CrossMark

Studies in Second Language Acquisition **42** (2020), 625–651
doi:10.1017/S0272263120000108

Research Article

PROMOTING GRAMMATICAL DEVELOPMENT THROUGH CAPTIONS AND TEXTUAL ENHANCEMENT IN MULTIMODAL INPUT-BASED TASKS

Minjin Lee *

Ewha Womans University

Andrea Révész

University College London

Abstract

This study assessed the extent to which captions, textually unenhanced and enhanced, can draw learners' attention to and promote the acquisition of a second language (L2) grammatical construction. A pretest–posttest–delayed posttest experimental design was employed. Seventy-two Korean learners of English were randomly assigned to an enhanced captions group, an unenhanced captions group, and a no captions group. Each group completed a series of treatment tasks, during which they watched news clips under their respective captioning condition. The target L2 construction was the use of the present perfect versus the past simple in reporting news. For the enhanced captions group, the present perfect and past simple forms were typographically enhanced using a different color. Eye-movement indices were obtained to examine attentional allocation during the treatment, and oral and written productive tests and a fill-in-the-blank test were used to assess participants' gains. A series of mixed-effects models found both captioning and textual enhancement effective in drawing learners' attention to and facilitating development in the use of the target construction. In addition, positive links were identified between attention to captions and learners' gains.

We would like to thank Aline Godfroid, Ana Pellicer-Sánchez, Carmen Muñoz, and Elke Peters for their insightful comments on previous versions of this manuscript. This research was supported in part by the International Research Foundation for English Language Education (TIRF) Doctoral Dissertation Grant and the Language Learning Dissertation Grant.

The experiment in this article earned an Open Materials badge for transparent practices. The materials are available at https://www.iris-database.org/iris/app/home/detail?id=york:937639

* Correspondence concerning this article should be addressed to Minjin Lee, Department of English Education, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea. E-mail: mlee.minjin@gmail.com

© *The Author(s) 2020. Published by Cambridge University Press*

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

INTRODUCTION

With task-based language teaching (TBLT) gaining prominence in both the fields of instructed second language acquisition and L2 pedagogy (e.g., Bygate et al., 2001; Ellis, 2003; Samuda & Bygate, 2008), the construct of task has been the subject of a growing amount of L2 research in recent years. Tasks are defined as activities "where meaning is primary; there is some communicative problem to solve; some sort of relationship with real-world activities; and the assessment of task is in terms of a task outcome" (Skehan, 1998, p. 95). Interest in tasks has been motivated by the fact that carrying out communicative tasks prepares learners for real-life activities and engages psycholinguistic processes that are thought to be beneficial for L2 learning (Long, 2000). Among the various dimensions along which tasks can be categorized, a key distinction is between output-based and input-based tasks. Output-based tasks require language learners to engage in production, either speaking or writing; whereas input-based tasks do not require learners to produce output (Ellis, 2013; Shintani, 2012). While the use of both output-based and input-based tasks is advocated in the TBLT framework (Ellis, 2009, 2013), input-based tasks have so far received comparatively little attention (Shintani, 2012). This constitutes an important gap in the TBLT literature, given that input-based tasks serve as an important source of rich and comprehensible input, which is essential to the success of second language learning (Shintani, 2016).

Input-based tasks are traditionally defined as involving either listening or reading (Ellis & Shintani, 2014). Input-based tasks, however, can also be conceptualized as multimodal entailing various modes, such as audio, written, and pictorial input. Within the TBLT framework, one way to operationalize multimodal input-based tasks is by the means of captioning, defined as adding "redundant text that matches spoken audio signals and appears in the same language as the target audio" (Vandergrift, 2007, p. 79). The role of captions in L2 comprehension and development has been the subject of much recent research, and a recent meta-analysis (Montero Perez et al., 2013) found that captions are beneficial for facilitating L2 verbal comprehension and acquisition of L2 vocabulary. So far, captions have rarely been investigated in the context of TBLT, most of the existing research has looked into the effectiveness of this technique in relation to comprehension-based activities rather than task-based work. It appears imperative to fill these gaps in instructed SLA research, as multimedia materials suitable for captioning (e.g., YouTube, DVDs, and podcasts) are more and more accessible and used by learners in both instructed and informal L2 contexts.

Against this background, the goal of this study was to assess the extent to which captions, textually enhanced and unenhanced, may promote development in L2 grammatical knowledge. Within the TBLT framework, our research is novel in that we investigated multimodal rather than unimodal input-based tasks using captioned videos. Also, few studies (e.g., Lee & Révész, 2018) have looked into the effects of captions on grammatical knowledge; most of the existing research has focused on vocabulary. Employing eye-tracking methodology, our intention was also to contribute to previous research by investigating whether attention allocated to target grammatical features is linked to L2 development (e.g., Godfroid et al., 2013), and whether this relationship may be moderated by type of captioning (Lee & Révész, 2018; Montero Perez et al., 2015).

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

BACKGROUND

CAPTIONING AND L2 DEVELOPMENT

In the field of instructed second language acquisition, much of the existing research on captioning has been concerned with the role of captions in promoting verbal comprehension (e.g., Chai & Erlam, 2008; Danan, 2004; Garza, 1991; Huang & Eskey, 2000; Rodgers & Webb, 2017; Winke et al., 2010) and acquisition of L2 vocabulary (e.g., Bird & Williams, 2002; Chai & Erlam, 2008; Danan, 1992; Markham, 1999; Markham et al., 2001; Sydorenko, 2010; Winke et al., 2010). As noted earlier, Montero Perez et al.'s (2013) meta-analysis has confirmed that captioning has a positive impact on L2 verbal comprehension and vocabulary learning. Of the 18 empirical studies included in the meta-analysis, 15 were used to estimate the effects of captioning on verbal comprehension, and 10 were involved in the analyses investigating the relationship between captioning and vocabulary development. The meta-analysis yielded a large effect size for both L2 verbal comprehension (g = .99) and vocabulary learning (g = .87).

In explaining the observed positive effects of captioning on verbal comprehension and vocabulary acquisition, researchers often referred to the assistance that captions provide in breaking down speech into words (Bird & Williams, 2002; Vanderplank, 1988). Once speech has been segmented into words, L2 users are expected to recognize words with greater ease (Bird & Williams, 2002; Markham, 1999). Word recognition, in turn, is generally regarded as a prerequisite for effective listening (Rost, 2011) as well as reading comprehension (Grabe, 2012). Increased success in word recognition is also likely to facilitate the process of identifying novel lexical items in the incoming speech and captions, and thereby foster attention to and acquisition of new lexical items (Winke et al., 2010).

It would appear that captions may also have the capability to facilitate development in the use of L2 grammatical features. As access to captions is expected to ease demands on word recognition processes, learners will probably have more attentional resources available to allocate to the grammatical features entailed in the input and, as a result, they will more likely learn the targeted grammatical constructions. To date, however, little direct evidence is available as to whether captioning may indeed promote development in L2 grammatical knowledge. A study by Lee and Révész (2018) was the first to explore the effects of different types of captions on the learning of L2 grammar (see the following text for details), but this research, in the absence of a no-captions group, provided no information about the usefulness of captions in facilitating development in the knowledge of L2 grammatical constructions. This limitation was addressed by Cintrón-Valentín et al. (2019), who used a no-captions group when investigating the effectiveness of textually enhanced captions on L2 vocabulary and grammar learning. However, in this study, the effects of captioning and textual enhancement were not isolated.

CAPTIONING, ATTENTION, AND L2 VOCABULARY DEVELOPMENT

Having established a positive relationship between captioning and L2 vocabulary development, some researchers have recently begun to seek direct evidence for the processes that may underlie the observed benefits of exposure to captioned materials. In particular, they have demonstrated a keen interest in assessing, by the means of eye-tracking, the

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

extent to which captions may have the capacity to direct learners' attention to target lexical constructions. Eye-tracking methodology is based on the assumption that the length, location, and order of an individual's eye movements reflect their attentional processes when they interact with visual information (Just & Carpenter, 1976). Thus, in studies of captioning, eye-tracking can be used to assess whether, how long, and how often learners view linguistic features included in captions.

