

Receptive and productive vocabulary acquisition: Effectiveness of three types of tasks. Results from
French students of Spanish as second language

Alicia San-Mateo-Valdehíta

Universidad Nacional de Educación a Distancia (UNED), España

Correo electrónico: asanmateo@flog.uned.es

Cecilia Criado-de-Diego

Universidad Nacional de Educación a Distancia (UNED), España

Correo electrónico: ceciliacriado@flog.uned.es

Fecha de recepción: septiembre de 2018

Fecha de aceptación: mayo de 2019

Abstract

This study presents the results of research on the vocabulary acquisition of French students of Spanish as second language. The aim is to know (1) which of these three vocabulary-learning tasks is more effective: definition-choosing, gap-filling or sentence-writing; and (2) which kind of knowledge, receptive or productive, the participants acquire with each vocabulary-learning task. The analysis shows that the most effective task is sentence-writing, then gap-filling, and in the third place, definition-choosing. Also all the three learning tasks trained students to complete above all activities which require a receptive knowledge of the words—definition-choosing and gap-filling tasks—, but they were much less prepared to carry out the sentence-writing task, which requires a productive knowledge of the words. The only productive task proposed—sentence-writing—is the one that allowed the students to learn receptively and productively a higher number of words. We associate our results with the *Involvement Load Hypothesis* and *Technique Feature Analysis*.

Keywords

Vocabulary learning, receptive and productive vocabulary, receptive and productive tasks, Spanish as second language, *Involvement Load Hypothesis*, *Technique Feature Analysis*

1. Introduction*

Lexical knowledge is decisive in second language (SL) acquisition (VonSydow, 2015; Hu and Nassaji, 2016; Shirzad et al. 2017; Webb and Nation, 2017). Among all the linguistic competences required to be proficient in a SL, the lexical competence is the first one to be mentioned in *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)* (Council of Europe, 2001). Despite this, recent studies have shown that the number of words that learners should know to be proficient in a SL (Nation, 2006) is higher than the number of words they actually know (Laufer, 2010). Naturally, researchers have been worried about establishing which the most effective task is for vocabulary learning in a SL.

The first matter that we must raise is understanding the meaning of ‘knowing a word’. As Richards (1976: 78-83; 1985: 176-188) explained, knowing a word entails knowledge of several aspects concerning the nature of lexical competence: (1) word frequency, (2) register, (3) collocation, (4) underlying forms and derivations, (5) word association, (6) semantic value and (7) different meanings associated with the word. Richards’ works were the starting point of many researchers who have addressed the issue. Nation (2001: 23-59) identified 18 kinds of knowledge grouped into three categories—Form, Meaning and Use. Form includes (1) spoken form, (2) written form and (3) word parts; Meaning includes (4) form and meaning, (5) concepts and referents, and (6) associations, and Use includes (7) grammatical functions, (8) collocations and (9) constraints on use, which include register, frequency... The receptive and productive knowledge applies to each of those aspects.

Receptive knowledge implies *recognizing* the form, meaning, and possible contexts of a word; *productive* knowledge implies *using* the word with its correct spelling and pronunciation, and its correct lexical, pragmatic and syntactic contexts. Both kinds of knowledge can be partial and divided in degrees. Here, it is of interest to know what kind of learning we need to acquire receptive and productive vocabulary. Recent studies have shown that ‘productive learning involves the knowledge needed for receptive use, whereas receptive learning may not involve the knowledge needed for productive use’ (Webb and Nation, 2017: 34).

* We thank the project “Grados de eficacia en ejercicios de incorporación de vocablos al lexicon de aprendices de español segunda lengua/lengua extranjera” (FFI2013-44117-P), supported by the Spanish Ministry of Economy, Industry and Competitiveness, and all members of our Research Group INVOLEX: M.^a Antonieta Andión Herrero and M.^a José Labrador Piquer, as well as all participants and teachers from École Nationale Supérieure des Mines, Université Paris-Dauphine, Université Sorbonne Nouvelle-Paris 3 and Instituto Cervantes in Paris.

Production tasks demand an active role, meaning that learners have to make a greater cognitive effort. For Hulstijn and Laufer (2001) this leads to positive results. They consider the three following factors to measure cognitive effort induced by a learning task: need, search and evaluation, and raised the *Involvement Load Hypothesis (ILH)*. *Need* is moderate when students need to understand or use the word because the teacher has asked them to do so, and therefore, task requirements are externally imposed; meanwhile, *need* is strong when learners feel the need to use or understand a word, thus, the task is self-imposed and intrinsically motivated.

There is *search* when the word or the definition of a word in the SL is not given to students, and as a result, they have to consult a dictionary or another authority, retrieve it from their memory or infer the meaning. In the original formulation of the hypothesis (Hulstijn and Laufer, 2001: 543-545; Laufer and Hulstijn, 2001: 14-15), only the presence or absence of *search* factor was mentioned, without degrees of prominence. Despite this, Laufer (2017a: 351) has recently distinguished *moderate search*, ‘a search for the meaning of a given word’, from *strong search*, ‘a search for word form to express familiar meaning’.

Evaluation is moderate if the process only implies comparison of the target word with other words, comparison of different meanings of a word, or comparison of several words in order to decide which one fits a context. *Evaluation* is strong if learners have to decide how new words will combine with others in an original context generated by themselves.

To describe a task in terms of an involvement load index, the absence of a factor is marked as 0, a moderate presence as 1, and a strong presence as 2 (Hulstijn and Laufer, 2001: 544). The *ILH* establishes that words processed with tasks that induce higher involvement will be retained better. On the basis of Hulstijn and Laufer (2001: 543-544) and Nation and Webb (2011: 3-5), the score obtained by each factor depending on the requirements of the tasks is shown in Table 1.

