Language Teaching (2024), 1–18
doi:10.1017/S0261444824000296

CAMBRIDGE UNIVERSITY PRESS

REPLICATION RESEARCH

The effects of enhancing L2 multiword items in captions: An approximate replication of Majuddin, Siyanova-Chanturia, and Boers (2021)

Elvenna Majuddin[1], Frank Boers[2] and Anna Siyanova-Chanturia[1]

[1]Te Herenga Waka – Victoria University of Wellington, Wellington, New Zealand and [2]Western University, London, Canada
Corresponding author: Frank Boers; Email: fboers@uwo.ca

(Received 9 November 2023; revised 8 June 2024; accepted 13 June 2024)

Abstract
Studies investigating the acquisition of multiword items (MWIs) from reading have furnished evidence that the likelihood of acquisition improves considerably if such items are typographically enhanced (e.g., bolded or underlined) in the texts. In the case of captioned audio-visual materials, however, an earlier study by the authors did not find such compelling evidence. In that study, indications of an effect emerged only when the same video was watched twice. Arguably, for learners to benefit more immediately from typographic enhancement in captions, they may need to be made aware of its purpose beforehand. The present article therefore reports an approximate replication of Majuddin et al. (2021), but this time the students were informed about the MWI-learning purpose of watching the video. As in the original study, the learners watched a video once or twice with standard captions, with captions in which MWIs were enhanced, or without captions. The positive effect of enhancement for MWI learning was clearer than in the original study, and it already emerged after a single viewing. On the downside, enhancement was found to have a negative effect on lower-proficiency learners' comprehension of the content of the video.

1. Background

1.1 Using typographic enhancement to help learners notice multiword items

The past couple of decades have seen a growing interest in lexical items larger than single words, such as idioms, collocations, phrasal verbs, prepositional phrases, and various other types of expressions (Boers & Lindstromberg, 2009; Schmitt, 2004; Siyanova-Chanturia & Omidian, 2020; Siyanova-Chanturia & Pellicer-Sánchez, 2019; Wood, 2010; Wray, 2002). It is now well recognized that knowledge of a large repertoire of multiword items (MWIs) is vital for learners' proficiency in a second language. Such knowledge has been shown crucial for learners' comprehension of second language (L2) discourse (e.g., Kremmel et al., 2017; Martinez & Murphy, 2011), as well as adequate production of L2 discourse (e.g., Bestgen, 2017; Boers et al., 2006; Crossley et al., 2015; Garner, 2022; Hou et al., 2018; Paquot, 2018, 2019; Saito, 2020; Siyanova-Chanturia & Spina, 2020; Stengers et al., 2011; Tavakoli & Uchihara, 2020). As a result, there has been a proliferation of studies gauging the effectiveness of various "interventions" to help learners master this important, phraseological dimension of language (see Pellicer-Sánchez & Boers, 2019, for a review). Typographic enhancement of MWIs is one of those interventions that has attracted a lot of interest in this strand of research.

Typographic enhancement (also known as textual enhancement and visual enhancement) of language features or items in texts is a simple intervention to make features or items more salient for learners (Sharwood Smith, 1993). Enhancement can take diverse forms such as underlining and

change of font (e.g., boldface). Directing learners' attention to certain language features or items this way is considered useful because there is a broad consensus in the literature that attention plays a vital role in L2 acquisition (Schmidt, 2001; VanPatten, 1996). Early studies on the effectiveness of typographic enhancement concerned the acquisition of grammar features (see Lee & Huang, 2008, for a meta-analysis, and Boers, 2021, pp. 53–61, for a narrative review), but more recently researchers have turned to the lexical dimension of language as a target for typographic enhancement (e.g., Vu & Peters, 2022a), with several studies focusing on MWIs such as collocations (i.e., word partnerships, such as *conduct research; slim chance; on purpose*). Directing learners' attention to the lexical makeup of collocations is useful, especially when it differs from a counterpart in the learners' first language (L1) (e.g., Eyckmans et al., 2016; Laufer & Girsai, 2008; Peters, 2016; Terai et al., 2023; Zhang & Graham, 2020). For example, the Dutch counterpart of the English *make an effort* is "do an effort", and the Dutch counterpart of *running water* is "streaming water." Furthermore, many collocations include components that go unnoticed because they are highly familiar words, such as high-frequency verbs (e.g., *do, have,* and *make*) and prepositions (e.g., *in* and *on*). These factors help to explain why even advanced learners produce inaccurate L2 word combinations according to analyses of learner language (e.g., Laufer & Waldman, 2011; Nesselhauf, 2003; Siyanova-Chanturia & Spina, 2020).

1.2 Enhancing multiword items in reading texts

This article reports a replication of a study that investigated the effects of typographically enhancing MWIs in video captions. However, most research on the benefits of enhancement for the acquisition of MWIs so far has concerned reading texts (and occasionally reading while listening to a text; Jung & Lee, 2023). Some used texts without deliberately manipulating the number of occurrences of the MWIs. In these cases, learners met most of the target MWIs only once. An example is Boers et al. (2017), where learners read texts in which MWIs were either enhanced (underlined) or unenhanced. Overall, the participants were more likely in a post-test to recognize the MWIs that had been enhanced than those that had been unenhanced. Choi (2017) had learners read a text containing MWIs in either an enhanced (boldface) or unenhanced version while their eye movements were recorded. The eye-movement data confirmed that the enhancement drew the learners' attention to the MWIs. In a two-week delayed recall test, the learners who had read the enhanced text recalled on average twice as many MWIs as the comparison group who had read the unenhanced version. Interestingly, the learners who read the enhanced version of the text were less able than the comparison group to remember segments of the text that had been left unenhanced. It is likely that they interpreted typographic enhancement as an effort on the part of the teacher/researcher to flag the important elements of a text, and so they felt less need to focus on the remainder of the text. A recent study, by Vu and Peters (2022b), applied enhancement in a longitudinal fashion where learners read a series of stories over nine weeks with MWIs either enhanced (underlined) or left unenhanced. A cued recall test showed a significant effect of enhancement: enhanced MWIs were, on average, recalled more than three times better than unenhanced ones.

Studies have also examined the effect of encountering enhanced MWIs multiple times (Jung et al., 2022; Puimège et al., 2023; Sonbul & Schmitt, 2013; Szudarski & Carter, 2016; Toomer & Elgort, 2019; Toomer et al., 2024). Post-tests almost invariably show greater learning gains for enhanced items if they are encountered several times. It needs to be conceded, however, that embedding multiple instances of the same MWIs in texts (e.g., Webb et al., 2013) requires considerable creativity and it risks compromising the authenticity of the texts, and so it is probably a less appealing choice for busy teachers and course designers. In the case of authentic audiovisual materials, which we turn to next, embedding extra instances of words or phrases is even less straightforward owing to technical obstacles.

