Do activities in graded readers promote vocabulary learning? A Technique Feature Analysis study

¿Las actividades en las lecturas graduadas promueven el aprendizaje de vocabulario? Un estudio de Technique Feature Analysis

Diego Alins Breda
University of California, Davis
dalinsbreda@ucdavis.edu

ABSTRACT

Vocabulary acquisition through reading has been repeatedly studied in second language learning research and several recent studies conclude that incidental learning through reading needs to be complemented with intentional learning activities that target specific words. These learning opportunities are potentially present in graded readers (i.e., books adapted to the proficiency level of the learners). Seeking to contribute to this growing area of scholarship, this study analyzes the activities in 10 elementary Spanish graded readers. The framework used to measure the activities' quality is Nation and Webb's (2011) *Technique Feature Analysis* (TFA). Results show that graded readers include only a low number of activities explicitly focused on vocabulary. In addition, TFA scores were extremely low for most of the activities. Overall, the study indicates the necessity of increasing the quantity and quality of vocabulary activities in these books.

Keywords: Intentional vocabulary acquisition, Technique Feature Analysis, Spanish graded readers, Incidental learning

RESUMEN

La adquisición de vocabulario a partir de la lectura se ha estudiado repetidas veces en la investigación sobre aprendizaje de segundas lenguas y varios estudios demuestran la importancia de complementar el aprendizaje incidental en la lectura con actividades intencionales de aprendizaje de léxico. Un tipo de lectura que brinda esta oportunidad son las lecturas graduadas (es decir, libros adaptados al nivel de los estudiantes). Tratando de contribuir a esta área de estudio, este trabajo evalúa las actividades incluidas en 10 lecturas graduadas de nivel elemental de español. El marco teórico utilizado para medir la calidad de las actividades es Nation y Webb (2011): *Technique Feature Analysis* (TFA). Los resultados muestran que las lecturas graduadas incluyen un número limitado de actividades explícitas de vocabulario y que la mayoría tienen un TFA bajo. Estos datos confirman la necesidad de incrementar la cantidad y calidad de las actividades de vocabulario en estos libros.

Palabras clave: Adquisición de vocabulario intencional, Technique Feature Analysis, Lecturas graduadas en español, Aprendizaje incidental

1. INTRODUCTION

As Horst (2010) demonstrates, the classroom environment does not provide enough lexically rich input for students to learn the necessary vocabulary to understand most words in a given text in their L2. Concretely, when she analyzed the words used by a teacher in a language classroom in every session of a 9-week course, she discovered that many words that would be useful for the students were either not introduced at all or repeated only a few times, making it difficult for incidental learning to take place. In this context, reading, and extensive reading in particular, can be seen as a welcome addition to provide learners with the additional input that they need to develop a broad lexicon. The practice of extensive reading has thus been considered as a complement to other language learning practices that students engage in during L2 classes. Indeed, reading has a demonstrated impact on reading comprehension (Martinez, 2017; Yamashita, 2008) and on the development of a positive attitude towards L2 reading (Brantmeier, 2005; Martinez, 2017; Saito, Horwitz & Garza, 1999; Sellers, 2000; Yamashita, 2004; Yamashita, 2013).

When it comes to vocabulary learning, there is currently a broad consensus that it is possible to achieve some, although limited, incidental vocabulary learning through reading (Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Schmitt, 2008; Teng, 2014). However, the extent of such learning is very variable across studies and learners, and many researchers are now proposing to use a more explicit approach to vocabulary learning/teaching, even in the context of reading activities (Eckerth & Tavakoli, 2012; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Schmitt, 2008; Teng, 2014). Specially, the idea is to complement non vocabulary-focused reading comprehension exercises, which do not explicitly tap into any determined set of target words, with intentional vocabulary learning exercises focused on specific word-learning goals. This combination of incidental and intentional vocabulary learning opportunities aims to tap into both explicit and implicit lexical learning, thus increasing the chances of learning new target words.

Graded readers can potentially satisfy these requirements. By using graded readers, students have the chance to complement incidental vocabulary learning through reading with activities focused on specific words. However, while graded readers' structure seems promising, the exercises' effectiveness will affect the amount of lexical learning that can realistically happen when using these books.

The present study is designed to examine the quality of the activities in a series of graded readers. The framework Technique Feature Analysis, proposed by Nation and Webb (2011), is used to measure these exercises. No previous study has used this framework before to evaluate the quality of the activities in these types of books. In other words, the aim of this article is to evaluate the activities included in ten graded readers for beginner learners of Spanish to analyze the effectiveness of these books in developing lexical knowledge.