Table 1. Score of vocabulary task requirements according to *Involvement Load Hypothesis* criteria

ILH Factor	Score	Task requirements
Need	0	Students do not need to know the word to complete the task.
	1	The teacher asks students to learn the words.
	2	Students feel they need to learn the word when reading or writing.
Search	0	The needed words or the needed meanings are given.
	1	Students must look for or retrieve the meaning of the word.
	2	Students must look for or retrieve the word form.
Evaluation	0	Students do not have to make choices between words or meanings.
	1	Students must recognise differences between words or between several meanings of a word in a given context.

	2	Students use the words in a context and provide that context.
--	---	---

The *ILH* has been decisive for most of the subsequent publications on vocabulary learning. The different studies confirmed, refuted or complemented this hypothesis. One of the most recurrent reviews is the lack of a time factor. Webb (2005), Folse (2006) and Keating (2008) experimented with various types of tasks and concluded that the effectiveness in tasks with different load involvement is reduced or disappears when the time to resolve them is the same. However, the divergence of the results in experiments (see Laufer, 2017b) means that the discussion is still on-going and no definitive conclusions have been made.

Another interesting contribution based on the *ILH* was a study conducted by Zou (2017). With 147 participants, he tested the effectiveness of three tasks: cloze-exercises, sentence-writing and composition-writing. He concluded that the three tasks implied effective learning, but the best results were from participants who worked with composition-writing, followed by students who practised with sentence-writing task, and in the third place were those who used cloze-exercises. He questioned why there are different degrees of effectiveness in writing tasks if both have the same load involvement. Thus, he proposed a new method of scoring the *evaluation* factor. Instead of the moderate-strong binomial, he suggested the following trinomial: moderate evaluation (given context, phrase level); strong evaluation (created context, sentence level); and very strong evaluation (created context and coherently associated, composition level).

Ten years after *ILH* publication, Nation and Webb (2011) proposed a new methodology to measure effectiveness. They considered that Hulstijn and Laufer’s criteria were not sufficient and could cause us to reach the wrong conclusions about tasks. They suggested a new system—*Technique Feature Analysis (TFA)* based on five factors (motivation, noticing, retrieval, generation and retention) and different questions to establish its existence and degree. The factors and the questions are detailed in Table 2. If the answer of each question is positive, the task gets 1 point; if it is negative, it obtains 0 points.

Table 2. *Technique Feature Analysis* factors and questions to assess vocabulary activities (Nation and Webb, 2011: 7)

TFA criteria	TFA Questions
Motivation	<ol style="list-style-type: none"> 1. Is there a clear vocabulary learning goal? 2. Does the activity motivate learning? 3. Do the learners select the words?
Noticing	<ol style="list-style-type: none"> 4. Does the activity focus attention on the target words? 5. Does the activity raise awareness of new vocabulary learning? 6. Does the activity involve negotiation?

Retrieval	<p>7. Does the activity involve retrieval of the word?</p> <p>8. Is it productive retrieval?</p> <p>9. Is it recall?</p> <p>10. Are there multiple retrievals of each word?</p> <p>11. Is there spacing between retrievals?</p>
Generation	<p>12. Does the activity involve generative use?</p> <p>13. Is it productive?</p> <p>14. Is there a marked change that involves the use of other words?</p>
Retention	<p>15. Does the activity ensure successful linking of form and meaning?</p> <p>16. Does the activity involve instantiation?</p> <p>17. Does the activity involve imaging?</p> <p>18. Does the activity avoid interference?</p>

Nation and Webb (2011: 11-15) provided some examples that compare the scores of tasks according to *ILH* and *TFA*. For instance, according to *ILH*, the fill in the blanks task¹ was more effective (they matched 4 out of 6) than word cards task (3 out of 6); however, according to *TFA*, word cards task was more effective (11 out of 18) than fill in the blanks task (8 out of 18).

Although Nation and Webb (Nation and Webb, 2011: 6-15; Webb and Nation, 2017: 235-237) suspected that their evaluation tasks method achieved the objective better than the *ILH* criteria, they did not conduct any experiment to demonstrate it. Hu and Nassaji (2016) showed that *TFA* was more effective at anticipating results than *ILH*. They analysed four tasks (Hu and Nassaji, 2016: 32): reading a text with multiple-choice questions (task 1), reading a text and choosing definitions (task 2), reading plus fill in the blanks (task 3), and reading a text and sentence rewording (task 4). In their study, the task with a lower score according to *ILH*, but a higher score according to *TFA*, i.e. task 3: reading plus fill-in, was the second most effective task. They also dismissed the *ILH* presumption that tasks with a similar load involvement would lead to similar results, since tasks 1, 2 and 4 led to different results despite having the same score.

Considering these and other recent vocabulary studies, Laufer (2017b) compared the results and concluded: (1) incidental learning based on input acquired by reading is effective when the text is interesting or includes images; (2) intentional learning is always effective, especially when participants

¹ Nation and Webb (2011: 321-323) described the activities they mentioned. With regards to *filling in the blanks*, they said: “A list of words is provided, followed by a certain number of sentences with a blank space where a word needs to be inserted. There are more words than blank spaces”. *Word cards* task is described as follows: “The learners choose unknown words and phrases that are useful for them. They make word cards featuring the L2 word on one side and the L1 translation on the other”.

are not explicitly asked to memorize the target words; and (3) learning with a high load involvement is the most effective, mainly if the evaluation is strong or very strong (cf. Zou, 2017).

Our main interest here is to verify whether *ILH* and *TFA* criteria are adequate to predict a successful learning, and to compare them to verify if one is more useful to forecast the effectiveness of several vocabulary-learning tasks.

2. Present study

This research intends to focus on analysing which of the three vocabulary-learning tasks is more effective and facilitates the incorporation of a greater number of words into the mental lexicon of students of, in this case, Spanish as second language. Secondly, the aim of the study is to determine the kind of knowledge that participants acquire with each vocabulary-learning task, that is, receptive or productive knowledge of words. In order to see how the three proposed tasks: (1) definition-choosing, (2) gap-filling, and (3) sentence-writing work in relation to the *ILH* and *TFA* criteria, we assess them below.