Montero Perez et al. (2015) is one of the first studies that has investigated L2 learners' attentional processes during exposure to captioned videos. The purpose of the study was to examine whether type of captioning (full vs. keyword captioning) and test announcement (presence vs. absence of it) might influence attentional allocation to and learning of target lexis. The participants, Dutch-speaking learners of L2 French, were randomly assigned to four experimental conditions: full captioned video plus test announcement, full captioned video minus test announcement, keyword captioned video plus test announcement, and keyword captioned video minus test announcement. A form recognition, meaning recognition, meaning recall, and clip association test (assessing the ability to associate target lexis and corresponding videos) were employed to assess learners' gains in vocabulary knowledge. To assess the amount of attention that participants paid to the target words, three eye-tracking measures were used: gaze duration (i.e., the sum of fixation durations before the target word was left), an index of initial processing (Rayner, 1998); second-pass reading time (i.e., the sum of fixation durations after the target word area was left), a measure of rereading, indicating reanalysis; and total fixation duration (i.e., the sum of all fixations on the target word area). Keyword captions led to longer gaze durations and better performance on the form recognition test than full captions, and, when test announcement was present, keyword captioning also resulted in higher second-pass reading times and total fixation durations. Interestingly, however, significant associations between the eye-gaze and developmental measures were only attested for the full-captions groups. In the presence of test announcement, higher total fixation time and second-pass reading times were related to higher vocabulary gains when full captions were available. However, when learners in the full captions group were not made aware of the forthcoming test, vocabulary gains had a positive association with gaze durations, and higher second-pass reading times were linked to lower gains on the form recognition test.

The results of Montero Perez et al. (2015) overall suggest that, when the physical salience of target words is enhanced in captions, L2 learners will more likely pay attention to and learn new L2 vocabulary items. These findings are also consistent with the earlier work of Montero Perez and colleagues (Montero Perez et al., 2014), who found greater vocabulary gains under conditions where the visual salience of target lexis was enhanced. From a theoretical perspective, both of these studies confirm Sharwood Smith's (1991, 1993) proposal that making target linguistic constructions visually salient in the input will attract learners' attention and thereby promote subsequent L2 development (Sharwood Smith, 1991, 1993).

CAPTIONING, ATTENTION, AND L2 GRAMMATICAL DEVELOPMENT

Although research investigating the effects of captioning on the acquisition of L2 grammar is still scarce, some empirical studies already exist that explore how increasing the physical salience of targeted grammatical constructions in captions may influence

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

learners' attention to and/or gains in L2 grammar. Among these are the previously mentioned studies by Cintrón-Valentín et al. (2019) and Lee and Révész (2018). Cintrón-Valentín et al. (2019) examined the effects of textually enhanced captioned videos on L2 vocabulary and grammatical development. A number of grammatical constructions were targeted, including the Spanish preterite and imperfect forms, copula and *gustar*-type verbs, and the subjunctive. Participants were randomly assigned to three groups: no captions, captions with enhanced vocabulary, and captions with enhanced grammar. Recognition and production tests were employed to assess participants' gains in the target grammar and vocabulary. While textually enhanced captions clearly facilitated performance on the vocabulary tests, they only yielded an advantage for some of the targeted grammatical forms (*gustar*-type verbs, subjunctive) on the productive test. The authors interpreted this finding as suggesting that the salience of grammatical forms might have influenced the effectiveness of textually enhanced captions. The results of the study, however, need to be interpreted with caution, as no pretest was included to control for learners' prior knowledge of the targeted grammatical features. Also, as pointed out earlier, the design did not allow for teasing out the effects of textual enhancement and captioning in the absence of an unenhanced captions group.

Lee and Révész examined the separate impact of textual enhancement in captions on participants' development in the use of a grammatical feature, pronominal anaphoric reference. This study also investigated how textually enhanced captions affect attentional allocation at the targeted grammatical feature. The researchers employed a pretest–posttest experimental design, with three treatment sessions. The participants were Korean learners of L2 English, who were randomly assigned into a captions and an enhanced captions group. The captions were added to a listening activity accompanied with static images. Under the enhanced condition, both the antecedents and personal pronouns in the pronominal anaphoric reference construction were boldfaced in the captions. Learners' attention to the target antecedents and pronouns were assessed with four eye-tracking indices: first-pass reading time or gaze duration, second-pass reading duration, total fixation duration, and number of visits. Participants' gains were gauged by a written and an oral grammaticality judgment test. Textual enhanced captions, as compared to unenhanced captions, were found more successful in directing learners' attention to the anaphora antecedents and in generating gains in receptive knowledge of pronominal anaphora. Similar to Montero Perez et al. (2015), significant relationships between attention and L2 gains were only observed in the unenhanced captions group. A possible explanation for this pattern may be that participants under the enhanced condition may have differed in the amount of higher level of processing they engaged in (Godfroid, 2019; Lee & Révész, 2018; Montero Perez et al., 2015), that is, they may have differed in degree of cognitive effort, level of analysis, and intake elaboration (Leow, 2015).

Lee and Révész's (2018) findings pattern well with some of the previous research investigating the role of textual enhancement in unimodal activities. Some empirical work has found that learners paid greater attention to grammatical features under enhanced conditions (Issa & Morgan-Short, 2019; Simard & Foucambert, 2013; Winke, 2013), but other studies identified no effects of textual enhancement on attentional allocation (Indrarathne & Kormos, 2017; Issa et al., 2015; Loewen & Inceoglu, 2016). Similarly, a meta-analysis by Lee and Huang (2008) only yielded a marginal positive impact of textual enhancement on grammar learning. Factors that have been suggested to account

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

for the mixed results include differential prior knowledge (e.g., Han et al., 2008; Lee & Huang, 2008; Park, 2004; Winke, 2013) and the varied salience of different forms of textual enhancement (e.g., underlining, boldfacing) utilized in the studies (Indrarathne & Kormos, 2017). Clearly more research is needed to disentangle these relationships.

THE PRESENT STUDY

The present study builds and expands on Lee and Révész's (2018) work. As noted earlier, one limitation of Lee and Révész (2018) was the lack of inclusion of a no-captions group in the design. In the current study, besides an unenhanced captions and enhanced captions group, we added a group who were not exposed to captions. This enabled us to examine whether the provision of captions, unenhanced or enhanced, had an impact on attentional allocation and L2 development. Another improved feature of the current design is that, instead of using static images and non-task-based activities, the treatment utilized multimodal input-based tasks operationalized as video-based listening activities. Considering the putative benefits of TBLT and the fact that many language learners watch news, movies, and/or dramas to improve their L2 proficiency, investigating the use of tasks incorporating video clips was considered more valuable from a pedagogical perspective. Finally, unlike Lee and Révész (2018), we included a delayed posttest to investigate the longer-term effects of captioning, enhanced and unenhanced, on L2 grammatical development.

RESEARCH QUESTIONS

We formulated the following research questions:

1. To what extent do multimodal input-based tasks without captions, with unenhanced captions, and enhanced captions affect development in L2 grammatical knowledge?
2. To what extent do textually unenhanced versus enhanced captions in multimodal input-based tasks draw learners' attention to the target linguistic construction?
3. To what extent does learner attention allocated to the target linguistic construction relate to development in L2 grammatical knowledge? Is this relationship influenced by whether learners are exposed to unenhanced or enhanced captions?

METHODOLOGY

OVERALL DESIGN

This study employed a pretest–immediate posttest–delayed posttest experimental design. We initially recruited 93 Korean university students. From among these students, 21 participants were excluded: 4 students failed to complete the delayed posttest and 17 students' eye-movement data were not suitable for further analysis due to loss of eye-gaze movements or technical issues during recording. Seventy-two Korean university students were included in the final participant pool. They were randomly assigned into three groups: a no-captions group (n = 24), a captions group (n = 24), and an enhanced captions group (n = 24). All three groups were administered a proficiency test, a pretest, a series of treatment tasks, an immediate posttest, a delayed posttest, and an exit questionnaire. Each test included an oral production test, a written production test, and a fill-in-the-blank test.

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

TABLE 1. Descriptive statistics for task completion on experimental task

			95% CI	
	M	*SD*	*Lower*	*Upper*
No captions ($N = 24$)	20.79	2.15	19.89	21.70
Unenhanced captions ($N = 24$)	20.42	2.90	19.19	21.64
Enhanced captions ($N = 24$)	21.42	3.03	20.14	22.70

Max. score = 24

PARTICIPANTS

Of the 72 participants, 45 were female and 27 were male. They were all native speakers of Korean learning English as a foreign language. The mean age was 21.86 ($SD = 1.42$). The students' proficiency was at level C1 and above according to the Common European Framework for Reference, as determined by their total scores on the Oxford Placement Test (OPT) (see Table 1 for the descriptive statistics in the supplementary information online). A one-way ANOVA found no significant difference in the three groups' performance on either the listening, F (2, 69) = 1.23, $p = .23$, $\eta^2 = .03$, or grammar, F (2, 69) = 1.12, $p = .33$, $\eta^2 = .03$, section of the OPT.

