Research Article

INVESTIGATING TEXTUAL ENHANCEMENT AND CAPTIONS IN L2 GRAMMAR AND VOCABULARY

AN EXPERIMENTAL STUDY

Myrna C. Cintrón-Valentín ⓘ *

University of Michigan

Lorenzo García-Amaya ⓘ

University of Michigan

Abstract

To probe the limits of attention raising through form-focused instruction, second-language research must adapt to the needs of a technologically driven learning environment. In this study, we used a randomized control design to investigate the effect of captioned media on the learning of vocabulary and grammar in L2 Spanish ($n = 369$ learners). Through four data-collection sessions, participants were presented with a grammar-lesson video and a multimodal video with one of three captioning formats: textually enhanced target vocabulary, textually enhanced target grammar, or no captioning. Results show strong immediate effects of captioning on target vocabulary, with additional effects of captioning on some, but not all, target-grammar structures. The findings demonstrate that (a) the learning of some grammatical structures is more conducive to captioning than others, and (b) there is space for future investigation into the factors that may influence the effectiveness of multimodal interventions, such as prior knowledge or frequency of use.

INTRODUCTION

One core theoretical question that permeates much of the second-language (L2) literature is the role of learner attention, namely whether the low perceptual salience of certain input features (e.g., verb inflectional morphology, grammatical particles) yields challenges for L2 acquisition (Ellis, 2017; Gass et al., 2017; Goldschneider & DeKeyser, 2001). As such, one key area of inquiry within L2 research is how to enhance learner attention to

*Correspondence concerning this article should be addressed to Myrna C. Cintrón-Valentín, Department of Psychology, 530 Church Street, Ann Arbor, MI 48109. E-mail: mcintron@umich.edu

commonly unattended input features (e.g., Schmidt, 2001). In particular, studies have examined the role of form-focused instruction (FFI) techniques, such as Textual Enhancement (TE) and explicit grammar instruction, in rendering target structures more salient (Norris & Ortega, 2000; Terrell, 1991). Empirical work that probes the limits of attention raising through FFI is shifting to work within a technologically driven language-learning setting (Blake, 2013; Cintrón-Valentín et al., 2019; Lee & Révész, 2018; Plass & Jones, 2005). With the increased availability of multimedia language-learning materials, FFI research can more deeply scrutinize the role of multimodal input (i.e., aural, written, and visual) in facilitating L2 development.

One promising multimodal technique is that of captioned video[1] (e.g., Montero Perez et al., 2013; Vanderplank, 2010). The effects of captioned media on L2 comprehension and vocabulary learning are well studied (e.g., Montero Perez et al., 2014; Muñoz, 2017; Winke et al., 2013), and researchers are now turning their attention to the role of salience raising through captioned video on the learning of L2 grammar (e.g., Cintrón-Valentín et al., 2019; Lee & Révész, 2018). However, much work remains before we can fully understand the benefits of captioning on grammar learning. For instance, it is not clear whether captioned media is reliably effective for all grammar structures (Lee & Révész, 2018), or whether the learning of some structures may require a greater degree of instructional support to reap the full benefits of multimodal input (Cintrón-Valentín et al., 2019). Additionally, we do not know if captioned media facilitates the salience of morphological forms whose semantic meaning derives from the surrounding discourse (e.g., Bardovi-Harlig, 1998), or whether any initial positive gains experienced through TE captions are maintained over time (Ellis, 2012). This article responds to such questions by examining the impact of TE in captioned video, alongside explicit grammar instruction, on the L2 acquisition of Spanish vocabulary and morphosyntax. We focus on four grammar structures that pose known challenges for English-speaking L2 learners of Spanish: the preterite/imperfect contrast, *gustar*-type verbs, the subjunctive in noun clauses, and the conditional. Lastly, we consider the practical implications (Larsen-Freeman, 2003; Vanderplank, 2010) and theoretical importance (Gass et al., 2017) of this study given the limitations of previous research.

BACKGROUND

SLA WITH MULTIMEDIA AND CAPTIONING

Within the past 30 years, technological advances have made it possible to integrate multimedia materials into the L2-classroom environment (Blake, 2013; Plass & Jones, 2005). SLA with multimedia can be defined as "the use of words and pictures [either static or dynamic] to provide meaningful input, facilitate meaningful interaction with the target language, and elicit meaningful output" (Plass & Jones, 2005, p. 469). Webb and Nation (2017) discuss how the use of elaboration techniques can "provide a memorable image of the meaning and context of a word" (p. 73), thereby facilitating acquisition. Captioned media is one of many multimedia materials available to L2 learners and instructors (e.g., Chun & Plass, 1997; Jones & Plass, 2002). This technique has garnered attention in recent years given its demonstrated benefits in facilitating L2 comprehension and vocabulary acquisition (e.g., Montero Perez et al., 2013; Vanderplank, 2010).

Winke et al. (2010) attribute the usefulness of captioned media to matters of attention, suggesting that this medium draws learners' attentional focus to unknown forms and promotes subsequent noticing and learning through repeated exposure. This hypothesis is consonant with foundational theories in SLA that stress that attention is central to successful L2 acquisition (e.g., Gass et al., 2017; Schmidt, 2001). Schmidt's (2001) Noticing Hypothesis, for instance, holds that conscious attention to linguistic forms in the input is an important precondition to learning (but see Truscott & Sharwood Smith, 2011). Vanderplank's (2016) model of language acquisition through captioned media similarly emphasizes how the "taking out" of language from captioned videos promotes learners' attention to language and allows them to shift their attentional focus, thereby meeting their learning goals through a process of adaptation.

PERCEPTUAL SALIENCE, FORM-FOCUSED INSTRUCTION, AND CAPTIONING IN L2 LEARNING

Research shows that the low perceptual salience of grammatical forms is one key factor underlying the challenges posed during L2-grammar acquisition (e.g., Ellis, 2017; Gass et al., 2017). Although it is also known that salient elements can, at times, fail to be learned rapidly (Ellis, 2006), there is evidence that the low perceptual salience of inflectional suffixes contributes to L2 learners' difficulty in acquiring them (Cintrón-Valentín & Ellis, 2016; Goldschneider & DeKeyser, 2001). Within L2 pedagogy and research, this challenge may be counteracted by providing learners with FFI (see VanPatten, 1996, for alternative interventions). FFI encapsulates a wide range of instructional activities that draw learners' attention to linguistic forms in the input that might otherwise be ignored (Ellis, 2012; Spada & Tomita, 2010).

Two common FFI methods are explicit-grammar instruction (EGI) and TE (Han et al., 2008; Lee & Huang, 2008; Norris & Ortega, 2000; Sharwood Smith, 1993). Regarding EGI, Terrell (1991, p. 53) defines this method as "the use of instructional strategies to draw the students' attention to, or focus on, form and/or structure." With the goal of increasing salience, instructors implementing EGI first point out the commonly ignored feature, explaining its structure, and then provide meaningful input containing many instances of the meaning-form relationship. TE, however, uses visual manipulations such as the color-coding, boldfacing, and underlining of the structure to provide a less obtrusive means of increasing learners' awareness of nonsalient forms (Sharwood Smith, 1993). Given the increased reliance on multimedia materials in L2 teaching and learning (see the "SLA with Multimedia and Captioning" section), it is of interest to investigate how FFI principles can guide the elaboration of multimodal research and pedagogical materials.

Recent studies in the vocabulary-learning literature have begun to implement such designs. Montero Perez et al. (2014) found significant advantages in vocabulary learning based on caption type in an experimental design using (a) the absence of captions, (b) standard captioning with full captions, (c) full captions plus highlighted keywords, and (d) keyword-only captions (see also Pujadas & Muñoz, 2019). In spite of such advances, Montero Perez et al. (2013) underscore that it is imperative to explore whether there are long-term effects of captioning on vocabulary retention (see also Ellis, 2012; Lee & Huang, 2008). Additionally, as new research methods are being proposed, it is

necessary to explore the extent to which past findings can be replicated through current methodologies (see Marsden et al., 2018). Our study is designed to respond to these gaps of knowledge, relating specifically to the effects of captioning in promoting L2-vocabulary knowledge.