1.3 Enhancing multiword items in captions

There has been a steep increase in the number of studies on the use of audiovisual input for L2 learning in general (e.g., Montero Perez, 2022; Peters, 2019; Pujadas & Muñoz, 2019; Teng, 2021). A

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

substantial number of these studies have assessed the benefits of captions for text comprehension and vocabulary acquisition (see Montero Perez, 2022, for a review), but only a few have focused on MWIs in captions, let alone with a specific interest in the use of typographic enhancement of MWIs. Puimège et al. (2024) examined eye movements to ascertain that typographically enhanced MWIs in captions attract more attention and are better recalled than unenhanced ones. They found that the enhancement indeed drew the learners' attention, and that the amount of attention they gave to MWIs predicted successful recall. The latter held true also for unenhanced MWIs, however, such that the presence of enhancement *per se* was not a statistically significant predictor of recall according to mixed effects modelling. In Puimège et al.'s (2024) study, the learners watched the video once and since each MWI occurred only once, there were no repeated encounters. A way of increasing the number of encounters is to have learners watch the same materials twice. This is what was done in a study by Majuddin et al. (2021). There were six viewing conditions: watching a video once with standard captions, enhanced captions, or without captions, and watching the same video twice with standard captions, enhanced captions, or without captions. Both the standard captions and the enhanced captions benefited MWI recall more than watching the video without onscreen text, but there was no difference between the two caption conditions when participants watched the video only once. The descriptive statistics did show larger learning gains from enhanced captions when the video had been watched twice, but altogether the effect of enhancement was not statistically significant according to mixed effects modeling. Viewing the video twice instead of once was found to be a stronger predictor of learning gains.

The findings of both Majuddin et al. (2021) and Puimège et al. (2024) suggest that, compared with its attested usefulness in reading texts, typographic enhancement of MWIs in captions has only a modest impact. A likely explanation is that viewing captioned video is more demanding as regards attention allocation than reading a text. There are more stimuli to attend to (images, sound, and captions) and (unless the video is paused) there is very little time to dwell on language items, regardless of whether they are typographically enhanced. Because authentic audiovisual materials are used in "real time" and require fast processing, learners also have less opportunity to reflect on the purpose and focus of typographic enhancement. If they see several MWIs enhanced in a text during self-paced reading, they may go back and forth between them, recognize what the enhanced items have in common (e.g., they are all word partnerships), and understand the intention of the instructor or materials designer is to direct their attention to this specific language feature. Discovering the purpose and focus of enhancement will likely take longer in the case of captioned video, because watching a captioned video in real time requires sequential processing of lines of text that disappear from the screen straightway. These factors may explain why enhancement made no difference in Majuddin et al. (2021) when the video was watched only once. Arguably, steps are needed to make learners aware of the purpose and focus of the intervention beforehand. It is this possibility that is explored in the approximate replication we report on below.

Majuddin et al. (2021) also assessed the learners' comprehension and recollection of the content of the video. Captions were found to support comprehension, which is in keeping with conclusions from other studies (Montero Perez et al., 2013, 2014), but this benefit was reduced in the caption condition with typographic enhancement. This is reminiscent of the findings of Choi (2017), which we mentioned earlier, and of some studies about enhancing grammar features as well (e.g., Lee & Huang, 2008). Enhancement may focus learners' attention on specific language features or items in the materials, but this may come at the cost of their engagement with other aspects of the input. Such a trade-off seems likely also when learners are made aware of the purpose and focus of the enhancement.

1.4 Justifications for replication

Majuddin et al. (2021) and the other studies on typographic enhancement mentioned previously are situated in the realm of "incidental" learning, where participants are not informed that a test about certain language items will follow (see, Webb, 2020, for a discussion). If a test is announced in

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

such studies, this usually concerns content comprehension, and it is expected participants will then focus primarily on the content of the text (or video), as they would in "real" life. Of course, learners may suspect that a language-oriented test will follow a certain activity even if the test was not announced beforehand, or they may consider it likely that certain language items or patterns in course materials will be included in the end-of-term summative assessment. A study by Montero Perez et al. (2018) on the benefits of various types of captions (but not typographically enhanced ones) for vocabulary learning found no significant impact of test announcement on learning gains, but it was recognized that, regardless of whether or not they were forewarned of a test, many students may have considered the materials as a source for intentional language learning. An additional reason why test announcement may not make a big difference is that students may not know precisely what kind of test to expect. For example, announcing a "vocabulary test" will not help learners much to determine which specific items in the input materials are likely to be included in the test. Typographic enhancement may help them in this regard, at least if they understand its purpose. If learners know that a test about MWIs will follow, and if they then see MWIs typographically enhanced in texts or in captions, this should be an incentive for them to focus on these items.

This approximate replication study was an integral part of a project comprising two studies that were planned at the same time. The replication part was conducted shortly after the initial study at the same research site and with the same participants. A few years have passed since the initial study was published (in March 2021), allowing an evaluation of its impact so far. At the time of writing the present report (in June 2024), it had been cited 64 times according to Google Scholar. The observation that the initial study was attracting a good amount of attention served as an additional incentive to report the findings of this replication.

Following Porte and McManus (2019, Chapter 5), we consider the new study an APPROXIMATE replication rather than a CLOSE replication, because two variables differ from the original (whereas in a close replication there should be only one). The two differences are (a) the video material and (b) the learners' awareness of the intended language-learning outcome.

The research questions were identical to those of Majuddin et al. (2021), except that this study investigated learning outcomes under what we shall (for simplicity's sake) call intentional learning as opposed to incidental learning:

1. Is there an effect of caption condition (i.e., enhanced captions, standard captions, no captions) on intentional learning of MWEs?
2. Is there an effect of caption condition (i.e., enhanced captions, standard captions, no captions) on content comprehension?
3. Is there an effect of repeated viewing on the learning of MWEs under the various caption conditions?
4. Is there an effect of repeated viewing on content comprehension under the various caption conditions?

As in the initial study, the same learner-participants, the same six viewing conditions (albeit applied to a different video), and the same types of tests were used. However, unlike the initial study, the students were made aware that the viewing activity served the purpose of MWI learning and that a test would follow with a focus on the MWIs encountered in the video.

Throughout the remainder of the article, we adhere as much as possible to the guidelines for reporting replication research outlined in McManus (2022) and Porte and McManus (2019).

2. Method

2.1 Participants

The participants were 126 students (48 males) in higher education in Malaysia, aged 17 to 22. Apart from four students, they had all previously participated in the study that is replicated here. They were first-year and second-year students in fields such as Hospitality Management and Information

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

Technology Systems. Their L1 was Malay or Mandarin. All the participants had learned English at school for at least ten years. Their vocabulary size test (VST) (Nation & Beglar, 2007) scores indicated receptive familiarity with 4,000- to 9,000-word families, with a mean of 6,600. The participants were in the same six intact classes as in Majuddin et al. (2021), and these intact classes were assigned to one of six viewing conditions, which was the same as in the initial study. Two of the classes had the same number of participants as the original study, while the others had one more. There was no significant difference between the classes in their VST scores (Kruskal-Wallis $\chi 2(5) = 4.57$; $p = 0.47$). Schmitt et al.'s (2001) Vocabulary Levels Test (VLT) (Version 2) was also administered. All six classes obtained a mean score of ≥ 25 out of the 30 test items belonging to the 2,000 most frequent word families and of ≥ 22 on the next frequency band (see Appendix 1).

2.2 Materials

As in Majuddin et al. (2021), the video used in this study was an episode from an American sitcom, but it was different from the sitcom selected for the original study. We could not re-use the same video (an episode of *Fresh off the Boat*) because the students had already watched it as participants in the original study. The *Fresh off the Boat* episode is 20-minutes long, and the story's premise centers around how a stern Asian mother uses her "tiger mother" ideals and values to deal with her family business and children. For the present study, an episode of *Raising Hope* was chosen (Episode 7, Season 1). *Raising Hope* tells the story of an American family raising a little girl they named Hope. The selected 21-minute episode is about events around the day of Thanksgiving.

