprehension and gap-filling. G2 was provided with the same paragraphs and questions which were used for task 1. However, in this case target words did not appear in the text. They were replaced by gaps which students had to fill in with the correct word. In order to carry out the task, a list with 27 L2 words was presented to the students, from which 18 were target words and 9 were not to be used. All of them appeared with their L1 equivalents. Given these characteristics, task 2 has an involvement load of 2 points, involving moderate need (1), no search (0) and moderate evaluation (1), since the selection of the adequate term for each gap had to be done from among several options.
- c. Task 3: Writing sentences with the help of glosses. G3 had to write a sentence with every target word. Similar to tasks 1 and 2, these words appeared in a list together with their L1 equivalents. Students were asked to write sentences that somehow could prove the understanding of the target word. This task represents an involvement load of 3, as it includes moderate need (1), no search (0) and strong evaluation (2), because words have to be used in a linguistic context created by the students themselves.
- d. Task 4: Writing sentences with the help of a bilingual dictionary. In a similar fashion to G3, G4 was asked to write a sentence with each of the target words. In contrast with the other tasks, in this case word meaning was not provided and participants had to look it up in the dictionary. Students had access to the bilingual dictionary *Collins First Time* (2009), a dictionary for beginners. It is important to highlight that the participants in this study had already used a bilingual dictionary and knew the mechanics. This task has an involvement load of 4, as it includes moderate need (1), search (1) and strong evaluation (2).

Task	Description	Involvement Load			
		Need	Search	Evaluation	Total
T1	Reading comprehension with marginal glosses	1	0	0	1
T2	Reading comprehension and gap-filling	1	0	1	2
T3	Writing sentences with the help of glosses	1	0	2	3
T4	Writing sentences with the help of a bilingual dictionary	1	1	2	4

Table 4. Tasks involved in the study with the involvement load contained in each of them.

3.2.3. Data collection instruments. Different alternatives were considered as for data collection instruments: the picture naming test, the free production test or the yes/no test. However, all of these alternatives were eventually ruled out for several reasons. With regard to the picture naming test, the most important drawback was the difficulty in representing some words graphically. Words such as lie, believe or follow could have led to confusion on the part of the participants, as their graphic representation would have been far from accurate. The free production test was discarded because of the participants' low level. A yes/no test was considered insufficient as the information provided by this instrument is far from specific.

Two tests were distributed in order to measure the participants' L2 vocabulary learning. They were designed ad hoc adopting the L1-L2/L2-L1 translation format. Two types of knowledge were tested: productive and receptive knowledge.

The productive test consisted of a list of 18 Spanish terms corresponding to the L2 target words studied. They were arranged alphabetically and each of them was accompanied by a dotted line where the L2 equivalent was expected to be provided by the students. The aim of this type of test is to check whether learners can recall a series of L2 word forms.

The receptive test presented a list of the 18 target words in English contained in the tasks. The receptive test measured the students' recognition of words and consisted of providing the L1 equivalent to each of them. Similarly to the productive test, words were alphabetically listed and were accompanied by a dotted line to provide the Spanish equivalent.

According to Takala (1984: 146), "the best pay-off between validity, reliability and practicality is shown by test types which ask students to write L2 or L1 equivalents to written decontextualized stimulus words." Tests based on providing equivalents to words in another language are considered to be reliable instruments for L2 vocabulary acquisition (Alcaraz-Mármol 2011; Read 2000; Takala 1984). In fact, Nation (2001) encourages the use of the L1 in the vocabulary testing on participants with a low level of L2, as it is this case. The author states that the attitude of rejecting translation for vocabulary assessment "is quite wrong [because] translation is one of a number of means of conveying meaning and in general is no better or worse than the use of pictures, real objects, definitions, L2 synonyms and so on" (Nation 2001: 351). Thus, the use of the



second language in a vocabulary test increases difficulty in the learners' performance. Consequently, it was decided to resort to the L1 as a stimulus (Nation 2001).

3.2.4. Data collection procedures. As regards the procedure, the tests were administered to subjects in an intact fifth-year class group by the present researchers in the presence of the class teacher. It is worth noting that tests were timed and that the provided resources were printed materials. Specifically, the four groups were given 30 minutes for the fulfillment of their respective tasks, and after their accomplishment, the materials were collected. Then, the productive test was administered; after 15 minutes, this test was collected and the participants were provided with the receptive test.

