



Vocabulary Acquisition in Beginner L2 Learners

Julia Charske, Andrea Landaeta and M. Gabriela Puscama

College of Liberal Arts and Sciences, University of Florida

Dr. M. Gabriela Puscama, Department of Spanish and Portuguese Studies

Abstract

This project focuses on how different vocabulary teaching methods may develop an efficient lexicon in second language (L2) learners. Vocabulary that frequently co-occurs in the environment becomes associated in the lexicon (e.g., *food + sports*) and it strengthens the representation of vocabulary. This experiment tests the effect of co-occurrence presentation and thematic list presentation on vocabulary processing. The goal is to inform vocabulary teaching methods. L2 learners completed a pre-test, four sessions of vocabulary training, and an immediate post-test. During training sessions, participants saw images on a computer and reviewed Spanish vocabulary, under one of two conditions: one word presented at a time in thematic lists (e.g., *food*), or words displayed in a visual scene or context (e.g., *dinner party*). During testing, participants completed a picture naming task to assess how different types of vocabulary exposure impacts L2 processing. Other individual differences such as working memory and proficiency were collected as well. The results from the picture naming task show that effectiveness of training in thematic list (TL) versus visual scene (VS) depended on initial proficiency and working memory. Participants with a higher initial proficiency and working memory improved more when trained with the VS condition, while those with lower proficiency and working memory had greater success when trained via the TL condition.

Keywords: L2 vocabulary, L2 Spanish, second language acquisition, co-occurrence, vocabulary training, semantic network

Introduction

In the era of expanding global connectivity, the importance of language learning is more vital than ever. However, numerous students enrolled in second language (L2) classes encounter challenges in mastering these languages. One prominent hurdle in L2 learning is the acquisition of vocabulary. Without an understanding of the words that make up a language (vocabulary), language mastery is impossible. Much research has been motivated by the critical role of vocabulary acquisition in second language learning (Akpınar et al., 2015; Nation, 2001; Vermeer, 2001).

Vocabulary knowledge comprehension can be divided into two aspects: breadth and depth (Read, 2004). Vocabulary breadth is the quantity of words known. In contrast, vocabulary depth refers to the quality of understanding. How well are these words known? Breadth can be gauged experimentally through tasks such as category fluency, where participants are tasked to name as many words as they can in a category under a time limit. On the other hand, vocabulary depth is more complex, thus more difficult to assess experimentally. Words known in depth are represented strongly in the lexicon. The meaning, contextual use, and nuances of the word are also mastered. When a word has a strong lexical representation, it can be retrieved more easily in tasks such as picture naming.

Though learners are typically exposed to numerous words in their L2, they frequently struggle with lexical retrieval, indicating potential weaknesses in lexical representation (less depth of knowledge). This may stem from how vocabulary is taught in schools: semantically related sets (e.g., clothing, food) (Finkbeiner & Nicol, 2003). Thematic methods of teaching aim to enhance apprehension of similar words and the nuances between them (Seal, 1991). While this teaching method has become widespread, empirical evidence for its support is limited (Finkbeiner & Nicol, 2003). Research suggests that learning in thematic lists could lead to slower translation times compared to those learned in a random sequence (Finkbeiner & Nicol, 2003). For instance, Bolger and Zapata (2011) found that learners who encountered unrelated vocabulary within the context of a story (e.g., visiting the zoo) performed better on word-verification tasks than those exposed to thematically related words (e.g., animals). Further research on different teaching methodology of L2 vocabulary suggests teaching vocabulary

through associative meanings (e.g., how a word fits into context) can enhance vocabulary acquisition (Zhang & Lu, 2015).

While the results from (Zhang & Lu, 2015) indicate that varying teaching methods impact L2 vocabulary acquisition, other factors, such as individual differences like working memory and proficiency also influence the effectiveness of teaching vocabulary to specific learners (Kim & Webb, 2022). Thus, the present research aims to analyze the effectiveness of different vocabulary presentations (thematic lists vs. co-occurrence), considering these individual differences.

