Does the Timing of Visual Support Affect Sentence Comprehension? An Eye-Tracking Study

Miriam Rivero-Contreras[1], Paul E. Engelhardt[2], Pablo Delgado[1], and David Saldaña[1]

[1]*Universidad de Sevilla, Sevilla, Spain*

[2]*University of East Anglia, Norwich, UK*

Address for correspondence: mrivero5@us.es
Miriam Rivero-Contreras

Author Note:

Miriam Rivero-Contreras is now at the Department of Psychology and Anthropology, University of Extremadura, Spain.

Does the Timing of Visual Support Affect Sentence Comprehension? An Eye-Tracking Study

Purpose: Recent research suggests that visual elements improve sentence processing for students, even at the university level. However, few studies have systematically examined the timing of visual support in reading. **Method**: We examined the impact of visual support and its timing on sentence comprehension in a sample of 40 typically developing university students. Across 60 sentences, half with images and half without, participants either viewed images simultaneously with sentences or before sentences. Word frequency was also manipulated. **Results**: Results showed that visual support facilitated sentence processing and that participants who viewed images before sentences exhibited a lower probability of regressions. **Conclusion**: In conclusion, incorporating images with text can benefit language comprehension. Moreover, the results suggest implications regarding the timing of visual support.

Keywords: visual support; image timing; sentence processing; eye movements

Introduction

Visual Support and Reading

Many reading books for young children consist of a single (simple) sentence accompanied by an image described by the sentence. Pictures in these books, aimed at pre-readers and young children, play an important role in enhancing text comprehension (Strouse et al., 2018), improving vocabulary learning (Wasik et al., 2016), and fostering parent-child interaction during shared reading (Dowdall et al., 2020; Fletcher & Reese, 2005). However, although pictures do not disappear from books for older children and adults, the role of visual support for narrative text comprehension beyond early primary school becomes increasingly smaller. In fact, most pre-teen reading books do have illustrations and pictures but, in most cases, they are considered text “enrichment” rather than text “support”. However, illustrations and pictures can serve to provide knowledge to older students who are expected to read about things that are not part of their daily experience (Hibbing & Rankin-Erickson, 2003). Visual support has also been shown to facilitate reading in cases of low-frequency or unfamiliar words, and when the text contains complex syntactic structures (Montag et al., 2015). In one particularly influential study, Brookshire et al. (2002) found better reading comprehension for books with text and illustrations, compared to illustrations with no text or text with no illustrations. More recently, Takacs and Bus (2018) showed that a congruent image contributed substantially to children's story retelling accuracy, suggesting that co-present visual representations lead to deeper/better encoding of linguistic information (Mayer, 2009).

Visual Support for Adults

It is generally believed that visual support is not necessary for understanding written information in adults (outside learning materials specifically designed in a multimedia format), because they are (in most cases) experienced readers. However, some recent research suggests that images do facilitate lexical access even in adults (Qu et al., 2016). A recent meta-analysis examined 39 experimental studies (21 with college students or older adults) published between 1985 and 2018 (Guo et al., 2020). The authors found that the inclusion of graphs had a moderate overall positive effect (*Hedges' g* = 0.39) on students' reading comprehension regardless of grade level (elementary, secondary, and undergraduate students and above), suggesting that visual elements improve reading comprehension for all students. Likewise, they found no significant differences between pictures, pictorial diagrams, and flow diagrams. In addition, in two prior studies from our lab, we found that visual support facilitated sentence processing in groups of adult readers with reading difficulties (dyslexia and a low level of education) and without them, even at university level (Rivero-Contreras et al., 2021, 2023).

Many studies carried out within the framework of the cognitive theory of multimedia learning also support the view that people learn better from words and images than from words alone (Mayer, 2009). However, cognitive limitations must be carefully considered (Mayer & Moreno, 2003). Multiple information sources, such as a text and an image, being processed in the same (visual) modality must be carefully designed to avoid cognitive overloading, split attention, and unnecessary incidental processing. But most studies under this model have focused on materials with substantial and converging overlap in content between different visual and auditory content. In our case, we explore the role of pictures as used in many narrative texts, in which the visual information rather provides a *context* for comprehension of written information.