3.2.5. Data analysis. In order to compare test results for each of the groups, the parametric test of the analysis of variance or ANOVA was run. In this case, the independent variable is the degree of involvement in the tasks. The ANOVA was completed with the post-hoc analysis Tukey's HSD (Honestly Significant Difference) for the identification of the significant inter-group differences. According to the research norm in Linguistics, differences were considered significant when  $p < 0.05$ .

#### 4. RESULTS AND DISCUSSION

As can be seen in Table 5, the critical level associated to F (.000) is much lower than .05 both in the receptive (R) and the productive test (P). Consequently, we can reject the null hypothesis that the population means are equal. Put another way, as could be expected, the four tasks accomplished in the study do make a difference in the students' degree of vocabulary acquisition. Furthermore, the F-ratio in the productive test is higher, which means that there is a mean difference broader than in the receptive test.

		Sum of squares	df	Quadratic mean	F	Sig.
R	Inter-groups	144.857	3	48.286	27.592	.000
	Intra-groups	42.000	24	1.750		
	Total	186.857	27			
P	Inter-groups	152.714	3	50.905	34.208	.000
	Intra-groups	35.714	24	1.488		
	Total	188.429	27			

Table 5. ANOVA test.

The results from the post-hoc analysis indeed reflect significant differences among the scores of most groups (Table 6). As regards the receptive test, it can be observed how all the groups differ except for the second one and the fourth one, which means in practice that both groups of participants have a similar level of receptive lexical

knowledge. On the other hand, G1 and G3 substantially differ, both between them and with the remaining groups. The former produces the expected performance, but the latter achieves the best score, despite the fact that it is not the group with the highest degree of involvement.

Dependent variable	(I) mode	(J) mode	Means difference (I-J)	Typical error	Sig.	Confidence interval 95%	
						Limit inferior	Limit superior
R	1	2	-3.429*	.707	.000	-5.38	-1.48
		3	-6.429*	.707	.000	-8.38	-4.48
		4	-3.286*	.707	.001	-5.24	-1.34
	2	1	3.429*	.707	.000	1.48	5.38
		3	-3.000*	.707	.002	-4.95	-1.05
		4	.143	.707	.997	-1.81	2.09
	3	1	6.429*	.707	.000	4.48	8.38
		2	3.000*	.707	.002	1.05	4.95
		4	3.143*	.707	.001	1.19	5.09
	4	1	3.286*	.707	.001	1.34	5.24
		2	.143	.707	.997	-2.09	1.81
		3	3.143*	.707	.001	-5.09	-1.19
P	1	2	-1.000	.652	.434	-2.80	.80
		3	-6.000*	.652	.000	-7.80	-4.20
		4	-3.571*	.652	.000	-5.37	-1.77
	2	1	1.000	.652	.434	-.80	2.80
		3	-5.000*	.652	.000	-6.80	-3.20
		4	-2.571*	.652	.003	-4.37	-.77
	3	1	6.000*	.652	.000	4.20	7.80
		2	5.000*	.652	.000	3.20	6.80
		4	2.429*	.652	.005	.63	4.23
	4	1	3.571*	.652	.000	1.77	5.37
		2	2.571*	.652	.003	.77	4.37
		3	-2.429*	.652	.005	-4.23	-.63

\*Means difference is significant at .05 level.

Table 6. Post-hoc test.

When it comes to the productive test, G1 and G2 score similarly; thus the post-hoc test does not detect significant differences among them. The third group scores markedly better than the fourth one, which parallels the results from the receptive test.

Accordingly, the hypothesis on the degrees of involvement was only partially confirmed. It was observed that the group with the lowest degree of involvement obtained the worst score, both in the receptive and in the productive tests. This proves the need for exercises which ‘involve’ the students, understanding ‘involve’ in the specific sense of the present study. The challenge is to determine the appropriate kind of involvement, as well as its influence on the learner’s lexical acquisition.