2. LITERATURE REVIEW

2.1. Reading and vocabulary learning

Hulstijn (2001) defined incidental vocabulary learning as the learning of vocabulary produced by any activity not explicitly focused on vocabulary. A number of studies on incidental word acquisition have focused on learning through reading (Eckerth & Tavakoli, 2012) and have specifically looked at the number of exposures to a target word necessary for students to learn said word. Findings in this area do, however, show significant variation across studies given the multifaceted nature of word knowledge. For instance, five exposures could be sufficient for orthographic recognition of a new word but not enough to develop a strong association between the word's form and its meaning (Brown, Waring & Donkaewbua, 2008; Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Webb, 2008). As an example of such studies, Pellicer-Sánchez & Schmitt (2010) measured the acquisition of the spelling, word-class, and meaning of 34 words both at the receptive and productive level when reading a novel. The authors concluded that ten exposures allow students to learn spelling and recognition aspects. However, more than ten repetitions are needed to correctly recall a new word and know its word class. For this reason, they highlighted the importance of explicit vocabulary learning activities to consolidate the potential of incidental learning that takes place through mere exposure, a conclusion that was shared by Pigada & Schmitt (2006). The latter authors evaluated the acquisition of word meaning, spelling, and grammatical behavior in one French student reading four different graded readers. They examined whether one month of extensive reading enhanced knowledge of a total of 133 target words. Results demonstrated that even after 20 repetitions the learner had trouble retaining some of the words' meanings, indicating that repetition alone is an insufficient learning tool. This is particularly clear when considering that the student involved in this case study was a highly motivated learner, thus the incomplete learning cannot be due to limited motivation. The obvious conclusion is again that incidental vocabulary learning needs to be supported by additional intentional focus on target words. On the same page, Teng (2014) proposes that explicit word-recall tasks should be included in any vocabulary learning endeavor, even when reading is involved.

In conclusion, available evidence shows that reading alone can slightly enhance a variety of word knowledge aspects when repeated encounters are ensured, but that, at the same time, explicit vocabulary learning tasks are expected to dramatically improve the learning rate of specific target words from the text (Hulstijn, 2001; Schmitt, 2008). One type of book that provides opportunities to meet both requirements is graded readers. On the one hand, they provide level-appropriate reading materials to the learners, by containing a majority of words that students are expected to know at a specified level, while also providing exposure to lower-frequency words relevant to the story. On the other hand, most of these books include activities that tap into different skills, such as grammatical knowledge, vocabulary learning, or reading comprehension. Finally, graded readers are known for positively affecting L2 learners' attitudes towards reading as they cultivate confidence, foster enjoyment, and decrease reading anxiety (Martinez, 2017; Rodrigo, 2016; Rodrigo, 2011).

Therefore graded readers appear to be an excellent tool to complement incidental vocabulary learning through reading with activities focused on specific words. However, before we get to any conclusion, it is essential to measure the quality of the activities to understand the amount of lexical learning that students can learn when using these books.

The present study aims to analyze the activities (both those explicitly aimed at vocabulary learning and those that do so implicitly) included in a series of graded readers aimed at beginner learners of L2 Spanish. The next section thus addresses the different models that have been used over time to evaluate the effectiveness of several types of exercises in developing lexical knowledge.

2.2. Ways to assess the effectiveness of reading activities for vocabulary learning

Two models have been used so far to determine pedagogical activities's usefulness for the purposes of lexical learning: Laufer & Hulstijn' (2001) Involvement Load Hypothesis (ILH) and Nation and Webb's (2011) Technique Feature Analysis (TFA). ILH builds a motivational-cognitive construct that is based on the assumption that deeper processing results in better retention (Craik and Lockhart, 1972). The model consists of three components: need, search, and evaluation. Need is the motivational component and can be moderate (1 point), if an external agent imposes the activity, or strong (2 points), if the task is self-imposed by the student. The two other components constitute the cognitive dimension. Search is the attempt to find the meaning or the form of an unknown L2 word (e.g., by consulting a dictionary or a teacher). Evaluation entails comparing the different meanings of a word, or comparing similar words to choose the one that best fits in a specified context. If the task demands students to compare the difference between words' meanings or choose between several meanings of a word (as in a fill-in-the-blanks activity where the target words are provided), it is considered moderate. If the evaluation requires a decision about combining several words with the target word (as in an original sentence or text), it is considered strong. Each of the three factors can be absent (0 points) or present to a higher or lower degree, receiving 1 or 2 points respectively. The sum of the scores obtained for need, search, and evaluation represents the total ILH of a given task, with six being the highest ILH possible. Hulstijn and Laufer (2001) provide two examples of tasks to explain how the involvement loads can vary. In task one, students have to write sentences with words provided by the teacher. The task induces a moderate need, as it was imposed by an external agent, no search, because the words were given, and strong evaluation, as students have to write an original text. The final score of this activity would thus be 3 (1 + 0 + 2). In the second task, students have to read a text with relevant words glossed, and answer comprehension questions. Need is moderate because it is imposed by the teacher, and search and evaluation are not present in this task because the meaning of the new words are provided. This activity's involvement load is of 1 (1 + 0 + 0), meaning that this activity should be less effective to learn vocabulary than task one.