Although there exists a considerable amount of research focused on the enhancing effect of the presence of visual support on text comprehension, as noted in a review by Eitel and Scheiter (2015), research in this field has yielded inconclusive results. Some studies have found that presenting a picture prior to textual information enhances comprehension, whereas other found the opposite. In addition, little work has directly experimentally examined whether the timing of presentation of pictorial support plays any moderating role in the enhancing effect of picture on text comprehension. Except for some studies focused on complex textual and pictorial materials in which the textual information is needed to understand a graph or figure, or vice versa, in the context of multimedia learning, to date, there is a lack of studies that have manipulated experimentally the timing in which supporting pictures are presented (i.e., whether the picture precedes the text or both elements appear concurrently. An exception is a study by Eitel et al. (2013), in which the information was presented to participants in six conditions. One condition contained only text. Four conditions varied the length of picture presentation before the text: 150 milliseconds, 600 milliseconds, 2 seconds, or participant determined (i.e., self-paced). Finally, there was a self-paced concurrent presentation of text and picture. The authors found that the self-paced before and self-paced concurrent conditions led to better memory performance, better comprehension, and faster processing compared to the text alone condition. In addition, they found that presenting the image 600 milliseconds or 2 seconds before the text improved comprehension and resulted in faster reading. Therefore, images which are presented before the text for even a short period of time can facilitate processing of verbal information, and thus, promote better comprehension.

Nevertheless, Eitel et al. (2013) used offline measures, which do not provide an accurate measure of text processing itself. In this regard, as also noted by Eitel and Scheiter (2015), there is a need for analysing processing data to explore the interaction between text and picture. Also except for the study of multimedia learning (for a review, see Alemdag & Cagiltay, 2018), most of the studies in the use of visual support to understand text have examined offline data, typically comprehension outcomes. Processing data is crucial not only to examine the existence of any effect, but also to provide an explanation for it. To fill this gap, Eitel and Scheiter called for increasing research analysing readers' eye movements during the reading course. Two exceptions to the lack of this type of research are the studies by Ferreira et al. (2013) and, more recently, by Wassenburg et al. (2020).

Ferreira et al. (2013) established that a preview of objects is critical to the observance of garden-path effects in the visual world paradigm. In their study, participants were often misled by *Put the apple on the towel in the box,* into initially thinking that they must move the apple onto a towel (as evidenced by frequent looks to the towel in the object array). However, this pattern of eye movements was only observed when objects were previewed before the sentence was heard, but not in a concurrent (or simultaneous) viewing condition. The authors argued that the preview of visual information (consecutive presentation) allowed comprehenders to generate certain expectations or predictions regarding the upcoming linguistic information (i.e., an effect of one information source on another). In contrast, when both information sources have to be processed simultaneously, there is competition for attention, and the incremental processing that would normally occur for either visual or linguistic information is disrupted. Incremental processing for linguistic input (visual or auditory) typically occurs one word at a time and after each word the reader/listener derives an interpretation of the input to that point (Altmann & Kamide, 1999).

Importantly, the study by Ferreira et al. (2013) did not explore the processing of texts but of auditory information. In the study by Wassenburg et al. (2020), the authors examined whether a picture presented before reading facilitates comprehension of a 404-word narrative text by fostering situation model construction, as compared to the same text without the picture. The authors reported that the eye-movement patterns of the participants who visualized the supporting picture reflected increased effort in processing information to generate inferences, whereas those who did not visualize the picture were more focused on processing literal information. Although there were no differences in text comprehension outcomes between both conditions (for tentative explanations see Wassenburg et al., 2020), this result indicates that visualizing a supporting image shapes subsequent reading behaviour.

Building on the studies above, we examined the impact of lexical frequency (high vs. low) and visual support that was presented simultaneously with the sentence (image vs. no image) on sentence processing, specifically in individuals with low educational levels and with dyslexia (Rivero-Contreras et al., 2021, 2023). Word frequency is a lexical variable that has been extensively studied using eye tracking (e.g., Ashby et al., 2005; Hyönä & Olson, 1995; Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Rayner & Fischer, 1996; Rayner et al., 1996; Staub et al., 2010). In short, more frequent words (high frequency) are read/processed faster than low frequency words. Consistent with this, our prior work showed a large effect of word frequency, particularly between less skilled readers (low education and dyslexics) and typically developing controls; and increased lexical access and word integration speed, along with a reduction in the number of regressions, facilitated by visual support (Rivero-Contreras et al., 2021, 2023).