This study replicates another carried out in L1 area by Coomber et al. (1986) with speakers of English, and by Matanzo (1991) and Reyes (1995) with speakers of Spanish. With students of Spanish as SL it was replicated by San-Mateo (2005, 2012, 2013). In L1 area, conclusions indicated that sentence-writing task was the most helpful to learn new words; and in SL the results were not definite, so it is necessary to investigate this issue more deeply with a larger number of subjects. This is what we are doing here.

Under the *ILH* (Hulstijn and Laufer, 2001), the three tasks have a moderate need, in other words, motivation is extrinsic because task requirements are externally imposed and students need to understand or use the word as the instructor has asked them to do so, and none of these three tasks induces search. Only the third task (sentence-writing) has a strong evaluation where students make up a context of use, since (2) the gap-filling task only entails recognizing differences between given contexts in order to decide which word fits in the space (moderate evaluation). However, (1) definition-choosing task induces no evaluation (see Table 3). None of them achieves the third level of evaluation (composition level) proposed by Zou (2017).

Table 3. Score of the three tasks used according to the *Involvement Load Hypotheses* criteria

<i>ILH</i> criteria	(1) Definition-choosing task	(2) Gap-filling task	(3) Sentence-writing task
Need	1	1	1
Search	0	0	0
Evaluation	0	1	2
Score	1/6	2/6	3/6

To assess the tasks according to the *TFA* (Nation and Webb, 2011: 6-11; Webb and Nation, 2017: 235-237), we answered the questions proposed by the authors and we obtained the score presented in Table 4.

Table 4. Score of the three tasks used according to the *Technique Feature Analysis* criteria

TFA criteria	(1) Definition-choosing task	(2) Gap-filling task	(3) Sentence-writing task
Motivation	2	2	1
Noticing	2	2	2
Retrieval	3	3	3
Generation	0	2	3
Retention	2	2	2
Score	9/18	11/18	11/18

Regarding *motivation*, all the three tasks have a clear vocabulary learning goal (question #1). But only definition-choosing (*cloze activity*) and gap-filling (*multiple choice*) tasks motivate learning (question #2), since they are, in terms of Nation and Webb (2011: 8), “activities that present a challenge to learners”. None of them allows learners to select the words (question #3).

In relation to *noticing*, the three tasks focus on the target word (which is found in uppercase and bold) (question #4) and they raise awareness about new vocabulary learning (question #5): in (1) definition-choosing task, learners select the correct word from a number of choices; in (2) gap-filling task, students meet words in original sentences; and in (3) sentence-writing task, learners meet words in original sentences and they use words in context (Nation and Webb, 2011: 8). However, none of the three tasks involve negotiation (question #6).

Concerning *retrieval*, all the tasks involve retrieval of the word meaning (question #7). Since students do not need to retrieve the word form, receptive retrieval (recognition) is always involved (question #8). In addition, there is no recall (question #9), but only recognition, since they need to retrieve the meaning by looking at a number of choices in the provided glossary. Although participants could not check the glossary in the second part of the task, they know that the solution is one of the meanings they have seen before in the glossary. Lastly, all the tasks offer two retrievals of the words (question #10), and these retrievals are spaced apart; in other words, the two items referring to the same words do not next to each other of a word (question #11).

In regard to *generation*, only (2) gap-filling task and (3) sentence-writing task involve productive generative use (questions #12 and #13). The first one leads to a low degree of generation, since students have to make “small grammatical or inflectional changes” (Nation and Webb, 2011: 9), whereas sentence-writing task leads to a high degree of generation, since learners must broaden the meaning by

using new collocations and applying derivational affixes (question #14). Thus, only in the sentence-writing task there is “a marked change that involves the use of other words” (cf. Nation and Webb, 2011: 9).

Regarding *retention*, all the tasks ensure successful linking of form and meaning (question #15). In the two parts of the learning tasks, students receive auditory feedback of word form and its meaning. In addition, in the first part they receive visual information of form and meaning when they look up words in the glossary. However, the three tasks do not involve instantiation (question #16) or imaging (question #17). Lastly, the tasks avoid interference because target words are unrelated, meaning that there are no members of a lexical set (question #18).

Therefore, according to *ILH*, (3) sentence-writing task is more likely to result in effective vocabulary learning than (1) definition-choosing or (2) gap-filling tasks, and task 2 will be more effective than task 1 too. Whereas according to *TFA*, (2) gap-filling and (3) sentence-writing tasks will obtain better results, with no differences between them, than (1) definition-choosing task.

The present study addresses the following research questions:

(1) To what extent do proposed tasks, (1) definition-choosing, (2) gap-filling, and (3) sentence-writing, contribute to SL vocabulary learning?

(2) To what extent do the *Involvement Load Hypothesis (ILH)* and *Technique Feature Analysis (TFA)* predict the effectiveness of the three vocabulary-learning tasks?

(3) What kind of knowledge (productive or receptive) of the word does each task encourage?

3. Method

3.1. Participants

The learners were 308 ($n = 78$ males; 230 females) students of Spanish as SL enrolled in a course at any of the following institutions: École Nationale Supérieure des Mines, Université Paris-Dauphine, Université Sorbonne Nouvelle-Paris 3 or Instituto Cervantes in Paris. 85% were between 15 and 30 years old; 7% were between 31 and 54 years, and 8% were older than 55. On the background questionnaire, all indicated French as their native language, 68% of the participants learned another language in addition to Spanish, 30% learned two languages, and 2% studied only Spanish as SL. Their level of Spanish was at least intermediate, B1-B2 according to *CEFR* (Council of Europe, 2001).