Several studies have examined the efficacy of ILH. Hulstijn & Laufer (2001) measured the retention of ten words by adult English as a Second Language (ESL) students in Israel and Netherlands after the completion of three different tasks: reading comprehension with marginal glosses, with an involvement index of 1 (moderate need, no search, and no evaluation), reading comprehension plus fill-in-the-blanks, with an involvement index of 2 (moderate need, no search, and moderate evaluation), and writing a composition using target words, with an involvement index of 3 (moderate need, no search, and strong evaluation). The findings related to the students in Israel were in line with the predictions of the ILH. The composition group scored higher than the reading plus fill-in students, and this group scored higher than the reading with glosses group. However, students of the Netherlands obtained higher results in the composition group but the two remaining groups did not show significantly different results. This latter set of data, thus, only partially

supports the ILH's predictions. Following Hulstijn & Laufer (2001), Kim (2008) used the same three tasks (i.e., reading with marginal glosses, reading plus fill-in, and writing a composition). The results were similar to those obtained by the Netherlands students in Hulstijn & Laufer (2001) experiment. Writing a composition, the task with the highest ILH, had the best score and, as in Hulstijn & Laufer (2001), there was no significant difference between reading with marginal glosses and reading plus filling-in. Finally, Keating (2008) measured retention after completing different tasks with beginning learners of Spanish. In line with the predictions of the ILH, retention was highest in the task with the highest amount of involvement (sentence writing using target words), lower in the reading plus filling task, and lowest in the reading with glosses. In sum, ILH has received support in the literature, based on evidence of its usefulness, but critics have also arisen who defend that other factors may be more (or just as) important than the three aspects taken into account in ILH in how words are ultimately remembered in the long term.

For example, Keating (2008) and Kim (2008) measured the role of time on task and partially criticized ILH because it does not consider this important factor. Likewise, Folse (2006) selected three tasks: one fill-in-the-blank exercise, three fill-in-the-blank exercises (i,e., the same words were practiced three different times), and one original sentence writing exercise. From the perspective of ILH, the third task was supposed to have the best effect on learning but, when time on task was controlled, students doing activity 2 outperformed learners doing the other two tasks. Based on these results, the author claimed that "the important feature of a given L2 vocabulary exercise is not the depth of word processing but the number of word retrievals required" (Folse, 2006, p.273). Moreover, Eckerth & Tavakoli (2012) worked with advanced learners of L2 English to investigate the effects of the variables' word exposure frequency' and 'elaboration of word processing' on word retention. Results showed that the elaboration with which unfamiliar words are processed is more relevant to word retention than repeated encounters with the same words. Finally, two more recent studies partially support ILH. Zou (2017) compared a cloze task, a task that required writing a sentence, and another that asked students to write a composition, and found that the first one (1+0+1) promoted less learning than both writing tasks (1+0+2), which supports the ILH. However, the composition-writing outperformed the sentence-writing even though they were accorded the same load. Yang et al. (2017) showed that the task with the lowest involvement load (comprehension only) was outperformed by sentence writing and cloze task as predicted by the ILH. Nevertheless, no difference was found between the students' performance on the cloze task (1+0+1) and the sentence-writing task (1+0+2).

In 2011, Nation and Webb created another theoretical framework: Technique Feature Analysis (TFA). TFA introduced more criteria to operationalize depth-of-processing than those in the ILH. This framework has five components: noticing, retrieval, generation, motivation, and retention. These factors are formed of different criteria, and each of these criteria can have a score of one, if it is present, or zero if it is absent (Nation & Webb, 2011). As seen in Table 1, the highest possible score for an activity is 18, and, as in ILH, the higher the score, the better the activity.