Thus, supporting images seem to positively affect text processing. However, both in the study by Wassenburg et al. (2020) and in those by Rivero-Contreras et al. (2021, 2023), the presentation timing of the pictures was not manipulated, as they were presented only prior or simultaneously to the text, respectively. To fill this gap in research and provide cumulative evidence on online text processing, we conducted a study in which we experimentally compared the effect of presenting supporting images prior vs concurrent with the text on readers' eye movements.

Current Study

In the present study, we aimed to build on the lines of research presented above (Eitel et al., 2013; Eitel & Scheiter; 2015; Rivero-Contreras et al., 2021, 2023); Wassenburg et al., 2020) We manipulated word frequency (high vs. low) and visual support (image vs. no image) in a sentence-reading task, which involved eye tracking. Importantly, we also manipulated a third between-subjects variable related to the timing of the images. Half of the participants viewed the image (for 1 second) prior to the sentence (consecutive condition), and half of the participants viewed the image simultaneously with the sentence (concurrent condition). The manipulation of the timing of visual support was based on the idea that previewing a scene may impact differently on text comprehension processes than having it present at the same time as written content, as found by Eitel et al. (2013) in offline comprehension outcomes.

We suggest that if linguistic and contextual visual information must be processed simultaneously, then incremental processing could be disrupted, leading to a more superficial and piecemeal processing strategy. It is important to understand that in our case pictures provide a context for interpretation, rather than specific converging content. If linguistic and visual information are processed consecutively, then there is much less risk of unnecessary incidental processing during reading the text, and performance might be enhanced. In short, we hypothesized that the timing of the presentation of visual support (the between-subjects variable) would lead to significantly shorter sentence reading times and a lower probability of regressions out in consecutive compared to concurrent visual support.

In the current study, we recorded eye movements of typically developing university students while they read four blocks of 15 sentences (60 in total for the whole experiment). As we have indicated, the three variables (word frequency, visual support, and timing) were manipulated in a 2 × 2 × 2 design. To examine visual support, half the sentences contained an image related to the sentence, while the other half did not contain an image. To examine word frequency (lexical simplification), half the sentences contained a high-frequency target word and half a low-frequency target word. (We expected faster reading times on high-frequency words). Visual support (image vs. no image) and lexical frequency (high vs. low) were manipulated within subjects. To examine the impact of the timing of visual support, half the participants viewed the image simultaneously with sentence (concurrent presentation), while the other half were presented with the image before the sentence (consecutive presentation). The effect of timing was only evaluated in the sentences including an image (i.e., 30 items in total). The possible confounding effects of participants' age, level of receptive vocabulary, and reading comprehension skill were controlled by including them as covariates when they showed to be significant predictors of any of the dependent variables (i.e., eye-movement measures).

Two final points need to be addressed. The first concerns the dependent measures. We examined the reading times and regressions for the entire sentence, which we refer to as the global text processing measures, and we also examined reading times and regressions for the manipulated (high vs. low) target word within each sentence, which we refer to as local text processing measures. The second is to further explicate the mechanism by which we expected the timing of visual support to facilitate processing. Given a 1-second preview of the visual scene, participants can determine the objects in the scene and their relations. Moreover, they very likely access the names of at least some of the objects in the scene, which is accomplished in part by the activation of semantic knowledge. There is also the possibility of constructing a partial situation model or event representation based on proposition-level knowledge (Kintsch, 1998; McRae et al., 2021). The extent of this processing will naturally vary by item and individual. However, the more representations that are activated/constructed upon viewing the picture will naturally lead to more predictions/expectations about the content of the upcoming sentence (i.e., there should be a facilitation effect of visual information on sentence processing).

Consistently with previous studies, and our expectations with respect to the timing in the sentences with an image, we expected shorter reading times both in global and local measures, less regressions, and greater accuracy for sentences with high-frequency target words (versus low frequency), images (versus no image), and consecutive (versus concurrent) presentation of the image when it was included.

Method

Participants

Forty typically developing university students took part in this study. Participants were assigned to two equal sized groups, which related to which visual support timing condition they were in (concurrent or consecutive). We assessed vocabulary for all participants to ensure that there were no vocabulary differences between groups (see Table 1). Reading experience was also evaluated using the author recognition test (ART). All participants were native speakers of British English with normal or corrected-to-normal vision. They were recruited through a Research Participation System at the University and compensated with psychology research-pool participation credits for taking part in the study.