According to the initial hypothesis, the higher the degree of involvement, the better the lexical acquisition. However, the group working with the highest degree of involvement did not achieved the best scores in none of the lexical post-tests. Curiously enough, the group working with degree of involvement 3 outperformed the other groups both in the receptive and in the productive tests (see Tables 7 and 8 and Figure 1). This fact makes us think of two questions, both closely interrelated. The first matter concerns the involvement hypothesis proposed, as regards the L2 level of the students involved. It might be possible that in former studies conducted on intermediate or advanced learners the involvement hypothesis is confirmed. Nonetheless, a possibility also exists that certain degrees of involvement are too complex or are beyond the capacity of elementary learners. This may explain why, for instance, the group working with degree of involvement 2 has scored slightly better in the receptive test than the group with degree 4. However, when it comes to the productive test the situation is altogether different. The group working with degree of involvement 4 is the one which outperforms group 2. The explanation may lie in the nature of the tests and the tasks: the second exercise dealt primarily with receptive skills, whereas the fourth one was eminently productive. Thus, the fact that G2 outperformed the rest in the receptive test and G4 did the same in the productive one may be partially explained by the kind of knowledge enhanced with the tasks. Studies such as Hui-Fang (2003) or Folse (2006) show that tasks can exert a significant effect on learning, since they can favour or hinder different aspects of L2 learning. Focusing more specifically on vocabulary acquisition, Yaqubi, Rayati and Gorgi’s study (2010) warns about the possible task effect on the involvement load. That is to say, task type and not involvement may be an influential factor on vocabulary retention. Furthermore, the authors insist that the operationalization of the levels of processing and the components of involvement needs reconsideration, especially when it comes to evaluation.

mode	N	Subset for alpha = .05		
		1	2	3
1	7	1.29		
4	7		4.57	
2	7		4.71	
3	7			7.71
Sig.		1.000	.997	1.000

Table 7. Homogeneous subsets for the receptive test (Tukey’s HSD).

mode	N	Subset for alpha = .05		
		1	2	3
1	7	.00		
2	7	1.00		
4	7		3.57	
3	7			6.00
Sig.		.434	1.000	1.000

Table 8. Homogeneous subsets for the productive test (Tukey's HSD).

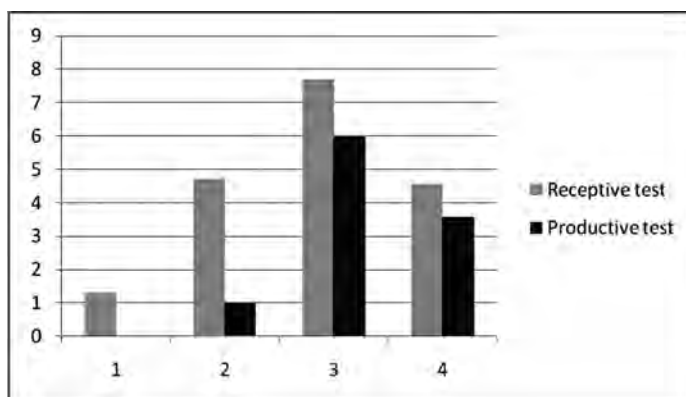


Figure 1. Results for the receptive and the productive tests.

The second question arises from the poor results of G4 as compared with other groups such as G3. G4 learners exercised the vocabulary by means of the activity with the highest degree of involvement. However, their results were not the best ones. That task involved the use of a resource external to the individual and to the exercise: the dictionary. Its efficiency as a useful tool when we face a foreign language is widely acknowledged. Studies like Knight (1994) and Hulstijn, Hollander and Greidanus (1996) show that dictionary use positively influences the incidental learning of vocabulary in a second language. Nevertheless, it is worth noting that the participants in both studies were adults with an advanced level of the language. As mentioned above, the profile of the participants in our study is clearly different. We have observed that dictionary use in task 4 has not exerted the positive effects found in the former studies. This finding suggests that the effect of dictionary use is not the same in all educational contexts.

All in all, it could be stated that the dictionary is a double-edged sword. This tool involves certain skills which go beyond the linguistic capacity. If the learner has not developed those skills yet, there is a possibility that dictionary use is not only not beneficial,

but even counterproductive in certain cases. Apart from the factor of the elementary level, it is also worth mentioning the fact that our individuals' cognitive abilities were still developing. For this reason, monitored dictionary use might be the solution to the problems posed by this tool at elementary foreign language levels and in early cognitive stages.

