




3,000 words in Spanish L2 basic language courses: A reachable goal?

Nausica Marcos Miguel, Ghent University 

Joshua R. Edge, Austin College

Mary Elizabeth Beaton, Denison University 

Claudia H. Sánchez-Gutiérrez, University of California, Davis 

Abstract

While studies on lexical development in English L2 abound, less is known about how learners develop their lexicons in other L2s and how their developmental paths relate to lexical frequency counts. To fill this gap, this longitudinal study tracks the receptive lexical knowledge of students who progress through three semesters of Spanish L2 in a US university. Using an online receptive vocabulary test taken at the end of each semester, this study explores what percentage of the 3,000 most frequent Spanish words (overall and by frequency band) these learners recognized. Factors influencing outcomes such as whether the students had Spanish courses before the university, or whether they spoke Spanish outside of class were also examined. Results are consistent with English L2 research. Moreover, as L2 learners' proficiency increased, less additional vocabulary was learned. Previous experiences and use of Spanish outside of class positively influenced scores. On average, learners could recognize around 65% of the most frequent 3,000 words by the end of the third semester. These findings have practical implications for designing the vocabulary component of language courses during and after the first three semesters.

Keywords: *vocabulary, Spanish L2, vocabulary size, classroom instruction, vocabulary frequencies*

APA Citation: Marcos Miguel, N., Edge, J. R., Beaton, M.E., & Sánchez-Gutiérrez, C. H. (2023). 3,000 words in Spanish L2 basic language courses: A reachable goal? *Second Language Research & Practice*, 4(1), 17–47.
<http://hdl.handle.net/10125/69877>

Introduction

Most universities in the United States set their second language requirements at two or three semesters of language study (Lusin, 2012). Because of a lack of empirical research in Spanish classes, it is not known how much vocabulary growth and retention learners can experience semester to semester in this context. However, a learner is expected to know the 2,000 most frequent words in a language to reach an Intermediate-Low level. This relationship can be visualized in [Table 1](#), which shows estimated vocabulary sizes in relation to language proficiency and hours of study based mostly on studies of English L2 and English Foreign Language (EFL). [Table 1](#) suggests that a student will need four years to reach knowledge of the 3,000 most frequent words and beyond. Ideally though, learners would recognize the 3,000 most frequent words at the end of their language requirement as these words have been the threshold suggested to understand most texts (see Vilkaitė-Lozdienė & Schmitt, 2019). Indeed, in Spanish, these 3,000 most frequent words offer 94% coverage in oral texts, approximately 90% coverage in fictional written texts and 88% of non-fictional written texts (Davies, 2005), which comes close to the 95–98% lexical coverage needed to understand a text (see Webb & Rodgers, 2009; Schmitt & Schmitt, 2014).

Knowing what percentage of these 3,000 words students can recognize after a basic language program will be helpful for Language Program Directors (LPDs) and instructors alike. That is, since some students go on to pursue a language major, LPDs and instructors who have a clearer understanding of their students' vocabulary levels after the language requirement could help better articulate the transition to the major. Since vocabulary knowledge is closely correlated with other language skills (e.g., Stæhr, 2008), assessment of vocabulary knowledge can give a picture of how well students exiting the language requirement can function in their L2 overall (see Table 1 for estimates of vocabulary size and proficiency levels). The higher the receptive vocabulary knowledge, the better the reading and listening comprehension of the learners will be. That is, our findings will be helpful for guiding LPDs and instructors as they set vocabulary learning goals for students in the language requirement and beyond.

This study also contributes to the literature on vocabulary frequencies in Spanish L2 textbooks (Davies & Face, 2006; Sánchez-Gutiérrez, Marcos Miguel, & Olsen, 2018). Although there is information on the percentages of general frequency word lists that are present in Spanish textbooks published in the United States, research is lacking on learners' vocabulary size. Knowing what vocabulary students develop in classrooms where these textbooks are used could broaden future understanding of how textbook input and learners' knowledge are related (see Newton, 2021 for a call to increase this line of research in English L2).

In brief, this study seeks to establish (a) what would be a feasible percentage of the 3,000 most frequent words to be receptively learned at the end of a typical university Spanish L2 basic language program (i.e., the three semesters of the language requirement), (b) what kind of increase in vocabulary recognition can be expected from semester to semester, and (c) whether word frequency is related to vocabulary learning in Spanish L2. Individual factors affecting students' word recognition such as previous instruction before college and learners' use of Spanish beyond class were also explored to better interpret the main findings.

Vocabulary Profiles of L2 Learners: Cross-Sectional Studies

There is a robust body of research on vocabulary growth mostly with cross-sectional data from EFL learners. This line of research, which also includes other FL and L2 studies, has established a close relationship between vocabulary size and learners' proficiency (e.g., Benigno & de Jong, 2019; Hacking et al., 2019; Milton & Alexiou, 2009) as well as between vocabulary size and the number of hours of instruction learners engaged in (e.g., Canga Alonso, 2013b; Iglesias Diéguez & Martínez Adrián, 2017). In this section, we outline the correspondence between proficiency levels, hours of instruction, and knowledge of the 1,000–3,000 frequency bands.¹

Proficiency Levels for the 1,000–3,000 Frequency Bands

In the *Common European Framework of Reference for Languages (CEFR)* (Council of Europe, 2001), about 2,000 and 3,000 words are general targets for intermediate levels A2–B1 (Benigno & de Jong, 2019; Milton & Alexiou, 2009). However, Milton and Alexiou (2009) suggested that vocabulary thresholds might vary for achieving A2–B1 proficiency in different languages as they compared data from acquiring French, English, and Greek (see Table 1 for a comparison of CEFR and ACTFL proficiency levels). Milton and Alexiou suggest that the different textual coverages of the 3,000 most frequent words in each language are the reason that fewer words are needed to reach the same level in French L2 than in English L2, whereas more words are needed in Greek L2 than in English L2. The higher the textual coverage of the 3,000 words, the more useful they are.

Hours of Instruction for the 1,000–3,000 Frequency Bands

Research on hours of instruction has revealed patterns in the relationship between instructional time and vocabulary size. Thus, to some extent, vocabulary size can be predicted based on the number of hours of instruction. Previous research on English in EFL contexts shows four general patterns. First, a minimum of 400 hours seems necessary to reach a 1,000-word vocabulary size in K–12. Second, studies on L2 English vocabulary sizes from primary and secondary school learners of varied L1s showed diverse results (see Canga Alonso, 2013a for a review of such studies). Specifically, L1 effects need to be considered when

measuring vocabulary sizes. Third, L1 Spanish speakers who learn English L2 in elementary school do not reach the 1,000-word mark after 500–1,000 hours of instruction (see Canga Alonso, 2013b; Iglesias Diéguez & Martínez Adrián, 2017). Fourth, most high school students receiving between 500 and 1,000 hours of instruction can still not master the 2,000-word goal (Canga Alonso, 2013a; Staehr, 2008).

Therefore, most L2 learners in primary and secondary schools do not reach a vocabulary size of 3,000 words. This lack of knowledge will carry over to university settings. For example, Tschirner (2004) showed that, even though the knowledge of the 5,000 most frequent words was set as a goal for the last year of high school in a German state, only half of the students in their first year of English studies at the university receptively recognized the 3,000 most frequent words in Nation's (2001) Vocabulary Levels Tests. In another study with Spanish L2 university students in the United States (Hacking, Rubio, & Tschirner, 2019), vocabulary was receptively measured with the Institute for Test Research and Development (ITT) Vocabulary Test, modeled on Nation's test. In this study, only two students out of 19 showed mastery of the 2,000-word band after four semesters (considering an 80% criterion of correct answers per band when scoring the test).

This literature suggests that there is a relationship between vocabulary size, time of study, and proficiency. These studies, however, have focused mostly on L2 English and primary and secondary school students. Based on this information, [Table 1](#) summarizes the estimates of proficiency and time of study to assess what vocabulary breadth learners could develop in Spanish L2 at a university level.

Table 1. *Estimates of Proficiency, Time of Study, and Vocabulary Size in Lower-Level Language Courses and Language Majors at U.S. Universities*

Proficiency Level	Years of Study	Time of Study (hours)	Estimated Vocabulary Size ^a
A2 equivalent to ACTFL Intermediate Low (IL) for productive skills and Intermediate Mid (IM) for receptive skills (ACTFL n.d.)	Most university students in the US after two years of Spanish study (see Winke et al., 2020) ^b	500 (+)	1,000–2,000
B1 equivalent to ACTFL IM and Intermediate High (IH) for productive skills; and IH and Advance low (AL) for receptive skills (ACTFL n.d.)	Most university students in the US after four years of Spanish study (see Winke et al., 2020) ^b	1,000 (+)	3,000+

^a Based on articles reported in the introduction and [Appendices A](#) and [B](#). ^b Winke et al. (2020) estimated proficiency based on speaking, reading, and listening in standardized tests.

Vocabulary Profiles of L2 Learners: Longitudinal Studies

Most studies on vocabulary bands have followed a cross-sectional design when collecting data and focused on EFL (see Pellicer-Sánchez, 2019). Longitudinal studies, because of their logistical difficulties, are less abundant (e.g., Coxhead & Boutorwick, 2018; Dóczy & Kormos, 2016; Terrazas Gallego & Agustín Llach, 2009; Sylvén & Ohlander, 2019; Webb & Chang, 2012). Doubtless, longitudinal research is highly recommended to show change over time (Ortega & Iberri-Shea, 2005).