Table 1. Summary of participants' age and scores on PPVT-4, ART, and Inferential comprehension by Timing conditions.

	Concurrent timing (*n* = 20)		**Consecutive timing (*n* = 20)**		***t* value**
	M (SD)	**Min-Max**	***M (SD)***	**Min-Max**	
Age[1]	19.90 (1.25)	18-22	19.05 (1.05)	18-21	2.36*
PPVT-4[2]	101.45 (8.99)	79-117	99.05 (9.91)	79-120	0.80
ART	9.70 (7.50)	0-30	6.35 (3.47)	0-13	1.81·
ICS (%)	84.49 (7.94)	64.82-94.57	84.25 (5.52)	72.75-94.39	0.11

Note. PPVT-4: Peabody Picture Vocabulary Test; ART: Author Recognition Test; ICS: Inferential Comprehension (true-or-false statement from experimental task). [1]In years. [2]Standard scores. ·$p < .10$, *$p < .05$.

Materials

Peabody Picture Vocabulary Test - 4 (PPVT-4)

The PPVT-4 (Dunn & Dunn, 2007) assesses receptive vocabulary. It has two forms, of which we used Form A. Participant were asked to select one of four images best illustrating a target word verbally presented by the researcher. The test was administered individually and took around 15 minutes to complete. The manual reports a reliability range from .89 to .97 for Form A.

Author Recognition Test (ART)

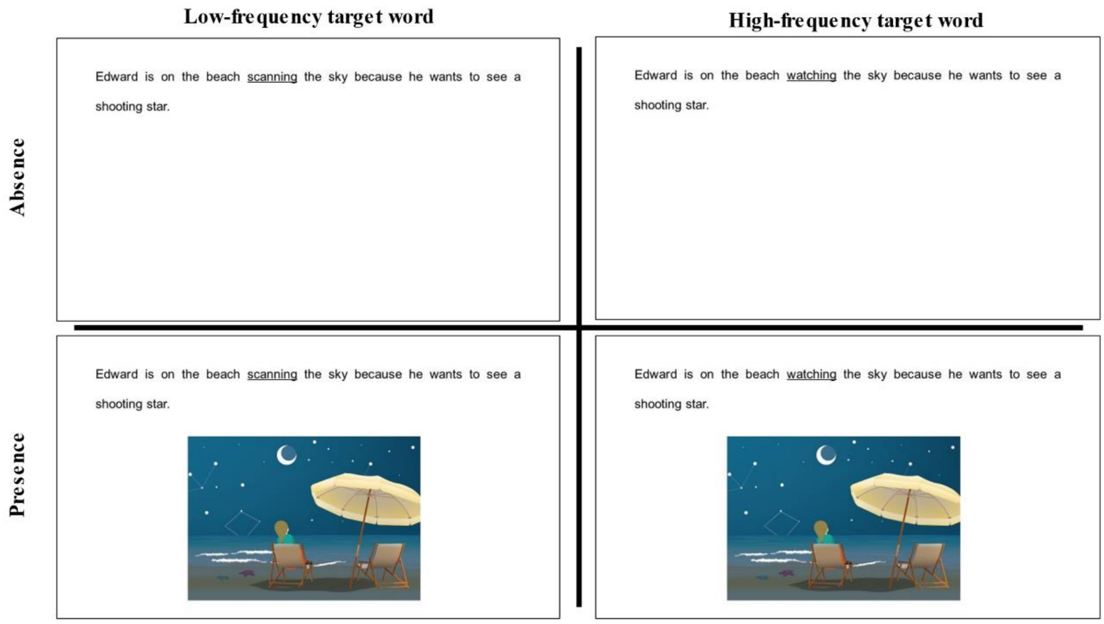
The ART is an indicator of reading experience, which is strongly related to reading skill. It contains 65 literary author names from the Acheson et al. (2008) version of the test, along with 65 additional names, which do not refer to known authors. ART scores have been shown to predict the speed with which university students decode words during reading (Moore & Gordon, 2015). The participant has to identify the names of authors listed in alphabetical order of surname. The test was administered on paper, and participants were asked to mark with a cross the names they recognized as authors. They were informed that to mark non-authors was penalized with one point for each error. The administration of the task typically took around 5 minutes. The internal reliability for the ART ranging from .75 to .89 (Mol & Bus, 2011).