Research Question

Do different methods of teaching vocabulary (i.e., co-occurrence vocabulary vs. thematic lists) to beginner language learners affect lexical representations, thus influencing their ability to recall vocabulary?

Hypothesis

Based on studies suggesting that learning L2 vocabulary in thematically related lists may hinder word representation and retrieval times (Finkbeiner & Nicol, 2003, and Bolger & Zapata, 2011), the authors hypothesize that learning vocabulary in the co-occurrence condition will result in higher accuracy and faster word retrieval during the picture naming task compared to the thematic list condition.

Methodology

Participants

In this study, 16 L2 learners were recruited from beginning-level Spanish classes at a large university (age range 18-23 years old., $M = 19.87$, $SD = 1.55$). The L2 learner group consisted of 10 participants who self-identified as female, 5 as male, and 1 who preferred not to disclose.

The L2 learners targeted in this study were beginning Spanish learners without prior immersion in their L2. Participants were asked to complete a language use and history questionnaire through Qualtrics (Qualtrics, Provo, UT), adapted from López-Beltrán Forcada (2021). Two participants reported having at least one Spanish-speaking caregiver. Upon closer examination, one participant used Spanish frequently at home during childhood and had a significantly higher rate of accuracy (65.92%) in the pre-test compared to the beginners, making

them ineligible for the study. The other participant indicated no Spanish use at home and responded similarly to other beginner learners (18.99%), thus they were included in the study. During coding, another participant with no reported Spanish-speaking background performed with a high degree of accuracy, not typical of a beginning learner (76.54%), so they were also excluded. The 14 remaining participants grew up in an English-speaking home, except one participant whose first language (L1) was Mandarin. One of the participants did not complete the individual difference tasks and therefore was excluded, leaving a total of 13 participants.

Materials and Tasks

The L2 learners were divided equally into two experimental groups (Visual Scene and Thematic List).

vocabulary training.

The training materials and procedures replicated the Puscama (2024) study by selecting 180 Spanish words based on the responses of native Spanish speakers to a word association task. For each word, an image that pictured the object was created using the platform Story Board That (Story Board That, 2022). To norm the pictures and make sure that the images evoked the right concepts for English speakers, 100 native English speakers were recruited on the crowdsourcing site Prolific (www.prolific.co). They were asked to name, in English, the pictures designed for the training. Their responses were rated for accuracy, coding as 1 for a response that matched the intended concept in the picture or 0 if the name given was not the meaning intended. The average accuracy in this norming task was 92.66%.

Spanish words were recorded individually by a female native Spanish speaker in a sound-attenuated booth, using a USBPre2 connected to a Dell desktop computer in Praat (Boersma & Weenink, 2017), with a Shure SM58 microphone and a stand. Using Praat scripts, the audio files' intensity was normalized to 65dB. The beginning and final 5ms was faded. The same sound files of the words were used for each vocabulary training condition.

Thematic list condition—this training condition mimics the vocabulary presentation found in L2 textbooks, where words are taught in thematic lists per chapter (food, professions), with a progression in theme complexity. The 180 words selected were classified

into themes, following different Spanish textbooks (Tamariz, 2000; Terrell et al., 2001; González-Aguilar and Rosso-O’Laughlin, 2005; Potowski et al., 2012; Olivella de Castells et al., 2013, 2015; Cubillos, 2015; Mir and Bailey de las Heras, 2015; Lumen Learning 2018–2019), then organized into an increasing word complexity level, as outlined in the textbooks. For example, words like *lápiz* 'pencil' and *reloj* 'clock' from the theme 'classroom items' were introduced early in the training, while *vidrio* 'glass' and *reciclaje* 'recycling' from the theme 'environment' appeared later in the sequence.