A RANGE (Nation & Heatley, 2002) analysis of the lexical profile of the two episodes did not show substantial differences between them. The most frequent 2,000-word families provided 94.87% and 91.42% coverage (including proper nouns and marginal words) of the total running words of the script for the video in Majuddin et al. (2021) and the video used in the present replication study, respectively. Knowledge of the most frequent 3,000-word families provided 96.37% cumulative coverage of the initial video's script and 94.25% coverage of the video used in the present study. In the case of reading texts, 95% coverage appeared necessary for adequate text comprehension (e.g., Laufer, 2020; Laufer & Ravenhorst-Kalovski, 2010), but for audiovisual input such as TV, the threshold for adequate comprehension was probably lower (e.g., Durbahn et al., 2020).

The original study had 18 target final items. In the present study, 23 MWIs were selected from the episode that were considered highly likely to be unfamiliar to the students. Each occurred once in the episode. According to the pre-test (see below), three of these were already known by over half of the students, and so these three MWIs were excluded from data analysis. The remaining 20 MWIs were (in order of appearance in the episode): *unsung hero, (someone's) day in the sun, play the […] card, make the most of, invest in, bring out the big guns, whip up, spread the word, the dust has settled, in the bag, screw with, take […] lying down, step […] up, have a beef with, small potatoes, have a run-in with, rat out, hopped up, hit the streets* and *[be] caught up in*. As in the original study, this represents a mix of MWIs, varying in formal and semantic characteristics, which is to be expected in authentic discourse. Neither the original study nor the approximate replication intended to examine the learnability of one specific type of MWIs or to compare the learnability of different types.

The format of all the test instruments in the present study was identical to that of the original study. MWI knowledge was gauged by means of form-recall tests. A sentence-based cued gap-fill format was used for the pre-test. For example: "Let's wait until the d_____ has s_____ before we decide what to do. It's best to wait until the situation has calmed down." For each MWI, only the content words had missing letters. The function words were kept intact. The test items were run through RANGE to make sure that all the words used in the sentences belonged to the 3,000 most frequent word families in English.

The cued gap-fill format was also used in the immediate post-test, but this time within the broader verbatim contexts in which the MWIs had been encountered in the sitcom episode, so the learners could possibly make use of episodic memory to recall the expressions. For example: "Virginia

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

[narrating]: When the d_____ has s_____ and the candy was gone, it was time to count up the earnings and see who was going to ride the float."

The delayed post-test, administered two weeks later, was the same as the pre-test, except for the order of the test items (see Appendix 2 for the test items for all 20 target MWIs).

A content comprehension test was given to the students shortly after they had watched the video. The questions were created by drawing on established listening constructs as well as taxonomies of listening skills. One of the most widely accepted descriptions of listening involves the notions of top-down and bottom-up processing (e.g., Buck, 2001). Skills such as identifying facts and other local points of information constitute the ability to perform bottom-up processing. The ability to perform top-down processing, on the other hand, is observed through global skills such as listening for gist and making inferences about contexts and speakers' attitudes. As both types of processing are vital for successful comprehension, we created a comprehension test intended to capture both. All words used in the questions belonged to the 3,000 most frequent word families in English. An initial set of comprehension questions was trialed with three postgraduate students. They first watched the video from start to end, proceeded to answer the questions, and then commented on the difficulty and clarity of the test items. Based on this feedback, the test was revised and trialed with one L1 and one L2 English postgraduate students, both of whom answered all the test items correctly and pointed out no further issues with clarity. The final version of the test consisted of nine multiple-choice questions and 11 true/false questions (see Appendix 3).

Appendix 4 presents a side-by-side comparison of the original study and the present study regarding participant groups, materials, and instruments.

2.3 Procedures

Prior to the start of the two-study project, the participants gave their informed consent, and they took the VLT, the scores on which helped to ascertain that the lexical profiles of the materials in both studies were matched to the participants' vocabulary knowledge. As in the original study, there was a two-week lapse between the MWI pre-test and the treatment, and between the treatment and the post-tests. The VST was also administered right after the MWI pre-test, which included the 20 target MWIs in the present study. Analogous to Majuddin et al. (2021), the students were, in their intact classes, assigned to one of six viewing conditions two weeks after the MWI pre-test: viewing once with enhanced captions ($n = 20$), viewing once with standard captions ($n = 24$), viewing once without captions ($n = 15$), viewing twice with enhanced captions ($n = 25$), viewing twice with standard captions ($n = 23$), and viewing twice without captions ($n = 19$). In the enhanced-captions conditions, the target MWIs were bolded and underlined. After watching the video (either once or twice), the participants first answered the 20 content comprehension questions and then proceeded with the immediate post-test on the MWIs. Two weeks later, they took the delayed MWI test.

2.4 Analysis

Two raters independently scored the responses in the MWI recall tests. Scoring was approached the same way as in the initial study. For test items with only one word missing, a full point was awarded also when the supplied word included a minor mistake, such as using a singular form when it should have been plural. For test items with two missing words, partial credit (0.5) was awarded when one of the supplied words was correct or exhibited only a minor mistake. In three-gap responses, two of the words needed to be correct (or exhibit only a minor mistake) for partial (0.5) credit. Inter-rater agreement was high (0.96, 0.94, and 0.97 for the pretest, immediate post-test, and delayed post-test, respectively), and so the average score awarded by the two raters was used in the statistical analyses. This yielded five possible accuracy levels: 0, 0.25, 0.5, 0.75, or 1. As the pre-test data was non-normally distributed with unequal variance, this data was first transformed using Tukey Ladder of Powers before running a one-way non-parametric ANOVA (Kruskal-Wallis) test. This revealed no significant differences between the groups' pre-test scores (Kruskal-Wallis $\chi 2\ (5) = 4.58$, $p = 0.47$).

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

Two cumulative link mixed models (*clmm*) (Christensen, 2019) were built in R (R Core Team, 2018) for the immediate and delayed post-test data, with performance on the pre-test at the item level included as one of the fixed effects in the statistical modelling. The following were the other fixed effects: caption condition, number of viewings (once or twice), and VST score. The latter was included given that learners with a relatively large vocabulary have an advantage in acquiring new lexical items from reading/listening/viewing activities (e.g., Elgort & Warren, 2014; Majuddin et al., 2021; Montero Perez et al., 2018; Montero Perez, 2020; Noreillie et al., 2018; Peters & Webb, 2018; Pujadas & Muñoz, 2019). The VST scores were centered prior to analysis. No issues of multicollinearity were found between pre-test scores and VST scores. The model comparison procedure for the post-test data was identical to Majuddin et al. (2021).

Starting with a full model that included all the fixed effects as well as the two- and three-way interactions, each factor was incrementally removed. The full model was then compared with the reduced model using likelihood ratio tests to determine the significance level of the removed factor. The comparison returned likelihood ratio statistics with a chi-square distribution. This is reported in the form (LRT $\chi 2$ ($n1$) = $n2$, $p < n3$), where $n1$ = degrees of freedom, $n2$ = likelihood ratio statistic, and $n3$ = p-value. To correct for multiple testing, Bonferroni adjustment was applied so that the level of SIGNIFICANCE was $p < 0.025$ rather than $p < 0.05$. The *emmeans* function in the *emmeans* package was used to locate the differences between the caption conditions (Lenth, 2018).