TARGET LINGUISTIC CONSTRUCTION

The target linguistic construction was the use of the English present perfect versus the past simple to report news. In news reports, the present perfect is often used to introduce a topic, whereas subsequent details are provided using the past simple (Eastwood, 1994). Such aspectual properties are considered difficult to master if, as in the case of Korean and English, morphosemantic discrepancies exist between the first and second language (e.g., Bardovi-Harlig, 2001; Gabriele, 2009). In Korean, the past suffix can denote meanings associated with both the English past simple and present perfect; and the corresponding difference in meaning can typically be derived from either the discourse context, the time adverbial, or other time-indicating word. Korean students often use the past simple form when the present perfect is expected in English (Han & Hong, 2015).

EXPERIMENTAL TREATMENT TASK

We operationalized multimodal input-based tasks in the form of a captioned video task, incorporating audio, visual, and/or textual input. The task was contextualized in an imaginary scenario in which the participant played the role of an editor in a newsroom, whose job was to categorize news items based on their content (see Figure 1). As part of the task, participants had first viewed a news clip, then they were asked to make a judgment about the appropriateness of a given title and category for the news item. If they considered both the title and category as appropriate, they were asked to press "z" on the keyboard, and when they felt that either the title or the category was inappropriate, they were instructed to press "m." In this way, we obtained a measure of task completion, that is, information about how participants performed in terms of the nonlinguistic outcome of the task. Of the total

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

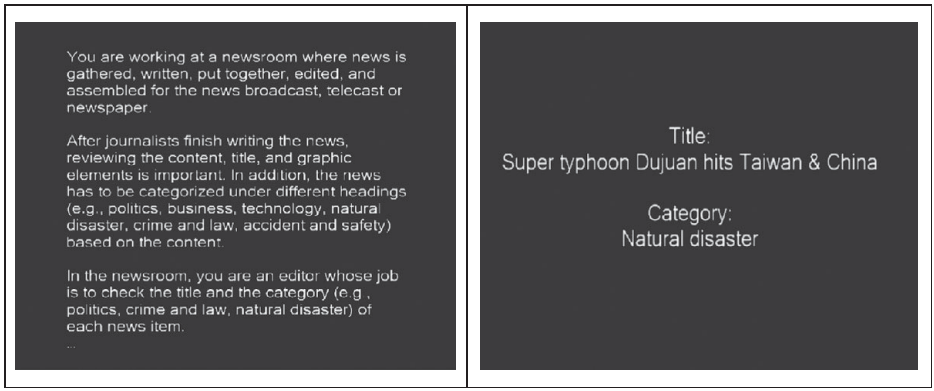


FIGURE 1. Experimental treatment task.

24 multimodal input-based tasks included in this study, half had matching titles and categories while the other half had mismatching titles and categories. Participant received one point for each correct response. Cronbach's α for the task completion index was found to be acceptable (.66). As shown in Table 1, participants, on average, selected the correct response more than 85% of the time in each group. A one-way ANOVA revealed no significant difference among the groups, F (2, 69) = .83, p = .44, η^2 = .002.

A total of 24 multimodal input-based tasks were developed using news clips on a variety of topics. The clips were collected from online news channels, each lasting 20–50 seconds. In all the clips, the present perfect introduced the topic, then the past simple tense was used to give details. The clips were selected in such a way that they contained equal instances of active and passive uses of the present perfect. For the captions and the enhanced captions groups, the news clips were modified with the help of the software Camtasia 8.0. For the unenhanced captions group, we added nonmanipulated captions to the news clips. For the enhanced captions group, the target constructions (present perfect and past simple) were additionally enhanced using yellow fonts with the program Subtitle Edit. Figure 2 illustrates the format of the videos for the three groups.

COLLECTION AND ANALYSIS OF EYE-TRACKING DATA

To capture participants' eye-movements during the treatment, a Tobii X2-60 remote eye-tracker with a temporal resolution of 60 Hz was employed. The eye-tracker was mounted on a 15-inch-screen laptop, with the participants being seated about 60 cm from the laptop screen. The visual angle was approximately 22 degrees. A nine-point calibration procedure was used to calibrate the eye-tracking system; this was repeated before each set of eight treatment tasks. The experiment was designed and conducted using Tobii Studio 3.3.1 software (Tobii Studio, 2015).

To analyze the eye-movement data, two types of interest areas were defined in the captions: one including the present perfect and another including the past simple construction (see Figure 3). We utilized four measures to gauge the amount of attention participants paid to the target linguistic constructions: first-pass reading time, second-pass

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press



FIGURE 2. No captions, unenhanced captions, and enhanced captions.

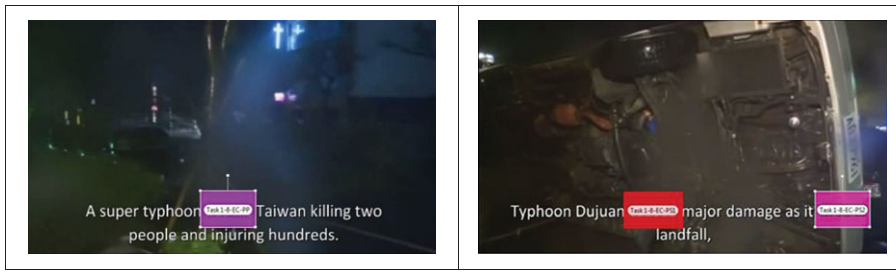


FIGURE 3. Areas of interest.

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

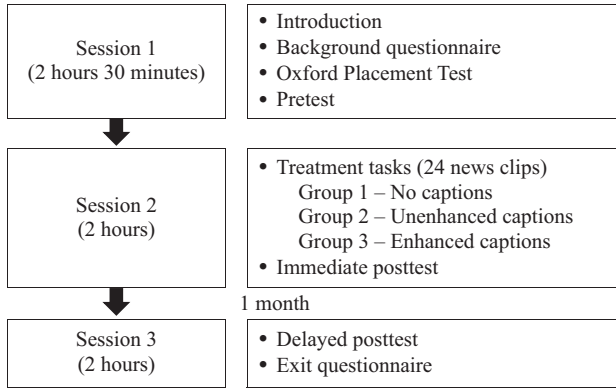


FIGURE 4. Experimental schedule.

reading time, number of visits, and skipping rate. First-pass reading time is defined as the sum of all the fixation durations during an initial visit to an interest area. This index is considered as a measure of initial processing. Second-pass reading time is the sum of all fixation durations made during the second visit to an interest area. That is, second-pass reading time reflects rereading in the area of interest; hence this measure is associated with reanalysis. A visit refers to the period when an individual's eyes first enter an area of interest until they leave. Finally, skipping rate is defined as the proportion of words that were skipped during first-pass reading (Conklin et al., 2018).

Our expectation was that participants in the enhanced caption group would exhibit longer first-pass reading times, longer second-pass reading times, make more visits to the target constructions, and show lower skipping rate. For first-pass reading times, this prediction might not seem straightforward. As a measure of lexical access (Conklin et al., 2018), no difference between the two conditions might be anticipated, as the lexical items in the target constructions are expected to be familiar to the participants. However, visual attention is also driven by cues such as saliency (Conklin et al., 2018), thus textual enhancement, which was realized through using a color contrast in the present study, would be expected to draw learners' attention to the targeted forms.

The data generated were cleaned before being submitted to further analyses (Conklin & Pellicer-Sánchez, 2016). First, fixation durations shorter than 80 ms were removed. Skipped areas of interest, which were recorded as 0 ms, were excluded in the fixation duration analyses. Next, mean fixation durations and SDs were calculated for each measure per participant. Fixation durations that differed from a participant's mean by more than three standard deviations were considered as outliers. Outliers were trimmed to three standard deviations above the mean: .87% of first-pass reading (unenhanced captions group: .7%, enhanced captions group: 1.04%) and .17% of second-pass reading times (unenhanced captions group: .17%, enhanced captions group: .17%) for the present perfect and .26% of first-pass reading (unenhanced captions group: .35%, enhanced captions group: .17%) and .26% of second-pass reading times (unenhanced captions group: .17%, enhanced captions group: .34%) for the past simple.