Sentence Comprehension Task

The experiment consisted of 60 short sentences, each containing a key target word. After participants read each sentence, they were presented with an auditory inferential comprehension statement, which they needed to respond to by indicating where the statement was true or false. This was intended to assess comprehension of each sentence.

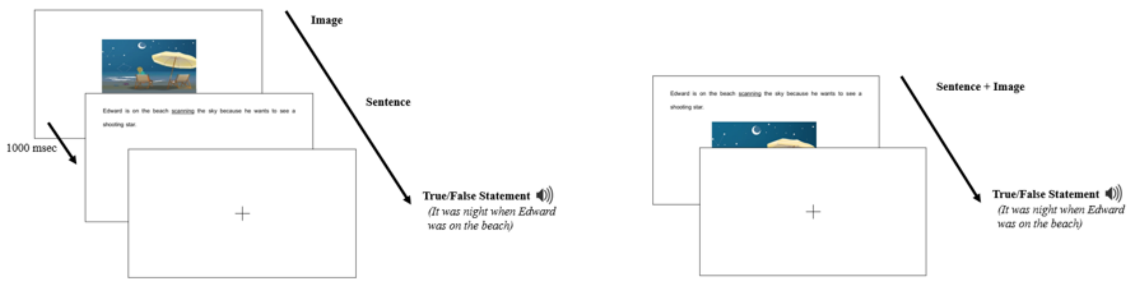
Within-Subject Variables: Half of the sentences were accompanied by an image, which was related to the content of the sentence, and half the sentences were not accompanied by an image. Half the sentences contained a low-frequency target word, and half a high frequency target word. The low- and high-frequency target words were similar in length ($n \pm 1$). We used the Word Frequencies in Written and Spoken English, based on the British National Corpus, to check the frequencies of these word (Leech et al., 2001). The frequency of the low-frequency words was 14.10 ($SD = 55.78$), and the frequency of the high-frequency words was 94.60 ($SD = 275.04$). The difference was statistically significant $t(59) = -2.74$, $p < .01$, $d = 0.41$. Images and target words were rotated in a Latin Square Design resulting in four lists of items. The four within subject conditions were presented in a blocked design with 15 trials in each block (60 items in total; see Figure 1).

Figure 1. Example item showing each of the four conditions. The underlined word was the target word, but was not underlined in the experiment.



Between-Subject Variable: The between-subjects manipulation varied the timing of the image. In the concurrent condition, participants were presented with the sentence and image at the same time (simultaneously), and in the consecutive condition, participants were presented with the image before the sentence. In the concurrent condition (Concurrent Picture or CP), the image was positioned below the sentence. In the consecutive condition (Picture Before or PB), the image was shown for 1 second, and then it disappeared. The sentence then appeared on the top of the screen (see Figure 2). Note that in the concurrent condition, there was no time limit on viewing the sentence or the image (i.e. it was self-paced). However, in the consecutive condition the viewing time of the image was limited to 1 second, but the viewing time of the sentence was self-paced.

Figure 2. Examples of the presentation orders in both versions of the experiment. Abbreviations: PB, Picture Before; CP, Concurrent Picture.



PB CP

Apparatus

Eye movements were recorded with an EyeLink 1000 eye tracker, sampling at 1000 Hz (SR Research, Ontario, Canada). Viewing distance was 70 cm from eyes to a 45-cm computer monitor, and at this distance, 1.0° of visual angle subtended 1.22 cm. This apparatus allows recording of eye movements through a camera with an infrared tracking system while the person is carrying out the experimental task. Head movements were minimised with a chin rest, and eye movements were recorded from the right eye. The sentences were presented in 20-pt. arial font on a white background, and pictures were in colour. A 9-point calibration procedure and the standard Eyelink error tolerances for "good" were used.

Design and Procedure

Participants completed two practice trials, followed by the 60 experimental trials. Trials were presented in random-ordered blocks of 15 items of the same condition (No picture and Low frequency; Picture and Low frequency; No picture and High Frequency; Picture and High Frequency). Within each block, trials were randomised. Each participant was assigned one of four lists, so each participant only read each sentence in one condition, and across participants, sentences were presented in all conditions. Half the participants were assigned to the concurrent condition and half were assigned to the consecutive condition.