As regards the content comprehension test, as in Majuddin et al. (2021), all responses were coded as correct or incorrect. The data were analyzed by means of the *glmer* function of the *lme4* package (Bates et al., 2015). The fixed effects included caption condition, number of viewings, and VST score. Model comparisons were performed in the same way as the *clmm* analyses.

We will consider findings from Majuddin et al. (2021) as replicated in this approximate replication when the same factors emerge as significant predictors of the learners' performance on the MWI post-tests and the content-comprehension test.

3. Results and discussion

3.1 Multiword item recall

Table 1 displays the scores on the cued MWI recall tests in this study and in the original study. Similar to the latter, very little learning occurred in the absence of captions. Also similar to the original study, learning was better in the captioned conditions, although (as in the original study) the learning gains diminished by the time the students took the delayed post-test. The best scores on the immediate and the delayed post-tests were obtained by the group who watched the video twice with enhanced captions. This held true only for the immediate post-test in the original study. Different from the original study is the indication that the enhanced captions brought about greater immediate learning gains than the standard captions after a single viewing of the video.

The results of the *clmm* analysis revealed that caption condition, ($\chi 2$ (2) = 30.6, $p < 0.001$), VST score ($\chi 2$ (1) = 26.8, $p < 0.001$), and pre-test score ($\chi 2$ (1) = 64.6, $p < 0.001$) predicted participants' immediate post-test scores. Unlike what was found in Majuddin et al. (2021), repeated viewing did not emerge as a significant predictor, and so that finding is not replicated here. Table 2 shows the output of the best-fit model for the immediate post-test in the present study, including the odds ratios (ORs) as indicators of effect sizes. As for the effect of caption condition, comparisons using the *emmeans* function in the *emmeans* package revealed that enhanced captions were much more likely to yield a score that is one level higher (e.g., a score of 1 instead of 0.75 for the item) compared with the uncaptioned video ($p < 0.0001$). The standard captions were also more likely to bring about a score that is one level higher in the immediate post-test compared with the uncaptioned video ($p = 0.004$), but the odds ratios indicate that the difference was not as pronounced. Importantly, the enhanced captions led to significantly better test performance than the standard captions ($p = 0.01$), which is different from the original study, where the advantage of enhanced captions relative to standard ones did not reach significance. Table 3 shows the model from the original study, for comparison.

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

Table 1. Mean scores (first row), standard deviations (second row), and 95% confidence intervals (third row) per condition for the MWI tests

	The present study			Majuddin et al. (2021)		
	Pre-test	Post-test 1	Post-test 2	Pre-test	Post-test 1	Post-test 2
Enhanced captions x1	2.53 (2.18)	6.30 (4.22)	3.36 (2.26)	1.79 (1.30)	6.15 (3.51)	3.38 (2.41)
	1.22, 2.36	4.61, 7.69	2.32, 4.43	1.20, 2.37	4.57, 7.72	2.29, 4.46
Standard captions x1	3.72 (3.80)	5.75 (4.53)	4.29 (4.42)	3.82 (3.61)	7.34 (4.81)	5.54 (4.28)
	2.39, 5.26	5.43, 9.25	3.83, 7.25	2.34, 5.19	5.37, 9.30	3.78, 7.29
Uncaptioned x1	2.27 (3.22)	3.35 (3.85)	3.53 (3.64)	2.78 (3.41)	3.55 (3.85)	3.72 (3.82)
	1.092, 2.93	7.73, 11.23	2.34, 5.99	1.05, 4.50	1.60, 5.49	1.79, 5.65
Enhanced captions x2	2.63 (1.91)	8.43 (4.87)	5.19 (4.20)	2.01 (2.34)	9.48 (4.47)	4.17 (4.66)
	1.76, 3.59	6.66, 10.13	2.89, 6.02	1.09, 2.92	8.08, 11.58	2.34, 5.99
Standard captions x2	2.89 (2.33)	6.29 (4.55)	3.54 (2.88)	2.68 (2.24)	8.40 (4.25)	4.46 (3.82)
	1.76, 3.59	6.66, 10.13	2.89, 6.02	1.74, 3.61	6.62, 10.17	2.86, 6.35
Uncaptioned x2	2.80 (2.95)	2.93 (2.87)	3.29 (2.89)	2.92 (2.47)	3.89 (3.21)	3.24 (2.91)
	1.81, 4.02	2.44, 5.33	1.93, 4.54	1.78, 4.05	2.40, 5.37	2.07, 4.76

Note: The maximum scores were 20 and 18 in the present study and in Majuddin et al. (2021), respectively. Post-test 1 = immediate; post-test 2 = delayed.

Table 2. Output of best-fit model for the immediate post-test in the present study

Parameter	Estimate	*SE*	*z*	*p*	*OR*	95% CI
Caption condition (Standard captions)	0.95	0.30	3.19	*<0.01*	2.58	1.43, 4.65
Caption condition (Enhanced captions)	1.75	0.30	5.79	*<0.001*	5.75	3.19, 10.35
Pre-test score	0.35	0.04	7.96	*<0.001*	1.42	1.31, 1.54
VST score (centered)	0.04	0.01	5.33	*<0.001*	1.04	1.02, 1.06

Note: Intercept levels: caption condition = uncaptioned; number of viewings = once.

Table 3. Output of best-fit model for the immediate post-test in Majuddin et al. (2021)

Parameter	Estimate	*SE*	*z*	*p*	*OR*	95% CI
Number of viewings (twice)	0.99	0.21	4.83	*<0.001*	2.69	1.77, 4.09
Caption condition (Standard captions)	1.48	0.26	5.73	*<0.001*	4.39	2.66, 7.25
Caption condition (Enhanced captions)	2.05	0.27	7.72	*<0.001*	7.77	4.70, 12.84
Pre-test score	0.47	0.05	9.39	*<0.001*	1.60	1.44, 1.78
VST score (centered)	0.06	0.01	8.83	*<0.001*	1.06	1.04, 1.08

Note: Intercept levels: caption condition = uncaptioned; number of viewings = once.

Table 4. Output of best-fit model for the delayed post-test in the present study

Parameter	Estimate	*SE*	*Z*	*p*	*OR*	95% CI
VST score (centered)	0.04	0.01	5.85	*<0.001*	1.04	0.02, 0.06
Pre-test score	0.79	0.04	17.6	*<0.001*	2.20	0.71, 0.87
Caption condition (Standard captions)	−0.05	0.24	−0.22	*0.82*	0.95	−0.51, 0.41
Caption condition (Enhanced captions)	0.55	0.24	2.32	*0.02*	1.73	0.08, 1.02

Note: Intercept levels: caption condition = uncaptioned; viewings = once.

As to the delayed post-test, strong predictors of successful recall were VST score ($\chi 2$ (1) = 32.3, $p < 0.001$) and pre-test score ($\chi 2$ (1) = 341.2, $p < 0.001$). Enhancement was as an additional predictor (Table 4), but standard captioning was not, which is different from the original study. Table 5 presents the model from the latter study for comparison.

Summing up, the results indicate that typographic enhancement had a positive impact on MWI learning in the present study, while repeated viewing did not emerge as a significant predictor. This is different from the study by Majuddin et al. (2021), where repeated viewing rather than typographic enhancement made the greater impact on MWI learning; this was likely because the enhancement only began to make a difference when the students watched the video twice. In short, this finding of the initial study is not replicated here.