Longitudinal studies on EFL have shown that secondary and university students need more than five years of study to master the 3,000 most frequent words (see Coxhead & Boutorwick, 2018; Webb & Chang, 2012; Zhang & Lu, 2014). EFL learners seemed to improve more homogeneously whereas the incidental nature of study abroad contexts showed larger differences in vocabulary growth between individual learners (see Dóczy & Kormos, 2016). In general, differences in instructional practices and students' motivation affect vocabulary learning. For example, when comparing students in content-based instruction with EFL students (e.g., Agustin-Llach & Canga Alonso, 2016; Sylvén & Ohlander, 2019), students in content-based instruction showed higher scores of their vocabulary size. Similarly, in Webb and Chang (2012), students whose academic track focused on English studies and had more in-classroom time also showed larger vocabulary sizes than students who were not following a language-studies track and had fewer hours of English instruction.

Despite these differences resulting from the instructional contexts, exposure to frequent words will happen in any communicative instructional setting. For instance, most classroom materials expose learners to frequent words. Coxhead and Boutorwick (2018) showed how the content-based materials their participants used included 80% or more of the most frequent 3,000 words. Glossaries of beginner and intermediate Spanish L2 textbooks published in the United States tend to include at least 60% of these frequent words (Sánchez-Gutiérrez, Marcos Miguel, & Olsen, 2018).

English L2 learners usually exhibit differences in vocabulary knowledge by frequency band (e.g., Dóczy & Kormos, 2016; Zhang & Lu, 2014). That is, a band is not completely learned before a learner begins learning vocabulary words in other frequency bands. Indeed, although general frequencies are helpful in determining learning (i.e., higher frequency words are generally learned faster across instructional contexts such as content-based instruction, EFL or study abroad contexts), there are other factors that influence vocabulary learning such as the processing burden of the word, its memorability, its frequency in the input, the learner's characteristics, etc. (see Plonsky & Loewen, 2013; Sánchez-Gutiérrez, Marcos Miguel, & Olsen, 2018). Therefore, a student's vocabulary development of the frequency bands will be influenced by general word frequencies, but full knowledge of a band should not be expected before knowing words in the other bands (see Ozturk, 2015, 2016; Schmitt et al., 2001). Because of the variety of instructional contexts and L1s, [Appendix A](#) summarizes eight longitudinal studies and includes the setting, the tests utilized, the number of times tested, and main findings related to receptive English L2 vocabulary knowledge. There is no study to our knowledge that has offered longitudinal data on Spanish L2.

Two laws influence vocabulary learning in any instructional setting. The first, Zipf's Law, states that the order of word frequency relates to their frequency of occurrence. Specifically, "the most frequent word of a language will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, and so on²" (Dóczy & Kormos, 2016, p. 30). The second law pertains to the Power Law of Learning and states that "the rate of improvement is reduced as practice continues" (Fitts & Posner 1967, p. 18, cited in Dóczy & Kormos, 2016).

Zipf's Law explains why lower frequency words are less frequent, while the Power Law of Learning explains why increased study does not result in similar levels of improvement after a certain amount of language practice. In brief, learners will be exposed to high frequency words and learn faster at the beginning of the L2 learning process, but the pace of learning will then slow because of these two laws. Both an S-shaped curve "with a plateau followed by further development" (Dóczy & Kormos, 2016, p. 30) or a power-law curve can be used to describe vocabulary growth.

The longitudinal studies in [Appendix A](#) suggest that it is advantageous to start data collection in the initial stages of L2 acquisition to better measure learning. The lack of research on beginner learners in university settings can be explained by the prevalence of English in primary and secondary education around the world. In the United States, it is possible to study learners that start learning an L2, such as Spanish, at the university level since most universities offer courses for beginners as many students do not have previous experience with additional languages from high school.

Attrition, Retention, and Growth in Vocabulary Learning

In any model of vocabulary learning, the fluctuation between attrition and growth needs to be accounted for (see Dóczy & Kormos, 2016). Several studies combining cross-sectional and longitudinal data illustrate this point (see Appendix B for a detailed description of these studies). For example, Terrazas Gallego and Agustín Llach (2009) carried out a longitudinal study with EFL learners, speakers of L1 Spanish from 4th to 7th grade. Although learners had larger vocabulary sizes by grade, the total number of words learned each year was not constant. These findings have been replicated in other studies (e.g., Orosz, 2009; Webb & Chang, 2012). Interestingly, because of the way the data was reported in the studies in Appendix B, there is little information on the percentage of words that were forgotten from year to year. However, it can be concluded that the percentages of retention tend to be higher than the percentages of new growth or attrition in those studies. Based on the increase in vocabulary sizes, there is a relationship between growth and retention as growth is also built upon retention.

All of the studies in Appendix B analyzed English L2, except for one study focusing on Spanish L2 university students in the United States. Robles García (2020a) measured receptive learning of the 3,000 most frequent words in Spanish in a cross-sectional design. English L1 speakers could receptively recognize 1,292 words in the first year of Spanish language courses at the university level, 347 additional words in the second year, and another 375 words during the remainder of the four-year language major.

In brief, for learners in a university setting, an average learning of 400-500 words per year of the 3,000 most frequent words seems reasonable with the understanding that this growth is faster at the beginning levels (Milton 2009; Ozturk, 2016; Webb & Zhang, 2012). As stated in the introduction, the 3,000 most frequent words will offer a coverage of around 90% of most texts. Although knowing the 3,000 words will not ensure that learners can fully understand most texts since a knowledge of 95%–98% of the words is required for such a goal, this knowledge will facilitate understanding. Furthermore, because of fluctuations in vocabulary knowledge, i.e., a word can be forgotten and/or be partially learned and forgotten, maintaining word knowledge should also be a pedagogical goal. Given that learners are expected to know 2,000 words to achieve the A2/IL level in two years (see Table 1), learners would theoretically need to learn 1,000 words each year, more than what previous literature has found to be possible in English L2.

From K–12 studies we know that children and teenagers in most EFL settings do not learn the 3,000 most frequent words at the end of their schooling. From university students in EFL courses, with or without some study abroad, we know that these words can be learned after four years of study, i.e., at the end of the language major. However, we do not know whether this information from L2 English is also applicable to other L2s. A previous study on Spanish L2 (Robles García, 2020a, 2020b) that has examined learners in a U.S.-language program provided a framework for reference suggesting fewer than 2,000 words after a Spanish language major. However, this study did not explore individual learners' growth, retention, and attrition since it was cross-sectional.

The program described in this paper is illustrative of language study programs in the United States that offer two years of language courses followed by two years of content courses in the L2 as part of a major or minor. Thus, this study provides a snapshot of the vocabulary knowledge of the 3,000 most frequent Spanish words in a basic language program. The findings can help LPDs consider their vocabulary selection by semester of study as it shows growth and retention of the first three

frequency bands.

Research Questions (RQs)

This study explores receptive vocabulary knowledge of Spanish L2 learners enrolled in a three-semester sequence that satisfies a university-level language requirement. The research questions are:

1. What percentage of the 3,000 most frequent words taken from Davies (2006) are learners able to recognize
 - a. at the end of each semester of the language requirement?
 - b. by frequency band (1k, 2k, and 3k) by semester (1st, 2nd, and 3rd)?
2. What are learners' overall percentages of vocabulary growth per semester?
3. Are there additional factors influencing the learners' recognition, namely:
 - a. Does having studied Spanish before college influence their learning?
 - b. Do learners who claim to speak Spanish outside of the classroom outperform their classmates in lexical recognition?

Methods

Participants and Instructional Context

This study was carried out at a Liberal Arts College (LAC) that requires all students to complete a General Education (GE) language requirement in a language of their choice. Students with no previous experience with the language they choose must take two semesters of that language. Students who have already taken courses in high school are required to complete three semesters of the language, but they often begin their study in second or third-semester courses because of high scores on an obligatory placement test. After taking the placement test, students could be placed in any of the courses independent of previous experiences. All students with previous study in the language will finish with the third semester course.

At the time of this study, the curriculum of the basic language program was organized around two textbooks: *Mosaicos* (de Castels et al., 2015), for first and second semester, and *Sueña* (Blanco & Tocaimaza-Hatch, 2015), for the third semester. Classroom instruction focused on the three modes of communication proposed by ACTFL (National Standards, 2013): interpersonal (learners communicate between each other), interpretive (learners interact with written and oral discourse), and presentational (learners present using both written and oral discourse). Instructors followed a communicative language teaching approach (Richards, 2006).

Students met four times a week for 50 minutes for 15 weeks, for a total of 50 contact hours per semester. Learners were expected to work autonomously for at least one additional hour per hour of class contact on regularly assigned homework in the online textbook. Although it is not possible to know the exact number of hours each student spent studying outside of the classroom, after three semesters, learners should have been exposed to the language for 300 hours. [Table 2](#) depicts the estimates of time in three semesters.

The online textbook included explicit practice of the chapter vocabulary, for example through fill-in-the-blanks and matching exercises, as well as opportunities for incidental learning when the activities focused on reading, listening, and writing. During class time, students also completed vocabulary activities from the textbook and were exposed to incidental learning of the textbook vocabulary and potentially to other words used by the instructor. Explicit vocabulary teaching targeted the thematically organized word lists provided in the textbook chapters (e.g., the university, food, clothing, etc.), but students did not receive targeted instruction specific to the 3,000-word frequency bands. Learners were productively assessed on these words in chapter tests. This approach to vocabulary teaching, which does not establish clear goals based on lexical frequency lists, is not specific to this institution and has been documented in other

universities and colleges across the United States (e.g., Marcos Miguel & Cubas Mora, 2023; Sánchez-Gutiérrez, Robles-García, & Pérez Serrano, 2022).

Table 2. *Cumulative Number of Hours in Three Consecutive Semesters*

	First Semester	Second Semester	Third Semester
Hours of class contact	50	100	150
Work outside the class	50	100	150
Total	100	200	300

Table 3 summarizes the number of participants per semester. 120 students participated in this study across three semesters.