First, the participant read a set of instructions with the details of procedure. The participant was instructed to read a series of sentences and told that they could be accompanied by a picture, either before or with the sentence. In addition, the participant was instructed to press the space bar after reading the sentence to hear the true/false statement, and to answer it with a key on the keyboard ("green key" for true; "red key" for false). After the instructions, two practice trials were completed, and nine-point calibration and validation procedures were carried out before starting with the first set of critical items. The participant had to look at these nine dots, which were placed at the same distance from each other forming a rectangle on the screen. Three were on the left side, three in the middle and three on the right side. In turn, in these three parts, one point was in the upper zone, one in the middle zone and one in the lower zone. Calibration and validation procedures were repeated before each block, allowing the participant to rest. The testing session for each participant lasted around 35 minutes in total, of which 20 minutes corresponded to the sentence comprehension task. The Ethics Committee of the University of East Anglia approved the study. Before the study, informed written consent was obtained from all participants, and all were debriefed at the end of the study.

Data Screening

Fixations less than 80 ms and longer than 1200 ms were excluded from the dataset. Data from each sentence were reviewed. Those not recorded or with excessive blinks were excluded from the analyses, resulting in data loss of 1.13 %. Data points for each eye-movement measure greater than 3 SDs from the mean were defined as outliers.

Outliers (4.52 % of data) were eliminated.

Target Word (Local) Eye-movement Measures

Four target-word measures were analysed. Gaze duration is the sum of duration of all fixations on a word from first entry to exit. Reading time is the sum of the duration of all fixations on the word, including regressions back to the word. These fixation duration measures are indicators of lexical access of word recognition (Inhoff & Rayner, 1986; Morton, 1969; Whaley, 1978) and are related to processing and integration of lexical meaning of words. Regressions are backward eye movements to previous parts of the text and are considered an indicator of difficulties of reading (Reichle et al., 2003). Regression path duration is the sum of all fixations from first entering a region until moving to the right of that region and is another indicator of processing difficulty (Hyönä et al., 2003).

Text (Global) Eye-movement Measures

There were three text-level measures (text gaze duration, text reading time, and probability of regressions out). Text gaze duration and Text reading time were calculated as the sum of all word reading times and word gaze durations on the sentence, respectively. To calculate the probability of Text regressions out, we summed all probabilities of regressions out of the sentence. Then, we divided the total probability of regressions out of each sentence by the number of words in each sentence.

Analyses

We analysed the effects of Picture, Frequency, and Timing on Text/Target gaze duration, Text/Target-word reading time, Text/Target-word regressions out, and on

Target-word regression path duration by means of linear mixed-effect models (LMM) using the “lmer” function from “lme4” package v1.1–27.1 for R (Bates et al., 2021). Given that the factor Timing was nested to the sentence-with-picture condition, we built two models for each eye-movement measure. One of them included Picture and Frequency as fixed factors, whereas the other one only included Timing. Participant and Item were entered in all models as random effects with random intercepts, whereas decisions on their slopes (i.e., fixed or random) were empirically based on goodness of fit comparisons between the null models (i.e., models including only the random effects).

We performed a priori statistical power analyses using the “R2power” function from “mixedpower” package v0.1.0 for R (Kumle et al., 2018) following the procedure described in Kumle and colleagues (2021; Scenario 2). These analyses were performed by simulating power for different combinations of the two random effects (i.e., number of participants and number of items) on a dataset from a previous similar study by Rivero-Contreras et al. (2021) that also manipulated the presence of picture and word frequency to examine their effects on the same eye-movement measures. Although the study was performed on a sample of young adults with typical development and young adults with dyslexia, given that no differences were found between both groups, we used the whole dataset to simulate power.

Simulations were performed separately for each eye-movement measure to determine the a priori statistical power in the present study for each variable based on the effects found in Rivero-Contreras et al. (2021). Different combinations of both random effects were performed as follows. Firstly, the number of items were fixed at 60 to determine the minimum number of participants to reach a beta value of .20 when analysing the effect of Picture and Frequency on each eye-movement measure. As can be seen in Table 2, the appropriate minimum sample size in the cases of the effect of Frequency on Text regressions out and the effect of Picture on Word gaze duration exceeded 40. Therefore, we will avoid concluding on the absence of these effects.