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

ASSESSMENT TASKS AND SCORING

To assess different types of knowledge of the target construction, three assessment tasks were developed: an oral production test, a written production test, and a fill-in-the-blank test. Three versions of each test were designed, which were counterbalanced across participants in the pretest, posttest, and delayed posttest.

Except for modality, the oral and written production tests had the same format. These tests were designed to test participants' ability to apply the targeted use of the present perfect in a less controlled context. Participants were asked to view a series of news clips in Korean, and their task was to report what they had seen in English. In the oral production test, the participants were asked to break the news to their friends in the oral mode, whereas, as part of the written production test, they were required to post the news on their Social Networking Service (SNS). Five news clips were included in both the oral and written production tests. The news clips entailed no captions and were similar in length to the clips used during the treatment. There was no word limit for the responses. The tasks were piloted with English–Korean bilinguals, and the data confirmed that the tests, as expected, succeeded in creating obligatory contexts for the two constructions.

To assess the learners' performance on the oral production and written production tests, a partial scoring procedure was employed. For each obligatory context of the present perfect, the maximum score was 2 points. Suppliance of the correct form was awarded a score of 2, and 1 point was given for the use of a partially correct form (e.g., correct use of *have/has* with incorrect form of the past participle). The majority of errors involved the use of the past simple form in present perfect contexts, thus only a very small number of partial scores were awarded (oral production data: .40%, written production data: 1.20%). In light of this, we decided to recode the data into a dichotomous scale (correct: 1 point, incorrect: 0 point). For the past simple, the number of obligatory contexts varied among participants, thus we calculated rate of accurate suppliance in obligatory contexts to evaluate participants' performance (Pica, 1983). We also applied a partial scoring system when assessing responses in past simple obligatory contexts, awarding 2 points for correct and 1 point for partially correct (e.g., *hurted*) forms. We also checked the responses for overuse of the present perfect in past simple contexts but found no evidence for this.

The aim of the fill-in-the-blank test was to gauge participants' ability to use the target construction in a controlled context. The participants were asked to complete sentences by filling in blanks. There were 10 target items and 30 distractors. Each item included two blanks. In the target items, one blank targeted the use of the present perfect and one the past simple. For the present perfect, half the target items required the active voice and the other half the passive voice. In the distractors, the two blanks were designed to elicit verb forms associated with if/unless conditionals (10 items), time clauses (10 items), and subjunctives (10 items). To assess participants' performance on the test, we originally used the same partial scoring system as for the oral and written production tests. However, the data were again recoded into a dichotomous scale given the small number of partial scores awarded (7.73%). Thus, the maximum total score for the target items was 20 points for both the present perfect and the past simple items. The internal consistency reliability for the three versions of the test was in the acceptable range (version A: $\alpha = .66$, version B: $\alpha = .68$, version C: $\alpha = .75$)

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

DATA COLLECTION PROCEDURE

As shown in Figure 4, each participant was required to take part in three individual sessions. In the first session, informed consent was obtained (15 minutes), then a background questionnaire (10 minutes), the OPT (40 minutes), and the pretest (80 minutes) were administered in this order. As part of the pretest, participants first completed the oral production test, followed by the written production and the fill-in-the-blank test. Responses on the oral and written production test were recorded using a voice recorder and word-processing software, respectively. The duration of both the oral and the written production test was 15–18 minutes. The fill-in-the-blank test took the form of a paper-and-pencil test lasting approximately 40 minutes. The procedure was the same for the immediate and delayed posttest. In the second session, which took place 2 days after the first session, the participants completed 24 multimodal input-based tasks, followed by the immediate posttest. While performing the treatment tasks, participants' eye-movements were recorded. The 24 treatment tasks took 13–15 minutes to complete. Session 3 took place a month later; the participants were asked to complete a delayed posttest and an exit questionnaire.

STATISTICAL ANALYSES

To address research questions 1 and 2, we carried out a series of mixed-effects models using the lme4 package in the R statistical environment (R Development Core Team, 2016). For models with binary dependent variables, we constructed logistic mixed-effects models using the *glmer* function. For models with continuous dependent variables, we employed linear mixed-effects models relying on the *lmer* function. In the case of continuous data (past simple scores and eye-tracking data), the variables were transformed into a natural logarithm scale as they did not meet the normality assumption. Each model included group and time as fixed effects, and intercepts for participants and items served as the random effects. By-participant and by-item random slopes for the fixed effects (time as a random slope by participant and group as a random slope by item) were also added to achieve a maximum model structure (Barr et al., 2013). However, if the maximal model failed to converge, the random effect that accounted for the least variance was removed until convergence was achieved (Blom et al., 2012). An alpha level of $p < .05$ was set for all tests. For the linear mixed-effects regressions, effect size estimates were calculated with the command "r.squared GLMM" from the "MuMin" package. To address research question 3, a series of Spearman correlation analyses were employed. An alpha level of $p < .05$ was also set for the correlational analyses, and r values of .25, .40, and .60 were considered to be small, medium, and large, respectively (Plonsky & Oswald, 2014).

RESULTS

PRELIMINARY ANALYSES

To test whether the three groups were comparable in terms of their performance on the oral production, written production, and fill-in-the-blank pretests, we conducted a series of mixed-effects analyses. We used logistic mixed-effects regressions for the present perfect scores and linear mixed-effects regressions for the past simple scores. In each model,

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

TABLE 2. Descriptive statistics for oral production test—present perfect

				95% CI	
	M	Mean Gain	*SD*	*Lower*	*Upper*
No captions (*N* = 24)					
Pretest	.92	–	1.14	.44	1.40
Immediate posttest	1.54	.62	1.59	.87	2.21
Delayed posttest	.1.12	.21	1.19	.62	1.63
Unenhanced captions (*N* = 24)					
Pretest	.83	–	1.01	.41	1.26
Immediate posttest	1.87	1.04	1.70	1.16	2.59
Delayed posttest	1.50	.67	1.61	.82	2.18
Enhanced captions (*N* = 24)					
Pretest	.87	–	.88	.48	1.22
Immediate posttest	3.43	2.57	1.30	2.89	3.98
Delayed posttest	4.00	3.13	1.32	3.44	4.56

Max. score = 5

TABLE 3. Descriptive statistics for written production test—present perfect

				95% CI	
	M	Mean Gain	*SD*	*Lower*	*Upper*
No captions (*N* = 24)					
Pretest	.79	–	1.05	.35	1.24
Immediate posttest	.81	.02	.86	1.50	3.00
Delayed posttest	.79	<.01	.88	.42	1.16
Unenhanced captions (*N* = 24)					
Pretest	.96	–	.27	.39	1.52
Immediate posttest	2.25	1.29	.36	1.50	3.00
Delayed posttest	2.12	1.26	1.89	1.32	2.92
Enhanced captions (*N* = 24)					
Pretest	.83	–	.22	.37	1.30
Immediate posttest	4.00	3.17	1.35	3.43	4.57
Delayed posttest	3.25	2.42	1.77	2.50	4.00

Max. score = 5

group served as the fixed effect, the random effects were participant and item, and the dependent variable was participants' score on the test. As shown in the Tables 2 and 3 in the supplementary information online, none of the analyses yielded a significant difference among the three groups for either the present perfect items or the past simple items. This means that the three groups had comparable scores on the three pretests.

EFFECTS OF NO CAPTIONS, UNENHANCED CAPTIONS, VERSUS ENHANCED CAPTIONS ON L2 GRAMMATICAL DEVELOPMENT (RQ1)

To address the first research question, we ran another series of mixed-effects models. In each model, the fixed effects were time, group, and their interaction, the random effects

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

TABLE 4. Descriptive statistics for fill-in-the-blank test—present perfect

	M	Mean Gain	*SD*	95% CI *Lower*	95% CI *Upper*
No captions ($N = 24$)					
Pretest	1.21	–	.36	.46	1.95
Immediate posttest	1.37	.17	.28	.79	1.96
Delayed posttest	1.40	.19	.30	.78	2.01
Unenhanced captions ($N = 24$)					
Pretest	1.42	–	1.45	.80	2.03
Immediate posttest	3.04	1.62	.48	2.05	4.03
Delayed posttest	2.60	1.19	2.63	1.49	3.72
Enhanced captions ($N = 24$)					
Pretest	1.40	–	1.61	.72	2.08
Immediate posttest	6.60	5.21	2.27	5.64	7.56
Delayed posttest	6.44	5.04	2.25	5.49	7.39

Max. score = 10

were participant and item, and the dependent variable was participants' performance on one of the three assessment tasks (see Tables 4–16 in the supplementary information on line for the full models and results).