Table 3. *Participating Students by Semester*

Semester of Study	Total Participants	Same Participants
1st semester (Fall 2017)	75	
2nd semester (Spring 2018)	88	53
3rd semester (Fall 2018)	55	16 (1st–3rd) 29 (2nd–3rd)
After Language Requirement (ALR) (Spring 2019)	17	9 (1st–ALR) 8 (2nd–ALR)

Participants were asked to complete a background questionnaire. Students reported their year of study at the university, number of years studying Spanish before the university, and whether they spoke Spanish outside of the classroom (see [Appendix C](#)). **Table 4** summarizes this information. Not all participants completed the background questionnaire and/or all questions in the questionnaire.

Table 4. *Participants' Background Information*

Background Information	First Semester (n = 67)	Second Semester (n = 78)	Third Semester (n = 48)
Year at school	First-Years 34 Sophomores 26 Juniors 5 Seniors 2	First-Years 45 Sophomores 24 Juniors 2 Seniors 7	First-Years 14 Sophomores 22 Juniors 9 Seniors 3
Number of years of previous Spanish instruction (average)	1.5 (2.26)	2.59 (3.26)	4.5 (3.25)

Percentage of students who indicated they spoke Spanish in class and outside	(n = 67)	(n = 77)	(n = 48)
	18%	29%	35%

Yes/No Vocabulary Test

A Yes/No Vocabulary Test format (Meara & Buxton, 1987) was chosen to develop an instrument that would be easy to administer and complete for beginner learners. This test format, where learners see a list of words and pseudowords³ and need to decide which ones they know and which they do not, is not without limitations since it can produce an overestimation of learners' knowledge. Knowledge of the meaning of the words is not assessed since this test measures word form recognition. However, this test format has been favored in placement tests and research studies in L2 Spanish since it allows for testing a long list of words in a short time span, and high scores on the test correlate with increased proficiency (see Fairclough, 2011; Lam, 2010; Robles García, 2020b).

The test was completed on the online survey software Qualtrics with randomized question order. The words were all listed on a screen where the learners selected whether they knew the meaning of each word or not. Some students took the test during class time in a computer lab whereas others took it on their own. The first screen showed the following instructions: "For each word: if you know what it means, check 'yes.' If you aren't sure, check 'no.' Do not guess! There are fake words!"

To select the words to be included in the test, the first of every 30 words in Davies' (2006) frequency dictionary was chosen, which created a test with a total of 151 words (see [Appendix D](#)). The only word that was skipped was "eh" (2,520), as it is an oral interjection that might not have been considered a word by the L2 learners. The 2,505th word, "estrategia," was chosen instead of "eh." The next word chosen was "descripción" (2,250). Subsequently, the selection of every 30 words was resumed. Cognates were included in the test as they are recommended to be included in tests of vocabulary knowledge (see Cobb, 2000; Nation & Webb, 2011). Following this selection procedure, the majority of the words tested in the 1,000 and 2,000-bands appeared in the first and second semester textbook (97% and 84%, respectively). The 3,000-band had a balance of words appearing and not appearing in the textbook (56% appeared in the textbook, 44% did not appear). There were 50 pseudo-words taken from the list that Carreiras and Perea (2004) utilized in their study. These words resemble Spanish words in their syllable structure and phonology (see [Appendix D](#)).

Since the frequency list used (Davies, 2006) employed lemmas⁴ as the unit of analysis, this same unit was utilized for the test. This decision was also reinforced by three additional factors:

1. Research in L2 Spanish has shown that knowing a word form does not ensure knowledge of the whole word family (e.g., Marcos Miguel, 2018; Morin, 2006);
2. Most available tests for Spanish L2 preferred the lemma as the counting unit (e.g., ITT Vocabulary Test; Izura, et al., 2014; Robles García, 2020b; Rodríguez Sánchez, 2021), this consistency in the counting unit serves as a comparison across Spanish L2 studies; and
3. Research in English L2 also suggests that the lemma is a suitable unit for vocabulary tests (see Stoeckel et al., 2020; also see Webb, 2021 for another perspective). To our knowledge, there is no current test based on word families in Spanish L2.

For the recognition score, the true hit rate was utilized. The true hit rate is defined by Anderson and Freebody (1983, quoted in Pellicer-Sánchez & Schmitt, 2012) as the $h-f$, where h is the percentage of words correctly identified and f is the percentage of pseudowords incorrectly identified. Although there are different ways of calculating a recognition score, the true hit rate has the benefit of being uncomplicated and has shown to be highly correlated to more sensitive measures such as reaction times to score Yes/No vocabulary tests (Pellicer-Sánchez & Schmitt, 2012).

Procedures

Data was collected for three semesters in the last three weeks of the semester. Students who started classes in the fall semester took the test for the first time in the first semester class and then at the end of each of the following two semesters for a total of three consecutive semesters. However, new students were also added every semester based on their placement test score or because they did not take the language requirement consecutively. As a result, some students only took two semesters of language courses instead of three. All participants received extra participation points or extra credit depending on their instructor's choice.

Students who took the test three (i.e., first, second, and third semester) or two consecutive times (i.e., second and third semester) were invited to retake the test the semester after completing the language requirement (see Table 3). For these students, a \$10 gift card was offered as compensation. At that point, only two of the students were in a language class. Table 5 describes the semester-by-semester data collection procedure of the vocabulary test and background questionnaire.

Table 5. *Test Taking Procedures*

Instrument	End of First Semester	End of Second Semester	End of Third Semester	End of Fourth Semester
Vocabulary Test	In class	In class	In class	On their own ^a
Background Questionnaire	In class	In class (if not taken already)	In class (if not taken already)	No

^a Only students who had taken Spanish from 1st to 3rd semester or 2nd and 3rd semester

Statistical Analysis

Linear Mixed Models (LMMs) were run with the lme4 package in R (Bates et al., 2015). Because of the longitudinal nature of the study (i.e., many participants took the vocabulary test multiple times), LMMs were preferred (see Coxhead & Boutorwick, 2018, and Walker et al., 2019, for use of LMMs in developmental vocabulary research). LMMs are similar to traditional linear models, but they delineate variables into fixed and random effects. Fixed effects are constant across individuals, and random effects vary (see Kreft & de Leeuw, 1998). For this analysis, we assumed that, on average, students in the same course learned approximately the same amount, but they had started the course with different baseline scores. Thus, the individual student served as a random effect and the course as the fixed variable. That is, LMMs control for individual differences in participants and allow for a larger, more representative sample of students to be included for analysis than other methods such as, for example, repeated measures ANOVA.

Repeated measures ANOVA requires non-missing values for every measurement being tested. In this study, we would need a test score from each student at the end of the first, second, and third semester. Of course, students taking Spanish often are not required to take all three semesters, and some begin in the second or third semester. Thus, we needed a statistical tool that allows missing values for some semesters. Doing so is sound statistically for two reasons. First, it makes it possible to increase the sample size by allowing us to include students who did not take all three semesters of Spanish. Secondly, it avoids biasing the data as the population of students who take all three semesters of Spanish are likely different than the population of students who do not take all three semesters. Because of language requirements and prior knowledge, some students begin the language sequence in first-semester Spanish while other students begin in third-semester Spanish. Since the type of student who begins in first-semester Spanish has different characteristics than those who begin in third-semester Spanish, the missing values for the student who begins in third-semester Spanish (i.e., the missing recognition scores that would have been recorded after taking first-semester and second-semester Spanish) must be considered missing-not-at-random. Thus, when

a study includes observations that contain data that is missing-not-at-random, an LMM is most appropriate.

Results

To visualize the results of the Vocabulary Test, the true hit rate, which is expressed as a percentage, is reported in all graphs and figures in this results section. The terms *recognition score* or *scores* used in this section refer to this rate.

Percentage of Recognized 3,000 Words (by Semester and by Frequency Band)

Using the scores of the Vocabulary Test, this section answers the first RQ: *What percentage of the 3,000 most frequent words taken from Davies (2006) are learners able to recognize?* Table 6 shows the average recognition score of all the 3,000 words at the end of each semester (RQ1a). Figure 1 also shows the average recognition score of the frequency bands but this time by semester (RQ1b).

In Table 6, the recognition scores of the 3,000 words increased from the first to the third semester. Specifically, students in the third semester recognized more words than those in the first semester.

Table 6. Descriptive Statistics of the Vocabulary Test (Recognition Scores of the 3,000 Words) by Semester

Descriptive Statistic	First Semester (n = 75)	Second Semester (n = 88)	Third Semester (n = 55)	Re-take After Language Requirement (ALR) (n = 17)
Mean	39.66%	49.03%	64.69%	64.33%
Standard Deviation	12.87%	13.76%	11.20%	13.26%
Range	14.21%-69.25%	1.56%-84.12%	39.60%-86.14%	30.12%-82.06%

The average score in the first semester showed that a learner could recognize close to 40% of the 3,000 words. By the second semester, they could recognize close to 50% of the words on average, and, at the end of the required sequence of language courses, they could be close to 65%. However, there was a high range of scores from the lowest to the highest rated participant as well as a relatively high standard deviation (SD) of over 11.20% for all semesters. That is, the students were not homogeneous.

Figure 1 shows the average recognition score by semester by frequency band. The horizontal axis shows the frequency bands clustered by the three semesters whereas the vertical axis shows the average percentage participating students recognized each semester. The same pattern evident in Table 6 is replicated by semester by frequency bands. Learners tended to recognize a higher percentage of words across frequency bands as they progressed through the Spanish sequence; words of higher frequency tended to be recognized more often than words of lower frequency in all semesters.