CRITERIA	SCORES	
Motivation		
Is there a clear vocabulary learning goal?	0	1
Does the activity motivate learning?	0	1
Do the learners select the words?	0	1
Noticing		
Does the activity focus attention on the target words?	0	1
Does the activity raise awareness of new vocabulary learning?	0	1
Does the activity involve learning?	0	1
Retrieval		
Does the activity involve retrieval of the word?	0	1
Is it productive retrieval?	0	1
Is it recall?	0	1
Are there multiple retrievals of each word?	0	1
Is there spacing between retrievals?	0	1
Generation		
Does the activity involve generative use?	0	1
Is it productive?	0	1
Is there a marked change that involves the use of other words?	0	1
Retention		
Does the activity ensure successful linking of form and meaning?	0	1
Does the activity involve instantiation?	0	1
Does the activity involve imaging?	0	1
Does the activity avoid interference?	0	1
Maximum score	18	

Table 1. Technique Feature Analysis (adapted from Nation and Webb, 2011, p. 7)

Motivation is related to whether the activity has a clear vocabulary goal and motivates learning. Noticing concerns whether the activity focuses attention on the target words or not and if the activity involves negotiation. Regarding retrieval, the authors differentiate between receptive and productive retrieval, the number of retrievals, and spacing between them. They also determine if the activity focuses attention on recall or recognition of the target words. Generation involves encountering (receptive generation) or using (productive generation) a word in a novel context. Fill-in-the-blanks exercises with previously taught words entail receptive generation, while sentence production activities involve productive generation. Finally, retention focuses on whether the activity ensures successful linking of form and meaning and if the activity involves instantiation or imaging, and whether or not it avoids interference.

Nation and Webb (2011) made a scoring comparison between ILH and TFA on different vocabulary activities, and they found some disagreements between these two frameworks. For example, upon calculating the final score for a fill-in-the-blanks activity, this task had a low TFA score (8 out of 18) but a high ILH score (4 out of 6). These diverging results entail different predictions depending on what framework is used when evaluating an activity. Therefore, a few studies have empirically examined and compared such predictions. Hu and Nassaji (2016), for example, contrasted TFA and ILH in four different tasks (i.e., reading a text with multiple-choice items, reading a text and choosing definitions, reading plus fill in the blanks, and reading and rewording the sentences) with adult ESL students who aimed to learn the meaning of 14 unknown words. These learners

were divided into four groups and completed one of four vocabulary activities ranked differently by TFA and ILH. The results showed two important conclusions: (1) TFA was more accurate than ILH in predicting vocabulary learning gains during task performance and in pretest to post-test vocabulary gains, and (2) form-focused tasks (i.e., activities directly related to the target words) were more productive to gain lexical learning. Gohar, Rahmanian, and Soleimani (2018) also worked with ESL students to assess the diverging predictions of TFA and ILH into three different activities (i.e., sentence making, composition, and reading comprehension). This investigation concludes that, although none of the vocabulary frameworks predicted vocabulary learning entirely, TFA was better than ILH at predicting the pretest to post-test score change, but was not as useful to predict students' learning during the task. Finally, Zou et al. (2018) compared four different vocabulary tasks to evaluate the TFA predictions' effectiveness. The post-test results were consistent with the TFA checklist, and the authors approved the reliability of this framework to measure word learning activities. The present study will rely on TFA, rather than ILH, as results from previous literature seem to indicate that it is more precise and accurate to determine the quality of a specific vocabulary learning task.

3. RESEARCH QUESTIONS

This investigation aims to use the TFA framework to assess the usefulness of the exercises included in ten graded readers for L2 Spanish beginner learners and, therefore, answer the two following research questions:

(1) How many explicit vocabulary exercises are included in each book's total number of activities, as compared to other types of reading comprehension or grammar exercises?

(2) What TFA scores do all the books' activities have, both those that explicitly tap into lexical learning and those that may do so more implicitly?

4. METHODS

4.1. Corpus

Ten graded readers were randomly selected from a pool of books from a variety of publishers from Spain. It needs to be said here that locating published graded readers from publishers that were not from Spain was extremely difficult, thus explaining the selection bias towards Spanish books. As a consequence, six of the stories take place in Spain, while only four take place in three different Latin American countries (Costa Rica, Cuba, and two of them in Mexico). Based on CEFR standards (Common European Framework Reference for Languages), the level of all the graded readers analyzed was between A1 and A2, ranging from *El secreto de su nombre*, an introductory A1 level book, to *La llamada de la Habana* with a high A2 level. In other words, these graded readers are focused on beginner learners but have different publication years, editorials, and activities. The reason for this diversity is to have a sample as much heterogeneous as possible to replicate the graded readers market at the elementary level of Spanish. By doing so, we will be able to understand what type of activities are proposed at this specific level.