Table 2 presents the descriptive statistics for the present perfect items on the oral production test. The logistic mixed-effects model carried out to examine the participants' development in the use of the present perfect on the oral production test yielded statistically significant time-by-group interaction effects. Given that time-by-group interaction effects were revealed, post-hoc models with the same structure were constructed, each comparing two groups' pretest–posttest or pretest–delayed posttest scores at a time. For the present perfect, the results revealed no significant interaction between the no-captions and unenhanced captions groups (pretest–posttest: estimate = .66, SE = .49, p = .17; pretest–delayed posttest: estimate = .78, SE = .51, p = .12). However, a significant interaction effect emerged when the performance of the no-captions group was compared with that of the enhanced captions group (pretest–posttest: estimate = 1.95, SE = .49, $p < .001$; pretest–delayed posttest: estimate = 3.17, SE = .52, $p < .001$). There were also significant interactions found for the comparisons between the unenhanced captions group and the enhanced captions groups (pretest–posttest: estimate = 1.27, SE = .48, p = .008; pretest–delayed posttest: estimate = 2.52, SE = .53, $p < .001$). Taken together, the enhanced captions group showed greater pretest–posttest and pretest–delayed posttest gains in the use of the present perfect than the unenhanced captions and no captions group.

Table 3 provides the descriptive statistics for the present perfect items on the written production test. The logistic mixed-effects model, which was conducted to gauge participants' development in the use of the present perfect on the written production test, generated significant interaction effects. All three pair-wise post-hoc tests, which compared two groups' pretest–posttest or pretest–delayed posttest performance at a time, identified a significant, small-size interaction effect. That is, there was a significant difference found between the scores of the no-captions and unenhanced captions groups (pretest–posttest: estimate = 1.69, SE = .53, p = .002; pretest–delayed posttest: estimate =

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

TABLE 5. Descriptive statistics for oral production test—past simple

	M	Mean Gain	*SD*	95% CI	
				Lower	*Upper*
No captions ($N = 24$)					
Pretest	4.72	–	.37	4.57	4.88
Immediate posttest	4.77	.04	.44	4.58	4.95
Delayed posttest	4.64	−.09	.43	4.46	4.82
Unenhanced captions ($N = 24$)					
Pretest	4.63		.41	4.45	4.80
Immediate posttest	4.80	.17	.43	4.61	4.98
Delayed posttest	4.78	.15	.35	4.63	4.92
Enhanced captions ($N = 24$)					
Pretest	4.60	–	.75	4.28	4.92
Immediate posttest	4.73	.13	.44	4.54	4.92
Delayed posttest	4.78	.18	.36	4.27	4.93

Max. score = 5

1.52, $SE = .53$, $p = .004$), the no-captions and enhanced captions groups (pretest–posttest: estimate = 4.00, $SE = .61$, $p < .001$; pretest–delayed posttest: estimate = 2.88, $SE = .55$, $p < .001$), and the unenhanced and enhanced captions groups (pretest–posttest: estimate = 2.57, $SE = .60$, $p < .001$; pretest–delayed posttest: estimate = 1.61, $SE = .55$, $p = .004$). These results indicate that access to captions, regardless of textual enhancement, facilitated participants' development in the use of the present perfect, as measured by the written production test. However, textually enhanced captions proved more effective than unenhanced captions in promoting knowledge of the present perfect.

Table 4 provides the descriptive statistics for the present perfect items on the fill-in-the-blank test. The logistic mixed-effects model, designed to test the extent to which participants developed in the use of the present perfect on the fill-in-the-blank test, found significant time-by-group interaction effects. The post-hoc tests, which assessed whether there were differences in pretest–posttest or pretest–delayed posttest scores between any of the two groups, yielded a significant interaction effect for the pretest–posttest and pretest–delayed posttest comparisons for the no-captions and enhanced captions groups (pretest–posttest: estimate = 2.53, $SE = .59$, $p < .001$; pretest–delayed posttest: estimate = 2.52, $SE = .61$, $p < .001$), and the unenhanced and enhanced captions groups (pretest–posttest: estimate = 1.78, $SE = .49$, $p < .001$; pretest–delayed posttest: estimate = 2.12, $SE = .52$, $p < .001$). Taken together, participants benefited from enhanced captions, as compared to no captions and unenhanced captions, in developing their knowledge of the present perfect, as measured by their performance on the fill-in-the-blank test.

Moving on to the result for the past simple, Tables 5–7 give the descriptive statistics for the three assessment tasks. The linear mixed-effects models, which were carried out to assess participants' development in the use of the past simple tense, yielded no significant interaction effects for either the oral production test, the written production test, or the fill-in-the-blank test. These results indicate that the presence of captions, irrespective of whether they were enhanced or not, had no statistically significant effect on learner gains in the use of the past simple tense on any of the three assessment tasks.

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

TABLE 6. Descriptive statistics for written production test—past simple

				95% CI	
	M	Mean Gain	*SD*	*Lower*	*Upper*
No captions ($N = 24$)					
Pretest	4.76		.46	4.57	4.95
Immediate posttest	4.77	.01	.39	4.60	4.94
Delayed posttest	4.50	−.26	.97	4.09	4.91
Unenhanced captions ($N = 24$)					
Pretest	4.60		1.03	4.16	5.04
Immediate posttest	4.78	.18	.35	4.63	4.92
Delayed posttest	4.56	−.04	.83	4.21	4.91
Enhanced captions ($N = 24$)					
Pretest	4.31	–	1.40	3.72	4.90
Immediate posttest	4.64	.34	.54	4.42	4.87
Delayed posttest	4.26	−.05	1.02	3.82	4.69

Max. score = 5

TABLE 7. Descriptive statistics for fill-in-the-blank test—past simple

				95% CI	
	M	Mean Gain	*SD*	*Lower*	*Upper*
No captions ($N = 24$)					
Pretest	15.62	–	2.43	14.60	16.65
Immediate posttest	16.71	1.08	1.92	15.90	17.52
Delayed posttest	16.75	1.12	2.09	15.87	17.63
Unenhanced captions ($N = 24$)					
Pretest	16.33	–	2.41	15.32	17.35
Immediate posttest	17.25	.92	2.33	15.27	18.23
Delayed posttest	17.46	1.12	1.95	16.63	18.28
Enhanced captions ($N = 24$)					
Pretest	16.21	–	2.39	15.20	17.22
Immediate posttest	17.04	.83	2.35	16.05	18.03
Delayed posttest	17.42	1.21	2.16	16.50	18.33

Max. score = 20

EFFECTS OF UNENHANCED CAPTIONS VERSUS ENHANCED CAPTIONS ON ALLOCATION OF ATTENTION (RQ2)

To address the second research question, we ran another series of mixed-effects models. Linear mixed-effects regressions were conducted for all measures; the only exception was skipping rate, for which the data were submitted to a logistic mixed-effects regression. In each model, group was included as a fixed effect, and participant and item were specified as crossed random effects. The dependent variable was one of the four eye-gaze measurements: first-pass reading time, second-pass reading time, number of visits, or skipping rate (see Tables 17–20 in the supplementary information online for the full models and results).

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

TABLE 8. Descriptive statistics for attention measurements—present perfect

				95% CI	
	N	*M*	*SD*	*Lower*	*Upper*
First-pass reading					
Unenhanced captions	24	131	62	105	158
Enhanced captions	24	175	52	153	197
Second-pass reading					
Unenhanced captions	24	90	76	57	122
Enhanced captions	24	270	82	235	304
Number of visits					
Unenhanced captions	24	1.62	.68	1.33	1.91
Enhanced captions	24	2.20	.47	2.01	2.40
Skipping rate					
Unenhanced captions	24	.24	.24	.14	.34
Enhanced captions	24	.07	.17	<.01	.14

TABLE 9. Descriptive statistics for attention measurements—past simple

				95% CI	
	N	*M*	*SD*	*Lower*	*Upper*
First-pass reading					
Unenhanced captions	24	237	205	150.11	323.38
Enhanced captions	24	354	199	270.40	438.31
Second-pass reading					
Unenhanced captions	24	109	92	70.27	148.06
Enhanced captions	24	198	141	138.16	257.27
Number of visits					
Unenhanced captions	24	2.83	1.76	2.09	3.57
Enhanced captions	24	3.95	1.73	2.86	3.92
Skipping rate					
Unenhanced captions	24	.35	.30	.23	.48
Enhanced captions	24	.19	.23	.09	.28

Table 8 presents the descriptive statistics for the eye-gaze measures for the areas of interest defined for the present perfect. The mixed-effects models revealed that there were significant differences between the two groups in terms of three eye-movement indices (second-pass reading: estimate = .49, SE = .08, $p < .001$; number of visits: estimate = 1.09, SE = .30, $p < .001$; skipping rate: estimate = −2.20, SE = .61, $p < .001$). These results mean that, as compared to unenhanced captions, textually enhanced captions were more effective in drawing learners' attention to the present perfect construction.