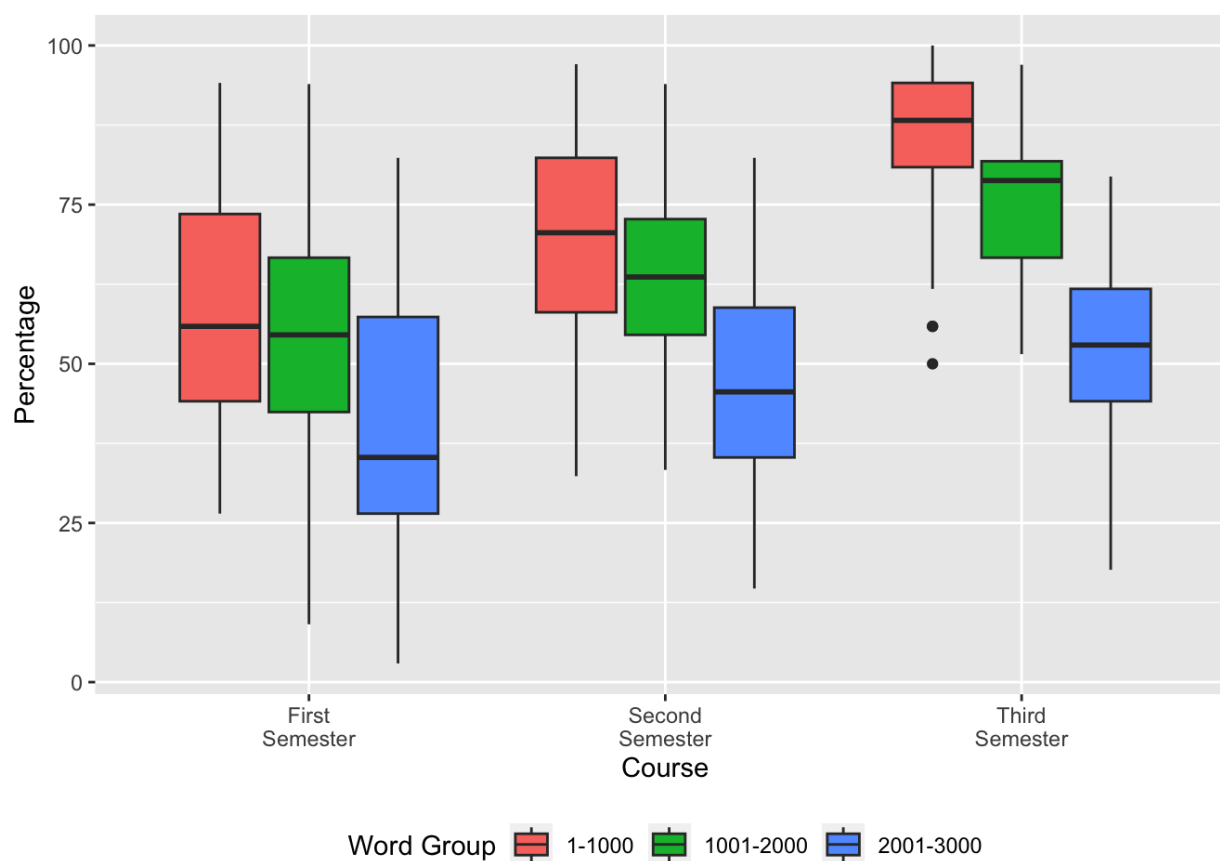


Figure 1. Recognition Scores by Semester and by Frequency.

Three LMMs were used to determine if there was a statistically significant difference in the percentage of words recognized from each frequency band (1–1,000; 1,001–2,000; and 2,001–3,000) by semester. Each model below represents one of the three pairwise comparisons (1–1,000 to 1,001–2,000; 1–1,000 to 2,001–3,000; and 1,001–2,000 to 2,001–3,000) of the percentage of words recognized in those frequency bands. Because many learners were tested multiple times, this method was utilized to add *participant* as a random effect to account for the repeated measures.

The data obtained from the Vocabulary Test had 235 observations of three variables: an anonymized number assigned to each student; the percentage of words recognized; and the semester in which the test was taken (see [Tables 3](#) and [5](#)). A factor analysis made it clear that the data was bi-dimensional: one dimension representing actual words and the other representing pseudo-words. Thus, we ran Cronbach's alpha test on both the real words ($p = .902$) and pseudo-words ($p = .946$), which indicated internal consistency.

To compare recognition scores by frequency bands, a Bonferroni correction was made to the p value in the LMMs. Because the probability of a Type I error rises dramatically as the number of comparisons increases, the p value correction was utilized to account for this fact. Even after this correction, there was a statistically significant difference between each pair of frequency bands, namely the 1–1,000 and 1,001–2,000 frequency bands ($p < 0.001$), the 1,001–2,000 and 2,001–3,000 frequency bands ($p < 0.001$), and the 1–1,000 and 2,001–3,000 frequency bands ($p < 0.001$). This implies that every semester, students had higher recognition of words that were more frequent. The details of the LMMs are found in [Tables 7](#), [8](#), and [9](#) below. Note the marginal R-square values (0.754, 0.564, and 0.643). These values represent the correlations between frequency bands not including the correlations inherent in data with repeated measures. These high

correlations indicate that a students' ability to acquire higher frequency vocabulary can also result in learning less frequent words.

Table 7. LMM Analyzing Recognition Scores of 1-1,000 and 1,001-2,000 Frequency Bands

<i>Predictors</i>	<i>Estimates</i>	Recognition Score	
		<i>CI</i>	<i>p</i>
Intercept	10.75	6.60 – 14.90	<0.001
1-1,000 Recognition Score	0.75	.070 – .81	<0.001
Random Effects			
σ^2	55.90		
τ_{00} participant	9.90		
ICC	0.15		
N _{participant}	132		
Observations	235		
Marginal R ² / Conditional R ²	0.754/ 0.791		

Table 8. LMM Analyzing Recognition Scores of 1-1,000 and 2,001-3,000 Frequency Bands

<i>Predictors</i>	<i>Estimates</i>	Recognition Score	
		<i>CI</i>	<i>p</i>
Intercept	-1.90	-7.80 – 4.01	0.529
1-1,000 Recognition Score	0.75	.62 – .78	<0.001
Random Effects			
σ^2	85.74		
τ_{00} participant	46.82		
ICC	0.35		
N _{participant}	132		
Observations	235		
Marginal R ² / Conditional R ²	0.564/ 0.718		

Table 9. *LMM Analyzing Recognition Scores of 1,001-2,000 and 2,001-3,000 Frequency Bands*

<i>Predictors</i>	Recognition Score		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Intercept	-7.01	-12.47 – 155.	0.012
1-1,000 Recognition Score	0.85	.77 – .93	< 0.001
Random Effects			
σ^2	81.61		
τ_{00} participant	24.91		
ICC	0.23		
$N_{\text{participant}}$	132		
Observations	235		
Marginal R^2 / Conditional R^2	0.643/ 0.726		

Students' Vocabulary Recognition Growth by Semester

Although the average score of recognized words for the 1–3,000 range rises every semester (see [Table 6](#) and [Figure 1](#)), there appear to be diminishing returns as students enter more advanced courses in the sequence. This section analyzes students' percentages of vocabulary growth in all 3,000 words per semester (see [Table 6](#) for the students' averages) and whether the growth was statistically significant when comparing semesters. That is, RQ2, *What are students' overall percentages of vocabulary growth per semester?*, is examined.

First, individual students' scores in their second and third semesters are compared to their previous attempt. In [Figure 2](#), data points falling above the line represent students who did better compared to their previous attempt, while students falling at or below the line did the same as or worse than the previous semester. When the data point for student performance is above the line, it shows growth, whereas when the data point for student performance is at the line, it shows retention. If the data point for student performance falls below the line, then it shows attrition. In [Figure 2](#), most points are above the line or very close to it showing growth and retention.



Figure 2. Scatterplot Comparing Students' Recognition Scores in Consecutive Semesters (1st and 2nd; 2nd and 3rd-semester).

Second, an LMM was performed to quantify the apparent differences in [Figure 2](#). After controlling for the participant, on average, a student would be expected to increase their recognition score by 11.15% ($p < 0.001$) after each subsequent semester. Therefore, we will expect with 95% confidence, that a student will have gained 277 to 393 words per semester of Spanish taken. So, if a student knew 1,000 words in the first semester, they would know 1,277-1,393 at the end of the second semester.

In this analysis, the random effect value of the individual student indicates the necessity of a mixed model (i.e., the scores are influenced by the students taking them). For the fixed effects, the small p -value for the variable indicates a statistically significant relationship between the semester and the recognition score. The results of this LMM are presented in [Table 10](#).

Table 10. LMM Analyzing Recognition Scores Gains per Semester

<i>Predictors</i>	Recognition Score		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Intercept	28.84	24.64 – 33.05	<0.001
Course Number	11.15	9.23 – 13.08	<0.001
Random Effects			
σ^2	90.66		

τ_{00} participant	75.29
ICC	0.45
$N_{\text{participant}}$	132
Observations	218
Marginal R^2 / Conditional R^2	0.307 / 0.621

In Figure 3, scores of students who were re-tested after their third-semester (ALR) course are compared to their third-semester score. As in Figure 2, data points falling above the line represent students who scored higher when re-tested and showed growth. Learners whose points fall at or below the line scored lower when re-tested.



Figure 3. Scatterplot Comparing Students’ Recognition Scores after the GE Sequence (3rd and 4th semester).

A paired t-test did not reveal any statistically significant difference between L2 learners’ scores after their third semester of Spanish and the subsequent retake a semester later ($M = 3.74$, $SD = 10.32$, $t(16) = 1.49$, $p > .077$).