Regarding grammar, to our knowledge only two studies (Cintrón-Valentín et al., 2019; Lee & Révész, 2018) have investigated the role of FFI in combination with captioned media on enhancing learner attention to L2 grammatical forms. Lee and Révész (2018) investigated the effects of TE-captioned media on the learning of pronominal-anaphoric reference for L1-Korean learners of English. The researchers presented learners with multimodal input (i.e., audio and written text) in a story narration. The design included the narrative script (with a TE group seeing text in bold) and pictures in a slide show. However, the pictures did not guide the narrative as the imagery in a video would (as recognized by the authors, see p. 574). To follow up on this point, Cintrón-Valentín et al. (2019) investigated how captioned video could serve as a useful tool for advancing grammar learning in L2 Spanish. The authors developed four original multimodal videos centered around grammar structures known to pose persistent challenges in L2-Spanish acquisition. The findings revealed significant effects of TE captions on some, but not all, target-grammar forms. However, two methodological limitations impacted the interpretability of their findings: (a) the authors did not include a pretest prior to conducting their study, making it difficult to tease apart any confound of preexisting knowledge on the experimental gains; and (b) all captioned videos were fronted by an explicit grammar lesson, making it difficult to determine whether the use of captioning was the single contributing factor to any learning effects.

RESEARCH QUESTIONS AND RATIONALE

The current study had three guiding aims, each with specific research questions.

(1) To examine the effects of full captions + TE vocabulary on improving learner knowledge of vocabulary in L2 Spanish:
   a. What is the relative effect of full captions + vocabulary TE, full captions + grammar TE, or no TE on **vocabulary recognition**?
   b. What is the relative effect of full captions + vocabulary TE, full captions + grammar TE, or no TE on **vocabulary production**?
   c. Are any initial gains on **vocabulary production** maintained over time?

(2) To examine the effects of full captions + grammar TE on improving learner knowledge of grammar in L2 Spanish:
   a. What is the relative effect of full captions + vocabulary TE, full captions + grammar TE, or no TE on **grammar production**?
   b. Are any initial gains on **grammar production** maintained over time?

(3) To investigate if the effects of full captions + grammar TE are equally facilitative in the absence of explicit instruction:
   a. What is the relative effect of lesson + grammar TE compared to no lesson + grammar TE on grammar production?
   b. Are any initial gains on **grammar production** maintained over time?

We included research aim 1 to assess the replicability of previous findings of captioning on vocabulary acquisition. As mentioned in the "Perceptual Salience, Form-Focused Instruction, and Captioning in L2 Learning" section, Montero Perez et al. (2013) underscore the need for experimental designs that consider the long-term effects of captioning on vocabulary retention through delayed posttests. We included research aim 2 to investigate the effects of captioning on L2 grammar, specifically for four grammar structures in L2 Spanish, under explicit instruction conditions. Critically, the inclusion of a pretest–posttest design examines how durable the effects of FFI might be on L2-grammar development (Ellis, 2012; Lee & Huang, 2008). We included research aim 3 to distinguish the individual effects of captioning from the presentation of a previous grammar lesson; the results will indicate which types of constructions might be better assisted by captioning, and which might require additional instructional support (see Tolentino & Tokowicz, 2014).

Regarding the materials for this study, we used the same videos from Cintrón-Valentín et al. (2019), which are multimodal materials designed for testing captioning effects on L2-Spanish learning. To our knowledge, previous research has not gone to such lengths to develop multimodal materials for L2 vocabulary or grammar learning with specialized software (www.nawmal.com), offering the opportunity to develop engaging plots designed around target lexical items and grammar structures. These videos additionally respond to the need for controlling the selection and frequency of occurrence of individual target items (Montero Perez et al., 2015), which is especially relevant when considering the importance of frequency in L2-grammar development (Ellis, 2006).

Our study also responds to the two key limitations from Cintrón-Valentín et al. (2019) identified in the "Perceptual Salience, Form-Focused Instruction, and Captioning in L2 Learning" section. First, we included a pretest of the targeted grammar forms to discern any effects of prior knowledge from those of the experimental treatment. Second, we included an experimental grammar group that did not receive explicit instruction prior to viewing the multimodal videos. These methodological differences between Cintrón-Valentín et al. (2019) and the current study will allow us to better assess the effectiveness of captioned videos in improving learner knowledge of L2 grammar within the L2-Spanish classroom setting.

METHOD

PARTICIPANTS

A total of 369 English-speaking L2 learners of Spanish were recruited from a Spanish grammar course at a large midwestern university in the United States. They were fifth-semester intermediate learners of Spanish who participated in the study for credit as one of their course requirements. The course had 21 sections, which we quasirandomly assigned to one of four groups: a Lesson + No Salience group (Lesson + Control); a Lesson + Salience on Vocabulary group (Lesson + SV); a Lesson + Salience on Grammar group (Lesson + SG); and a No Lesson + Salience on the grammatical features group (No Lesson + SG) (see Table 1 for descriptive statistics).

TABLE 1. Descriptive statistics for background information

		Age range			Sex	
Group	N subjects	Minimum	Maximum	Mean age (SD)	Females	Males
Lesson + Control	89	17	35	18.74 (1.96)	58	27
Lesson + SV	84	17	28	18.74 (1.45)	53	27
Lesson + SG	88	17	29	18.59 (1.65)	63	23
No Lesson + SG	108	17	24	18.44 (1.10)	70	31

Note: SV = Salience on Vocabulary; SG = Salience on Grammar. These were several participants who did not report their sex (Lesson + Control = 4; Lesson + SV = 4; Lesson + SG = 2; No Lesson + SG = 7). Of the initial 369 participants, 63 were excluded from the study because they spoke an L1 other than English, had been exposed to the Spanish language before age 6, or had participated in a Spanish study-abroad experience for longer than 2 months.

WRITTEN INSTRUMENTS

Language History Questionnaire

Participants completed a Language History Questionnaire (LHQ; Li et al., 2014), which included questions about demographics and previous language-learning experiences.

Spanish Vocabulary-Proficiency Test

The Lextale-ESP (Izura et al., 2014), a 90-item (60 words + 30 nonwords) Spanish vocabulary proficiency test, was administered to all learners. Learners were asked to select words they recognized as Spanish words. The test was scored using the following formula, which penalized for guessing behavior:

$$\text{Score} = N_{\text{"yes" to words}} - 2 * N_{\text{"yes" to nonwords}}.$$

To control for any possible familiarity of the target-vocabulary items, we included the 23 target-vocabulary words alongside foils in this test (the foils were added so that participants would be less inclined to select all words as "seen" in the multimodal video). The target-vocabulary words were coded and scored separately. Participants received one point for each target vocabulary word they recognized as Spanish, for a total of 23 maximum points.

Elicited Imitation Task

Participants completed an Elicited Imitation Task (EIT), originally developed by Ortega et al. (1999), which we use as a proxy to measure global Spanish proficiency. Specifically, we used the revised EIT from Bowden (2016). Participants' utterances were scored on a 0–4 scale: a minimum score of 0 points was given for instances of silence, unintelligible productions, or minimal repetitions; a maximum score of 4 points was given for exact repetitions. Each EIT audio was scored independently by two raters, and any discrepancies were resolved prior to statistical analysis.

It should be noted that this study was not designed to gauge how learners from distinct proficiency levels respond to the multimodal interventions. We included the instruments mentioned in the "Language History Questionnaire" section through the "Elicited Imitation Task" section as a way of controlling for unexpected proficiency differences within the same grammar course, and also to control for previous knowledge of the target vocabulary items.

Grammar Pretest

Participants completed a grammar pretest that included a representative sample of each of the target-grammar structures. The test contained 51 items, where the learners were asked to translate target verbs from English to Spanish (see Supplementary Materials; Section A).

Immediate Posttests

Vocabulary-Recognition Test.   Participants were tested on their recognition of target vocabulary (see Supplementary Materials; Table B.1). They were presented with a series of written words and were asked to select "True" if they recalled being exposed to that word in the experimental session, or "False" if they did not recall the word. We tested all 23 target words, as well as the 23 foils. A score of 1 was given for each correctly identified target word.

Vocabulary-Translation Test.   Our translation test required learners to provide the Spanish translation of English words. Each correct translation was given a score of 1; synonyms or other related words not presented in the movie were scored as incorrect to ensure that we measured only the recall of target words.

Grammar-Translation Test.   Our translation test presented participants with sentences in English and asked them to provide the appropriate Spanish translation. The responses were scored based on the provision of the correct target inflection (e.g., participants were expected to distinguish the usage of the two past forms for the preterite/imperfect). For each response, participants received either a score of 1 for a correct inflection or a score of 0 for an incorrect inflection.

Delayed Posttests

Approximately 2 weeks after each of the four experimental sessions, similar grammar and vocabulary translation tests were administered during learners' regular class time. For the vocabulary portion, the delayed posttests included all target-vocabulary items presented in the multimodal video as well as foil words that appeared in the multimodal video. For the grammar portion, the delayed posttests included the same verb items the learners had been tested on in the immediate posttests but in different sentential contexts. We did not include the vocabulary-recognition instrument in the delayed posttest due to time restrictions.