TITLE	YEAR OF PUBLICATION	PUBLISHER	LEVEL	WORDS USED	NUMBER OF ACTIVITIES
Muerte en Valencia	1987	Edelsa	Level 1	n/a	0
La sombra de un fotógrafo	1991	Santillana	Level 1	Less than 400	46
El secreto de su nombre	2014	Santillana	Introductory level	Less than 300	36
Me llamo Yago, Yago Ayala	2014	EnClaveELE	A1- A2	500 words	22
La llamada de la Habana	2003	Difusión	A2 +	n/a	12
Fantasmas en la escalera	2009	Difusión	A1	n/a	3
Dos semanas con los ticos	2008	Difusión	A1- A2	n/a	14
Ojalá que te vaya bonito	2014	Difusión	A1- A2	n/a	17
Amnesia	2002	Edinumen	Initial Level	n/a	26
Historia de una distancia	2001	Edinumen	Initial Level	n/a	23

Table 2. Graded Readers

4.2. Activities in the books

In addition to the text itself, most graded readers include activities to work on listening and reading comprehension, oral and writing production, and vocabulary and grammar acquisition. Only one of the books, *Muerte en Valencia*, has no activities and *Fantasmas en la escalera* has only three. On the other hand, *La sombra de un fotógrafo* and *El secreto de su nombre* have the most activities with 46 and 36, respectively. The rest of the graded readers include around 20 activities each.

ACTIVITIES	FINAL SCORE	EXAMPLE
Word cards	11	An activity that uses cards to help students to remember the meaning of the words.
Completing a word parts table	9	An activity that asks students to use prefixes to form new words.
Writing using target words	8	An activity that involves oral or written production using specific words.
Using the keyword technique	8	An activity that requires students to use what a word sounds like to visualize something memorable that will help them later recall the definition.
Filling in the blanks	8	An activity that requires students to complete a sentence with a word retrieved from memory.
Finding the words in the text	8	An activity that asks students to find sentences in the text about a specific topic.

Reading plus fill in	7	An activity that involves completing a text or a sentence with words from a list.
True/ False statements	6	An activity that involves students to answer comprehension questions about a given text using true and false statements.
Multiple-choice items about a given text	6	An activity that asks students to answer comprehension questions about a specified text with multiple-choice options.
Rewording sentences	6	An activity that asks students to rewrite a sentence.
Reading the words, choosing the definition	6	An activity that requires students to join different expressions with their meanings
Reading with glosses	5	An activity that focuses attention on the reading of a text with the help of glosses.

Table 3. TFA applied to 12 different activities (adapted from Nation and Webb, 2011)

As an illustration of how the TFA framework could help define the characteristics of L2 pedagogical activities from the perspective of vocabulary learning, Nation and Webb (2011) calculated the TFA scores of 12 different activities shown in Table 3. Most of the graded readers' activities were listed in those TFA analyses, but some exercises from the books were not considered in the original table presented by the authors. To evaluate all the activities in the books, three different Spanish teachers were asked to assess them using the TFA framework. In the first round of ratings, two teachers attributed a TFA score to each of the tasks and reached 80% of inter-rater agreement. In a meeting, the differences in the tasks' scores were discussed to reach an agreement. Based on those agreements, they wrote a detailed document with instructions on how to calculate TFA scores. A third rater was then invited to use the document to rate the 20% of activities where a consensus had not been initially reached. The final inter-rater agreement between the three final raters for those activities reached 100%. Finally, the activities that did not match any of the activities in Table 3 were organized in Table 4. The description, final TFA scores, and examples of the types of activities that were not in the original study by Nation and Webb (2011) are provided in this table.

ACTIVITIES	FINAL SCORE	EXAMPLE
Look for words in the dictionary	8	An activity that focuses attention in a word to look up the meaning.
Find information related to a topic	7	An activity that asks students to look up information on the internet.
Look up target words and use them in oral or written production	9	An activity that asks students to find words in a text and write a text with them.
Join columns based on a given text	6	An activity that focuses attention on reading comprehension and asks students to link the information in the columns.
Identify similar words.	7	An activity that focuses attention on target words and asks students to link the information in the columns.
Reading and listening plus fill in about a text	6	An activity that requires students to read or listen to a story and answer questions about it.