Visual scene condition—for the visual scene training, six visual scenes were created using the platform Story Board That (2022). Each scene depicted at least two characters in each of these situations: shopping, dinner party, cleaning, sporting event, traveling, and hospital visit. Each scene included 30 vocabulary items for the participants to review, for a total of 180 words as in the list-by-theme training condition. Because of the density of items, each scene was divided into two panels, with 15 items in each one. This allowed the objects to be spatially distributed, easier to distinguish, and to show a story progression, like a comic strip (Figure 1).



Figure 1. Sample Visual Scene for training

picture naming.

For this task, 180 color-drawing pictures that depicted Spanish words were created using the platform Story Board That (2022). These were the same images used for the thematic list condition vocabulary training.

In addition to the experimental tasks detailed above, participants completed two individual differences measures, a verbal fluency task to assess language proficiency, and a digit span task to gauge working memory.

verbal fluency.

This task is adapted from Baus, Costa, and Carreiras (2013), and Linck, Kroll, and Sunderman (2009) involved participants naming as many items as possible within a category in thirty seconds. The task was presented on a video and oral responses were recorded using a Marantz PMD661MKII recorder attached to a Shure SM86 microphone. Eight categories were used, divided into two groups and counterbalanced between the languages. One group of categories included *clothing, animals, vegetables, and musical instruments*, and the other group included *colors, fruits, body parts, and furniture*. The order of presentation of languages was English first and then Spanish. The task began with a practice category which was not analyzed. Participants earned points for each new item mentioned within a category, and these points were averaged across all categories to calculate an overall fluency score per participant per language.

digit span.

During this task, participants saw a sequence of digits on the screen, and they were asked to remember and recall orally in English each sequence in the same order. The task was presented via video on a laptop computer and oral responses were recorded using a Marantz PMD661MKII recorder attached to a Shure SM86 microphone. The sequences increased progressively from 2 to 9 digits, and there were two sequences of each length, for a total of 16 trials. Participants received one point for each sequence recalled correctly, with the maximum possible score being 16. Participants did not receive points after failing to recall both sequences of the same length.

Procedure

Participants answered a language history questionnaire online and then completed two in-lab sessions of 60-80 minutes each.

in-lab sessions.

The in-lab sessions were approximately 1 week apart, and the task order varied in each session. Table 1 summarizes this order, and the procedure of each task is detailed below.

Table 1. Experimental Procedure Summary

Session 1	Session 2
Picture Naming	Training 2
Training 1	Picture Naming
Verbal Fluency	
Digit Span	

Training—this task was completed on Experiment Builder (SR Research, 2004–2015b), participants saw either individual pictures (TL condition) or visual scenes like those shown in Figure 1, above (VS condition). In the TL condition, participants viewed and heard the name of each object in Spanish, they were instructed to repeat it aloud, and then press any button to continue to the next object. In the VS condition, participants saw 6 visual scenes, sequentially, and were instructed to press any button, so that one of the objects on the scene was highlighted with a red circle and enlarged in a bubble, as they heard and read the name of the object. Participants were instructed to repeat the name aloud before moving on to the next objects. They repeated this procedure for each of the 6 scenes. Each training session included two repetitions of each condition, totaling four vocabulary trainings across sessions 1 and 2. Both training conditions used the same audio. In each case, participants were trained in 180 words (30 per scene in the VS condition).

Analysis

The picture naming responses were coded for accuracy (1 or 0). If a participant provided a response not intended but that was reasonable, the researchers discussed and reached a consensus on whether to consider a response correct or incorrect. Accuracy was analyzed with generalized linear mixed-effects models (GLME) using the Buildmer (Voeten, 2020) package in R (R Core Team, 2013). The main model had accuracy as a binomial outcome (0 or 1), and it included fixed effects of session (session 1 or pre-test vs. session 2 or posttest), training condition (thematic list vs. visual scene), Spanish proficiency as a continuous variable, and working memory as a continuous variable. The model also included random effects of participant and target item on

the intercept. Session 1 and TL training were the baseline factors against which all comparisons were made.