GRAMMAR-LESSON VIDEOS

For each grammatical structure, a short grammar-lesson video was created using Microsoft PowerPoint and Camtasia. Each video summarized how the relevant target form is conjugated in Spanish, provided learners with detailed discussions on two to three rules, and offered multiple practice exercises. Each video lasted approximately 10 minutes (see Supplementary Material; Figure C.1 for sample slides from the conditional mood video).

MULTIMODAL VIDEOS

The multimodal videos for the preterite/imperfect, *gustar*-type verbs, and subjunctive were the same as those presented in Cintrón-Valentín et al. (2019); a new video was created for the conditional mood (see Supplementary Material; sections D and E for Spanish and English versions of the video script). For each target structure, there were three versions of the video. Each version differed only in the focus of their captioning lines (No Captions, Salience on Vocabulary, or Salience on Grammar; see Supplementary Material; Figure C.2 for example slides).

Vocabulary Content

The multimodal videos created for each lab session included 23 target-vocabulary words spread across the four sessions (see Supplementary Materials; Table B.1). The target vocabulary chosen for the experiment were either low-frequency words taken from the NIM Frequency database (Guasch et al., 2013) or regional vocabulary words. For each video, there were as many unique target-vocabulary words and target-grammar rules. Each target-vocabulary word was presented four times per video, and though the unique items were spread across each script, all repetitions of each word were placed one after the other in consecutive sentences (i.e., they were massed).

Grammar Content

Session 1: Preterite and Imperfect. The standard usage of the Spanish past-tense system requires that learners understand the aspectual distinction between the preterite and imperfect (Comajoan, 2013). Preterite forms characterize past actions as having a definitive beginning and endpoint (e.g., *caminé* "I walked"), whereas imperfect forms characterize past habitual actions or states in progress (e.g., *caminaba* "I was walking/I used to walk"). As noted in Liskin-Gasparro (2000), tense-aspect morphological forms differ in their frequency distribution in the input received by L2 learners of Spanish, leading to infrequent exposure of the contrast of these forms. As a motivating point for our study, Blyth (2005, p. 213) argues that, although there can be unintended consequences, pedagogical interventions that render surface forms more frequent and salient can allow learners to focus on form in a meaningful way. In our study, we manipulated the frequency of appearance of the preterite and imperfect forms so that both would have an equal chance of being attended to by the learners. We additionally enhanced the physical salience of both forms using distinctive highlighting with the aim of facilitating learner differentiation of these forms within our tailored narrative contexts (see Bardovi-Harlig,

1998, regarding the importance of narrative context in determining how the two aspectual choices are used).

For the preterite/imperfect, three rules for each form, and one rule that contrasted their usage, were included in the respective animated video. Each rule was represented through four different verb instances within the video script. We additionally controlled for lexical aspect in the selection of the preterite and imperfect verbs (Bardovi-Harlig, 2000).

Session 2: Gustar-Type Verbs.   L2-learners' mastery of the *gustar*-type verb construction is challenging given its difference from the English counterpart "to like."

Despite their closeness in meaning, these predicates exhibit a divergent syntactic behavior: whereas "like" codes as subject the entity that experiences a certain feeling, and as object the stimulus responsible for that feeling, gustar expresses the experiencer though an indirect object (or dative) and the stimulus through the subject. (Vázquez Rozas, 2006, p. 1)

Previous literature on the acquisition of *gustar*-type verbs relates to the processing and use of the clitic pronoun preceding the verb (e.g., Lee & Malovrh, 2009). In our study, we focus on an additional challenge, namely the agreement between verb morphology and its subject. We included six verbs—*gustar* "to like," *encantar* "to love," *interesar* "to be interested," *importar* "to care," *molestar* "to be bothered," and *quedar* "to be left"—each presented four times: twice in the singular and twice in the plural.

Session 3: Subjunctive in Noun Clauses.   The Spanish subjunctive mood is typically used in sentences with multiple clauses, in which the subject of the main clause exerts influence or will on the subject of the subordinate clause (Gudmestad, 2012). The subjunctive in L2 Spanish is often described as a "late-emerging item in both first and second language learners," due to a combination of its low frequency and the low salience of the subjunctive inflection in the input (DeKeyser & Prieto Botana, 2013, p. 454). Critically, studies have shown that breaking down the syntactic and inflectional components of this structure can facilitate its acquisition regardless of learners' readiness (e.g., Collentine, 2013). In the current study, both the verb in the main clause, which acts as a cue to the subjunctive, as well as the subordinated subjunctive verb were made salient to facilitate learners' understanding of the rules underlying subjunctive usage.

Session 4: The Conditional Mood.   Conditional sentences are considered to be highly complex structures in L1 and L2 acquisition due to their morphosyntactic complexity and the semantic nuance involved in input processing (e.g., López Ornat, 1994). The Spanish conditional is generally used to express probability or hypotheses about the past, present, or future (Areizaga Orube, 2009). In our study, we focus on one usage of the conditional: the expression of speculation or probability about the past, using the "must have + verb construction" (e.g., Where was John last night? He wasn't at home. He **must have been** in the lab./*¿Dónde estaba John anoche? No estaba en casa.* ***Estaría*** *en el laboratorio*). We targeted a low-frequency usage of the Spanish conditional, deviating from the usage included in the learners' course syllabus. In doing so, we aimed to explore the extent to which there are TE-captioned media effects on improving learner knowledge for a structure with minimal prior exposure.[2]

TABLE 2. Summary of Captioning + Textual Enhancement manipulations per grammar topic

Grammar topic	Experimental groups		
	Lesson + Control	Lesson + Salience on vocabulary	Lesson + and No Lesson + Salience on grammar
Preterite-imperfect	n/a		target verb is underlined; preterite target morpheme is bold and orange; imperfect target morpheme is bold and yellow
Gustar-type verbs	n/a		target verb is bold and underlined; target morpheme is coded in yellow; all other plurality markers are coded in yellow, but not in bold
Subjunctive in noun clauses	n/a	Target vocabulary word visually enhanced in bold and yellow	main clause verb is bold and orange, followed by an orange bold arrow; conjunction *que* "that" in bold and white; subjunctive verb is underlined, and target subjunctive morpheme is in bold and yellow
Conditional	n/a		target verb is underlined; conditional target morpheme is bold and yellow

Captioning Content and Textual Enhancement Manipulations

The effect of TE on vocabulary and grammar within the captioning line was investigated through four experimental groups, summarized in Table 2.[3]

DATA-COLLECTION PROCEDURE

The present study used a randomized control design to investigate the effect of captioned media on the learning of vocabulary and grammar in L2 Spanish (Hudson & Lorena, 2015).[4] A complete, chronological list of the data-collection procedure is indicated in Table 3. On the first day of class of the 15-week semester, the members of the research team attended all 21 course sections and administered the Spanish vocabulary-proficiency test and the grammar pretest. During the first week of class, all learners filled out the web-based LHQ. Additionally, the EIT was administered throughout the first month of class, and all learners were tested individually in a quiet room. We used a Marantz Pmd620 digital recorder and Shure WH20 head-mounted microphones to conduct these recordings.

The lab phase of the study took place over four sessions spaced throughout the semester in the order presented in the course syllabus: (1) preterite/imperfect; (2) *gustar*-type verbs; (3) subjunctive in noun clauses; and (4) conditional mood. On average, approximately 2 to 3 weeks separated each lab session. During each session, the experimenters met with the learners in a preassigned computer classroom. The experimental protocol was computerized and made available to each participant through the Canvas Learning Platform (https://www.instructure.com/canvas/), which allows for the creation of multimedia surveys.

TABLE 3.   Overview of procedure

Phase	Test	Time
Preexperimental Phase	Grammar pretest	First day of class
	Vocabulary proficiency test	
	Language history questionnaire	First week of class
	Elicited imitation task	First Month of class
	Grammar video lesson	
Experimental Phase	Animated video	Experimental session (4 times)
	Immediate vocabulary recognition	
	Immediate vocabulary translation	
	Immediate grammar translation	
	Delayed vocabulary and grammar translation test	Delayed in-class posttest (4 times)

Note: The Experimental Phase took place during eight different time points across the 15-week semester. Students saw the animated videos and took the immediate posttests for each of the four structures on their assigned class day. Two weeks after each experimental session, participants were tested on their production of the target vocabulary and grammar.

During each experimental session, learners from the first three groups were presented with the grammar-lesson video about the target form prior to watching the corresponding multimodal video manipulated per group: no captioning (Lesson + Control), target vocabulary was highlighted using TE (Lesson + SV), or grammatical features were highlighted using TE (Lesson + SG). For the fourth group (i.e., No Lesson + SG), learners saw the grammar-lesson video after watching the corresponding multimodal video.