Read plus verify the information	6	An activity that asks students to read a story and verify the information previously listened
Writing or speaking without target words	7	An activity that asks students to talk with other students about a specific topic
Writing or speaking based on an image	8	An activity that asks students to describe a photo
Writing or speaking specific vocabulary based on an image	10	An activity that requires students to write the correct word to identify each illustration
Identify cognates	6	An activity that asks students to recognize the words that are cognates.
Crosswords	9	An activity that requires students to complete a crossword based on definitions or in a picture.
Write a list of words about a specific topic.	9	An activity that asks students to think about a topic and write words related to it.
Inflectional morphology	6	An activity that asks students to write the plural of a word or its conjugation.

Table 4. TFA scores and description of the activities that were not listed in Nation and Webb (2011)

5. RESULTS

The number of activities included in the analyzed graded readers varied greatly across books. While most of the books analyzed had between 15 to 25 activities, two of them included several tasks much higher than the rest: *El secreto de su nombre* (36 activities) and La sombra de un fotógrafo (46 activities), both published by Santillana. On the other hand, *Fantasmas en la escalera* included only three activities, and *Muerte en Valencia* did not propose any activities. In this context, not all books presented the same potential in terms of how much vocabulary could be learned through explicit activities. However, a book with few activities focused on vocabulary learning may be more efficient in developing lexical knowledge than a book with multiple activities that do not tap into specific vocabulary at all, or do so with a very low TFA.

Therefore, all books' activities were classified into (1) explicit vocabulary activities and (2) other activities that did not explicitly tap into any specific target words to learn. This second set of activities, however, could potentially require readers to use new vocabulary or, at least, understand it, in order to answer reading comprehension questions, for example. Thus, the main difference between both was that the former type of exercises would explicitly aim at practicing and using a specific set of target words whereas the latter would not specifically target any word in an explicit way. Table 5 provides a list of the total number of activities in each book and the number of vocabulary-focused and non vocabulary-focused activities. As evidenced in the table, vocabulary is rarely at the center of the activities. Only in *Fantasmas en la escalera* are all the activities, albeit limited in number, explicitly focused on vocabulary acquisition. The rest of the books include less than 40% of vocabulary oriented activities. For instance, in *Me llamo Yago, Yago Ayala* only 4.5% of the total activities explicitly address lexical learning, *La llamada de la Habana* includes 12 activities, but none of them focus on vocabulary learning, and only 8.7% of the activities in *La sombra de un fotógrafo* are geared towards the learning of specific words. This last graded reader deserves a special mention because it includes 46 activities, and only four of

them are focused on vocabulary, which indicates that the inclusion of an overall high number of exercises does not necessarily entail an increase in vocabulary activities. In the same way, a clear relation between publishers and vocabulary activities was not found. In other words, graded readers from the same publisher had different proportions of vocabulary activities with no apparent preference among publishers for more, or less, vocabulary-oriented activities. For example, *La llamada de la Habana, Ojalá que te vaya bonito*, and *Fantasmas en la escalera* are published by Difusión, but the first book includes 0% of vocabulary activities while the second includes a total of 41.2% and the latter a total of 100% of such activities.

	EXPLICIT VOCABULARY ACTIVITIES	OTHER ACTIVITIES	TOTAL NUMBER OF ACTIVITIES
Ojalá que te vaya bonito	6 (35.3%)	11 (64.7%)	17 activities
Dos semanas con los ticos	4 (28.6%)	10 (71.4%)	14 activities
Historia de una distancia	8 (34.8%)	15 (65.2%)	23 activities
La llamada de la Habana	0 (0%)	12 (100%)	12 activities
Me llamo Yago, Yago Ayala	1 (4.5%)	21 (95.5%)	22 activities
Amnesia	9 (34.6%)	17 (65.4%)	26 activities
La sombra de un fotógrafo	3 (6.5%)	43 (93.5%)	46 activities
El secreto de su nombre	7 (19.4%)	29 (80.6%)	36 activities
Fantasmas en la escalera	3 (100%)	0 (0%)	3 activities
Muerte en Valencia	0 (0%)	0 (0%)	0 activities

Table 5: Number of vocabulary focused vs. non-vocabulary focused activities

In order to address the second research question, we further explored the TFA of all the activities in the graded readers. Table 6 lists the average TFA scores of the activities in each graded reader and the number of activities in each book that pertain to the following ranges of TFA scores: 0-5, 6, 7-8, 9-10, 10+. As a reminder, a higher TFA indicates that the activity is expected to better serve the students in learning new vocabulary, and the maximum TFA score is 18.