3.2 Content comprehension

Let us now turn to the results of the content comprehension test. As expected, the mean scores were higher in the captioned conditions than in the condition without captions (see Table 6), but this advantage was less pronounced in the case of enhanced captions. This suggests that the presence of typographic enhancement partly compromised the supporting role of captions for content

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

Table 5. Output of best-fit model for the delayed post-test in Majuddin et al. (2021)

Parameter	Estimate	*SE*	*Z*	*p*	*OR*	95% CI
VST score (centered)	0.05	0.01	6.94	*<0.001*	1.05	0.03, 0.07
Pre-test score	1.05	0.10	11.57	*<0.001*	2.86	0.86, 1.24
Caption condition (Standard captions)	0.52	0.27	1.97	*0.04*	1.68	0.00, 1.04
Caption condition (Enhanced captions)	0.84	0.27	3.14	*<0.01*	2.32	0.31, 1.37
Pre-test score x Standard captions	−0.01	0.12	−0.11	*0.92*	1.00	−0.25, 0.24
Pre-test score x Enhanced captions	−0.38	0.12	−3.10	*<0.001*	0.68	−0.62, −0.15

Note: Intercept levels: caption condition = uncaptioned; viewings = once.

Table 6. Descriptive statistics for the content comprehension test

	The present study		Majuddin et al. (2021)	
Condition	Mean (*SD*)	95%CI	Mean (*SD*)	95% CI
Enhanced captions x1	13.47 (2.65)	12.28, 14.66	17.44 (2.28)	16.44, 18.43
Standard captions x1	15.30 (2.55)	14.26, 16.34	17.87 (2.36)	16.93, 18.81
Uncaptioned x1	12.47 (3.02)	10.94, 13.99	16.67 (1.11)	16.10, 17.23
Enhanced captions x2	13.72 (4.08)	12.11, 15.32	18.55 (2.61)	17.53, 19.57
Standard captions x2	15.96 (2.77)	14.80, 17.12	18.00 (2.51)	16.97, 19.02
Uncaptioned x2	11.05 (3.39)	9.48, 12.62	17.57 (1.91)	16.71, 18.43

comprehension. In Majuddin et al. (2021), there was some indication of this side effect as well, but it was confined to the condition where participants watched the video only once.

In the present replication, by contrast, repeated viewing did not appear to make up for the distracting effect of enhancement. It is likely that the participants' awareness of the language-learning goal of the activity made them attend specifically to MWIs at the cost of attending to content both times they processed the material. That said, the *glmer* analysis of the content-comprehension data did not show a significant difference between the caption conditions, which is similar to Majuddin et al. (2021). The analysis did reveal a significant interaction between caption condition and VST score ($\chi 2\ (2) = 12.5$, $p < 0.01$). No such interaction emerged in the original study. To examine this interaction, we broke down the VST scores into ranges from the lowest (38) to highest (112) scores. The mean VST score was close to 70, and so we used ranges with the following values: 40, 50, 60, 70, 80, 90, and 100. The probabilities of correct responses in the comprehension test were then compared for these ranges (Figure 1).

The interaction suggests that, for content comprehension, participants with a low VST score did not benefit from captions that featured typographic enhancement. By contrast, content comprehension by participants with a high VST score benefited from captions regardless of whether they featured enhancement. This finding, when compared with the original study, suggests that learners at a relatively low proficiency level found it particularly hard to process the content of the video when they were presented with enhanced captions AND were aware of the intended learning outcome from the enhancement.

4. Conclusions

4.1 Summary of the findings

This approximate replication of Majuddin et al. (2021) examined the effect of the same six viewing conditions, but this time the English as a second language (ESL) learners were made aware of the

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

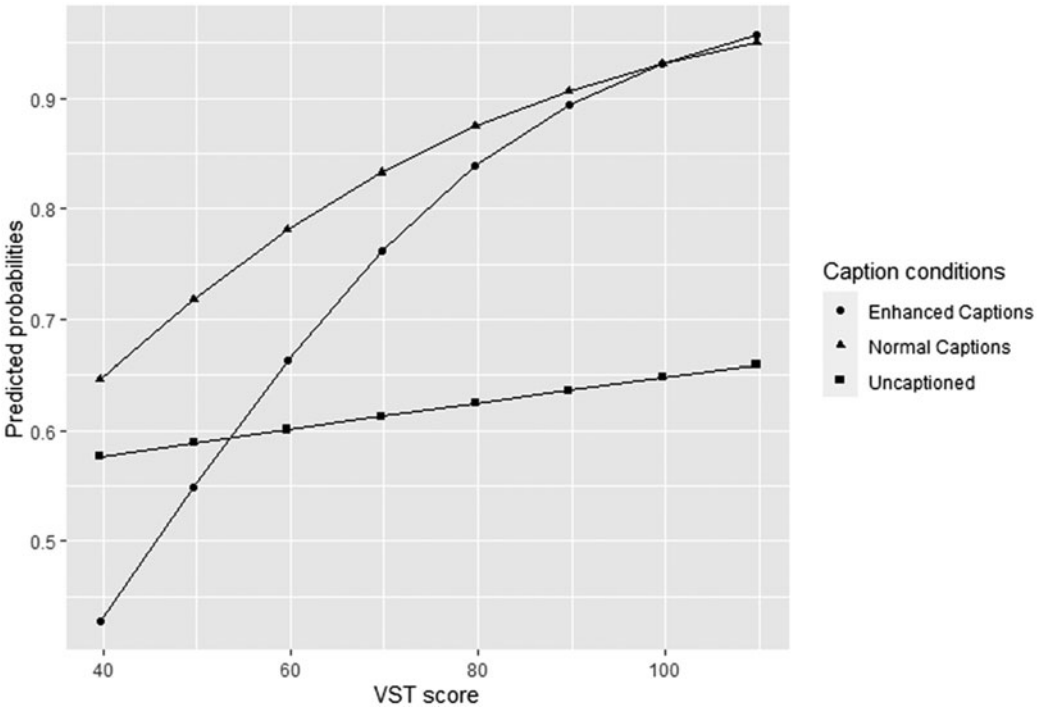

Figure 1. Content comprehension in relation to vocabulary size.

MWI focus of the activity and of the type of test that would follow. Trends found in the original study were replicated, but the strength of these trends differed, so that different factors stood out as significant predictors of test performance. Most notably, whereas repeated viewing was found to be a stronger predictor of learning gains than enhancement in the original study, enhancement emerged as a stronger predictor in the present study. This is probably because enhancement already led to better immediate recall after watching the video once, whereas in the original study the benefits of enhancement for MWI learning only emerged after the repeated viewing. In addition, the present study found indications that enhanced captions can negatively affect content comprehension even when the learners watched the video twice, while in the original study this seemed confined to the condition where the students watched the video only once. The present study further detailed that this downside of enhanced captions affected mostly students with comparatively poor proficiency (going by their VST scores). These students must have found it particularly challenging to process the content of the sitcom episode while they tried to attend to the enhanced MWIs in the captions. Learners with a comparatively large vocabulary size performed well on the content comprehension test regardless of whether captions were enhanced. However, learners with a comparatively poor vocabulary knowledge, and who by that token must have found the video challenging, benefited more from standard captions. The learners' prior vocabulary knowledge, as gauged by the VST, was a strong predictor of both MWI learning and content comprehension. This replicates Majuddin et al. (2021), and it accords with numerous previous studies (e.g., Montero Perez, 2020; Peters & Webb, 2018; Puimège & Peters, 2019; Pujadas & Muñoz, 2019).