Table 9 gives the descriptive statistics for the eye-gaze measures associated with the interest areas defined for the past simple. The linear mixed-effects models found significant effects for second-pass reading (estimate = .26, SE = .10, p = .01) and for skipping rate (estimate = −1.54, SE = .61, p = .01). Overall, these results show that, textually

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

enhanced captions were also more likely to direct learners' attention to the past simple construction than unenhanced captions.

RELATIONSHIPS BETWEEN ATTENTION AND L2 DEVELOPMENT (RQ3)

To investigate the third research question, we ran a series of Spearman correlational analyses for the unenhanced and enhanced captions groups separately. In particular, we examined whether there were significant relationships between the eye-gaze indices and participants' pretest–posttest gains and pretest–delayed posttest gain scores on the three assessment tasks.

As shown in Table 10, for the unenhanced captions group, only a few significant correlations were identified, there were large-size correlations between the number of visits and participants' pretest–posttest and pretest–delayed posttest gains in the written production test. That is, in the unenhanced captions group, participants who visited more frequently the areas of interest defined for the present perfect exhibited higher gains on the written production test.

The correlational analyses yielded more significant relationships for the enhanced captions group (see Table 10). Similar to the unenhanced captions group, however, all significant correlations involved gain scores in the use of the present perfect. None involved gains in the past simple. The oral production pretest–posttest and pretest–delayed posttest gains were found to have medium- to large-size relationships with participants' second-pass reading times, number of visits, and skipping rates. Medium- to large-size correlations were also identified between the participants' written production pretest–posttest gains and all the eye-tracking indices. Overall, these results indicate that, in the enhanced captions group, participants who fixated longer and more frequently on the present perfect construction were more likely to obtain higher gains on the oral and written production tests.

DISCUSSION

We asked three research questions regarding the relationships between captioning and L2 development, captioning and attentional allocation, and attention and L2 development. To facilitate the discussion, the results of the study are summarized in Table 11 with respect to each research question.

CAPTIONING AND DEVELOPMENT IN L2 GRAMMATICAL KNOWLEDGE (RQ1)

Our first research question asked the extent to which multimodal input-based tasks without captions, with unenhanced captions, and with enhanced captions affect development in L2 grammatical knowledge. The results revealed that the presence of unenhanced captions, as compared to the absence of captions, had a positive impact on learners' immediate and delayed posttest gains in the use of the present perfect on all tests. These positive effects, however, only reached significance for participants' gains on the written production test. Overall, these results indicate that captions cannot only facilitate the acquisition of L2 vocabulary (Montero Perez et al., 2013) but also have the capacity to promote development in L2 grammatical knowledge.

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

TABLE 10 Results of Spearman correlations between eye-tracking and developmental measures

		Oral Production		Written Production		Fill-in-the-blank	
		Pretest-Immediate	Pretest-Delayed	Pretest-Immediate	Pretest-Delayed	Pretest-Immediate	Pretest-Delayed
Unenhanced—present perfect							
First-pass reading	ρ	.27	.36	.31	.37	.38	.38
	p	.20	.08	.15	.07	.07	.07
Second-pass reading	ρ	.14	.21	.33	.41	.18	.27
	p	.51	.31	.12	.05	.39	.14
Number of visits	ρ	.23	.22	.70**	.71**	.33	.33
	p	.27	.31	<.01	<.01	.12	.11
Skipping rate	ρ	−.19	−.36	−.32	−.20	−.32	−.29
	p	.38	.09	.12	.35	.12	.16
Enhanced—present perfect							
First-pass reading	ρ	.51	.44	.76***	.27	.50	.08
	p	.10	.03	.00	.20	.82	.71
Second-pass reading	ρ	.67***	.49*	.70***	.36	.07	.12
	p	.00	.01	.00	.08	.73	.48
Number of visits	ρ	.52*	.47*	.61**	.23	−.01	−.01
	p	.01	.02	.00	.28	.95	.97
Skipping rate	ρ	−.46*	−.40*	−.52*	−.19	−.13	−.19
	p	.02	.05	.01	.38	.54	.37
Unenhanced—past simple							
First-pass reading	ρ	−.32	−.12	.20	.02	.14	.04
	p	.13	.57	.36	.92	.50	.86
Second-pass reading	ρ	−.28	−.11	.02	.03	.19	.10
	p	.19	.62	.93	.89	.37	.65
Number of visits	ρ	−.25	−.13	.09	.07	.12	.03
	p	.25	.55	.68	.76	.58	.88
Skipping rate	ρ	.24	.15	−.14	−.01	.05	.07
	p	.25	.50	.52	.94	.81	.76
Enhanced—past simple							
First-pass reading	ρ	.17	.08	.10	−.15	.01	.14
	p	.43	.71	.63	.49	.96	.51
Second-pass reading	ρ	.24	.04	.17	−.10	.06	.07
	p	.26	.85	.44	.65	.77	.76
Number of visits	ρ	.25	.16	.20	−.07	.12	.17
	p	.23	.46	.36	.75	.56	.42
Skipping rate	ρ	−.17	.01	−.13	.05	−.09	−.15
	p	.42	.98	.54	.82	.69	.48

$N = 48$ *** $p < .001$, ** $p < .01$, * $p < .05$

A question that arises, however, is why the positive effects of captioning were most pronounced on the written production test, reaching significance only on this test type. A possible way of explaining this finding may be that the unenhanced captions group had

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

TABLE 11. Summary of results

Research Question	Sig	Measures	Results
Captioning and L2 grammatical knowledge			
Present Perfect	Yes	Oral Productive	*Pretest–Posttest* No captions < Enhanced Unenhanced < Enhanced *Pretest–Delayed posttest* No captions < Enhanced Unenhanced < Enhanced
	Yes	Written Productive	*Pretest–Posttest* No captions < Unenhanced/ Enhanced Unenhanced < Enhanced *Pretest–Delayed posttest* No captions < Unenhanced/ Enhanced Unenhanced < Enhanced
	Yes	Fill-in-the-blanks	*Pretest–Posttest* No captions < Enhanced Unenhanced < Enhanced *Pretest–Delayed posttest* No captions < Enhanced Unenhanced < Enhanced
Past simple	No	–	–
Captioning and attention			
Present perfect	No	First-pass reading	
	Yes	Second-pass reading	Unenhanced < Enhanced
	Yes	Number of visits	Unenhanced < Enhanced
	Yes	Skipping rate	Unenhanced > Enhanced
Past simple	No	First-pass reading	
	Yes	Second-pass reading	Unenhanced < Enhanced
	No	Number of visits	
	Yes	Skipping rate	Unenhanced > Enhanced
L2 learning and attention			
Present Perfect			
Unenhanced	No	Oral Productive	–
	Yes	Written Productive	Number of visits (+)
	No	Fill-in-the-blanks	–
Enhanced	Yes	Oral Productive	Second-pass reading (+) Number of visits (+) Skipping rate (–)
	Yes	Written Productive	First-pass reading (+) Second-pass reading (+) Number of visits (+) Skipping rate (–)
	No	Fill-in-the-blanks	
Simple past			
Unenhanced	No		
Enhanced	No		

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

developed both their procedural and declarative knowledge as a result of the treatment, but it was primarily their declarative knowledge that they relied on during the tests. According to the skill acquisition approach, procedural knowledge is difficult to transfer across skills; transfer between skills is likely to occur through declarative knowledge of rules (DeKeyser, 2007). Hence, any gains in procedural knowledge were less likely to surface on the tests, given that all three tests required producing the target construction. The participants' superior performance on the written, as compared to the oral, production test might be attributed to the fact that the written task imposed lower time pressure, thereby enabling learners to deploy their declarative knowledge of the target construction to a greater extent. The lack of significant effects for captioning on the fill-in-the-blank test might have been an artifact of this task requiring the application of new knowledge in a context different from the treatment. According to the principle of transfer-appropriate processing, it is easier to recall information in contexts that are similar to those in which the information was initially encoded (Lightbown, 2008).