In summary, for these Spanish L2 learners, there was a statistically significant growth in word recognition of about 11.15% for every additional Spanish class taken, even after accounting for individual differences in these first three semesters ($p < 0.001$, 95% margin of error ± 1.93). Learners’ scores persisted even one semester after completing the last course in the Spanish sequence.

Individual Factors Influencing Recognition

This section provides some nuance on the previous data by examining whether two external factors influenced students' recognition (RQ3), namely years of instruction before the university (RQ3a) and use of Spanish outside of class (RQ3b).

Effects of studying Spanish before the university basic language program were examined with an LMM (see Table 11). From this model, for every additional year of high school study, students performed 1.84% points better (95% CI 0.89-2.79, $p < 0.001$) on the recognition test. Thus, a student who took four years of high school Spanish could expect a score that is 7.36% points higher on the recognition test than a student who did not study Spanish in high school.

Figure 4 shows the scores organized by number of years of instruction before college. The blue line serves as an indicator of how an average student for each number of years would do every subsequent semester. For example, the average student with one year of high school Spanish recognized 40% of words on the test. This percentage corresponds to an average of around 1,200 words in the 1-3,000 frequency bands recognized by the student. The average student who has taken five years of Spanish, on the other hand would be able to recognize approximately 1,800 words in the 1-3,000 frequency bands on average.

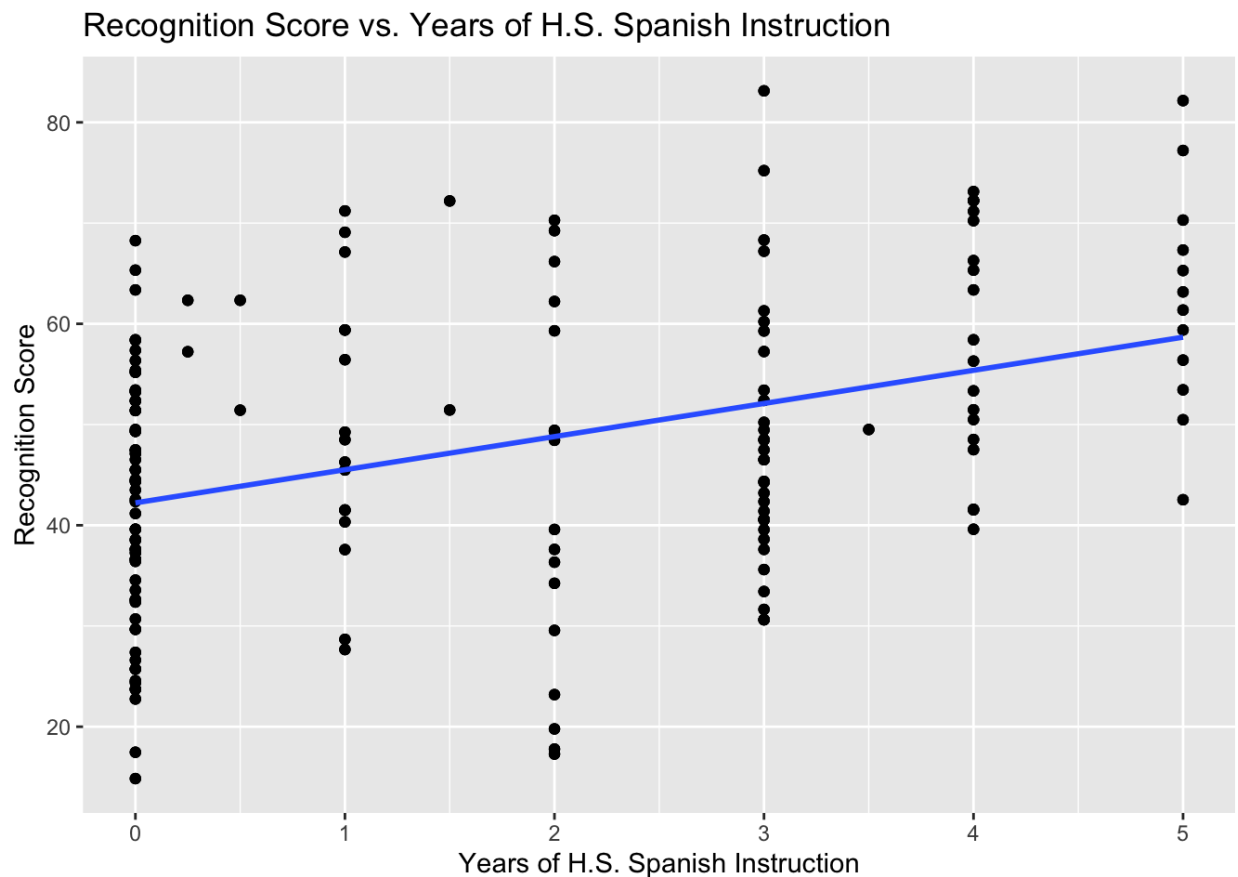


Figure 4. Recognition Scores versus Years of High School Spanish Instruction.

Table 11. *LMM Analyzing Effects of Years of Spanish in Recognition Scores*

<i>Predictors</i>	Recognition Score		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Intercept	45.41	42.49 – 48.33	<0.001
Years of Study in HS	1.84	.89 – 2.79	<0.001
Random Effects			
σ^2	61.32		
τ_{00} participant	134.80		
ICC	0.69		
$N_{\text{participant}}$	125		
Observations	576		
Marginal R^2 / Conditional R^2	0.048/ 0.702		

Next, an LMM examined whether speaking Spanish outside of class increased scores (see Table 12). From this model, on average, a student who spoke outside of the classroom would receive a recognition score that is 4.05% points higher (95% CI 1.21 – 6.89) than a student who did not.

Table 12. *LMM Analyzing Effects of Speaking Outside of Spanish Classes in Recognition Scores*

<i>Predictors</i>	Recognition Score		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Intercept	50.18	47.53 – 52.84	< 0.001
Outside of Class	4.05	1.21 – 6.89	0.005
Random Effects			
σ^2	53.56		
τ_{00} participant	194.93		
ICC	0.78		
$N_{\text{participant}}$	125		
Observations	576		
Marginal R^2 / Conditional R^2	0.643/ 0.726		

Discussion

After analyzing the data of students taking three semesters of Spanish, it is possible to answer RQ1a, *What percentage of the 3,000 most frequent words taken from Davies (2006) are learners able to recognize at the end of each semester of the language requirement?* The average score in the first semester showed that a learner could recognize 40% of the 3,000 words (i.e., around 1,200 words). By the second semester, they

could recognize close to 50% of the words (i.e., around 1,500 words) and finally they would be close to 65% (i.e., around 1,950 words) at the end of a three-semester sequence (see Table 6). When the recognition scores were calculated by frequency band and semester (RQ1b), the patterns were similar across semesters: the most frequent words were better known (see Figure 1). These scores show mostly growth and retention semester to semester.

This study also answered RQ2: *What are students' overall percentage of vocabulary growth per semester?* An average learner increased their recognition score by 11.15% ($p < 0.001$) after each subsequent course that they completed. This suggests that on average students acquired between 277 and 393 words in the 1–3,000 frequency bands for each additional course. These results seem consistent with studies on English L2 of adults (see Appendix A & B). A target of 400 words per year, as Webb and Chang (2012) recommended for L2 English, should be attainable by most learners.

When exploring the additional factors influencing recognition scores (RQ3), students with exposure to Spanish before the university (RQ3a) showed higher recognition scores. This finding is not surprising as more hours of language study correlate with larger vocabulary sizes. Moreover, students who spoke Spanish outside of class (RQ3b) also seemed to slightly outperform their classmates. Consistently, more hours of exposure facilitate acquisition. This extra exposure could also be conflated with motivation as students who speak outside of the classroom might be more motivated than those who do not.

The findings of the present study allow for comparisons with similar studies in English, Spanish, and other L2s. They are consistent with data reported, for example in Milton (2010), Milton and Alexiou (2009), Benigno and de Jong (2019) where learners at the A2 level in several L2s did not reach knowledge of the 3,000 most frequent words. The A2 level is roughly equivalent to the level of the L2 learners participating in this study (see Table 1). Moreover, higher frequency seemed to drive participants' vocabulary learning, which is also consistent across longitudinal studies (e.g., Coxhead & Boutorwick, 2018; Webb & Chang, 2012; Zhang & Lu, 2014).

When compared with K–12 learners of English L2 (see Canga Alonso, 2013a, 2013b; Iglesias Diéguez & Martínez Adrián, 2017), the learners in the present study showed higher receptive knowledge. Age differences can explain these disparities (see Center for Applied Language Studies, CASLS, 2011). The use of lemma instead of word family could also be a factor since English L2 studies mostly used word families whereas this study used lemmas as the unit of vocabulary knowledge. Indeed, knowledge of word families requires a higher number of lemmas being recognized by the learner since each word family contains various lemmas (e.g., the word family of *beber* will include at least the three lemmas *bebida* (noun, drink), *bebible* (adjective, drinkable), and *beber* (verb, to drink).

When comparing the current study findings with Robles García (2020a), both studies show consistent trends in their results. Nevertheless, the vocabulary gains in this present study are slightly higher than those displayed in Robles García. In this previous study, a similar population of Spanish L2 learners gained 1,292 words in the first year of Spanish language courses at the university level and 347 in the second year. Although the tests used in both studies followed a Yes/No format, Robles García's 3K–LEX does not include cognate words, which may have reduced the overall knowledge gains revealed in the test. As Milton and Alexiou (2009) suggest, effects of learners' L1 need to be taken into consideration when measuring vocabulary sizes. For example, Ozturk (2015) showed how scores were high in bands where cognate words (Turkish–English) were tested. For English speakers learning L2 Spanish, the multiple cognates in the first 3,000 frequency bands might benefit them. Another difference is that in this study the frequency list was taken from Davies (2006), whereas Robles García (2020b) used Davies and Davies (2017). Further studies should explore whether the differences in both tests are due to the tests, i.e., the cognates, the source of the frequency list, the scoring, etc., or other factors in the samples examined, such as the textbooks used or instructors' input.