Following the videos, participants completed the three written instruments (one vocabulary recognition, one vocabulary translation, and one grammar translation). Each lab session lasted approximately 50 minutes. Two weeks after each lab session, similar versions of the grammar- and vocabulary-translation tests were administered by the learners' instructors.

STATISTICAL ANALYSIS

Statistical analyses were conducted using RStudio version 1.0.143 (RStudio Team, 2015). The data were analyzed by generalized linear models and multilevel generalized linear regression models utilizing the glm() and glmer() functions within the lme4 package in R (Bates et al., 2015). Model diagnostics were based on plots of distributions of residuals, plots of residual versus fitted values, and checks for outlier values with high leverage. For all generalized-linear models used for vocabulary, we report odds-ratios (Exp(B)) as our effect-size statistic. Regarding the multilevel models used for grammar, to our knowledge there is not a clear agreement on whether effect sizes should be reported for such models (Rights & Sterba, 2019), thus we do not report values for these models.

Vocabulary Data

For the vocabulary recognition and translation analyses, we ran logistic-regression models on the pooled results (collapsing across all vocabulary sessions). The dependent

measures were proportion of trials correct, with GROUP (Lesson + Control, Lesson + SV, Lesson + SG, and No Lesson + SG) as the predictor of interest. The week 1 VOCABULARY PROFICIENCY test was included as a fixed variable to take into account individual differences in Spanish proficiency (see the "Proficiency Data" section, Table 4; this variable was mean centered before inclusion in the model).

Grammar Data

For the grammar-translation analysis, the dependent measures were proportion of trials correct, with GROUP, STRUCTURE (preterite/imperfect, *gustar*-type verbs, subjunctive, and conditional), and TIME (pretest, immediate posttest, and delayed posttest) as predictor terms, as well as random intercepts for SUBJECTS. The EIT was included as a fixed, mean-centered variable to take into account individual differences in Spanish proficiency.

Missing Data

Given that the learners received course credit for their participation in each of the lab sessions, they were allowed to attend a makeup session for any lab that they did not attend. If participants took a makeup after being presented with the lab material by their instructor, their data for that individual lab session was treated as missing. For the vocabulary-recognition data, any experimental word known at baseline was treated as missing for each participant.

RESULTS

PROFICIENCY DATA

Table 4 presents the group means, standard deviations, and confidence intervals for the Spanish vocabulary-proficiency test and the EIT. As can be seen, there are no obvious between-group differences on each measure.

TABLE 4.   Descriptive data for the vocabulary and EIT proficiency tests and the pretest recognition of target vocabulary

Group	Mean	SD	95% CI
Lextale-ESP vocabulary proficiency			
Lesson + Control	−8.338	6.019	[−8.830, −7.847]
Lesson + SV	−9.623	7.521	[−10.246, −9.000]
Lesson + SG	−8.591	6.686	[−9.111, −8.068]
No Lesson + SG	−7.253	6.412	[−7.744, −6.762]
Elicited imitation task Spanish proficiency			
Lesson + Control	66.515	19.638	[61.850, 71.178]
Lesson + SV	68.543	20.923	[63.640, 73.445]
Lesson + SG	68.819	18.102	[64.638, 73.000]
No Lesson + SG	67.908	22.430	[62.865, 72.950]

Note: SV = Salience on Vocabulary; SG = Salience on Grammar.

The vocabulary-proficiency test included 46 words that were used as experimental items in this study (23 target-vocabulary words and 23 foils). These 46 items were removed from the scoring of the proficiency test to separately assess learners' prior knowledge of these words.

VOCABULARY

Recognition

The vocabulary-recognition data are plotted in the left-hand panel of Figure 1 (see Supplementary Materials; Table B.2). The data pattern suggests an advantage of captioning over noncaptioned video, with all captioning groups scoring higher than the Lesson + Control group. The results also suggest an overall advantage for the Lesson + SV participants over the Lesson + Control and the two Grammar groups (i.e., Lesson + SG and No Lesson + SG).

The first iteration of the generalized linear model, with the Lesson + Control group as the reference level, revealed significant positive group effects for the Lesson + SV group ($\beta =$ 1.286, SE = 0.075, $p < 0.001$, Exp(B) = 3.619, 95%CI[3.127, 4.193]), the Lesson + SG group ($\beta = 0.755$, SE = 0.067, $p < 0.001$, Exp(B) = 2.127, 95%CI[1.864, 2.427]), and the No Lesson + SG group ($\beta = 0.756$, SE = 0.066, $p < 0.001$, Exp(B) = 2.129, 95%CI[1.871, 2.422]). Thus, all captioned groups were more accurate in their recognition than the control learners. The second iteration, with Lesson + SV as the reference level, revealed a significant negative effect in all comparisons: Lesson + SG group ($\beta = -0.532$, SE = 0.071, $p < 0.001$, Exp(B) = 0.588, 95%CI [0.511, 0.675]); and No Lesson + SG group ($\beta = -0.531$, SE = 0.070, $p < 0.001$, Exp(B) = 0.588, 95%CI[0.512, 0.675]). Thus, there was an overall advantage of the Lesson + SV group in their recognition accuracy.
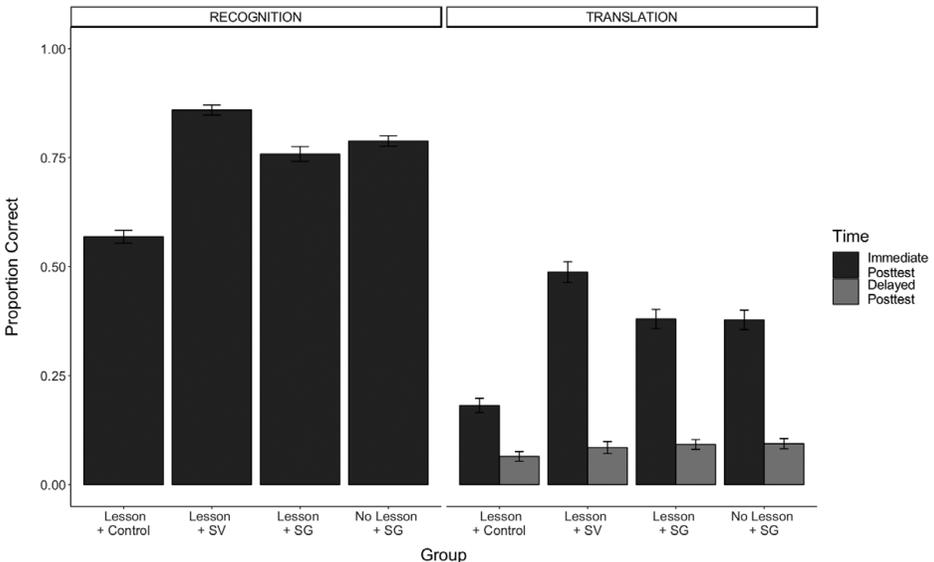
FIGURE 1.   Mean Accuracy Scores for vocabulary recognition and translation. Error bars are two standard errors long. SV = Salience on Vocabulary; SG = Salience on Grammar.

Translation

Immediate Posttest. As with the vocabulary-recognition results, the data pattern for the translation scores suggests an advantage of captioning over noncaptioned video, as well as an overall advantage for the Lesson + SV group over the Control and Grammar groups (see the right-hand panel of Figure 1 and Supplementary Materials; Table B.2). We implemented the same statistical analysis from the recognition data. The first iteration, with the Control group as the reference level, revealed a significant positive group effect for the Lesson + SV group ($\beta = 1.528$, SE = 0.099, $p < 0.001$, Exp(B) = 4.608, 95%CI [3.801, 5.602]); the Lesson + SG group ($\beta = 1.067$, SE = 0.098, $p < 0.001$, Exp(B) = 2.912, 95%CI[2.405, 3.538]); and the No Lesson + SG group ($\beta = 1.102$, SE = 0.098, $p < 0.001$, Exp(B) = 3.010, 95%CI[2.488, 3.655]). In other words, all captioned groups were more accurate in their production. The second iteration, with the Lesson + SV group as the reference level, revealed a significant negative group effect for all comparisons: Lesson + SG group ($\beta = -0.459$, SE = 0.079, $p < 0.001$, Exp(B) = 0.632, 95%CI[0.541, 0.737]); and the No Lesson + SG group ($\beta = -0.426$, SE = 0.079, $p < 0.001$, Exp(B) = 0.653, 95%CI[0.559, 0.762]). These results confirm our initial observation of the overall advantage of the Lesson + SV group in their translation accuracy.