In this model, the Buildmer package (Voeten, 2020) takes the maximal model specified by the user (Barr et al., 2013; Bates et al., 2015), and, if it does not converge, it automatically performs a backward stepwise elimination based on the significance of the change in log likelihood.

Results

The average working memory score (Digit Span) was 10.54 ($SD=1.98$). The average proficiency score (Verbal Fluency) indicates that participants were more proficient in English ($M=46.77$, $SD=7.78$) than in Spanish ($M=9.69$, $SD=5.63$).

Figure 2 and Figure 3 summarize the accuracy scores (y-axis) across sessions 1 and 2 (pre and posttest, x-axis) for each training group, TL (blue) and VS (orange). Figure 2 shows the participants divided by working memory score, and Figure 3 shows them divided by proficiency score (low on left and high on right). Note that the individual differences are visualized as median-split scores, but these variables were modeled as continuous predictors.

Overall, all participants showed an improvement from pretest to posttest, when comparing accuracy scores of sessions 1 and 2 in each panel of Figures 2 and 3. However, when taking into consideration the individual differences, some participants improved more than others in the two training conditions. Thus, working memory and proficiency modulates the effectiveness of different trainings, as confirmed with the statistical analysis.

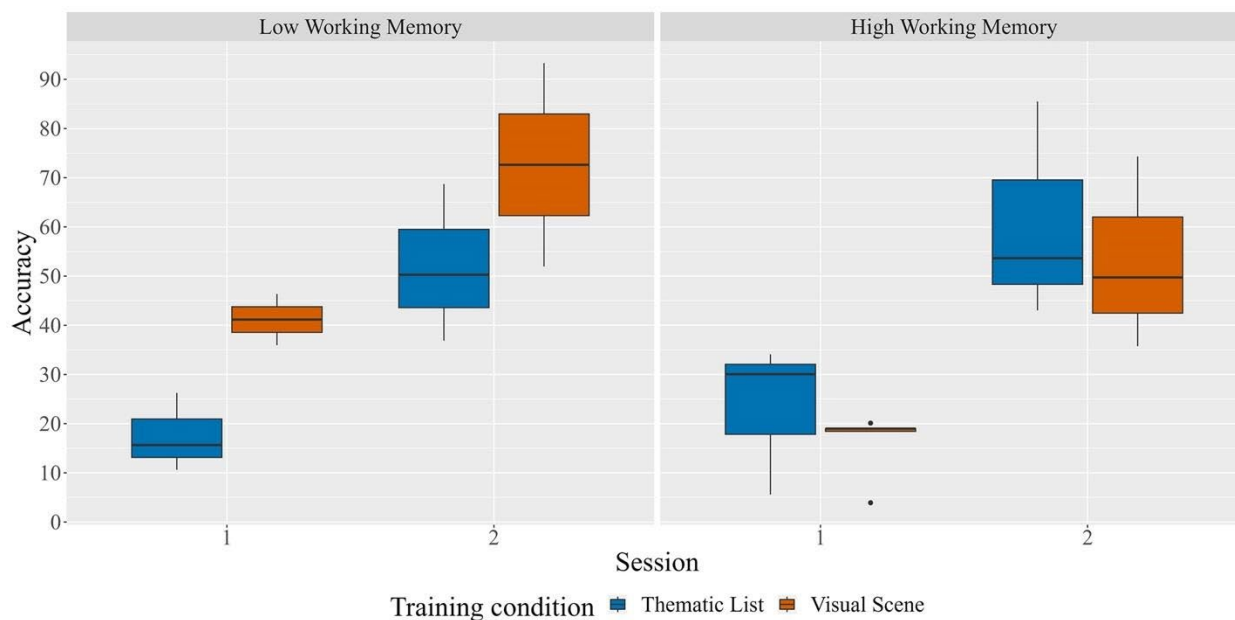


Figure 2. Picture Naming Accuracy by Session and Working Memory Score

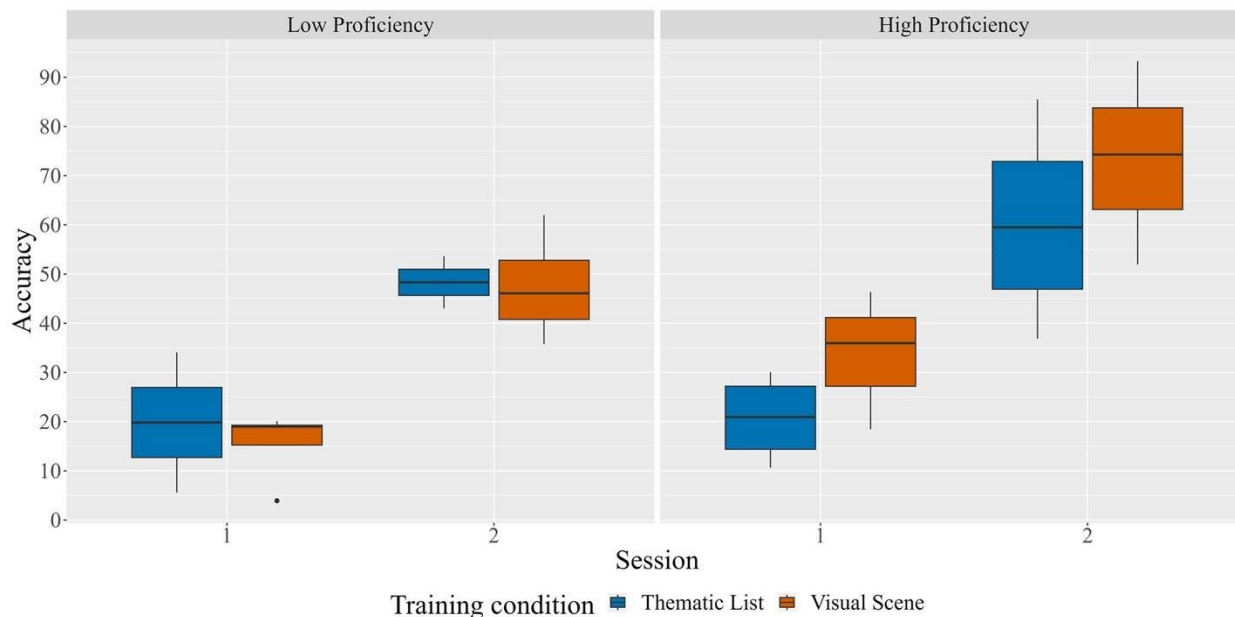


Figure 3. Picture Naming Accuracy by Session and Spanish Proficiency Score

Table 2 presents the results for the generalized linear mixed effects model. From the analysis, three main findings are relevant to the research question.

Table 2. GLME results

	<i>E</i>	<i>SE</i>	<i>z</i>	<i>p</i>
(Intercept)	-34.00	5.14	-6.62	<.001 ***
Session	15.73	3.00	5.24	<.001 ***
Proficiency	2.71	0.45	6.02	<.001 ***
Working memory	2.74	0.44	6.17	<.001 ***
Training condition	43.20	5.55	7.78	<.001 ***
Session*Prof.	-1.24	0.26	-4.71	<.001 ***
Prof.*WM	-.024	0.04	-6.15	<.001 ***
Session*WM	-1.27	0.26	-4.88	<.001 ***
Prof.*Training	-3.29	0.47	-7.01	<.001 ***
Session*Training	-23.82	3.26	-7.30	<.001 ***
WM*Training	-3.83	0.48	-7.95	<.001 ***
Sess.*Prof.*WM	0.11	0.02	4.91	<.001 ***
Sess.*Prof.*Train.	1.79	0.28	6.45	<.001 ***
Sess.*WM*Train.	2.08	0.28	7.32	<.001 ***
Prof.*WM*Train.	0.30	0.04	7.12	<.001 ***
Sess.*Prof.*WM*Train.	-0.16	0.02	-6.40	<.001 ***

First, the significant main effect of session indicates that participants in the TL condition in session 2 are more likely to name a picture correctly than in session 1.