A comparison with the findings of the original study needs to remain tentative because a different video and consequently also different target MWIs were used. For example, the average scores on the comprehension test were lower than in the previous study, which could be interpreted as evidence for a greater trade-off between a focus on the language code and a focus on content, but it is also possible that the learners simply found the sitcom episode in the original study easier to follow. The same word

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

of caution applies to a comparison of the MWI learning gains in the two studies – it cannot be ruled out that the target MWIs in one study were more challenging to remember than in the other. While bearing all this in mind, the findings do lend indirect support for the notion that awareness of the intended vocabulary-learning outcome and the anticipation of a test on MWIs influences learners' uptake of MWIs from enhanced captions.

4.2 Possible directions for further (replication) research

If it is true that typographic enhancement can negatively impact learners' intake of content, then a possible procedure to remedy this is to have learners first watch a video with normal captions (to support content comprehension), and subsequently watch it again with enhanced captions (to direct attention to specific language features or items). Future studies could put this scenario to the test by comparing this procedure with a condition where participants watch the same video twice with either standard captions or enhanced captions (see Wi & Boers, 2024, for an example, albeit not focused on enhancement). In a similar vein, more compelling evidence of the effect of test announcement could be obtained from between-participant designs where all learners are exposed to the same input materials but with presence/absence of test-announcement as the independent variable (e.g., Montero Perez et al., 2018). In addition, triangulation with eye-movement data would be helpful to ascertain not only if typographically enhanced items attract attention (e.g., Puimège et al., 2023) but also if the enhancement affects learners' attention to other elements (e.g., Choi, 2023).

Another future direction is to evaluate the effects of typographic enhancement when this intervention is integrated in a larger ensemble of instructional approaches to MWIs – and to other language items and features, for that matter (e.g., Chung & Révész, 2024). For example, an additional way of prompting learners to attend to certain language features or items in texts is to provide explicit instruction about them beforehand (e.g., Cintrón-Valentín & García-Amaya, 2021; Indrarathne & Kormos, 2017; Pellicer-Sánchez et al., 2020; Pujadas & Muñoz, 2019). The research agenda should also include longitudinal interventions that regularly raise learners' awareness of the importance of phraseology with a view to promoting autonomous engagement with MWIs in discourse (e.g., Boers et al., 2006, 2023; Bui et al., 2020; Jones & Haywood, 2004).

Within each of these lines of inquiry, replication research is needed to gauge the generalizability of findings and to identify factors for consideration in future applications. Replication studies are usually inspired by previously published studies that have over time attracted attention, but they can also be planned proactively as an integral feature of a multi-study project, as was the case here.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, *69*, 65–78. doi:10.1016/j.system.2017.08.004

Boers, F. (2021). *Evaluating second language vocabulary and grammar instruction: A synthesis of the research on teaching words, phrases, and patterns*. Routledge.

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Palgrave Macmillan.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, *10*(3), 245–261. doi:10.1191/1362168806lr195oa

Boers, F., Demecheleer, M., He, L., Deconinck, J., Stengers, H., & Eyckmans, J. (2017). Typographic enhancement of multiword units in second language text. *International Journal of Applied Linguistics*, *27*(2), 448–469. doi:10.1111/ijal.12141

Boers, F., Bui, T., Deconinck, J., Stengers, H., & Coxhead, A. (2023). Helping learners develop autonomy in acquiring multiword expressions. *Modern Language Journal*, *107*(1), 222–241. doi:10.1111/modl.12829

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Bui, T., Boers, F., & Coxhead, A. (2020). Extracting multiword expressions from texts with the aid of on-line resources: A classroom experiment. *ITL-International Journal of Applied Linguistics*, *171*(2), 221–252. doi:10.1075/itl.18033.bui

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye-movement study. *Language Teaching Research*, *21*(3), 403–426. doi:10.1177/1362168816653271

Choi, S. (2023). Visual saliency in captioned digital videos and learning of English collocations: An eye-tracking study. *Language Learning & Technology*, *27*(1), 1–21. https://hdl.handle.net/10125/73536

Christensen, R. H. B. (2019). *Regression models for ordinal data (Version 2019.12-10)* [R package]. Retrieved from http://www.cran.r-project.org/package=ordinal/

Chung, Y., & Révész, A. (2024). Investigating the effect of textual enhancement in post-reading tasks on grammatical development by child language learners. *Language Teaching Research*, *28*(2), 632–653. doi:10.1177/13621688211005068

Cintrón-Valentín, M. C., & García-Amaya, L. (2021). Investigating textual enhancement and captions in L2 grammar and vocabulary: An experimental study. *Studies in Second Language Acquisition*, *42*(5), 1068–1093. doi:10.1017/S0272263120000492

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*(5), 570–590. doi:10.1093/applin/amt056

Durbahn, M., Rodgers, M., & Peters, E. (2020). The relationship between vocabulary and viewing comprehension. *System*, *88*. doi:10.1016/j.system.2019.102166

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*(2), 365–414. doi:10.1111/lang.12052

Eyckmans, J., Boers, F., & Lindstromberg, S. (2016). The impact of imposing processing strategies on L2 learners' deliberate study of lexical phrases. *System*, *56*, 127–139. doi:10.1016/j.system.2015.12.001

Garner, J. (2022). The cross-sectional development of verb-noun collocations as constructions in L2 writing. *International Review of Applied Linguistics in Language Teaching*, *60*(3), 909–935. doi:10.1515/iral-2019-0169

Hou, J., Loerts, H., & Verspoor, M. H. (2018). Chunk use and development in advanced Chinese L2 learners of English. *Language Teaching Research*, *22*(2), 148–168. doi:10.1177/1362168816662290

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*(3), 401–430. doi:10.1017/S027226311600019X

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 269–300). John Benjamins.

Jung, J., & Lee, M. (2023). Incidental collocational learning from reading-while-listening and the impact of synchronized textual enhancement. *International Review of Applied Linguistics in Language Teaching*. Epub ahead of print. doi:10.1515/iral-2023-0029

Jung, J., Stainer, M. J., & Tran, M. H. (2022). The impact of textual enhancement and frequency manipulation on incidental learning of collocations from reading. *Language Teaching Research*. Epub ahead of print. doi:10.1177/13621688221129994

Kremmel, B., Brunfaut, T., & Alderson, J. C. (2017). Exploring the role of phraseological knowledge in foreign language reading. *Applied Linguistics*, *38*(6), 848–870. doi:10.1093/applin/amv070

Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, *54*(4), 1076–1085. doi:10.1002/tesq.3004

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*(4), 694–716. doi:10.1093/applin/amn018

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30. https://hdl.handle.net/10125/66648

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*(2), 647–672. doi:10.1111/j.1467-9922.2010.00621.x

Lee, S.-K., & Huang, H.-K. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, *30*(3), 307–331. doi:10.1017/S0272263108080479

Lenth, R. (2018). emmeans: Estimated marginal means, aka least-squares means [R package]. http://CRAN.R-project.org/package=emmeans

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual input: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, *43*(5), 985–1008. doi:10.1017/S0272263121000036

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*(2), 267–290. doi:10.5054/tq.2011.247708

McManus, K. (2022). Replication research in instructed SLA. In L. Gurzynski-Weiss & Y. Kim (Eds.), *Instructed second language acquisition research methods* (pp. 103–122). John Benjamins.