	ACTIVITIES AVERAGE SCORE (TFA)	0-5	6	7-8	9-10	+10
Ojalá que te vaya bonito (17 activities)	6.88	0	8 (47.1%)	9 (52.9%)	0	0
Dos semanas con los ticos (14 activities)	6.57	0	9 (64.3%)	4 (28.6%)	1 (7.1%)	0
Historia de una distancia (23 activities)	6.39	0	16 (69.6%)	7 (30.4%)	0	0
La llamada de la Habana (12 activities)	6	0	12 (100%)	0	0	0

Me llamo Yago, Yago Ayala (22 activities)	6.18	0	19 (86.4%)	3 (13.6%)	0	0
Amnesia (26 activities)	6.4	0	15 (57.7%)	7 (26.9%)	4 (15.4%)	0
La sombra de un fotógrafo (46 activities)	6.21	0	41 (89.1%)	4 (8.7%)	1 (2.2%)	0
El secreto de su nombre (36 activities)	6.64	0	22 (61.1%)	12(33.3%)	2 (5.6%)	0
Fantasmas en la escalera (3 activities)	8	0	0	3 (100%)	0	0
Muerte en Valencia (0 activities)	0	0	0	0	0	0

Table 6. Activities Score

As evidenced in Table 6, most activities present low TFA scores, with all books, except for one, having overall average scores that range between 6 and 7. *Fantasmas en la escalera* is the graded reader with the highest average score, even though eight is far from the maximum possible score of 18, and the book only includes three activities. Only four graded readers have at least one activity with a score of nine: *Dos semanas con los ticos* (one activity), *Amnesia* (four activities), *La sombra de un fotógrafo* (one activity), and *El secreto de su nombre* (one activity). Finally, *El secreto de su nombre* is the only graded reader with an activity that reaches a score of 10, which is the highest in this analysis. There are no activities with a score lower than six and no activities with a score higher than 10.

6. DISCUSSION

Previous literature highlights the importance of incorporating explicit vocabulary learning activities in otherwise incidental learning approaches based on repeated exposure to new words through reading (Pellicer-Sánchez & Schmitt, 2010; Pigada & Schmitt, 2006; Schmitt, 2008; Teng, 2014). Because of their structure, graded readers have the potential of being excellent tools to combine this type of complementary, incidental and intentional, vocabulary learning opportunities. Indeed, these proficiency-adapted books offer repeated exposures to target words and generally include activities at the end of the text. This analysis aimed to respond to two questions that explore the usefulness of 10 beginner L2 Spanish graded readers for vocabulary learning:
    (1) How many explicit vocabulary exercises are included in each book's total number of activities?
    (2) What TFA scores do all the books' activities have, both those that explicitly tap into lexical learning and those that may do so more implicitly?
    Upon analyzing the readers, it is evident that none of them focused their end-of-book activities on specific vocabulary learning goals. In all the books, except for two, vocabulary-oriented activities represented less than 40% of the total exercises proposed, with some books including zero activities aimed at developing lexical knowledge. These results indicate that graded readers' potential as tools for vocabulary learning is limited due to the lack of opportunities for explicit lexical learning. This means that learners have to rely almost exclusively on incidental learning through repeated exposure to learn new words from the text. However, not only activities that are explicitly focused on vocabulary learning

can promote such new word acquisition. For example, reading comprehension exercises can require that learners go back to the text and use the new vocabulary to respond to the questions, hence developing lexical knowledge in a more incidental way but with more attention than by simple exposure through reading. Therefore, the second question addressed the extent to which all activities, explicitly focused on specific target words or not, could promote lexical acquisition, as evidenced by their TFA scores.

While the maximum possible TFA score for an activity is 18, most of the exercises had a final score ranging between 6 and 8. This shows that the usefulness of those activities for lexical development is relatively low overall. Nonetheless, these results are similar to the ones presented by Nation and Webb (2011) when they analyzed a list of 12 different activities, as most of the activities they studied also scored between 6 and 8. The highest score found in their list was a word card activity with 11 points and a word part table activity with 9 points. Two important conclusions can be drawn from this analysis. First, in this study and in the examples proposed by Nation and Webb (2011) the results show that the final score of tasks explicitly focused on vocabulary are higher than the score on non-vocabulary focused activities. For this reason, this analysis claims for the inclusion of more vocabulary focused activities on graded readers to complement the incidental learning approach of reading and, therefore, increase the opportunities for lexical learning. Second, the fact that no activities, either in this study or in Nation and Webb's reach a score that is even close to the maximum may also mean that the TFA framework includes too many variables to actually reflect real activities.