Second, the significant interaction of session and training condition with a negative slope indicates that the effect of session is modulated by the type of training, with the VS group potentially benefiting less from the training than the TL group. This could be due to the VS group having initially higher accuracy than the TL group in session 1 (see main effect of training condition in session 1).

Finally, there are two significant three-way positive interactions between session, training condition, and proficiency, and session, training condition, and working memory, respectively. The combined effect of session, proficiency and training condition with a positive slope may indicate that participants with higher proficiency benefited more from the VS training. Similarly, participants with higher working memory scores showed more improvement in session 2 after the VS training.

Discussion

This research aimed to determine if learning vocabulary presented in visual scenes (VS) yielded a stronger representation in the lexicon, facilitating quicker word retrieval than vocabulary presented in thematic lists (TL). After both training conditions (TL and VS), accuracy increased for the picture naming task. During session 1, participants completed the

picture naming task (pre-test) prior to vocabulary training. In session 2 participants completed the same task (immediate posttest) after completing 4 training sessions (2 from session 1 and 2 from session 2). Scores of the posttest for both conditions (TL and VS) were significantly more accurate in the second session suggesting participant engagement and learning. Overall, both training conditions improved participant vocabulary accuracy.

When comparing the effectiveness of the TL and VS training conditions, the negative slope of the effect indicates participants from the TL training improved more from session 1 to 2 than VS condition. However, participants from the VS training started with higher initial scores in the picture naming task before training than the TL group, potentially limiting their room for improvement. This could explain the greater increase in accuracy for the TL group. Increasing the number of participants could help mitigate the initial group differences observed in the data. This interaction suggests that while TL training resulted in greater improvement for participants with lower initial scores, it does not necessarily imply that TL was a superior training method overall.

Participants with higher proficiency and working memory (WM) benefited more from the VS training than participants with lower proficiency and WM. This is indicated in the three-way positive interactions between both session, training condition, and proficiency, and session, training condition, and working memory. These findings suggest that participants with higher initial proficiency and WM could effectively allocate cognitive resources to learn vocabulary in a visually complex environment, such as the VS condition. Whereas participants with an initial lower proficiency could have been overwhelmed by the visual scene and would have learned better if the words were in isolation, as in the TL condition. Overall, the VS condition was most effective in participants who initially possessed higher proficiency and WM.

limitations and future directions.

Limitations of the study include the small sample size. To mitigate this effect, a sample of 40 participants would be preferred to reduce the effect of initial group differences and outliers. Another limitation is timing of the trainings and posttest. Participants completed 4 trainings across 2 weeks and were given an immediate posttest. To assess long term effects, another test could be administered a week following the trainings to examine sustained learning from different methods of teaching vocabulary.

This study demonstrates that there is no ‘one size fits all’ approach when it comes to L2 learning. Participants come from diverse backgrounds and learning styles. The hypothesis that the VS condition emulating an immersion environment would result in greater success in vocabulary accuracy was explored. Results indicated that there is not necessarily one superior training condition; rather, the effectiveness of training type depended on proficiency and working memory. Regardless of training condition, increased practice time improved vocabulary accuracy, highlighting its significance in improving language proficiency.

Acknowledgements

The authors gratefully acknowledge the CLAS Scholars Program for funding this research. Thank you to Dr. Jorge Valdés Kroff for providing lab equipment and lab space. Special thanks to Dr. Puscama, for her consistent support.