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, *42*(4), 749–773. doi:10.1017/S0272263119000706

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, *55*(1), 163–192. doi:10.1017/S0261444821000501

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, *41*(3), 720–739. doi:10.1016/j.system.2013.07.013

Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, *18*(1), 118–141. https://hdl.handle.net/10125/44357

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, *31*(1), 1–26. doi:10.1080/09588221.2017.1375960

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13. https://jalt-publications.org/files/pdf/the_language_teacher/07_2007tlt.pdf

Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts. Retrieved from http://www.vuw.ac.nz/lals/staff/paulnation/nation.aspx

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*(2), 223–242. doi:10.1093/applin/24.2.223

Noreillie, A. S., Kestemont, B., Heylen, K., Desmet, P., & Peters, E. (2018). Vocabulary knowledge and listening comprehension at an intermediate level in English and French as foreign languages: An approximate replication of Stæhr (2009). *ITL-International Journal of Applied Linguistics*, *169*(1), 212–231. doi:10.1075/itl.00013.nor

Paquot, M. (2018). Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly*, *15*(1), 29–43. doi:10.1080/15434303.2017.1405421

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, *35*(1), 121–145. doi:10.1177/0267658317694221

Pellicer-Sánchez, A., & Boers, F. (2019). Pedagogical approaches to the teaching and learning of formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 153–173). Routledge.

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2020). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning*, *71*(1), 162–203. doi:10.1111/lang.12430

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113–138. doi:10.1177/1362168814568131

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*(4), 1008–1032. doi:10.1002/tesq.531

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), 551–577. doi:10.1017/S0272263117000407

Porte, G. K., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Puimège, E., & Peters, E. (2019). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning*, *69*(4), 943–977. doi:10.1111/lang.12364

Puimège, E., Montero Perez, M., & Peters, E. (2023). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study. *Second Language Research*, *39*(2), 471–492. doi:10.1177/02676583211049741

Puimège, E., Montero Perez, M., & Peters, E. (2024). The effects of typographic enhancement on L2 collocation processing and learning from reading: An eye-tracking study. *Applied Linguistics*, *25*(1), 88–110. doi:10.1093/applin/amad003

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*(4), 479–496. doi:10.1080/09571736.2019.1616806

R Core Team. (2018). R: A language and environment for statistical computing (Version 3.4.4) [Computer software]. Retrieved from https://www.R-project.org

Saito, K. (2020). Multi- or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech. *Language Learning*, *70*(2), 548–588. doi:10.1111/lang.12387

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

Schmitt, N. (Ed.) (2004). *Formulaic sequences*. John Benjamins.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, *18*(1), 55–88. doi:10.1177/026553220101800103

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*(1), 165–179. doi:10.1017/S0272263100011943

Siyanova-Chanturia, A., & Omidian, T. (2020). Key issues in researching multi-word items. In S. A. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 529–544). Routledge.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. Routledge.

Siyanova-Chanturia, A., & Spina, S. (2020). Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study. *Language Learning*, *70*(2), 420–463. doi:10.1111/lang.12383

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121–159. doi:10.1111/j.1467-9922.2012.00730.x

Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2011). Formulaic sequences and L2 oral proficiency: Does the type of target language influence the association? *International Review of Applied Linguistics in Language Teaching*, *49*(4), 321–343. doi:10.1515/iral.2011.017

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*(2), 245–265. doi:10.1111/ijal.12092

Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning*, *70*(2), 506–547. doi:10.1111/lang.12384

Teng, M. F. (2021). *Language learning through captioned videos: Incidental vocabulary acquisition*. Routledge.

Terai, M., Fukuta, J., & Tamura, Y. (2023). Learnability of L2 collocations and L1 influence on L2 collocational representations of Japanese learners of English. *International Review of Applied Linguistics*. Epub ahead of print. doi:10.1515/iral-2022-0234

Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, *69*(2), 405–439. doi:10.1111/lang.12335

Toomer, M., Elgort, I., & Coxhead, A. (2024). Contextual learning of L2 lexical and grammatical collocations with and without typographic enhancement. *System*, *121*. doi:10.1016/j.system.2024.103235

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Ablex Publishing Corporation.

Vu, V. V., & Peters, E. (2022a). Learning vocabulary from reading-only, reading-while-listening, and reading with textual input enhancement: Insights from Vietnamese EFL learners. *RELC Journal*, *53*(1), 85–100. doi:10.1177/0033688220911485

Vu, V. V., & Peters, E. (2022b). Incidental learning of collocations from meaningful input a longitudinal study into three reading modes and factors that affect learning. *Studies in Second Language Acquisition*, *44*, 685–707. doi:10.1017/S0272263121000462

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Webb, S., Newton, J., & Chang, A. C.-S. (2013). Incidental learning of collocation. *Language Learning*, *63*(1), 91–120. doi:10.1111/j.1467-9922.2012.00729.x

Wi, I., & Boers, F. (2024). Sequential use of L1 and L2 captions: Exploring the benefits for vocabulary acquisition. *TESOL Quarterly*, *58*(1), 511–521. doi:10.1002/tesq.3243

Wood, D. (2010). *Perspectives on formulaic language: Acquisition and communication*. Continuum.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Zhang, P., & Graham, S. (2020). Vocabulary learning through listening: Comparing L2 explanations, teacher codeswitching, contrastive focus-on-form and incidental learning. *Language Teaching Research*, *24*(6), 765–784. doi:10.1177/1362168819829022

Appendix 1. Scores on the vocabulary size test (VST) and the vocabulary levels test (VLT)

VST scores (out of 140)

Condition	*N*	Mean (*SD*)
Enhanced captions x1	20	68.90 (11.3)
Normal captions x1	24	73.70 (18.27)
Uncaptioned x1	15	74.40 (16.63)
Enhanced captions x2	25	65.92 (16.75)
Normal captions x2	23	70.17 (12.06)
Uncaptioned x2	19	66.05 (13.70)

Note: x1 = one viewing; x2 = two viewings.

VLT scores (out of 30 per level)

	VLT 2000		VLT 3000		VLT 5000	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Enhanced captions x1	26.00	3.51	21.89	4.46	16.32	4.62
Normal captions x1	26.70	3.24	24.35	5.31	18.13	8.92
Uncaptioned x1	27.13	2.97	24.00	5.07	18.07	6.63
Enhanced captions x2	25.60	3.52	22.00	5.55	15.96	7.38
Normal captions x2	26.91	5.02	24.27	5.00	18.05	5.91
Uncaptioned x2	25.59	4.91	22.11	5.70	15.67	5.93

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

Appendix 2. Pre-test and delayed post-test items (target MWIs)

In each of the following questions, there is one phrase with missing letters. Look at the context and fill in the blanks with the missing letters.