Delayed Posttest. Similar to the immediate posttest, the pattern for the delayed posttest suggests an advantage of captioning over noncaptioned video, with all captioning groups scoring higher than the no-captions Control group. However, there is no longer an apparent advantage for the Lesson + SV group over the Grammar groups (see the right-hand panel of Figure 1 and Supplementary Materials; Table B.3). The first model iteration, with the Lesson + Control group as the reference level, revealed significant positive group effects, for the Lesson + SV group ($\beta = 0.464$, SE = 0.207, $p < 0.05$, Exp(B) = 1.590, 95%CI[1.046, 2.399]); the Lesson + SG group ($\beta = 0.488$, SE = 0.203, $p < 0.05$, Exp(B) = 1.629, 95%CI[1.099, 2.439]); and the No Lesson + SG group ($\beta = 0.563$, SE = 0.195, $p < 0.01$, Exp(B) = 1.756, 95%CI[1.206, 2.597]). In other words, all captioned groups were more accurate in their translation accuracy than the control group. The second iteration, with Lesson + SV as the reference level, did not reveal significant effects for the Salience + Grammar groups: Lesson + SG group ($\beta = 0.024$, SE = 0.181, $p = 0.895$, Exp(B) = 1.024, 95%CI[0.719, 1.461]); and No Lesson + SG group ($\beta = 0.099$, SE = 0.175, $p = 0.571$, Exp(B) = 1.104, 95%CI[0.784, 1.561]).

Thus, after 2 weeks, there was a sustained (albeit slight) advantage of the captioned-Vocabulary group when compared against the no-captions Control group, but not against the two captioned-Grammar groups. This suggests that the initial advantage of TE-on-vocabulary over TE-on-grammar was lost at the delayed posttest.

GRAMMAR: COMPARING EXPLICIT-GRAMMAR INSTRUCTION GROUPS

Figure 2 illustrates the group mean scores as well as the standard errors by structure for the grammar pretest, the immediate posttests, and the delayed posttests (see also Supplementary Materials; Tables B.4–B.6). The data pattern shows similar effects across structures, whereby all groups display an increase in their immediate posttest accuracy scores (when compared to their respective pretest scores) but no obvious differences between groups at
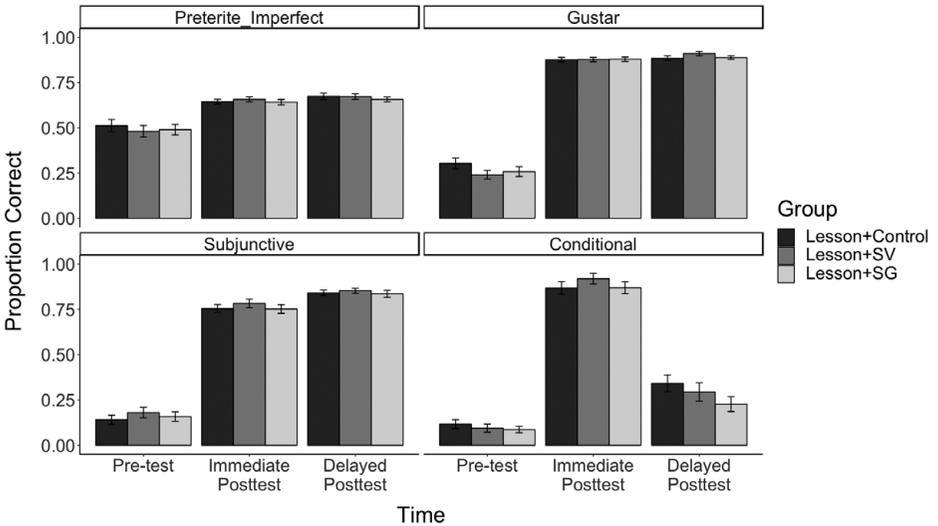
FIGURE 2.   Mean accuracy scores for grammar translation by structure, group, and time (lesson groups). Error bars are two standard errors long.

immediate posttest or at delayed posttest (see also Supplementary Materials; Table B.4). In the analyses that follow, we focus on group gains from pretest to immediate posttest, and from pretest to delayed posttest.

Immediate Posttest

The generalized linear mixed-effects model included the no-captions Control group and the preterite/imperfect structure as reference levels. We used the emmeans package (Length, 2018) to run pairwise Tukey tests comparing pretest/immediate-posttest gains by group within each structure.

Preterite/Imperfect.   The model returned a significant effect in group gains between the Lesson + Control and Lesson + SV groups, $\beta = 0.270$, SE = 0.126, $p = 0.033$; and nonsignificant differences between the Lesson + Control and Lesson + SG groups, $\beta = 0.228$, SE = 0.126, $p = 0.070$; and between the Lesson + SV and Lesson + SG groups, $\beta = 0.041$, SE = 0.123, $p = 0.735$. To summarize, only the Lesson + SV group led to greater translation accuracy from pretest to immediate posttest (compared to the Lesson + Control group), with no significant differences between the Lesson + SV and Lesson + SG groups.

Gustar-Type Verbs.   The models for the *gustar*-type verbs returned a significant difference between the Lesson + Control and Lesson + SG groups, $\beta = 0.418$, SE = 0.205, $p = 0.041$, but not between the Lesson + Control and Lesson + SV groups, $\beta = 0.309$, SE = 0.205, $p = 0.132$; or the Lesson + SV and Lesson + SG groups, $\beta = -0.109$, SE = 0.208, $p = 0.599$. Thus, only the Lesson + SG displayed greater translation accuracy from pretest to immediate posttest.

Subjunctive in Noun Clauses. The results for the subjunctive did not show significant differences in group gains between the Lesson + Control and Lesson + SV groups, $\beta = -0.259$, SE = 0.194, $p = 0.204$; the Lesson + Control and Lesson + SG groups, $\beta = -0.226$, SE = 0.195, $p < 0.05$; or the Lesson + SV and Lesson + SG groups, $\beta = 0.021$, SE = 0.189, $p = 0.912$. Thus, all groups displayed similar developmental patterns between pretest and immediate posttest.

Conditional Mood. The results for the conditional mood revealed a significant effect in group gains between the Lesson + Control and Lesson + SV groups, $\beta = 0.838$, SE = 0.387, $p < 0.05$; a nonsignificant effect between the Lesson + Control and Lesson + SG groups, $\beta = 0.252$, SE = 0.352, $p = 0.474$; and a nonsignificant effect between the Lesson + SV and Lesson + SG groups, $\beta = 0.586$, SE = 0.390, $p = 0.133$. Thus, only the Lesson + SV group displayed greater translation accuracy from pretest to immediate posttest when compared to the Lesson + Control group.

To summarize the immediate-posttest results, we uncovered significant between-group differences in favor of the captioning groups, but the effects were inconsistent regarding which TE format was more beneficial when combined with video captions. Specifically, there was an advantage of Lesson + SV for the preterite/imperfect and the conditional, but an advantage of Lesson + SG for *gustar*-type verbs.

Delayed Posttest

The results from the pairwise comparisons, comparing pretest and delayed posttest, revealed a significant effect in group gains between the Lesson + Control and Lesson + SV groups for *gustar*-type verbs only, $\beta = 0.630$, SE = 0.210, $p < 0.001$ (see Supplementary Materials; Table B.7). No other differences returned significant effects. Altogether, the implication is that any gains from initial pretest were lost at the delayed posttest (recall that for *gustar*-type verbs, there was an effect of Lesson + SG at immediate posttest).

COMPARING LESSON + TE CAPTIONS ON GRAMMAR VERSUS NO LESSON + TE CAPTIONS ON GRAMMAR

To tease apart the individual effects of captioning from the presentation of the initial grammar lesson, we compared the Lesson + SG group (who saw the grammar-lesson video before the multimodal-captioned video) to the No Lesson + SG group (who did not see the grammar-lesson video).

Figure 3 illustrates the group mean scores as well as the standard errors by structure for the grammar pretest, the immediate posttests, and the delayed posttests (see also Supplementary Materials; Tables B.4–B.6). The results display a general pattern whereby all groups show an increase in their accuracy when compared to their corresponding pretest scores. Taking a closer look at the data, from pretest to immediate posttest, the explicit Lesson + SG group shows a slight advantage over the No Lesson + SG group for the preterite/imperfect and for *gustar*-type verbs, a considerable advantage for the conditional mood, but no advantage for the subjunctive. However, any between-group differences do not appear to hold by the delayed posttest.
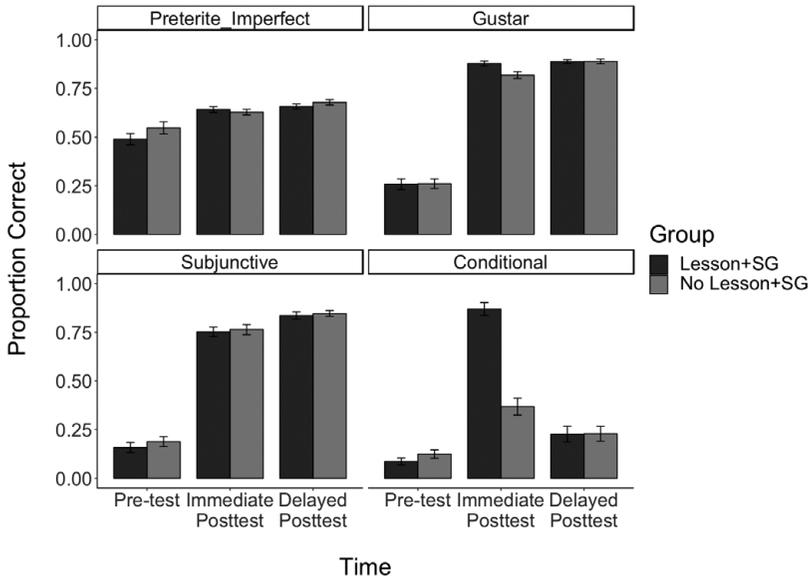
FIGURE 3.　Mean accuracy scores for grammar translation by structure, group, and time (Lesson + SG versus No Lesson + SG). Error bars are 2 standard errors long.