When compared with vocabulary presented in beginner and intermediate Spanish L2 textbooks (Sánchez-Gutiérrez, Marcos Miguel, & Olsen, 2018), the students' results are similar: most textbooks include 60%

of the 3,000 most frequent words (i.e., around 1,800 words), and words belonging to the first band are also the most frequent. Although there is no current estimate of Spanish L2 words that teachers and students orally produce in their classrooms, estimates from English L2 also suggest that the 1,000 most frequent words are present in most classrooms' input (Horst, 2009; Tang & Nesi, 2003). Therefore, even with incidental exposure, learners seem to acquire a good percentage of these frequent words, revealing that textbook content tends to correspond with learners' vocabulary knowledge.

Overall, these results are encouraging since frequent words are learned even without targeted teaching. Although learners were exposed to vocabulary instruction, it did not specifically target the 3,000 words, but rather the selection of words presented in the textbook (which only partially followed a frequency criterion). Given that textbook vocabulary content has an impact on learners' knowledge, it could be helpful if textbook authors and publishers consider these frequent words when planning their vocabulary selection and goals.

This study focused on a specific school where only volunteering students participated. As an example, out of a total of 94 students enrolled in the first semester course, 75 participated in this study, which is almost 80% of the possible participants. Because of the focus on a specific school, the findings can be compared to other instructional settings, but are not necessarily generalizable as other schools could have a different instructional approach and their student body might have different characteristics.

The question remains whether more targeted instruction could enhance students' scores. Moreover, because of the data collected, it is not possible to know whether certain words were targeted by the instructor, but we hypothesize that the students' results should be similar across language programs that focus on textbook vocabulary rather than targeting high frequency words. Replicating this study in other instructional settings with different L2s will contribute to establishing generalizability of its findings. Conducting such replications can contribute to vocabulary research as well as to curriculum development in language programs.

In addition to replications, further studies should explore what these results mean in terms of what the learners can understand when reading in Spanish and other L2s (see Schmitt et al., 2020). Future studies could include a reading comprehension test that all learners take as an external and shared proxy of proficiency. Moreover, studies on classroom lexical input and selection of vocabulary by Spanish L2 instructors would be helpful to assess the degree of consistency between the general list of word frequencies used (Davies, 2006) and classroom realities.

Furthermore, based on these findings, some recommendations can be given to LPDs and language instructors. First, although the data analyzed show vocabulary growth and retention, additional intentional activities can be added to target the 3,000 most frequent words. Including a vocabulary test at the beginning of the semester can also help instructors decide what vocabulary to favor for a specific class and for individual students. Intentional activities that include opportunities for repeated exposure and retrieval practice (see Webb, Yanagisawa, & Uchihara, 2020) should be used to complement textbook activities. For example, using flashcards designed by the instructor has been recommended as an activity to learn the 3,000 most frequent words (see Lei & Reynolds, 2022).

Second, given the observed differences in students' vocabulary knowledge, it can be helpful to include individually tailored activities. Vocabulary notebooks (Dubiner, 2017; Walters & Bozkurt, 2009) and flashcards can be specially recommended towards that goal.

Finally, because of the importance of both intentional and incidental vocabulary learning, including time for additional reading materials inside and outside of class can be helpful (see Dykstra, et al., 2023; Park, Isaacs & Woodfield, 2018). Graded readers that have considered frequency counts in their design and that include vocabulary activities can be a good resource for complementing the input provided by the textbook (see Alins Breda, 2021).

Conclusion

This study has shown a baseline for the number of words to be learned in a semester (i.e., learners can gain between 277 and 393 words per semester of the 3,000 most frequent words), and learners' vocabulary growth (i.e., continuous learning throughout the three semesters with diminishing gains with proficiency, and with effects due to the frequency bands of the words). Results in Spanish L2 are consistent with results in English L2. We hope this study will encourage LPDs and instructors to examine the vocabulary profiles of their students and to devote sufficient time to vocabulary instruction and practice in their classroom that target the 3,000 most frequent words. Based on previous literature, receptive knowledge of these frequent words will contribute to improving the learners' general comprehension of most spoken and written texts.

Acknowledgements

We thank the instructors and the students for their participation on their study. Without their contribution, this study would not have been possible. We also would like to thank the editors and peer-reviewers for their help. All remaining errors are our own.

Notes

1. When measuring knowledge of frequent words, researchers usually analyze knowledge by band. For example, when research is about the 3,000 most frequent words, the words will be analyzed as 1–1,000, 1,001–2,000, and 2,001–3,000. That is, there are three bands that are examined.
2. In Davies' (2006) frequency dictionary, the three most frequent words listed are: *el/la* (2,037,803, raw frequency), *de* (1,319,834) and *que* (662,653).
3. The inclusion of pseudowords in these tests responds to the need to assess whether learners are completing the test faithfully and with attention.
4. Lemmas represent all inflected forms such as the form *beber* (to drink) representing all its inflected verbal forms. For example, *bebo* (I drink), *bebí* (I drank), *beberé* (I will drink), etc. would all be classified as instances of *beber* (to drink).

References

- ACTFL. (2012). *ACTFL proficiency guidelines* [Electronic version]. <https://www.actfl.org/sites/default/files/guidelines/ACTFLProficiencyGuidelines2012.pdf>
- ACTFL. (n.d.). *Assigning CEFR ratings to ACTFL assessments*. [Electronic version]. https://www.actfl.org/sites/default/files/reports/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf
- Admiraal, W., Westhoff, G., & De Bot, K. (2006). Evaluation of bilingual secondary education in the Netherlands: Students' language proficiency in English. *Educational research and Evaluation, 12*(1), 75-93. <https://doi.org/10.1080/13803610500392160>
- Agustín-Llach, M. P., & Canga Alonso, A. (2016). Vocabulary growth in young CLIL and traditional EFL learners: Evidence from research and implications for education. *International Journal of Applied Linguistics, 26*(2), 211–227. <https://doi.org/10.1111/ijal.12090>
- Alins Breda, D. (2021). ¿Las actividades en las lecturas graduadas promueven el aprendizaje de vocabulario? Un estudio de Technique Feature Analysis. [Do activities in graded readers promote vocabulary learning? A Technique Feature Analysis study]. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas, 15*(30), 103–117. <https://doi.org/10.26378/rmlael1530430>

- Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B.A. Hutson (Ed.), *Advances in reading/language research* (pp. 132–255). JAI Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Blanco, J., & Tocaimaza-Hatch, C. C. (2015). *Sueña. Español sin barreras. Curso intermedio breve* (3rd Edition). Vista Higher Learning.
- Benigno, V., & de Jong, J. (2019). Linking vocabulary to the CEFR and the global scale of English: A psychometric model. In A. Hutha, G. Erickson & N. Figueras (Eds.) *Developments in language education. A memorial volume in honour of Sauli Takala* (pp. 8–29). University Printing House.
- Canga Alonso, A. (2013a). Receptive vocabulary size of secondary Spanish EFL learners. *Revista de Lingüística y Lenguas Aplicadas*, 8(1), 66–75. <https://doi.org/10.4995/rlyla.2013.1180>
- Canga Alonso, A. (2013b). The receptive vocabulary of Spanish 6th-grade primary-school students in CLIL instruction: A preliminary study. *Latin American Journal of Content and Language Integrated Learning*, 6(2), 22–41. <https://laclil.unisabana.edu.co/index.php/LACLIL/article/view/3138>
- Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language*, 90(1–3), 393–400. <https://doi.org/10.1016/j.bandl.2003.12.003>
- Center for Applied Second Language Studies (CASLS) (2011). Report: *How Do Proficiency Levels Compare Between K-12 and University Students?* CASLS, University of Oregon. <https://casls.uoregon.edu/wp-content/uploads/pdfs/tenquestions/TBQLevelofLanguage.pdf>
- Cobb, T. (2000). One size fits all? Francophone learners and English vocabulary tests. *Canadian Modern Language Review*, 57(2), 295–324. <https://doi.org/10.3138/cmlr.57.2.295>
- Coxhead, A., & Boutorwick, T.J. (2018). Longitudinal vocabulary development in an EMI international school context: Learners and texts in EAL, maths, and science. *TESOL Quarterly*, 52(3), 588–610. <https://doi.org/10.1002/tesq.450>
- Council of Europe. (2001). *Common European Framework of Reference for Languages (CEFR)*. Cambridge University Press.
- Davies, M. (2005). Vocabulary range and text coverage: Insights from the forthcoming Routledge frequency dictionary of Spanish. In D. Eddington (ed.), *Selected proceedings of the 7th Hispanic Linguistics Symposium*, (pp.106–115). Cascadilla Press. <http://www.lingref.com/cpp/hls/7/paper1091.pdf>
- Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. Routledge.
- Davies, M., & Davies, K.H. (2017). *A frequency dictionary of Spanish: Core vocabulary for learners, Second Edition*. Routledge.
- Davies, M., & Face, T. L. (2006). Vocabulary coverage in Spanish textbooks: How representative is it? In N. Sagarra & A.J. Toribio (Eds.), *Selected proceedings of the 9th Hispanic Linguistics Symposium* (pp. 132–143). Cascadilla.
- De Castells, M.O., Guzmán, E. E., Lapuerta, P., & Liskin-Gasparro, J. E. (2015). *Mosaicos: Spanish as a world language* (6th ed.). Pearson.
- Dóczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. Oxford University Press.