References

- Akpınar, K. D., Aşık, A., & Vural, A. S. (2015). The relationship between the effectiveness of vocabulary presentation modes and learners’ attitudes: Corpus based contextual guessing, dictionary use, and online instruction. *The Asian EFL Journal Quarterly*, 17(17), 90–115.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv*, 1506.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), pp. 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica*, 142(3), pp. 402-409.
- Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.36. <URL <http://www.praat.org/>>
- Bolger, P., & Zapata, G. (2011). Semantic categories and context in L2 vocabulary learning. *Language Learning*, 61(2), 614-646. <https://doi.org/10.1111/j.1467-9922.2010.00624.x>
- Cubillos, J. H. (2015). *Charlemos: Conversaciones Prácticas*. Pearson.
- Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in Second language word learning. *Applied Psycholinguistics*, 24(3), 369–383. <https://doi.org/10.1017/s0142716403000195>
- González-Aguilar, M., & Rosso-O’Laughlin, M. (2005). *Atando Cabos: Curso Intermedio de Español*. (2nd ed.). Pearson.
- Kim, S. K., & Webb, S. (2022). Individual difference factors for second language vocabulary. In: S. Li, P. Hiver, and M. Papi. (Eds.). *The Routledge Handbook of Second Language Acquisition and Individual Differences*. (pp. 282-293). Routledge.

- Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, 20(12), pp. 1507-1515.
- López-Beltrán Forcada, P. (2021). *Heritage speakers' online processing of the Spanish subjunctive: A comprehensive usage-based study*. [Doctoral dissertation, The Pennsylvania State University]. Electronic Thesis and Dissertations for Graduate School.
- Lumen Learning (2018-2019). *Introductory Spanish II*. <URL <https://courses.lumenlearning.com/wm-spanish2/>>
- Mir, M., & Bailey de las Heras, Á. (2015). *¡Qué me Dices! A Task-Based Approach to Spanish Conversation*. Pearson.
- Nation, I. S., & Nation, I. S. P. (2001). Learning vocabulary in another language (Vol. 10, pp. 126-132). Cambridge University Press.
- Olivella de Castells, M., Guzmán, E. E., Lapuerta, P., & Liskin-Gasparro, J. E. (2013). *Mosaicos: Course Materials for Spanish 2* (4th Custom ed.). Pearson.
- Olivella de Castells, M., Guzmán, E. E., Lapuerta, P., & Liskin-Gasparro, J. E. (2015). *Mosaicos: Course Materials for Spanish 3* (5th Custom ed.). Pearson.
- Potowski, K., Sobral, S., & Dawson, L. M. (2012). *Dicho y Hecho: Beginning Spanish*. (9th ed.). John Wiley & Sons.
- Prolific. (2014) <URL <https://www.prolific.co>>
- Puscama, M. G. (2024). Semantic network development in L2 Spanish and its impact on processing skills: A multi-session eye-tracking study. *Languages* 9(43). <https://doi.org/10.3390/languages9020043>
- Qualtrics. (2005). <URL <https://www.qualtrics.com>>
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <URL <http://www.R-project.org/>>
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? *Vocabulary in a second language*, 209.
- Schmitt, N. (1997). Vocabulary learning strategies. *Vocabulary: Description, acquisition and pedagogy*, 199227, 4-6.
- Seal, B. D. (1991). Vocabulary learning and teaching. *Teaching English as a second or foreign language*. Boston, MA: Heinle & Heinle.
- SR Research Experiment Builder 2.2.38 [Computer software]. (2004-2015). SR Research Ltd.
- Story Board That. (2022). <URL <https://www.storyboardthat.com/>>
- Tamariz, M. (2000). *Oxford Spanish Cartoon-Strip Vocabulary Builder*. Oxford University Press.
- Terrell, T. D., Andrade, M., Egasse, J., & Muñoz, E. M. (2001). *Dos Mundos*. (5th ed.). McGraw-Hill.

- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217–234. doi:10.1017/S0142716401002041
- Voeten, c. C. (2020). buildmer: Stepwise elimination and term reordering for mixed-effects regression. <URL <https://cran.r-project.org/web/packages/buildmer/index.html>>
- Zhang, X., & Lu, X. (2015). The relationship between vocabulary learning strategies and breadth and depth of vocabulary knowledge. *Modern Languages Journal*, 99(4), 740-753.