3.2. Materials

As this study is a replication, we adopted the three vocabulary-learning tasks from Matanzo (1991), Reyes (1995), and San-Mateo (2005, 2012, 2013), which are commonly included in textbooks: (1) identifying the correct definition of the target word; (2) filling the gap; and (3) writing sentences using a given word. Here is a sample of each one:

(1) Definition-choosing task:

¿Qué significa **PACTAR**? Escribe la palabra al lado de la definición que le corresponda.

1. _____ A propósito, con intención (carácter negativo).
2. pactar Acordar algo entre dos o más personas que están obligadas a cumplirlo.
3. _____ Caer gotas pequeñas de lluvia o de un líquido.
4. _____ Recipiente o vaso en que se conserva y transporta algo.

(2) Gap-filling task:

¿En qué ejemplo usarías la palabra **PACTAR**? Escribe la palabra en el hueco del ejemplo que le corresponda.

1. Me dijo que no lo hizo _____ ni con mala intención.
2. Creo que debemos ponernos de acuerdo y pactar sobre la cantidad de dinero que pagará cada uno.
3. Hay una mancha en el techo y _____ agua sucia.
4. Dame cualquier _____ para guardar lo que ha sobrado de la paella.

(3) Sentence-writing task:

Responde a la siguiente pregunta con una oración completa. Utiliza la palabra indicada en mayúsculas y subráyala.

Tu contestación debe demostrar que conoces el significado de dicha palabra.

¿Crees que es adecuado **PACTAR** con tu pareja una distribución de las tareas domésticas? ¿Por qué?

Sí, me parece muy práctico pactar quién va realizar cada trabajo de la casa porque así evitamos muchos problemas de convivencia.

On one hand, the learning task was composed of 12 separate items or questions (one for target word) of each kind of task (3 sets of 12 items). On the other hand, the vocabulary test had 3 sections of 12 questions each (3 x 12 = 36). In every section, one of the learning tasks was practised (gap-filling task, definition-choosing task, or sentence-writing task). The type of task from the learning phase and the vocabulary test was the same (see examples above). The test was administered after the learning phase.

The glossary consisted of 24 six-letter words without an accent mark: 12 low frequency words in Spanish (fillers or distracting words) and 12 pseudo-words (target words), which were set up following the type of syllable and phonological patterns of the language, as recommended by Nation and Webb (2011: 268). There were six words for each grammar category (noun, adjective, verb and adverb): three distracting items and three target words, and each one had only one meaning. These 12 target words were the words to be learned and measured, so only they were included in the learning tasks and the test. The glossary is provided in the Appendix.

3.3. Procedure

Data was collected at the aforementioned institutions in Paris (France) in October 2016. In each session, after signing the consent form, the participants were randomly divided into three groups, each one carried out one kind of vocabulary-learning task (definition-choosing, gap-filling, or sentence-writing), and they were presented with the glossary and the instructions. The learning phase was made up of two attempts. In the first, the 24-word glossary was available, so it did not purely involve trial and error and the subjects could consult it if so needed; in the second attempt, this was not allowed. In both cases, 45 seconds were provided to complete each of the 12 items, and when the time was up, the researcher read aloud the right definition of the word as feedback so that participants could check the chosen answer.

During the learning phase, the first group ($n = 93$) worked on the task of choosing the right definition for each of the 12 target words; the second group ($n = 105$) practised the gap-filling task; and the third group ($n = 110$) answered the questions in a sentence that included the target word.

To avoid recency effects that would favour the last word of the sequence (Gavett and Horwitz, 2012), a five-to-ten-minute delay was established after the learning phase, during which time participants filled out a background and personal information questionnaire.

Eventually, the vocabulary test was administered to allow participants to demonstrate their knowledge of the 12 target words in the three tasks. The test was the same for the three groups of participants and they worked at their own pace during 15 minutes without the glossary or any feedback. The students had to answer 12 questions of each learning task: first section of the test consisted of 12 definition-choosing items; second section, 12 gap-filling items, and third section, 12 sentence-writing items. In every section, each correct answer scores one point: 12 points is the highest possible score.

4. Results

For the purpose of facilitating the reading, we have organized results in two paragraphs. In the first one we mention the number of words that participants were able to retrieve in the vocabulary test, in order to answer the research question #1: To what extent do proposed tasks, (1) definition-choosing, (2) gap-filling, and (3) sentence-writing, contribute to SL vocabulary learning?, and question #2: To what extent do the *ILH* and *TFA* predict the effectiveness of the three vocabulary-learning tasks? In the second paragraph, we remark the number of words that participants were able to retrieve in each section of the vocabulary test. This allows us to explain the type of knowledge shown by the subjects after the vocabulary-learning task (research question #3: What kind of knowledge (productive or receptive) of the word does each task encourage?).

4.1. Results for vocabulary-learning task variable

The average number of words retrieved, that is, the average total number of correct answers in the test, was 9.36 out of every 12 words (maximum score), which is equivalent to 78% (see descriptive statistics in Table 5). This high figure demonstrates the efficacy of the three vocabulary-learning tasks used to help subjects incorporate new words into their mental lexicon and increase their vocabulary in the SL, regardless of *Involvement Load* and *Technique Feature Analysis* scales. Small standard deviation (2.5) means that the participants' ability to complete the tasks was homogeneous.

Table 5. Mean and standard deviation of test scores

N	Mean*	Standard Deviation	Standard Error
308	9.355	2.532	0.144

Note: *Maximum score = 12

The next step was to consider the results for each of the three tasks. The vocabulary-learning task was the independent variable, and the dependent variable was the number of words retrieved in the vocabulary test. As we see in Table 6, the group which practised (3) the sentence-writing task retrieved 10.34 words in the test, subjects who worked with (2) the gap-filling task retrieved 9.75 words, and the group trained with (1) the definition-selection task managed to retrieve just 7.75 words. This means the sentence-writing task, the only productive task of the three, was the most effective.