- Dubiner, D. (2017). Using vocabulary notebooks for vocabulary acquisition and teaching. *ELT Journal*, 71(4), 456–466. <https://doi.org/10.1093/elt/ccx008>
- Dykstra, S., Sánchez-Gutiérrez, C.H., Marcos Miguel, N., & Alins Breda, D. (2023). Reading and affect: University Spanish learners' perceptions of a reading program. *Foreign Language Annals*, 56(1), 191–213. <https://doi.org/10.1111/flan.12662>
- Fairclough, M. (2011). Testing the lexical recognition task with Spanish/English bilinguals in the United States. *Language Testing*, 28(2), 273–297. <https://doi.org/10.1177/0265532210393151>
- Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Brooks/Cole.
- Hacking, J., Rubio, F., & Tschirner, E. (2019). Vocabulary size, reading proficiency and curricular design: The case of college Chinese, Russian and Spanish. In P. Winke & S. M. Gass (Eds.), *Foreign language proficiency in higher education*, 25–44. Springer.
- Horst, M. (2009). Revisiting classrooms as lexical environments. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners* (pp. 53–66). Multilingual Matters.
- Iglesias Diéguez, K., & Martínez-Adrián, M. (2017). The influence of CLIL on receptive vocabulary: A preliminary study. *Journal of English Studies*, 15, 107–134. <http://hdl.handle.net/10810/21430>
- Institute for Test Research and Development. (n.d.) *Vocabulary Test Spanish Receptive*. http://www.it-leipzig.de/static/vltspanish_01r/index.html
- Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, 35, 49–66. <https://www.uv.es/revispsi/articulos1.14/3IZURA.pdf>
- Kreft, I. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Sage.
- Lam, Y. (2010). Yes/No Tests for foreign language placement at the post-secondary level. *Canadian Journal of Applied Linguistics/Revue Canadienne de Linguistique Appliquée*, 13(2), 54–72. <https://eric.ed.gov/?id=EJ944127>
- Lei, Y., & Reynolds, B. L. (2022). Learning English vocabulary from word cards: A research synthesis. *Frontiers in Psychology*, 5335. <https://doi.org/10.3389/fpsyg.2022.984211>
- Lusin, N. (2012). *The MLA survey of postsecondary entrance and degree requirements for Languages Other Than English, 2009–10*. https://www.mla.org/content/download/3316/81618/requirements_survey_200910.pdf
- Marcos Miguel, N. & Cubas-Mora, M. (2023). Interpreting the designated curriculum: Teachers' understanding of vocabulary instruction and adherence to the textbook. In L. Marques-Pascual & I. Checa-García (Eds.), *Current perspectives in Spanish lexical development* (p. 269–301). De Gruyter, Mouton. <https://doi.org/10.1515/9783110730418-011>
- Marcos Miguel, N. (2018). Analyzing the relationship and development of proficiency, derivational knowledge, and vocabulary size in Spanish L2 learners. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 31(1), 224–256. <https://doi.org/10.1075/resla.16027.mar>
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142–154. <https://doi.org/10.1177/026553228700400202>
- Meara, P., & Milton, J. (2003). *X_Lex, the Swansea Levels Test*. Express.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Bartning, M. Martin, & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections*

- between SLA and language testing research* (pp. 211–232). European Second Language Association.
- Milton, J., & Alexiou, T. (2009). Vocabulary size and the common European framework of reference for languages. In B. Richards, H. M. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 194–211). Palgrave Macmillan.
- Morin, R. (2006). Building depth of Spanish L2 vocabulary by building and using word families. *Hispania*, 89(1), 170–182. <https://doi.org/10.2307/20063269>
- Nation, I.S.P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13.
- Nation, I.S.P., & Webb, P. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.
- National Standards in Foreign Language Education Project. (2015). *World-readiness standards for foreign language learning*. ACTFL.
- Newton, J. (2021). Teacher and learner perspectives on vocabulary learning and teaching (VLT). In H. Mohebbi, & C. Coombe (Eds.), *Research Questions in Language Education and Applied Linguistics* (pp. 137–141). Springer.
- Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 26–45. <https://doi.org/10.1017/S0267190505000024>
- Orosz, A. (2009). The growth of young learners' English vocabulary size. In M. Nikolov (Ed.), *Early Learning of Modern Foreign Languages: Processes and Outcomes* (pp. 181–194). Multilingual Matters.
- Ozturk, M. (2015). Vocabulary growth of the advanced EFL learner. *The Language Learning Journal*, 43(1), 94–109. <https://doi.org/10.1080/09571736.2012.708053>
- Ozturk, M. (2016). Second language vocabulary growth at advanced level. *The Language Learning Journal*, 44(1), 6–16. <https://doi.org/10.1080/09571736.2012.708054>
- Park, A. Y., Isaacs, T., & Woodfield, H. (2018). A comparison of the effects of extensive and intensive reading approaches on the vocabulary development of Korean secondary EFL learners. *Applied Linguistics Review*, 9(1), 113–134. <https://doi.org/10.1515/applirev-2017-0025>
- Pellicer-Sánchez, A. (2019). Examining second language vocabulary growth: Replications of Schmitt (1998) and Webb & Chang (2012). *Language Teaching*, 52(4), 512–523. <https://doi.org/10.1017/S026144481800037X>
- Pellicer-Sánchez, A., & Schmitt, N. (2012). Scoring Yes–No vocabulary tests: Reaction time vs. nonword approaches. *Language Testing*, 29(4), 489–509. <https://doi.org/10.1177/0265532212438053>
- Richards, J. C. (2006). *Communicative Language Teaching today*. Cambridge University Press.
- Robles García, P. (2020a). *¿Cuántas palabras conocen los aprendientes de español como lengua extranjera? Competencia léxica y oportunidades de aprendizaje léxico en el aula de ELE*. [How many words do L2 Spanish learners know? Lexical breadth and vocabulary learning opportunities in the L2 classroom]. [Unpublished doctoral dissertation]. University of California, Davis.
- Robles-García, P. (2020b). 3K-LEx. Desarrollo y validación de una prueba de amplitud léxica en español. [3K-LEX: Development and Validation of a Vocabulary Levels Test in Spanish]. *Journal of Spanish Language Teaching*, 7(1), 64–76. <https://doi.org/10.1080/23247797.2020.1770465>
- Rodríguez Sánchez, I. (2022). Testing L2 Spanish vocabulary knowledge. In J. Barcroft & J. Muñoz-

- Basols (Eds.) *Spanish vocabulary learning in meaning-oriented instruction*, 130–149. Routledge.
- Sánchez-Gutiérrez, C. H., Marcos Miguel, N., & Olsen, M. (2018). An analysis of vocabulary coverage and lexical characteristics in L2 Spanish textbooks. In P. Ecke & S. Rott (Eds.) *Understanding vocabulary learning and teaching: Implications for language program development. American Association of University Supervisors, Coordinators, and Directors of Language Programs (AAUSC) Volume 2018* (pp. 78-98). Cengage. <http://hdl.handle.net/102015/69783>
- Sánchez-Gutiérrez, C. H., Robles-García, P., & Pérez Serrano, M. (2022, *published online*). L2 Spanish vocabulary teaching in US universities: Instructors' beliefs and reported practices. *Language Teaching Research*, <https://doi.org/10.1177/13621688221074443>
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. <https://doi.org/10.1111/lang.12077>
- Schmitt, N., Nation, I. S. P. & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, 53(1), 109–120. <https://doi.org/10.1017/S0261444819000326>
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1177/026553220101800103>
- Stæhr, L.S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36(2), 139–152. <https://doi.org/10.1080/09571730802389975>
- Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181–203. <https://doi.org/10.1017/S027226312000025X>
- Sylvén, L. K., & Ohlander, S. (2019). English receptive vocabulary. In L.K. Sylvén (Ed.), *Investigating content and language integrated learning: Insights from Swedish high schools* (pp. 101–116). Multilingual Matters.
- Tang, E., & Nesi, H. (2003). Teaching vocabulary in two Chinese classrooms: Schoolchildren's exposure to English words in Hong Kong and Guangzhou. *Language Teaching Research*, 7(1), 65–97. <https://doi.org/10.1191/1362168803lr113oa>
- Terrazas Gallego, M., & Agustín Llach, M. P. (2009). Exploring the increase of receptive vocabulary knowledge in the foreign language: A longitudinal study. *International Journal of English Studies*, 9(1), 113–133. <https://revistas.um.es/ijes/article/view/90681>
- Tschirner, E. (2004). Breadth of vocabulary and advanced English study: An empirical investigation. *Electronic Journal of Foreign Language Teaching*, 1(1), 27–39. <https://e-flt.nus.edu.sg/v1n12004/tschirner.pdf>
- Vilkaitė-Lozdienė, L., & Schmitt, N. (2019). Frequency as a guide for vocabulary usefulness: High-, mid- and low-frequency words. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 81–96). Routledge.
- Walker, E. A., Redfern, A., & Oleson, J.J. (2019). Linear Mixed-Model Analysis to examine longitudinal trajectories in vocabulary depth and breadth in children who are hard of hearing. *Journal of Speech, Language, and Hearing Research*, 62(3), 525–542. https://doi.org/10.1044/2018_JSLHR-L-ASTM-18-0250