Interestingly, the enhanced captions group outperformed the unenhanced caption group on all tests, not only on the written production test. Following the previous line of reasoning regarding the limits on transferability of skills, a possible way to account for this finding may be that the increased salience of the target construction prompted the participants in the enhanced captions group to reflect more on the target construction, that is, they had more opportunities to apply their declarative knowledge throughout task performance. As a result, they were able to automatize their explicit knowledge of the present perfect to a greater degree. This, in turn, could explain why the performance of the enhanced captions group was less affected by the time pressure imposed during the oral production test. The greater number of opportunities afforded to use declarative knowledge might have also better enabled the enhanced captions group to recall knowledge in contexts different from the ones experienced during the treatment.

Continuing with the comparison between the gains of the unenhanced and enhanced captions groups, our results are aligned with the findings of Montero Perez et al. (2015) and Lee and Révész (2018), who also observed an advantage for increasing the visual salience of target linguistic features in captions. The results obtained here are also consistent with theoretical proposals that claim that enhancing features in the input will facilitate the noticing and subsequent learning of L2 constructions (e.g., Sharwood Smith, 1991).

It is also worth noting that this study, similar to Lee and Révész (2018), yielded a greater advantage for textual enhancement than Lee and Huang's meta-analysis focusing on the role of textual enhancement in the context of reading. Unlike this study and Lee and Révész (2018), Lee and Huang (2008) only found marginal positive effects of textual enhancement on development in L2 grammatical knowledge. An explanation for the discrepancy in findings between the captioning and reading studies may be that textual enhancement together with captioning might have increased the salience of the target features to a greater degree than textual enhancement alone, leading to a greater depth of processing (Leow & Martin, 2017). Another explanation might lie in the potentially different skipping rates in unimodal versus multimodal conditions. Given that captions in multimodal input are redundant to the oral input, viewers might be more likely to skip them in the absence of enhancement, as compared to unenhanced, nonredundant text in unimodal input. Indeed, in the present study, we observed a significantly higher skipping

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

rate under the unenhanced condition. Other factors that might have contributed to the more positive outcomes for textual enhancement in the captioning studies include prior knowledge (e.g., Han et al., 2008; Park, 2004) and the relative salience of the targeted grammatical constructions (Gass et al., 2017). Both Lee and Révész (2018) and the present experiment targeted a perceptually salient construction, of which learners had some prior knowledge. Last but not least, instructed L2 learners tend to be better at reading than listening skills; therefore, in the auditory modality input enhancement techniques such as captioning and textual enhancement may have greater potential to have an impact.

Another noteworthy result of the present study is that textual enhancement only promoted development in participants' use of the present perfect; it had no significant impact on learners' knowledge of the past simple. This was probably due to a ceiling effect, as participants achieved considerably high mean scores on all three pretests in the use of the past simple, leaving little space for improvement. This was not an unexpected finding, given the high proficiency level of the participants.

CAPTIONING AND ATTENTION TO L2 GRAMMATICAL CONSTRUCTIONS (RQ2)

Our second research question was concerned with the extent to which textually unenhanced versus enhanced captions in multimodal input-based tasks can draw learners' attention to the target construction. As expected, textually enhanced captions were more effective in directing learners' attention to the present perfect construction, and, although to a smaller extent, textual enhancement also succeeded more in drawing learners' attention to the past simple. These results are consistent with those of Lee and Révész (2018), where participants were also found to allocate more attention to textually enhanced than unenhanced grammatical constructions in captions. Our findings are also partially parallel to the patterns observed in Montero Perez et al. (2015). This study yielded an advantage for increasing the visual salience of target lexis in captions, but the positive effects of enhanced captions on attentional allocation only emerged under the condition where participants had been made aware of a forthcoming vocabulary test.

It is also worth highlighting that both Lee and Révész (2018) and the present study found higher second-pass reading times and number of visits when captions were enhanced, but no significant difference emerged in first-pass reading times between the enhanced and unenhanced groups. The lack of significant results for first-pass reading times, although also attested in previous studies (e.g., Lee & Révész, 2018; Winke, 2013; see however, Alsadoon & Heift, 2015), is somewhat surprising. Textual enhancement constitutes a visual manipulation, which was expected to trigger effects also in early eye-tracking measures. Further research is needed to shed more light on this pattern.

It is also interesting to compare the findings obtained here with studies examining the effects of textual enhancement in the context of reading. As noted previously, existing results for the relationship between textual enhancement and attentional allocation in unimodal input are mixed. Some studies generated positive effects for textual enhancement (Simard & Foucambert, 2013; Winke, 2013), whereas others yielded no benefits for the provision of enhanced input (Indrarathne & Kormos, 2017; Issa et al., 2015; Loewen & Inceoglu, 2016). The more uniformly positive results observed for textual enhancement in captions might be due, as discussed earlier, to the greater salience of textual enhancement in captions than in unimodal reading activities (Leow & Martin, 2017).

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

RELATIONSHIP BETWEEN ATTENTION AND L2 DEVELOPMENT (RQ3)

Our third research question addressed the relationship between learner attention allocated to the target linguistic construction and development in L2 grammatical knowledge. We were also interested in exploring whether the presence of textual enhancement in the captions moderated this relationship. While significant positive correlations between attention and learner gains were observed for both the enhanced and unenhanced captions groups, we found considerably more significant associations for the enhanced captions group. In the unenhanced captions group, participants who paid more attention to the present perfect only exhibited higher gains on the written production test. In the enhanced captions group, however, participants who allocated more attention to the present perfect construction were more likely to obtain higher gain scores on both the oral and written production tests. No significant relationships emerged for participants' gains in the use of the past simple. This was probably due to a ceiling effect and a related lack of variation in scores at the pretest stage. This was not an unexpected finding given the proficiency level of the participants.

It is intriguing why, in the unenhanced captions group, a positive relationship between attention and learning was only found on the written production test. A possible reason may be that participants showed somewhat greater variance in their written than oral production posttest scores, which made it more likely that any relationships between attentional allocation and development would surface.

It is also worthwhile to evaluate our findings in relation to previous research exploring associations between textual enhancement and development in grammatical knowledge. Like the present study, some previous research found positive relationships between increased attention to target constructions and gains in grammatical knowledge (e.g., Godfroid & Uggen, 2013; Indrarathne & Kormos, 2017). Other research (e.g., Issa et al., 2015; Winke, 2013), however, yielded no such links. The contradictory findings across studies may be explained by the fact that eye-tracking measures may indicate different levels of processing (Godfroid, 2019). In studies where no relationships were found between attentional allocation and L2 learning, participants with higher gains might have engaged in greater depth of processing than their counterparts with lower gains. However, in the absence of triangulation with verbal protocol data, this explanation remains tentative.

LIMITATIONS AND FUTURE DIRECTIONS

In interpreting the findings obtained here, it is also important to take into account the limitations of the study. First, the study would have benefited from the inclusion of a group who only participate in the testing sessions. This would have allowed for gauging the effects of being exposed to audio-visual input versus no treatment.

A second limitation has to do with the nature of input enhancement. We could have made the distinction between the uses of the past simple and present perfect more salient by using different colors to enhance the two constructions. In future research, it would be interesting to explore whether using different colors would make the effects of textual enhancement more pronounced than the use of a single color.

A third, methodological weakness is that the eye-tracking measures were not triangulated with verbal protocol comments. Combining eye-tracking with verbal protocol data would have enabled us to gather information not only about learners' attentional

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

allocation but also their potential engagement in higher level of processing. This, in turn, would have made our interpretations less tentative. Future research would benefit from supplementing eye-tracking indices with data collected through verbal protocols (see e.g., Jung & Révész, 2018).