Immediate Posttest

The generalized linear mixed-effects model, using emmeans, included the Lesson + SG group and the preterite/imperfect structure as reference levels. The results revealed a significant effect in gains between the Lesson + SG and the No Lesson + SG groups for the preterite/imperfect, $\beta = -0.407$, SE $= 0.120$, $p = 0.001$; *gustar-type* verbs, $\beta = -0.689$, SE $= 0.199$, $p = 0.001$; and the conditional, $\beta = -3.020$, SE $= 0.314$, $p < 0.001$; but not for the subjunctive in noun clauses, $\beta = -0.238$, SE $= 0.182$, $p = 0.192$. These results are consistent with our initial observations, whereby the Lesson + SG group showed a greater advantage for all structures except the subjunctive.

Delayed Posttest

We did not uncover significant between-group differences in pretest versus delayed posttest group gains for any of the grammar structures (see Supplementary Materials; Table B.8). Thus, one important outcome of this study is that, although we uncovered positive effects of TE and captioning on the immediate posttests, the treatments did not lead to sizeable gains in terms of long-term effects. We address this discrepancy in greater detail in the "Discussion" section.

DISCUSSION

The goal of our experimental study was to investigate the role of TE captions in the learning of L2 vocabulary and grammar. Our research into the effect of captioning on

L2-vocabulary learning aligns with the need for replication studies in L2 research (Marsden et al., 2018). Our inquiry into the effect of captioned media on L2 grammar improves upon previous research by providing learners with multimodal input designed with specialized software and novel plots designed around each target structure. Our delayed-posttest design responded to the pressing need for achieving external validity of FFI research by examining the durability of instruction effects (Ellis, 2012; Lee & Huang, 2008). Our methodology additionally improved on previous work by directly investigating the effects of multimodal TE-captioned video with and without explicit instruction (cf. Cintron-Valentin et al., 2019; Lee & Révész, 2018).

RESEARCH QUESTION 1: VOCABULARY

The first aim of this study was to examine the effects of full captions + vocabulary TE on improving learner knowledge of target vocabulary. RQs 1a and 1b considered the relative effect of full captions + vocabulary TE, full captions + grammar TE, or no TE, on the recognition and production of L2-Spanish target vocabulary. The results showed robust, positive effects of captioning and of highlighting with TE on enhancing learner knowledge of vocabulary. Specifically, the vocabulary recognition and production results show that learners in all three captioning groups (Lesson + SV; Lesson + SG; No Lesson + SG) were more successful than noncaptioned control learners in improving their vocabulary knowledge.

RQ 1c asked if any initial gains on the production of vocabulary would be maintained over time. We tested participants' abilities to translate the target vocabulary words approximately 2 weeks after each lab session. Across all experimental groups, there was a noticeable reduction in learners' ability to produce the vocabulary words between the immediate and the delayed posttest. There was also an advantage for each captioned group (Lesson + SV; Lesson + SG; No Lesson + SG) against the Lesson + Control group, but no significant differences between the captioned groups. Overall, the findings of the immediate posttest support previous research demonstrating the role of captioning in promoting L2-vocabulary knowledge (e.g., Montero Perez et al., 2013). One additional illuminating outcome of our study is that we did not find any evidence in support of long-term retention patterns (see Neuman & Koskinen, 1992).

There are several possible explanations for the lack of robust retention effects. First, the target vocabulary selected for this experiment was of low frequency (to control for learner familiarity of the target vocabulary). Within L2 acquisition, vocabulary size is largely dependent on the relative frequency with which items are encountered in the input (Nation, 2006). Additionally, although the current design provided learners with frequent and meaningful encounters with the target words during the multimodal videos, the learners were not explicitly encouraged to use these words throughout the semester (cf. Pujadas & Muñoz, 2019). It is thus possible that the lack of additional opportunities to revisit the target vocabulary, in addition to the low frequency of the items, contributed to learners' reduced ability to produce them at the delayed posttest (see Webb & Nation, 2017, p. 63).

RESEARCH QUESTION 2: GRAMMAR

Our second research aim was to examine the effects of full captions + grammar TE on improving learner knowledge of target grammar. RQ 2a examined the relative effect of

full captions + vocabulary TE, full captions + grammar TE, or no TE on the production of target grammar. Based on the results of the translation task, captioned videos—either on vocabulary or grammar—showed an advantage over noncaptioned videos. However, this advantage was obtained for some, but not all, target structures. RQ 2b asked whether any initial gains on grammar were maintained over time. The delayed posttest revealed a significant difference in group gains between the Lesson + SV and the Lesson + Control groups for *gustar*-type verbs only.

We believe a combination of methodological and structure-specific factors could help explain our mixed findings on grammar (cf. Cintrón-Valentín et al., 2019). In the following subsections, we focus on the effects uncovered for each structure and consider the factors that may have impacted their saliency in the input.

Preterite/Imperfect

For the preterite/imperfect, we uncovered significant positive effects for Lesson + SV at immediate posttest but not at delayed posttest. An additional important finding is that all groups appeared to have more baseline knowledge of the preterite/imperfect than of the other structures included in this study (see Figure 3). Yet, this initial advantage did not result in greater learning gains following the captioning intervention. One possible explanation for the small gains observed for the preterite/imperfect may relate to the amount of structures being targeted during a single lab session. Regarding this possibility, Overstreet (1998) suggests that the lack of a TE effect on the acquisition of the preterite/ imperfect may be due to the difficulty of learning how two forms contrast within a specific semantic context. Specifically, the added TE on this structure may have distracted learners' attention from the surrounding discourse, which offers critical information regarding how the two aspectual choices are used (cf. Bardovi-Harlig, 1998). This could explain the positive effect for the Lesson + SV group, who received captions that did not include highlighting on morphological forms. Given the importance of the surrounding discourse in understanding how such forms are used in context (Bardovi-Harlig, 2000), it would be beneficial for future work to investigate if increasing the sources of explicit information at more strategic points during the captioned media would lead to more robust learning outcomes.

Gustar-Type Verbs

For *gustar*-type verbs, we uncovered positive significant effects for Lesson + SG at immediate posttest and for Lesson + SV at delayed posttest. This outcome suggests that learner knowledge of subject-verb agreement can be supported by TE + multimodal captioned media (see also Cintrón-Valentín et al., 2019). Notably, this was the only structure for which we found positive effects of captions + grammar TE on improving learner knowledge. We believe that our findings for *gustar*-type verbs may relate chiefly to learners focusing on the lexical learning of the verbs in question, rather than on more detailed grammar points. Differently from the other target forms, the *gustar*-type structure requires learning fewer inflectional endings than other grammar forms, and instead relies on (a) understanding the noncanonical mapping of thematic roles and (b) learning the particular lexical forms used in the construction. Within our study, the goal of the

grammar lesson was to provide learners with a general understanding of how the *gustar*-type structure works. During the experiment it is thus possible that learners reanalyzed the linguistic focus of the task such that they focused on the set of verbs that are unique to the *gustar*-type construction.

Subjunctive in Noun Clauses

The results for the subjunctive did not reveal significant learning differences between groups at either of the two posttest times. That is, although all groups showed a notable increase in their ability to produce the subjunctive from pretest to immediate posttest, they appeared to be performing at the same level at both posttest times. These results were unexpected given the findings of Cintrón-Valentín et al. (2019), who reported positive effects of captioning relative to the noncaptioned control condition for the same structure. The question that immediately arises is why the two studies yielded apparent contradictory outcomes. The divergent findings could relate to the different grammar TE manipulations implemented in each study. In Cintrón-Valentín et al. (2019), the main clause and the subordinate subjunctive verbs were highlighted in bold and yellow. Further, the current design incorporated an arrow indicating the relationship between the main clause and the subordinated subjunctive verbs (per Collentine, 2013) and also added color to the subjunctive clause for differentiation. It is thus possible that, given the short presentation time of the captions, the added TE may have served as a distraction to the Lesson + SG group, hence learners' similar performance across groups.