Table 6. Mean and standard deviation of test scores regarding vocabulary-learning tasks

Learning task	N	Mean*	Standard Deviation	Standard Error
(1) Definition-choosing task	93	7.7455	2.86128	0.29670
(2) Gap-filling task	105	9.7524	2.14836	0.20966

(3) Sentence-writing task	110	10.3364	1.84766	0.17617
---------------------------	-----	---------	---------	---------

Note: *Maximum score = 12

In order to determine if there were differences between the groups established according to vocabulary-learning task, a variance analysis² (ANOVA) was conducted. The results indicated that the difference between the scores obtained was significant to a confidence level of 95%: $F(2, 305) = 34.525$, $p = 0.000$. Post hoc multiple comparison procedures indicated which pairs of learning tasks had a statistically significant difference.

As variances were unequal, we conducted the Games-Howell test. It determined that, at a confidence level of 95%, the difference between the average achieved by the subjects trained by (1) the definition-choosing task and by the subjects that had practised with the other two tasks was significant ($p = 0.000$). However, the difference between the results achieved by the groups trained through (2) the gap-filling task and (3) the sentence-writing task was not significant ($p = 0.086$). These results indicate that sentence-writing and gap-filling tasks yielded significantly better scores than definition-choosing task; but the comparison between the sentence-writing and gap-filling tasks did not reach statistical significance, which means there was no difference in the outcome of learning new words if one of these two tasks was practised.

4.2. Results by sections of the vocabulary test

In order to answer research question #3: What kind of knowledge (productive or receptive) of the word does each task encourage, we present now the number of words retrieved in each section of the test. So that we will be able to analyse whether receptive learning-vocabulary tasks (definition-choosing and gap-filling) train participants to perform receptive and productive tasks; and whether the productive learning-vocabulary task (sentence-writing) trains participants to perform receptive and productive tasks.

First of all, subjects were able to recognise the meaning of most target words, specifically 10.47 meanings and words were matched (definition-choosing section of the test), and 9.24 words were inserted in the gaps (gap-filling section). Nevertheless, they were only able to use 8.35 words in a new sentence when answering the given questions (sentence-writing section). The general linear model confirmed that

² The analysis of variance is used to examine differences among group means in sample data. A result is statistically significant when it is considered that unlikely differences have occurred by chance. The significance level or probability threshold (p value) is 0.05, this means that only when this value is equal or less than 0.05 are the differences between means deemed statistically significant. Post hoc multiple comparisons are conducted to figure out between which groups differences are statistically significant, and the p value is also 0.05. Depending on equal or unequal variances several post hoc tests are applied (HSD Tukey, Scheffe, Tamhane, Games-Howell and so on).

the difference between these scores was statistically significant ($p = 0.000$). This means all the three learning-vocabulary tasks trained students to complete above all activities which require a receptive knowledge of the words (definition-choosing and gap-filling tasks). However, learners were much less prepared to the sentence-writing task, which requires a productive knowledge of the words.

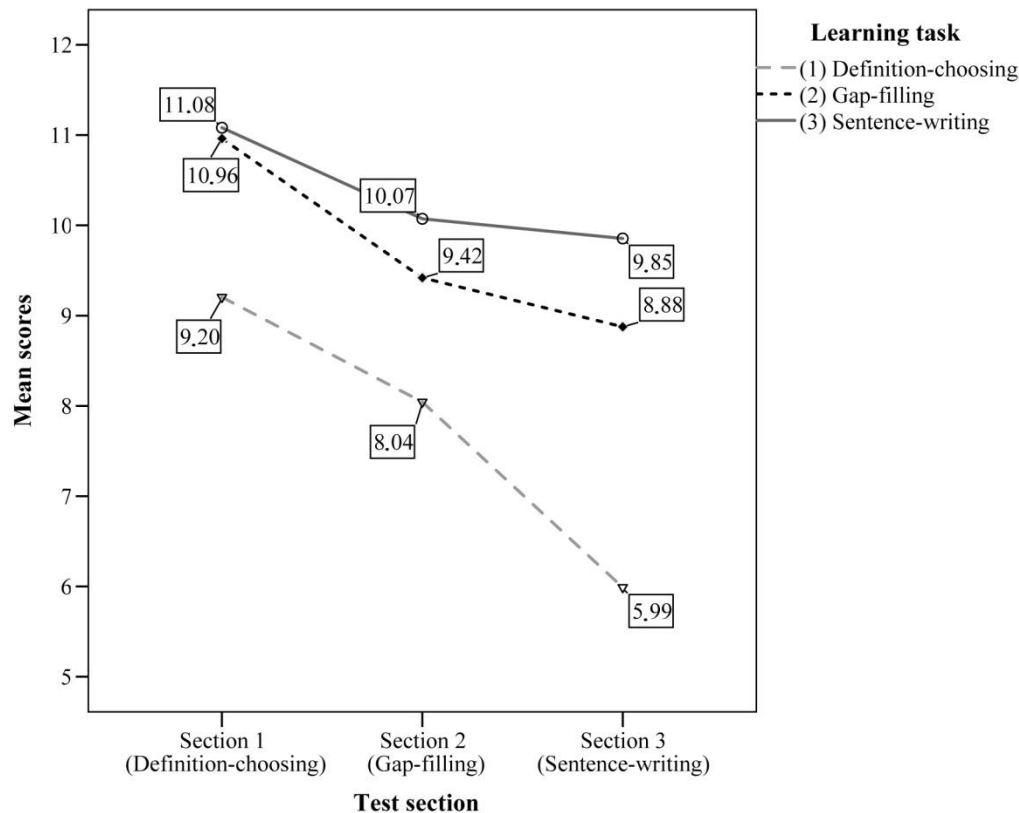
The next step is to consider the results in each section of the test regarding vocabulary-learning tasks to analyse the effect of each task. In Table 7 we can see that subjects trained with definition-choosing task were able to retrieve fewer words in every section of the test than the other two groups who practiced gap-filling and sentence-writing tasks. They were especially poorly enabled to carry out section 3 (sentence-writing, i.e. the productive task): only 5.99 words out of 12 were retrieved. Instead, subjects trained with sentence-writing task scored better than the other two groups in the three sections of the test. Although their training was productive, they performed the receptive tasks (and the productive one) better than the groups trained with receptive tasks.