## 5. FINAL REMARKS

The present study tested Hulstijn and Laufer's hypothesis on involvement as far as SL lexical acquisition is concerned. It was observed that this hypothesis was only partially confirmed on elementary learners. The task with the degree of involvement 4 did not yield the expected results. Quite the reverse, the third task was the one which outperformed all the remaining tasks, both in the receptive and in the productive tests. Our results run counter to the initial hypothesis. The reasons suggested above point to different factors which may have led to this situation. The first one is the linguistic level of the participants in our study. It is the first time this hypothesis was tested with elementary level students. Maybe this theory is only confirmed with students in higher L2 levels. This idea is somehow connected to the second possible reason for our results: our participants were between eight and nine years old. At this age their metacognitive skills are still developing, thus some tasks could have been too complex for them, such as the use of a bilingual dictionary.

In conclusion, this study shows that the Involvement Load Hypothesis seems to be confirmed in the three first stages, but not in the last one. This may be due to the learners' lack of cognitive maturity or lack of practice in dictionary use. Therefore, although the Involvement Load Hypothesis cannot be confirmed in all L2 learning contexts, it provides a theoretical basis for vocabulary teachers. Nevertheless, further studies are needed to improve upon the hypothesis and reach its full potential.

## NOTES

\* Correspondence to: Gema Alcaraz Mármol. Universidad Católica de Murcia. Av. de los Jerónimos. CP 30107. Guadalupe (Murcia). E-mail: galcaraz@ucam.edu.

## REFERENCES

- Alcaraz-Mármol, G. 2011. *Vocabulary Acquisition in the Context of Non-Systematic Input*. Saarbrücken: LAP Lambert Academic Publishing.
- Collins First Time*. 2009. Oxford: Random House Mondadori.
- Craik, F. and R. Lockhart. 1972. "Levels of processing: A framework for memory research". *Journal of Verbal Learning and Verbal Behaviour* 11(3): 671-684.
- Ellis, N. C. 1994. "Vocabulary acquisition: the implicit ins and outs of explicit cognitive mediation". *Implicit and Explicit Learning of Languages*. Ed. N. C. Ellis. London: Academic Press. 211-282.

- Ellis, R. and X. He. 1999. "The roles of modified input and output in the incidental acquisition of word meanings". *Studies in Second Language Acquisition* 21(2): 285-301.
- Escott, J. 2008. *A Pretty Face*. London: Dominoes Publishers.
- Folse, K. S. 2006. "The effect of type of written exercise on L2 vocabulary retention". *TESOL Journal* 40(1): 273-293.
- Hui-Fang, T. 2003. *Effects of Task-Induced Involvement on Incidental Vocabulary Learning in a Second Language*. Unpublished MA thesis. Department of Linguistics, National Tsinghua University (Taiwan).
- Hulstijn, J., M. Hollander and T. Greidanus. 1996. "Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words". *The Modern Language Journal* 80(2): 327-339.
- Hulstijn, J. and B. Laufer. 2001. Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning* 51(4): 539-558.
- Joe, A. 1998. "What effects do text-based tasks promoting generation have on incidental vocabulary acquisition?" *Applied Linguistics* 19(2): 357-377.
- Keating, G. D. 2008. "Task effectiveness and word learning in a second language: The involvement load hypothesis on trial". *Language Teaching Research* 12: 365-386.
- Knight, S. 1994. "Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities". *The Modern Language Journal* 78(1): 285-299.
- Laufer, B. and J. Hulstijn. 2001. "Incidental vocabulary acquisition in a second language: the construct of task-induced involvement". *Applied Linguistics* 22(1): 1-26.
- Lu, M. 2013. "Effects of four vocabulary exercises on facilitating learning vocabulary meaning, form, and use". *TESOL Quarterly* 47: 167.
- Newton, J. 1995. Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research* 11(1): 159-177.
- Piaget, J. 1936. *Origins of Intelligence in the Child*. London: Routledge & Kegan Paul.
- Read, J. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N. 1998. "Tracking the incremental acquisition of second language vocabulary: A longitudinal study". *Language Learning* 48(2): 281-317.
- Scholfield, P. 1991. "Vocabulary rate in coursebooks: Living with an unstable lexical economy". *Proceedings of 5th Symposium on the Description and/or Comparison of English and Greek*. Thessaloniki: Aristotle University. 11-32.
- Takala, S. 1984. *Evaluation of Student's Knowledge of English Vocabulary in the Finnish Comprehensive School*. Urbana (IL): University of Illinois at Urbana.
- Waring, R. 1997. "A study of receptive and productive learning from word cards". *Studies in Foreign Languages and Literature* 21(1): 94-114.
- Yaqubi, B., R. A. Rayati and A. Gorgi. 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* 1(1): 146-163.