Conditional Mood

The findings for the conditional revealed a significant difference in learning gains from pretest to immediate posttest between the Lesson + SV and Lesson + Control groups. For the delayed posttest, all groups showed learning gains, but unlike the other structures, there was a notable drop in learner performance. As mentioned previously, conditional sentences are highly complex structures for both L1 and L2 acquisition due to both their morphosyntactic complexity and the semantic nuance involved in learners' processing of this form (e.g., López Ornat, 1994). In addition, we targeted a low-frequency usage of the conditional whose analysis is largely dependent on the surrounding discourse. In our study, learners used contextual observations to (a) understand how the structure works from the presentation of the animated video and (b) provide the appropriate tense in the translation instrument. Similar to the preterite/imperfect, it is possible that TE on the grammatical forms might have distracted learners' attention from the key surrounding discourse. This might explain the slight advantage of the Lesson + SV group, whose TE manipulation only included highlighting of the target vocabulary, which never appeared in the same sentential contexts as the target grammar.

Altogether, our findings for the delayed posttest suggest a lack of sustained treatment effects. This outcome is consonant with prior findings in the grammar-learning literature, which report that significant short-term effects from grammar interventions are often diminished by the point of a delayed posttest (see for instance, Lightbown et al., 1980; Norris & Ortega, 2000). In their meta-analysis, Norris and Ortega (2000) conclude that such longitudinal declines could be due to "a loss of instructional effect on the part of

treatment groups and some degree of maturation on the part of control or comparison groups" (p. 478). In our study, both explanations are likely relevant, although to a different degree based on the target structure in question. For the conditional, it is worth noting that the grammar rule that we tested was not included in the learners' course curriculum, thus there may have been a substantial loss-of-instruction effect. For the remaining structures for which significant effects of full captions + grammar TE were found at immediate posttest (i.e., *gustar*-type verbs), it is likely that both factors mentioned by Norris and Ortega were responsible for the lack of sustained effects. This is because, for *gustar*-type verbs, there was not a substantial drop in Lesson + SG performance at the delayed posttest; rather, all groups slightly increased from immediate posttest to delayed posttest, likely due to maturation effects.

At the same time, it is difficult to assess any degree of maturation that occurred as a consequence of exposure to the target structures between the individual data-collection sessions. Although our study aimed to achieve high ecological validity through its classroom design, the experimental nature of our materials did not allow us to probe the treatment of the target structures in the day-to-day curriculum. Further research into the source of immediate, but not sustained, effects of grammar interventions would do well to disentangle any confounding effects of TE and instructional design on long-term grammar learning (see also Truscott, 2014).

RESEARCH QUESTION 3: COMPARING THE LESSON VERSUS NO-LESSON GRAMMAR GROUPS

Our third research aim was to examine whether the effects of full captions + grammar TE were equally facilitative in the absence of explicit instruction. RQ 3a considered the effects of explicit instruction, in addition to full captions + grammar TE, on improving learner production of grammar, and RQ 3b considered if any initial gains are maintained over time. At immediate posttest, the Lesson + grammar TE group showed a significant advantage for all structures except the subjunctive; however, any between-group differences were lost by the delayed posttest.

The advantage of the Lesson + grammar TE group over the No Lesson + grammar TE group at immediate posttest is not unexpected. In their meta-analysis of the effects of grammar instruction, Norris and Ortega (2000) showed that learners who received explicit types of L2 instruction outperformed learners who received implicit types, with stronger effect sizes reported at immediate posttest and substantial declines reported for all treatment groups at delayed posttest (see also, Truscott, 2004). As Truscott (2014) points out, explicit conceptual grammar knowledge is most active directly after instruction, leading to enhanced performance on target forms immediately following a grammar intervention. For long-term acquisition, however, the collective findings suggest that the nature of the form in question might determine the degree of instructional support required for successful acquisition (see also Spada & Tomita, 2010; Tolentino & Tokowicz, 2014). Our data support this idea, specifically with regard to the large between-group differences for the conditional mood, a structure for which learners did not have much prior knowledge and for which explicit instruction proved necessary even at the immediate posttest. For all other target structures, prior knowledge of the form-meaning mappings likely aided throughout the experimental sessions (even though learners generally

performed below chance at pretest). An important implication from this study, therefore, is that grammar development may be impacted by two interrelated factors: the target structure's frequency of usage, and learners' prior experience with the form in question (see Larsen-Freeman, 2009).

BROADER IMPLICATIONS

To motivate the need for more dynamic approaches to grammar teaching, Larsen-Freeman (2003) calls for an increased implementation of "grammaring." In short, grammaring is a pedagogical strategy in which students practice grammar use in situations that are analogous to those that they will encounter outside of the classroom. Of importance, grammaring requires that instructors tailor classroom practices to the nature of the learning challenge posed by a given grammar rule (Larsen-Freeman, 2009; p. 527). Although some structures might require little pedagogical intervention, others impose challenges due to complex morphology, meaning, or contextual use. For the four structures tested here, *gustar*-type verbs pose challenges due to their morpho-syntactic construction, whereas the preterite/imperfect, subjunctive, and conditional pose challenges in large part due to their use in discourse. In such cases, Larsen-Freeman (2009) argues that students must be placed in situations that force them to decide between the two forms contextually, even though they might have similar surface-level meanings.

Building on these ideas, we do not believe that our findings will necessarily be applicable to all linguistic constructions. One of the strengths of our study was the inclusion of a diverse set of grammar structures for which we uncovered varying effects of FFI and captioned media. As discussed previously, for a structure such as the conditional, it is possible that if learners are given limited support in what to pay attention to in the input, their attentional processes may still not be fully directed to the target feature, even if there are abundant examples of it in the text. This aligns well with the FFI literature that shows that different forms require different levels of explicitness and explanation (e.g., Spada & Tomita, 2010).

Adding to this point, it is unclear whether the results uncovered here would extend to other L2s. To our knowledge, only one study in the captioning literature has investigated the effects of captioning on the learning of multiple target languages. In that study, Winke et al. (2010) showed trends whereby captioning viewing was generally less beneficial for target languages with a greater orthographic distance from a learner's native language. Winke et al. (2010) suggest that for cases such as these, there may be a greater "reliance on listening because the written symbols are not well learned" (p. 80).

LIMITATIONS AND FUTURE DIRECTIONS

One limitation that warrants exploration in future research is the lack of additional comparison groups that did not receive explicit instruction. Specifically, we were not able to include a No Lesson + No Captions group or a No Lesson + unenhanced captions group due to issues of power (i.e., due to the limited number of grammar courses to which we had access). The inclusion of such groups would allow for more definitive conclusions regarding the effects of TE-captioned media on the structures in question. At the same time, the inclusion of a direct comparison between enhanced versus unenhanced

experimental conditions (i.e., captions without TE vs. captions with TE) would be advantageous in understanding unique contributions of TE in facilitating learner acquisition of the target grammatical forms (e.g., Leow & Martin, 2017).

A further limitation was the lack of additional outcome measures to assess the effects of our treatments. It is possible that the inclusion of more receptive measures of grammar competence would have resulted in different outcomes (e.g., Lee & Révész, 2018). Meta-analyses of the effects of instruction, for instance, demonstrate that the effectiveness of techniques vary as a result of explicitness of measure (Norris & Ortega, 2000).

An additional consideration for future research is the role of prior knowledge and its influence on the recognition and production of target structures. Studies that probe learners' prior knowledge in more detailed ways would allow researchers to gain insight into the degree of exposure needed for successful captions + TE interventions. In the current study, we surmised the role of prior knowledge to be relevant in the testing of the conditional mood. However, there were likely more nuanced differences within the three other target structures that we were unable to separate. Along these lines, we believe it is critical for future research to gauge how a learner's proficiency level may affect their ability to focus on textually enhanced forms in the input through FFI interventions (Lee & Huang, 2008).

CONCLUSION

The current study examined the role of textually enhanced captions on the learning of vocabulary and grammar in L2 Spanish. One key contribution of our design was the integration of principles derived from FFI into the elaboration of innovative multimodal research materials. For vocabulary, our findings replicate those of previous research demonstrating that captioning is reliably effective for vocabulary learning; at the same time, we have suggested that long-term effects (i.e., through our delayed posttest) are not as stable for low-frequency items. For grammar, TE captions, either on target vocabulary or grammar, led to immediate positive effects on production abilities for some structures (i.e., *gustar*-type verbs, the preterite/imperfect, and the conditional), but not others (i.e., the subjunctive). The findings on grammar contribute to the limited body of research on this topic by showing that multimodal pedagogical interventions can, in fact, lead to significant improvement in learners' production (even in the absence of explicit grammar instruction). Critically, future research is needed to understand the lack of sizeable long-term gains on grammar learning.