This means that a revision of the framework may be needed to make it easier and more useful for teachers. As a first step, it would be necessary to eliminate the points that are not present in most of the exercises. From all the activities analyzed in this project, there are five TFA criteria not presented in any of the activities and only presented in one of the exercises introduced by Nation and Webb (2011). In other words, a number of the criteria in the TFA framework are not relevant enough to evaluate the most important activities for lexical development. As a second step, future studies should be focused on the importance of each criterion. Although Nation and Webb (2011) give scores of one and zero to each factor, it would be necessary to analyze if activities with different factors involved but with the same TFA score have the same effectiveness. By doing so, language instructors could use the TFA framework to choose the most valuable activities.

7. CONCLUSION

When the goal is vocabulary learning, graded readers can be useful in the L2 classroom. These books provide an opportunity to learn new words incidentally through reading and have the opportunity to complement this reading with activities that tap into different skills. However, the exercises included in this study's graded readers did not focus primarily on lexical learning and, very few of the exercises presented a high TFA. Given that explicit vocabulary learning tasks score the highest in terms of TFA, the most useful way to complement incidental learning through reading would be for L2 teachers to design additional exercises that tap into the specific words that they want their students to learn. By letting students simply read the story and respond to the few available questions in the graded readers, it would be unreasonable to expect much lexical learning to take place.

One possible explanation for the lack of activities focused on vocabulary learning in the graded readers analyzed can be the students' level. At the elementary level, graded readers' tasks might have to be more focused on reading comprehension activities' to make sure students understand the text's main ideas. When the student's level increases, graded readers can dedicate more activities to work on grammatical knowledge and vocabulary learning. For this reason, in future studies it would be relevant to include a broader range of graded readers, which include books aimed at students with different levels of proficiency.

On the same page, a revision of the TFA framework is suggested by this study. Five of the 18 criteria are not present in any of the activities analyzed in the graded readers and are not relevant enough to measure vocabulary activities in the graded readers. Furthermore, the TFA framework gives the same value to the 18 components. A future study should analyze if a number of the criteria are more important than the others.

REFERENCES

Brantmeier, C. (2005). Anxiety about L2 reading or L2 reading tasks? A study with advanced language learners. *The Reading Matrix*, 5(2), 67 -85.

Brown, R., Waring, R., and Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20, 136– 163.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6), 671-684.

Eckerth, J. and Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research*, 16, 227 – 252.

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

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, 61, 355–382.

Hu, H. M., and Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load hypothesis versus technique feature analysis. *System*, 56, 28–39.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.

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

Jafari Gohar, M., M. Rahmanian, H. Soleimani. (2018). Technique feature analysis or involvement load hypothesis: estimating their predictive power in vocabulary learning. *Journal of psycholinguistic research*, 47(4), 859-869.

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

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

Martinez, L. (2017). Adult Hispanic ESL Students and Graded Readers. *The CATESOL Journal*, 29 (2), 41- 59.

Nation, P., and Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.

Pellicer-Sanchez, A., and Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, 22 (01), 31-55.

Pigada, M., and Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18, 1–28.

Rodrigo, V. (2011). El primer libro, la primera victoria. Actitud de los estudiantes hacia la lectura extensiva. In J. de Santiago Guervós, H. Bongaerts, J. Sánchez Iglesias, and M. Seseña Gómez (Eds.), *Del texto a la lengua: La aplicación de textos a la enseñanza aprendizaje del español L2–LE* (755–768). (Vol. 2). Spain: Universidad de Salamanca.

Rodrigo, V. (2016). Graded Readers: Validating Reading Levels across Publishers. *Hispania*, 99 (1), 66–86.

Saito, Y., T. J. Garza, and E. K. Honvitz. (1999). Foreign Language Reading Anxiety. *Modern Language Journal*, 83, 202-218.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–363.

Sellers, V. (2000). Anxiety and Reading Comprehension in Spanish as a Foreign Language. *Foreign Language Annals*, 33 (5), 512- 520.

Teng, F. (2014). Incidental Vocabulary Learning by Assessing Frequency of Word Occurrence in a Graded Reader: Love or Money. *LEARN Journal*, 7 (2), 36–50.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, 20, 232–245.

Yamashita, J. (2004). Reading attitudes in L1 and L2, and their influence on L2 extensive reading. *Reading in a Foreign Language*, 16 (1), 539-578.

Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, 36, 661–672.

Yamashita, J. (2013). Effects of extensive reading on reading attitudes in a foreign language. *Reading in a Foreign Language*, 25 (2), 248- 263.

Yang, Y., Shintani, N., Li, S., and Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System*, 70, 38-49.

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

Zou, D., Wang, F., Kwan, R., and Xie, H. (2018). Investigating the effectiveness of vocabulary learning tasks from the perspective of the technique feature analysis: The effects of pictorial annotations. *Technology in education: Innovative solutions and practices*, 3-15.