Table 7. Mean and standard deviation of test scores in each section regarding learning task

Learning task	Test					
	Section 1 (Definition-choosing) Mean (SD)		Section 2 (Gap-filling) Mean (SD)		Section 3 (Sentence-writing) Mean (SD)	
(1) Definition-choosing ($n = 93$)	9.2043	(3.11922)	8.0430	(3.05712)	5.9892	(3.83773)
(2) Gap-filling ($n = 105$)	10.9619	(1.83410)	9.4190	(2.87155)	8.8762	(3.23352)
(3) Sentence-writing ($n = 110$)	11.0818	(1.71412)	10.0727	(2.50764)	9.8545	(2.76557)

The ANOVA analysis indicated that the difference between the scores obtained in each section of the test considering vocabulary-learning task was significant in the three cases, because p value was lower than 0.05: in section 1 (definition-choosing), $F(2, 305) = 21.002$, $p = 0.000$; in section 2 (gap-filling), $F(2, 305) = 13.512$, $p = 0.000$; and in section 3 (sentence-writing), $F(2, 305) = 37.071$, $p = 0.000$. Post hoc comparisons confirmed that in sections 1 and 2, in which definition-choosing and gap-filling tasks were practiced, the learning effect of choosing the right definition and the other two tasks, filling the gap ($p = 0.000$; $p = 0.004$) and writing sentences ($p = 0.000$; $p = 0.000$), was statistically significant. However, the difference between the learning effect of writing sentences and filling the gap was not significant ($p = 0.874$; $p = 0.181$). In section 3 (sentence-writing), the effect of the three learning tasks was statistically significant ($p = 0.000$; $p = 0.000$; $p = 0.048$).

Figure 1. Mean scores in each test section regarding vocabulary-learning tasks



In summary, in the three test sections sentence-writing was the most effective learning task, followed by gap-filling, and by last definition-choosing (see Figure 1). Differences were significant between the three learning task in section 3 (sentence-writing), and only between learning task 1 and the other two (tasks 2 and 3), in sections 1 (definition-choosing) and 2 (gap-filling).

5. Discussion

In relation to the research question #1 we set out, we have shown that the three tasks: (1) definition-choosing, (2) gap-filling, and (3) sentence-writing, contribute to SL vocabulary learning. However, there are degrees of effectiveness between them: the most effective task was sentence-writing, then gap-filling, and in the third place, definition-choosing.

With respect to the research question #2—concerning the ability of *ILH* and *TFA* criteria to predict the effectiveness of the three vocabulary-learning tasks, the *ILH* predicted accurately the scale of effectiveness of our tasks: the task which induces a higher level of involvement load (cognitive effort), i.e. (3) sentences-writing, was the most effective in learning new words. On the contrary, the learning task which entails less involvement load, i.e. (1) definition-choosing, allowed students to learn fewer

words. Nevertheless, the statistical analysis determined no significant differences between (2) gap-filling and (3) sentence-writing tasks.

The *TFA* also predicted that (1) definition-choosing task would yield worse results; however, *TFA* did not foresee a different degree of effectiveness between gap-filling and sentence-writing tasks. In fact, in our study these two tasks have achieved similar results: when one of them was practiced in the learning phase, there was no difference in the outcome of learning new words. But that only happened when we analysed results for vocabulary-learning task variable (see Table 6). If we carefully analyse the results of each test section, we found that when the only productive task (sentence-writing) had to be completed, the effect of the three vocabulary-learning tasks was statistically significant (see Table 7), which means that in this case—when participants had to understand a question and answer it using the target word, there were differences in the outcome of learning new words depends on the vocabulary-learning task.

For this reason, we consider that differences between gap-filling and sentence-writing tasks could be established. Question #8 of *TFA*: “Is it productive retrieval?” (Nation and Webb, 2011: 9) could be itemized. Although in the three tasks retrieval always involves the meaning of the word (not the form), in (2) gap-filling task and (3) sentence-writing task learners must retrieve the different derivations of words³. In (2) gap-filling task, they must do so at a low level, provided that they must retrieve the word, changing it according to a given context; and in (3) sentence-writing task, they carry this out at a high level, provided that they must retrieve the word, changing it according to the context they have created. Therefore, in addition to the establishment of two degrees of *productive generation* (question #14), it is possible that two degrees of *productive retrieval* could also be set.

On the other hand, question #5 of *TFA*: “Does the activity raise awareness of new vocabulary learning?” (Nation and Webb, 2011: 8) could also be itemized. In (3) sentence-writing task the awareness of new learned words was raised in two ways: subjects met new words in the question raised, and they used target words in generated answers; meanwhile in tasks 1 and 2, they simply selected the correct meaning or context from a number of choices.

If these two expansions were applied, the sentence-writing task would rate higher than the gap-filling task, and the results obtained in productive task of this study (section 3 of the test) would support the *TFA*. As stated by Nation and Webb (2011: 7), their proposal is “a first attempt” to evaluate and design teaching techniques, and those developments are our contribution to the analysis.

³ Nation and Webb (2011: 8-9) considered this fact in the explanation of the criteria, but they did not give any points to the writing-sentence task in this item when they assessed this activity (Webb and Nation, 2017: 236).