Altogether, through the type of research conducted here, we are beginning to understand the array of factors that can have an impact on the effectiveness of multimodal research designs, such as the frequency of word exposure, the morpho-syntactic relations of a grammar structure in question, the surrounding discourse, learners' prior knowledge of a target structure, and the degree of instructional support. Although we were not able to assess these effects definitively, future research would do well to scrutinize any of these factors in greater detail.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263120000492.

NOTES

[1]Captioned video in this study refers to video that includes subtitles in the same language as the audio (Jung, 1990).

[2]Learners' difficulty with, and minimal knowledge of, the targeted conditional construction was confirmed through a small pilot study that included 31 learners.

[3]The No Lesson + Salience on the grammatical features group (No Lesson + SG) received the same type of TE as the Lesson + SG group. For ethical reasons, this group also received the grammar lesson, but *after* completing all the study questions at the end of each lab session.

[4]As noted in Hudson and Lorena (2015), a limitation in some studies within SLA quantitative research is the lack of attention to "how participants are allocated to different groups or conditions" (p. 86). In our study we account for this methodological consideration by implementing a randomized control design.

REFERENCES

Areizaga Orube, E. (2009). *Gramática para profesores de español como lengua extranjera*. Ediciones Díaz de Santos.

Bardovi-Harlig, K. (1998). Narrative structure and lexical aspect: Conspiring factors in second language acquisition of tense-aspect morphology. *Studies in Second Language Acquisition*, *20*, 471–508.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Blackwell.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Blake, R. J. (2013). *Brave new digital classroom: Technology and foreign language learning*. Georgetown University Press.

Blyth, C. (2005). From empirical findings to the teaching of aspectual distinctions. In D. Ayoun & R. Salaberry (Eds.), *Tense and aspect in romance languages* (pp. 211–252). John Benjamins.

Bowden, H. W. (2016). Assessing second language oral proficiency for research. *Studies in Second Language Acquisition*, *38*, 647–675.

Chun, D. M., & Plass, J. L. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology*, *1*, 1–35.

Cintrón-Valentín, M. C., & Ellis, N. C. (2016). Salience in second language acquisition: Physical form, learner attention, and instructional focus. *Frontiers in Psychology*, *7*, 1284.

Cintrón-Valentín, M. C., García-Amaya, L., & Ellis, N. C. (2019). Captioning and grammar learning in the L2 Spanish classroom. *The Language Learning Journal*, *47*, 439–459.

Collentine, J. (2013). Subjunctive in second language Spanish. In K. L. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 270–288). Wiley Blackwell.

Comajoan, L. (2013). Defining and coding data: Narrative discourse grounding in L2 studies. In R. Salaberry & L. Comajoan (Eds.), *Research design and methodology in studies on L2 tense and aspect* (pp. 309–356). De Gruyter.

DeKeyser, R., & Prieto Botana, G. (2013). Acquisition of grammar by instructed learners. In K. L. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 449–465). Wiley Blackwell.

Ellis, N. C. (2006). Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*, 1–31.

Ellis, N. C. (2017). Salience in usage-based SLA. In S. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 21–39). Routledge.

Ellis, R. (2012). Form-focused instruction and second language learning. *Language teaching research and pedagogy* (pp. 271–306). Wiley-Blackwell.

Gass, S. M., Spinner, P., & Behney, J. (Eds.). (2017). *Salience in second language acquisition*. Routledge.

Goldschneider, J. M., & DeKeyser, R. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1–50.

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, *45*, 765–771.

Gudmestad, A. (2012). Acquiring a variable structure: An interlanguage analysis of second-language mood use in Spanish. *Language Learning*, *62*, 373–402.

Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, *29*, 597–618.

Hudson, T., & Lorena, L. (2015). Design issues and inference in experimental L2 research. In J. M. Norris, S. J. Ross, & J. Schoonen (Eds.), *Improving and extending quantitative reasoning in second language research* (pp. 76–96). Wiley.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-ESP: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*, 49–66.

Jones, L., & Plass, J. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal*, *86*, 546–561.

Jung, U. O. (1990). The challenge of broadcast videotex to applied linguistics. *IRAL—International Review of Applied Linguistics in Language Teaching*, *28*, 201–220.

Larsen-Freeman, D. (2003). *Teaching language: From grammar to grammaring.* Thomson/Heinle.

Larsen-Freeman, D. (2009). Teaching and testing grammar. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 518–542). Blackwell.

Lee, J. F., & Malovrh, P. A. (2009). Linguistic and non-linguistic factors affecting OVS processing of accusative and dative case pronouns by advanced L2 learners of Spanish. In J. Collentine, M. García, B. Lafford, & F. Marcos Marín (Eds.), *Selected proceedings of the 11th Hispanic Linguistics Symposium* (pp. 105–116). Cascadilla Proceedings Project.

Lee, M., & Révész, A. (2018). Promoting grammatical development through textually enhanced captions: An eye-tracking study. *The Modern Language Journal*, *102*, 557–577.

Lee, S. K., & Huang, H. T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, *30*, 307–331.

Length, R. (2018). *Emmeans: Estimated marginal means, aka least-squares means // R package version 1.2.3.* https://CRAN.R-project.org/package=emmeans

Leow, R. P., & Martin, A. (2017). Enhancing the input to promote salience of the L2: A critical overview. In S. Gass, P. Spinner, & J. Behney (Eds.), *Salience in second language acquisition* (pp. 167–186). Routledge.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*, 673–680.

Lightbown, P., Spada, N., & Wallace, R. (1980). Some effects of instruction on child and adolescent ESL learners. In R. C. Scarcella & S. D. Krashen (Eds.), *Research in second language acquisition* (pp. 162–172). Newbury House.

Liskin-Gasparro, J. (2000). The use of tense-aspect morphology in Spanish oral narratives: Exploring the perceptions of advanced learners. *Hispania*, *83*, 830–844.

López Ornat, S. (1994*). La adquisición de la lengua española.* Siglo XXI de España Editores.

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews, and recommendations for the field. *Language Learning*, *68*, 321–391.

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*, 118–141.

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned Video: An eye-tracking study. *Modern Language Journal*, *99*, 308–328.

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*, 720–739.

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, *67*, 77–86.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59–82.

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly*, *27*, 94–106.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.

Ortega, L., Iwashita, N., Rabie, S., & Norris, J. M. (1999). *A multilanguage comparison of measures of syntactic complexity [Funded project].* University of Hawai'i, National Foreign Language Resource Center.

Overstreet, M. (1998). Text enhancement and content familiarity: The focus of learner attention. *Spanish Applied Linguistics*, *2*, 229–258.

Plass, J. L., & Jones, L. C. (2005). Multimedia learning in second language acquisition. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 467–488). Cambridge University Press.

Pujadas, G., & Muñoz, C. (2019). Extensive viewing of captioned and subtitled TV series: A study of L2 vocabulary learning by adolescents. *The Language Learning Journal*, *47*, 479–496.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*, 309–338.

RStudio Team. (2015). *RStudio: Integrated Development for R*. RStudio, Inc.: Boston, MA. http://www.rstudio.com/.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*, 165–179.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*, 263–308.

Terrell, T. (1991). The role of grammar instruction in a communicative approach. *Modern Language Journal*, *75*, 52–63.

Tolentino, L. C., & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Language Learning*, *64*, 279–309.

Truscott, J. (2004). The effectiveness of grammar instruction: Analysis of a meta-analysis. *English Teaching and Learning*, *28*, 17–29.

Truscott, J. (2014). Conclusion: Consciousness in second language learning. In D. Singleton (Ed.), *Consciousness and second language learning* (pp. 231–238). Multilingual Matters.

Truscott, J., & Sharwood Smith, M. (2011). Input, intake, and consciousness: The quest for a theoretical foundation. *Studies in Second Language Acquisition*, *33*, 497–528.

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, *43*, 1–37.

Vanderplank, R. (2016). "Effects of" and "effects with" captions. How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, *49*, 235–250.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Ablex.

Vázquez Rozas, V. (2006). Gustar-type verbs. In J. Clancy Clements & J. Yoon (Eds.), *Functional approaches to Spanish syntax. Lexical semantics, discourse and transitivity* (pp. 80–114). Palgrave MacMillan.

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Winke, P. M., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*, 65–86.

Winke, P. M., Sydorenko, T., & Gass, S. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *Modern Language Journal*, *97*, 254–275.