1. I can wh_____ up a meal in no time.
2. Teachers are the uns_____ he_____ of a great writer's success. They are often not noticed or praised for their hard work.
3. We are only in Paris for a day, so let's m_____ the m_____ of it. We should enjoy our day as much as possible.
4. We have arranged a meeting for next Thursday, so if you see anyone, do sp_____ the w_____. We have to inform everyone.
5. When the score got to 8-2, we knew the game was in the b_____. We knew we were going to be the champions.
6. Let's just wait until the d_____ has s_____ before we decide what to do. It's better to make a decision when the situation has calmed down.
7. Someone has been sc_____ with my computer, and now it doesn't work anymore.
8. The gap between the rich and the poor is wide, and the poor aren't going to t_____ it ly_____ d_____. They are going to be more violent.
9. After the recent bombing attacks, airports in major cities around the world have installed more cameras to st_____ up security.
10. I don't have anything against advertising, but I do have a b_____ with how many bad advertisements there are on TV.
11. We i_____ a lot in Tom, so we have every right to expect a lot from him. We devoted a lot of time and effort in training him to be a professional athlete.
12. He had a r_____ in with his boss. The argument caused him to lose his job.
13. My brother came home drunk so I r_____ him out to my mother. I told my mother that he had been sneaking out at night.
14. You have so much energy on your television shows, I always suspect you must be ho_____ up on energy drinks.
15. I was so c_____ up in my school work that I didn't realise what was happening with my sister.
16. The cooking competition is designed to give home cooks their d_____ in the s_____. The home cooks will finally get the attention they deserve.
17. The candidate pl_____ the race c_____, claiming that she received less attention than the Malay candidate simply because she is Chinese.
18. After failing to convince the IT department that new security passwords are needed, Mike felt it was time to br_____ out the b_____ g_____. So he called a meeting with the Head of the Company.
19. The new iPhone model is confirmed to h_____ the str_____ at the end of 2017.
20. Last week's rain was sm_____ pot_____ compared to the thunder storm we had two months ago.

Appendix 3. Content comprehension test

The following questions are about the video you have just watched. Circle the correct answer.

1. What did the high school donate to the Natesville Auxiliary (volunteer) Police?
   A) Radishes
   B) Whistles
   C) Chocolate
2. Why did Jimmy want Hope to become the little pilgrim?
   A) To boost her confidence for potty (toilet) training
   B) To make her the most popular child
   C) To prove that Virginia was not cheating
3. Why was young Jimmy not allowed in the pool?
   A) He was being naughty
   B) He did not sell any chocolate bars
   C) He did not win the contest
4. Why did young Virginia fail to sell the original chocolate bars?
   A) She did not try hard enough
   B) Rosa's family stopped her from selling the bars
   C) The original chocolate bars did not taste good
5. At first, Virginia did not want to help Hope win because
   A) She wanted Rosa's grandson to win
   B) She did not want to burden Maw Maw

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

C) She knew winning would make another child sad

6. When young Virginia just started working with Knock Knock Knock Housekeeping, young Jimmy also started his long-life affair with
   A) Chocolate
   B) Fire and explosives
   C) Growing radishes
7. What did Maw Maw do when Burt found the horse's head in his bed?
   A) She made her chocolate stronger
   B) She destroyed the Flores family's chocolate bars
   C) She sold her chocolate bars at a cheaper price
8. Why was Virginia angry at Sabrina?
   A) Sabrina started stealing Maw Maw's Magic Brown
   B) Sabrina did not help Jimmy win the contest
   C) Sabrina did not sell enough Maw Maw's Magic Brown
9. What did Barney want Jimmy to do?
   A) Beat his mother
   B) Wear a wire
   C) Confess to cheating

The following statements are about the video you have just watched. If the statement is true, circle the word True. If the statement is false, circle the word False.

10. Burt named the horse Clip Clop.	True/False
11. Maw Maw thought the cupboard she was hiding in was a toilet.	True/False
12. The Chance family played the violin badly.	True/False
13. Maw Maw's Magic Brown secret ingredient was cocoa powder.	True/False
14. The Carlos family hit Burt with water guns.	True/False
15. As an auxiliary (volunteer) officer, Barney had the legal right to arrest a person.	True/False
16. Virginia could not stop herself from eating Maw Maw's Magic Brown.	True/False
17. Hope was successfully potty trained after becoming the little pilgrim.	True/False
18. Dating Sabrina gave Jimmy the confidence to grow both Ricky Radish and Sabrina Squash.	True/False
19. Frank thinks that Barney will not do what he threatens to.	True/False
20. Barney was much fatter 20 years ago.	True/False

Appendix 4. Side-by-side comparison of participant groups and materials

	Majuddin et al. (2021)			This approximate replication			Notes
Number of participants and mean VST scores (out of 140). The students were the same in both studies, except for four students who were absent at the time of the original study. Intact classes	Condition	n	Mean (SD)	Condition	n	Mean (SD)	X1 = one viewing; X2 = two viewings
	Enhanced captions X1	19	68.37 (11.31)	Enhanced captions X1	20	68.90 (11.30)	
	Standard captions X1	23	73.04 (18.38)	Standard captions X1	24	73.70 (18.27)	
	Uncaptioned X1	15	74.40 (16.63)	Uncaptioned X1	15	74.40 (16.63)	

(*Continued*)

https://doi.org/10.1017/S0261444824000296 Published online by Cambridge University Press

(*Continued*)

	Majuddin et al. (2021)			This approximate replication			Notes
were assigned to the same conditions in both studies.	Enhanced captions X2	25	65.92 (16.75)	Enhanced captions X2	25	65.92 (16.75)	
	Standard captions X2	22	71.09 (11.49)	Standard captions X2	23	70.17 (12.06)	
	Uncaptioned X2	18	66.39 (14.02)	Uncaptioned X2	19	66.39 (13.07)	
Videos	Title: *Fresh off the Boat* (Season 1, Episode 2); 20 min			Title: *Raising Hope* (Season 1, Episode 7); 21 min			
Number of pre-test items	25			27			Includes filler items
Number of post-test items	20			23			Filler items excluded
Number of items retained for analysis	18			20			Familiar MWIs excluded
Number of comprehension questions	22			20			

Elvenna Majuddin currently serves as a Senior Learning Adviser at Te Taiako Student Learning, Te Herenga Waka–Victoria University of Wellington, New Zealand. She earned her Ph.D. in Applied Linguistics from the same institution. Her research focuses on the learning of multi-word expressions through video watching, exploring various caption conditions and their impact on both incidental and intentional learning settings. The first study undertaken as part of her doctoral research was published in *Studies in Second Language Acquisition*.

Frank Boers was a language teacher and teacher trainer in his home country, Belgium (1988–2010), before he joined Te Herenga Waka–Victoria University of Wellington, New Zealand (2010–2018) to teach courses in Applied Linguistics. He is now a Professor at the Faculty of Education of Western University, Canada. He publishes mostly on matters of second language education. His latest books are *Evaluating second language vocabulary and grammar instruction* (2021) and (with L. Zwier) *English L2 vocabulary learning and teaching* (2023), both published by Routledge.

Anna Siyanova-Chanturia is Associate Professor in Applied Linguistics at Te Herenga Waka–Victoria University of Wellington, New Zealand. Anna's research interests include psychological aspects of second language acquisition, bilingualism, usage-based approaches to language acquisition, processing and use, vocabulary and multi-word expressions, as well as quantitative research methods (e.g., corpus research and eye-movement research).

Cite this article: Majuddin, E., Boers, F., & Siyanova-Chanturia, A. (2024). The effects of enhancing L2 multiword items in captions: An approximate replication of Majuddin, Siyanova-Chanturia, and Boers (2021). *Language Teaching* 1–18. https://doi.org/10.1017/S0261444824000296