In regard to the third research question, first of all we must set out what kind of knowledge is encouraged by each task. Considering Nation (2001), the three tasks explore the connection between form and meaning but only the sentence-writing task does it productively. Use aspect (grammatical functions and collocations) is also covered by gap-filling and sentence-writing tasks but, again, the first one involves a receptive knowledge while the second task includes the productive one too. Observing our results, for all participants (1) definition-choosing task was the easiest one. Therefore, the task that requires just a receptive knowledge of the word is also the one that requires less mental effort to the learner. At the same time, the productive task, in other words, sentence-writing, was the most difficult for all learners, but also the most effective, considering that students trained with it retrieved the largest number of words in tasks that need receptive knowledge of the words—(1) definition-choosing and (2) gap-filling tasks, as well as in the task that needs productive knowledge—(3) sentence-writing. On the other hand, definition-choosing was the least effective vocabulary-learning task and least enabled students to carry out any other type of task with the target words, whether it were receptive or productive.

We conclude that receptive trainings can lead to productive knowledge (first group wrote sentences with 5.99 words; second group did so with 8.88), although productive training was more worthwhile than receptive training in both productive and receptive knowledge (third group got the maximum score in the three sections of the test). Therefore, words require different degrees of knowledge in order to be used in receptive or productive tasks. Given that when the productive task had to be completed the results from the group trained with definition-choosing task were also significantly different from the results of the group trained with gap-filling task, we can further conclude that it would be possible to set a proficiency scale even among tasks that require receptive knowledge.

Due to all these reasons, when teachers choose vocabulary-learning tasks, it is advisable to keep in mind the degree of knowledge of the word that they want students to achieve, since knowing a lexical unit involves different aspects, not all of which are always available and which are gradually and progressively incorporated into the mental lexicon of the learners.

6. Conclusions and future research

Findings of this research study are consistent with previous studies that confirm the effectiveness of writing as vocabulary-learning task (Hulstijn and Laufer, 2001; Browne, 2003; Webb, 2005; Keating, 2008; Agustín, 2009; Pichette et al., 2012; San-Mateo, 2012, 2013; Chacón and Labrador, 2016; Andión and San-Mateo, 2018). According to Meara, writing “is a good way to consolidate your knowledge of words”, and furthermore, it “does not put you under time pressure, so it lets you access and rehearse

vocabulary that you can then use later in speech” (Meara et al., 2005: 4-5). However, there are researchers as Folse (2006) and Barcroft (2004) who consider that writing tasks consume too many resources and therefore do not effectively enrich vocabulary.

In our case, after testing the three tasks with more than 300 students with an intermediate level of Spanish, we have concluded that the cognitive effort implied by writing, i. e. creating a context in which new words are included, has shown highly satisfactory results. This effect was predicted by *ILH* and *TFA* criteria, but the *ILH* was more accurate to forecast the scale of effectiveness of the three activities we proposed. Finally, according to our trial, receptive learning results in receptive and productive knowledge of words, although both are poorer than those provided by productive learning.

Further research with a larger sample of participants, with different L1 and level of Spanish (or even any other SL) is needed to explore the ability of *ILH* and *TFA* criteria to predict the effectiveness of different vocabulary-learning tasks. With respect to *ILH* criteria, the trinomial *evaluation* factor (moderate, strong and very strong) by Zou (2017), and the binominal search factor (moderate and strong) by Laufer (2017a) should be tested. On the other hand, with respect to *TFA* criteria, the two degrees of *productive retrieval* and the two ways of *the awareness of new learned words* proposed here could be explored.

7. Works cited

- AGUSTÍN, M.^a Pilar, 2009: “The effect of reading only, reading and comprehension, and sentence writing in lexical learning in a foreign language: some preliminary results”, *Revista Española de Lingüística Aplicada* 22, 9-33.
- ANDIÓN, M.^a Antonieta and Alicia SAN-MATEO, 2018: “Tres actividades de adquisición léxica. Resultados en aprendices brasileños de español como lengua extranjera”, *Calidoscópico* 16(3), 437-449.
- BARCROFT, Joe, 2004: “Effects of sentence writing in second language lexical acquisition”, *Second Language Research* 20(4), 303-334.
- BROWNE, Charles, 2003: Vocabulary acquisition through reading, writing and tasks: A comparison. PhD thesis, Temple University.
- COOMBER, James E., David A. RAMSTAD and Dee Ann SHEETS, 1986: “Elaboration in vocabulary learning: A comparison of three rehearsal methods”, *Research in the Teaching of English* 20(3), 281-293.
- COUNCIL OF EUROPE, 2001: *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*, Cambridge: CUP.