- Walters, J., & Bozkurt, N. (2009). The effect of keeping vocabulary notebooks on vocabulary acquisition. *Language Teaching Research*, 13(4), 403–423. <https://doi.org/10.1016/j.sbspro.2016.10.046>
- Webb, S. (2021). A different perspective on the limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(2), 454–461. <https://doi.org/10.1017/S0272263121000449>
- Webb, S., & Chang, A.C. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113–126. <https://doi.org/10.1177/0033688212439367>
- Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407–427. <https://doi.org/10.1093/applin/amp010>
- Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *The Modern Language Journal*, 104(4), 715–738. <https://doi.org/10.1111/modl.12671>
- Winke, P., Zhang, X., Rubio, F., Gass, S., Sonesson, D., & Hacking, J. (2020). The proficiency profile of language students: Implications for programs. *Second Language Research and Practice*, 1(1), 25–64. <http://hdl.handle.net/10125/69840>
- Zhang, X., & Lu, X. (2014). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics*, 35(3), 283–304. <https://doi.org/10.1093/applin/amt014>
- Zhang, S., & Zhang, X. (2020). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, 1–30. <https://doi.org/10.1177/1362168820913998>

Appendix A. Longitudinal Vocabulary Studies on Receptive Vocabulary Sizes (1,000–3,000 Word Frequencies)

Context	Test and Language	Times Tested	Receptive Vocabulary Knowledge
Content and Language Integrated Instruction (CLIL) versus non-CLIL in secondary education in the Netherlands (Admiraal, et al., 2006) (n = 584 and 721 students in the CLIL and control group)	EFL vocabulary Test (Meara, 1992)	Twice a year during a period of five years	Increase
Immersion school setting in Germany (Coxhead & Boutorwick, 2018) (n = 468)	VLT (English L1 and L2); test of mastery (See Schmitt et al., 2001)	Tested in 6 to 11 grades	All learners showed mastery by grade 10 (2,000 & 3,000)

EFL Taiwan, high school and university (Webb & Chang, 2012) (n = 222)	VLT (English L2); test of mastery	Tested during three years of high school and two at the university	Most motivated students at the end of study: 1,000 level: 73% 2,000 level: 33% 3,000 level: 5%
Intensive EFL in a secondary school in Hungary, estimated A1 level (Dóczy & Kormos, 2016) (n = 67)	A selection of 25 items from the Vocabulary Size Test (English L2) (Nation & Beglar, 2007).	Tested twice one year	Pre-test: 11% Post-test: 15% (2,000, 3,000, 5,000 and 10,000 frequency levels, together)
Post-secondary learners studying in the UK, estimated B1 level (Dóczy & Kormos, 2016) (n=116)	A selection of 25 items from the Vocabulary Size Test (English L2)	Tested twice one year	Pre-test: 15% Post-test: 17% (2,000, 3,000, 5,000 and 10,000 frequency levels, together)
CLIL versus non-CLIL in secondary education in Sweden (Sylvén & Ohlander, 2019) (Tested both times: CLIL: n = 57; non-CLIL: n = 52)	VLT (English L2)	Tested in first year of upper-secondary education (10 grade) and last year (12 grade)	All learners showed mastery by grade 10 of 2,000 level CLIL students showed mastery by grade 10 of 3,000 level Non-CLIL students still below the 80% mastery by grade 12
Post-secondary, EFL Turkey (Ozturk, 2015) (n = 17)	VLT (English L2); test of mastery	Tested in first and fourth year	Developmental plateau All learners showed mastery in first year (2,000 & 3,000)
Post-secondary, EFL China (at least six years of English before university) (Zhang & Lu, 2014) (n = 298)	VLT (English L2)	Tested three times over 22 months	2000: mastery on first test 3000: mastery on second test

Appendix B. Longitudinal and Cross-Sectional Studies on Vocabulary Growth Within a Language Program

Context	Test and Language	Times Tested	Growth (pattern)	Growth (words)
---------	-------------------	--------------	------------------	----------------

<p>Content and Language Integrated Instruction (CLIL) versus non-CLIL in primary education in Spain (Agustín-Llach & Canga Alonso, 2016)</p>	<p>2K VLT (Schmitt, Schmitt & Clapham, 2001) (English L2)</p>	<p>Tested from 4 to 6 grade (longitudinal)</p>	<p>General increase, no differences between groups in number of words increased per year</p>	<p>+ Around 200 words increase per year</p>
<p>(CLIL: n = 58; non-CLIL: n = 49)</p>			<p>Higher amount of total learning in the CLIL group (more class hours)</p>	
<p>EFL Spain, primary-secondary school (Terrazas Gallego & Agustín Llach, 2009)</p>	<p>2K VLT; test of mastery (English L2)</p>	<p>Tested from 4 to 7 grade (longitudinal)</p>	<p>General steady increase, but to different incremental degrees</p>	<p>+ 152 (average increase per year, 4–5, 5–6, and 6–7)</p>
<p>(n = 224)</p>			<p>More divergence between participants at higher proficiency level</p>	
<p>EFL Hungary, primary school (Orosz, 2009)</p>	<p>XLex test (English L2) (Meara & Milton, 2003)</p>	<p>Tested from 3 to 6 grade (cross-sectional)</p>	<p>General steady increase, but not constant</p>	<p>+370 (average increase per year, 3–4, 4–5, and 4–6)</p>
<p>(n = 253)</p>				
<p>EFL Taiwan, high school and university (Webb & Chang, 2012)</p>	<p>VLT test (English L2)</p>	<p>Tested from 1st to 5th year (1st to 3rd=high school) (4th to 5th=university)</p>	<p>General steady increase, differences by motivation and instruction time</p>	<p>+ 200–300 (average for the most motivated students)</p>
<p>(n = 222)</p>				<p>Authors recommend a target of +400</p>
<p>*(reported on</p>		<p>(longitudinal)</p>		

Appendix A)

per year

English majors, in a Turkish university (Ozturk, 2016) (n = 174)	VST (English L2)	Tested from 1st to 4th year (cross-sectional)	General steady increase, with attrition in their last year	+ 500 (estimated average per year, word-families)
Spanish L2 at a US university (Robles García, 2020a) (n = 1,121)	3K-LEx test (Spanish L2)	Tested from 1st to 4th year, in different courses (each course taught for 4 months) (cross-sectional)	General increase, with plateau at advanced courses followed by an increase	+ 542 (from first tested course to the second one) +103 (average word learning in subsequent courses)

Appendix C. Questions in the Background Questionnaire Examined in this Study

1. University level
Freshman, Sophomore, Junior, Senior
2. How long have you studied Spanish before starting at XX? (in years)
3. Where do you speak Spanish?

Appendix D. Vocabulary Test Words and Pseudowords

Words in the Vocabulary Test

1. el	2. ir	3. primero	4. siempre	5. tal
6. mano	7. embargo	8. mes	9. señor	10. bastante
11. ciento	12. nacer	13. escuchar	14. servicio	15. verdadero
16. miedo	17. alguien	18. faltar	19. conocido	20. consecuencia

21. vuelta	22. exigir	23. expresar	24. dispuesto	25. incluir
26. participar	27. caballo	28. perfecto	29. existencia	30. firmar
31. población	32. labor	33. mejorar	34. mínimo	35. plaza
36. región	37. novela	38. escaso	39. falso	40. médico
41. temprano	42. torno	43. particular	44. justicia	45. sorprender
46. inferior	47. delicado	48. nervioso	49. período	50. barato
51. mezclar	52. quejar	53. poderoso	54. padecer	55. raza
56. tienda	57. abogado	58. abuelo	59. traje	60. sueldo
61. alternativa	62. casado	63. primo	64. borrar	65. cuello
66. demanda	67. posiblemente	68. absurdo	69. rostro	70. aproximar
71. desempeñar	72. acceder	73. onda	74. casualidad	75. vestido
76. orientar	77. juntar	78. fase	79. plenamente	80. capitán
81. azar	82. posteriormente	83. herida	84. intercambio	85. estrategia

86. descripción	87. remontar	88. acento	89. determinación	90. geografía
91. cementerio	92. novio	93. visitante	94. realización	95. socialista
96. hierba	97. hueco	98. diseñar	99. coro	100. gigantesco
101. campana				

Pseudo-Words in the Vocabulary Test (taken from Carreiras & Perea, 2004)

1. gespa	2. borga	3. cinte	4. parbis	5. rinto
6. funte	7. busfa	8. dulta	9. restín	10. fungo
11. desbo	12. porbes	13. senfón	14. sombal	15. golna
16. conmes	17. sanco	18. rasben	19. cervo	20. tancón
21. cergie	22. tosmi	23. gerpa	24. mulges	25. bonga
26. ciste	27. risto	28. furte	29. bulfa	30. durta
31. fusgo	32. delbo	33. pombes	34. sosbal	35. lasvín
36. gozna	37. sasco	38. cesvo	39. tunquín	40. jaspi

41. sampen	42. tascón	43. cesgue	44. niti	45. nomi
46. fipe	47. ropi	48. nalu	49. junquen	50. cenvo

About the Authors

Nausica Marcos Miguel (PhD, University of Pittsburgh) is currently an Assistant Professor at Ghent University. In the United States, she has taught Spanish and courses on linguistics at Denison University and at the University of Pittsburgh. Her research focuses on L2 vocabulary learning and instruction, and on teaching and learning materials.

E-mail: nausica.marcosmiguel@ugent.be

Joshua R. Edge (PhD, Indiana University—Bloomington) is an Assistant Professor at Austin College where he teaches data science and statistics courses. His main research area is applied statistics.

E-mail: jedge@austincollege.edu

Mary Elizabeth Beaton (PhD, The Ohio State University) is an Assistant Professor at Denison University where she teaches Spanish and linguistics. Her main research areas are phonetics, phonology, and sociolinguistics.

E-mail: beatonm@denison.edu

Claudia H. Sánchez-Gutiérrez (PhD, Universidad de Salamanca, Spain) is an Associate Professor at the University of California, Davis, where she teaches Spanish and Applied Linguistics while also working as the First-Year Spanish program coordinator. Her research interests include L2 morphological awareness, vocabulary teaching and learning, L2 Spanish textbooks, and teacher cognition.

E-mail: chsanchez@ucdavis.edu