A further limitation concerns the relatively large spatial resolution (0.2 degrees) and low temporal resolution (60 Hz) of the eye-tracking equipment we used; these technical features might have affected the accuracy of the eye-tracking data we obtained. Spatial accuracy and precision might have suffered, as our areas of interests were relatively small (average angular size of the present perfect: $8.6° \times 3.0°$, average angular size of the past simple: $4.2° \times 2.8°$) and the spatial resolution of the eye-tracker was relatively large. This issue was, however, mitigated by the fact that, for each participant, we had a considerably large number of trials (24), decreasing the chance of error. Similar, the 60 Hz temporal was arguably acceptable because this study only included fixation analyses. According to Raney et al. (2014, p. 2), "the average temporal error will be approximately half the duration of the time between samples." Thus, a sampling rate of 60 Hz will result in an error of about 8 msec on average. As explained by Raney et al., while an 8 msec error might be too large to examine saccade durations, it is not too large to investigate fixation durations.

Finally, another shortcoming has to do with the frequency with which the present perfect is used to introduce news across various dialects of English. Although both British and American English appeared in the news items in the present study, the present perfect is more commonly used in British English than American English (Quirk et al., 1985). Considering that Korean learners of English are more often exposed to American English, selecting a target linguistic construction that is more widely used in American English might have been more relevant to the participating students. Future research might want to take this factor into account when selecting linguistic targets.

More generally, future research would benefit from investigating the effects of captioning on other linguistic targets. Investigating the acquisition of features that are less perceptually salient than the construction examined here are especially needed, given that such features are less likely to capture learners' attention in the absence of input enhancement. Further studies are also warranted to explore whether the findings obtained here would transfer to other genres (e.g., dramas and documentaries). Replication studies are additionally needed with other learner populations with different first languages, educational backgrounds, and proficiency levels. It would be particularly interesting to explore whether the findings would transfer to contexts where, unlike in Korea, films are usually dubbed rather than subtitled (Lindgren & Muñoz, 2013).

CONCLUSION

The main aim of this study was to help close the gap in current task-based research on input-based tasks by launching an investigation into the extent to which multimodal input-based tasks can promote learner attention to and subsequent development in the knowledge of L2 grammar. We operationalized multimodal input-based as tasks presenting learners with audio, video, and textual input simultaneously, with the textual input taking the form of captions with or without textual enhancement. In doing so, we also aimed to contribute to previous research examining the impact of visual enhancement on

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

attentional allocation to and learning of grammatical constructions. Last but not least, we intended to expand on existing research by exploring the link between attention and L2 development in grammatical knowledge.

As expected, we found that access to captions, with or without textual enhancement, facilitated the acquisition of grammatical knowledge. In addition, when captions were textually enhanced, participants paid more attention to and achieved greater gains in their knowledge of the targeted present perfect construction, as compared to when they were exposed to unenhanced captions. Finally, we observed positive links between attention and development for both the enhanced and unenhanced captioning conditions, but more and stronger relationships were found for the enhanced captions group.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263120000108.

REFERENCES

Alsadoon, R., & Heift, T. (2015). Textual input enhancement for vowel blindness: A study with Arabic ESL learners. *The Modern Language Journal*, *99*, 57–79.

Bardovi-Harlig, K. (2001). Another piece of the puzzle: The emergence of the present perfect. *Language Learning*, *5*, 215–264.

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*, 255–278.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Linguistics*, *23*, 509–533.

Blom, E., Paradis, J., & Sorenson Duncan, T. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular *–s* in child L2 English. *Language Learning*, *62*, 965–994.

Bygate, M., Skehan, P., & Swain, M. (2001). *Researching pedagogic tasks: Second language learning, teaching and testing*. Longman.

Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. *New Zealand Studies in Applied Linguistics*, *14*, 25–44.

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

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, *32*, 453–467.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning*, *42*, 497–527.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta*, *49*, 67–77.

DeKeyser, R. (2007). Situating the concept of practice. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1–18). Cambridge University Press.

Eastwood, J. (1994). *Oxford guide to English grammar*. Oxford University Press.

Ellis, R. (2003). *Task-based language teaching and learning*. Oxford University Press.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*, 221–246.

Ellis, R. (2013). Task-based language teaching: Responding to the critics. *University of Sydney Papers in TESOL*, *8*, 1–27.

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Routledge.

Gabriele, A. (2009). Transfer and transition in the SLA of aspect. *Studies in Second Language Acquisition*, *31*, 371–402.

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, *24*, 239–258.

Gass, S. M., Spinner, P., & Behney, J. (2017). Salience in second language acquisition and related field. In S. Gass, P. Spinner & J. Behney (Eds.), *Salience in second language acquisition* (pp. 1–18). Routledge.

Godfroid, A. (2019). Investigating instructed second language acquisition using L2 learners' eye-tracking data. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning*. Routledge.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*, 483–517.

Godfroid, A., & Uggen, M. S. (2013). Attention to irregular verbs by beginning learners of German. *Studies in Second Language Acquisition*, *35*, 291–322.

Grabe, W. (2012). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

Han, J., & Hong, S. (2015). The acquisition problem of English present perfect to Korean adult learners of English: L1 transfer matters. *English Language and Linguistics*, *213*, 141–164.

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

Huang, H., & Eskey, D. (2000). The effects of closed-captioned television on the listening comprehension of intermediate English as second language students. *Educational Technology Systems*, *28*, 75–96.

Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit learning conditions: An eye-tracking study. *Studies in Second Language Acquisition*, *39*, 401–430.

Issa, B., & Morgan-Short, K. (2019). Effects of external and internal attentional manipulations on second language grammar development: An eye-tracking study. *Studies in Second Language Acquisition*, *41*, 389–417.

Issa, B., Morgan-Short, K., Villegas, B., & Raney, G. (2015). An eye-tracking study on the role of attention and its relationship with motivation. *EUROSLA Yearbook*, *15*, 114–142.

Jung, J., & Révész, A. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, *40*, 755–780.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480.

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. (2006). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, *30*, 307–331.

Leow, R. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge.

Leow, R., & 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 SLA* (pp. 167–186). Routledge.

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han (Ed.), *Understanding second language process* (pp. 27–44). Multilingual Matters.

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, *10*, 105–129.

Loewen, L., & Inceoglu, S. (2016). The effectiveness of visual input enhancement on the noticing and L2 development of the Spanish past tense. *Studies in Second Language Learning and Teaching*, *6*, 89–110.

Long, M. H. (2000). Focus on form in task-based language teaching. In R. D. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy: Essays in honor of A. Ronald Walton* (pp. 179–192). Benjamins.

Markham, P. (1999). Captioned videotapes and second-language listening word recognition. *Foreign Language Annals*, *32*, 321–328.

Markham, P., Peter, L., & McCarthy, T. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals*, *34*, 439–445.

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.

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The 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.
Park, E. S. (2004). Constraints of implicit focus on form: Insights from a study of input enhancement. *Teachers College, Columbia University Working Papers in TESOL and Applied Linguistics*, *4*, 1–30.
Pica, T. (1983). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, *6*, 69–78.
Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.
Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A contemporary grammar of the English language*. Longman.
R Development Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/.
Raney, G. E., Campbell, S. J., & Bovee, J. C. (2014). Using eye movements to evaluate the cognitive processes involved in text comprehension. *Journal of Visual Experimentation*, *83*, e50780.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
Rodgers, M. P. H., & Webb, S. (2017). The effects of captions on EFL learners' comprehension of English language television programs. *CALICO Journal*, *32*, 20–38.
Rost, M. (2011). *Teaching and researching listening*. Longman.
Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Palgrave Macmillan.
Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learners. *Second Language Research*, *7*, 118–132.
Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*, 165–179.
Shintani, N. (2012). Input-based tasks and the acquisition of vocabulary and grammar: A process-product study. *Language Teaching Research*, *16*, 253–279.
Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. John Benjamins Publishing Company.
Simard, D., & Foucambert, D. (2013). Observing noticing while reading in L2. In J. M. Bergsleithner, S. N. Frota & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 207–226). National Foreign Language Resource Center, University of Hawaii at Mānoa.
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*, 50–73.
Tobii Studio. (2015). *User manual—Tobii Studio (Version 3.3.0)*. http://www.tobii.com/Global/Analysis/Downloads/User_Manuals_and_Guides/Tobii_X2-30_EyeTrackerUserManual_WEB.pdf
Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*, 191–210.
Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, *42*, 272–281.
Winke, P. (2013). The effects of input enhancement on grammar learning and comprehension: A modified replication of Lee (2007) with eye-movement data. *Studies in Second Language Acquisition*, *35*, 323–352.
Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, *14*, 65–86.

https://doi.org/10.1017/S0272263120000108 Published online by Cambridge University Press