- CHACÓN, Carmen and M.^a José LABRADOR, 2016: “Grados de eficacia en la incorporación léxica en ELE: Resultados en aprendices brasileños (Sao Paulo)” in *Actas del III Congreso Internacional Nebrija en Lingüística Aplicada a la Enseñanza de Lenguas*, Madrid, 23-25 June 2016, Madrid: Universidad Nebrija, 57-63.
- FOLSE, Keith S., 2006: “The effect of type of written exercise on L2 vocabulary retention”, *TESOL Quarterly* 40(2), 273-293.
- GAVETT, Brandon E. and Julie E. HORWITZ, 2012: “Immediate list recall as a measure of short-term episodic memory: Insights from the serial position effect and item response theory”, *Archives of Clinical Neuropsychology* 27(2), 125-135.
- HU, Hsueh M. and Hossein NASSAJI, 2016: “Effective vocabulary learning tasks: Involvement Load Hypothesis versus Technique Feature Analysis”, *System* 56, 28-39.
- HULSTIJN, Jan H. and Batia LAUFER, 2001: “Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition”, *Language Learning* 51(3), 539-558.
- KEATING, Gregory D., 2008: “Task effectiveness and word learning in a second language: the Involvement Load Hypothesis on trial”, *Language Teaching Research* 12(3), 365-386.
- LAUFER, Batia, 2010: “Form-focused instruction in second language vocabulary learning” in Rubén CHACÓN-BELTRÁN, Christian ABELLO-CONTESSÉ, M.^a del Mar TORREBLANCA-LÓPEZ and M.^a Dolores LÓPEZ-JIMÉNEZ (eds.): *Further insights into non-native vocabulary teaching and learning*, Bristol: Multilingual Matters, 15-27.
- LAUFER, Batia, 2017a: “The three “I”s of second language vocabulary learning: Input, Instruction, Involvement” in Eli HINKEL (ed.): *Handbook of research in second language teaching and learning*, Volume III, New York: Routledge, 343-354.
- LAUFER, Batia, 2017b: “From word parts to full texts: Searching for effective methods of vocabulary learning”, *Language Teaching Research* 21(1), 5-11.
- LAUFER, Batia and Jan H. HULSTIJN, 2001: “Incidental vocabulary acquisition in a second language: The construct of Task-Induced Involvement”, *Applied Linguistics* 22(1), 1-26.
- MATANZO, Gloria, 1991: *Vocabulario y enseñanza: Estudio de la relación existente entre los métodos empleados y la incorporación del léxico nuevo a la competencia lingüística de estudiantes universitarios puertorriqueños*. PhD thesis, Universidad de Puerto Rico.
- MEARA, Paul, Batia LAUFER and Paul NATION, 2005: “Ten best ideas for teaching vocabulary”, *The Language Teacher* 29(7), 3-6.

- NATION, Paul, 2001: *Learning vocabulary in another language*, Cambridge: CUP.
- NATION, Paul, 2006: "How large a vocabulary is needed for reading and listening?" *The Canadian Modern Language Review* 63, 59-82.
- NATION, Paul and Stuart WEBB, 2011: *Researching and analyzing vocabulary*, Boston: Heinle.
- PICHETTE, François, Linda de SERRES and Marc LAFONTAINE, 2012: "Sentence reading and writing for second language vocabulary acquisition", *Applied Linguistics* 33(1), 66-82.
- REYES, M.^a Josefa, 1995: *Enriquecimiento de la competencia léxica: análisis estadístico*. PhD thesis, Universidad de Las Palmas de Gran Canaria.
- RICHARDS, Jack C., 1976: "The role of vocabulary teaching", *TESOL Quarterly* 10(1), 77-89.
- RICHARDS, Jack C., 1985: "Lexical knowledge and teaching vocabulary" in *The context of language teaching*, Cambridge: CUP, 176-188.
- SAN-MATEO, Alicia, 2005: *Aprendizaje de léxico en español como segunda lengua. Investigación sobre tres métodos*. Master dissertation, Universidad Nacional de Educación a Distancia.
- SAN-MATEO, Alicia, 2012: *Aprendizaje de léxico en español como segunda lengua/lengua extranjera: investigación sobre la efectividad de tres tipos de actividades para aprender vocabulario*. PhD thesis, Universidad Nacional de Educación a Distancia.
- SAN-MATEO, Alicia, 2013: "El conocimiento productivo y receptivo del vocabulario en L2: Análisis de la efectividad de tareas de reconocimiento y producción", *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas* 14, 13-29.
- SHIRZAD, Maryam, Abbass E. RASEKH and Azizollah DABAGHI, 2017: "The effects of input and output tasks on the learning and retention of EAP vocabulary", *Theory and Practice in Language Studies* 7(2), 145-152.
- VONSYDOW, Albin, 2015: *Vocabulary acquisition. Possibilities within the task-based framework*, Gothenburg: Gothenburg University.
- WEBB, Stuart, 2005: "Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge", *Studies in Second Language Acquisition* 27(1), 33-52.
- WEBB, Stuart and Paul NATION, 2017: *How vocabulary is learned*, Oxford: OUP.
- ZOU, Di, 2017: "Vocabulary acquisition through cloze exercises, sentence-writing and composition writing: Extending the evaluation component of the Involvement Load Hypothesis", *Language Teaching Research* 21(1), 54-75.

8. Appendix

Glossary

Adrede (adv.): A propósito, con intención (carácter negativo).

Aprisa (adv.): Con rapidez o prontitud.

Barmil* (adj.): Se dice de un asunto muy problemático y que provoca discusiones.

Catilo/-a* (adj.): Se dice de una persona valiente, atrevida, que actúa con mucha decisión.

Diforo/-a* (adj.): Fundamental.

Dorsal (adj.): Relacionado con el dorso, la espalda o lomo.

Ecivar* (v.): Forzar u obligar a una persona a hacer algo.

Envase (n. masc.): Recipiente o vaso en que se conserva y transporta algo.

Gotear (v.): Caer gotas pequeñas de lluvia o de un líquido.

Hogaño (adv.): En esta época o año.

Inicuo/-a (adj.): Injusto, malvado.

Jotone* (adv.): Correctamente, sin errores.

Letaer* (v.): Prevenir, precaver.

Muleta (n. fem.): Apoyo de madera o de otro material que ayuda a caminar.

Nesoal* (adv.): Con esfuerzo e interés.

Pactar (v.): Acordar algo entre dos o más personas que están obligadas a cumplirlo.

Pecoso/-a (adj.): Se dice de la persona que tiene pequeñas manchas rojizas en la piel.

Ronoar* (v.): Aceptar una invitación de otra persona.

Sotiro* (n. masc.): Petición que se hace como favor y con humildad; súplica, ruego.

Talefa* (n. fem.): Desorientación; asombro o confusión por algo sorprendente e inesperado.

Urjale* (adv.): Inútilmente.

Ubicar (v.): Situar o colocar algo o a alguien en determinado espacio o lugar.

Viruta (n. fem.): Hoja delgada que se saca de la madera cuando se le pasa el cepillo.

Zienga* (n. fem.): Multitud. Gran cantidad de personas, animales o cosas.

*Target words.