Secondly, since the factor Timing was nested within the sentence-with-picture condition, such as only 30 items were involved, we fixed the number of items at 30, to simulate power to find an effect of this factor (on each eye-movement measure) like the one found in Rivero-Contreras et al. (2021). For those eye-movement variables that were affected by Picture in Rivero-Contreras et al. (2021), these were the effects used to simulate power because this factor is related closely to Timing (i.e., the moment in which the picture is presented). In the remaining cases, we based the simulations on the effects of Frequency. Results showed that the appropriate minimum sample size exceeded 40 in the cases of the possible effects of Timing on Text regressions out, Word regressions out, Word reading time, Word gaze duration, and Word regression path duration (see also Table 2). We will thus avoid arriving at conclusions, in any case, on the absence of these effects of Timing. The dataset and syntax used to perform all the a priori statistical power analyses can be found at https://osf.io/s28f5/?view_only=b5ce705485b04684bd4074cd61f7ee0a.

Table 2. Results of a priori statistical power analyses for each eye-movement measure based on the dataset and effects found in Authors (2021)

Factor	Eye-movement measure	Necessary sample size[1] to reach β = .20	
		60 items	30 items
Frequency	Text reading time	14	n/a[2]

	Text regressions out	60*	75*
	Target-word gaze duration	16	n/a[2]
	Target-word reading time	4	n/a[2]
	Target-word regressions out	33	
	Target-word regression path duration	25	n/a[2]
Picture	Text reading time	13	15
	Text gaze duration	10	12
	Target-word reading time	24	47
	Target-word gaze duration	46*	100*
	Target-word regression path duration	37	65*

Note. [1]In number of participants; [2]Statistical power simulation for 30 items was not performed on this effect because the effect of Picture on this measure was prioritised to simulate power for Timing (i.e., dataset with 30 items per participant); *Insufficient statistical power for our study's sample size.

The LMMs to examine the effects of the experimental factors on the eye-movement measures in the present study were built as follows. Prior to entering the fixed effects in each model, we tested for possible covariation between the eye-movement measures and participants' scores on PPVT-4, on ART, and on participants' age. To that end, these three variables were added separately as continuous fixed factors to the null model for each eye-movement measure. When any of them proved to be a significant predictor, it was kept in subsequent models including the experimental factors (i.e., Picture and Frequency or Timing). Thus, the LMM for each eye-movement measure was built by sequentially adding terms to the null model as follows: (1) null model showing best goodness of fit, (2) null model + covariates (if any), and (3) null model + covariates (if any) + fixed effects. The model that was finally selected in each case was also the one that showed best goodness of fit (see Table 5), which was tested using the "anova" function from "stats" package v.4.0.2 for R (R Core Team, 2020). All eye-movement measures were normalized and centered using the "normalize" function with standardization method included in the "BBmisc" package v.1.11 for R (Bischl et al., 2017).

Results

Covariates

The model estimates of participants' ART score, PPTV-4 score, and age for each eye-movement measure can be found in Table 3. ART score significantly predicted Target-word gaze duration, Target-word reading time, and Regression path duration in the full dataset, and only Target-word gaze duration and Target-word reading time in the data subset of models for Timing and Frequency. In addition, participants' age significantly predicted Text regressions out. Therefore, ART score and participants' age were included (centered; Cohen et al., 2013) as covariates in these cases. Thus, the marginal significant difference in ART score and the significant difference in age between Timing conditions (see Table 1) were controlled. Lastly, participants' scores on PPTV-4 did not significantly predict any of the eye-movement measures so that this covariate was not included in any of the models.

Table 3. Estimates of predicting models for ART score, PPTV-4 score, and participants' age on each eye-movement measure.

	Global eye-movement measures			Local eye-movement measures			
	Text gaze duration	***Text reading time***	***Text regressions out***	***Target-word gaze duration***	***Target-word reading time***	***Target-word regressions out***	***Regression path duration***
	Estimates for data included in the models for Picture and Frequency (k_{item} = 60)						
ART	-0.14	-0.11	-0.04	-0.11*	-0.11*	-0.01	-0.08*
PPTV-4	-0.12	-0.07	-0.07	-0.09	-0.08	-0.02	-0.05
Age	0.004	0.01	0.12*	-0.01	-0.05	0.03	0.06
	Estimates for data subset included in analyses of the effect of Timing (k_{item} = 30)						
ART	-0.12	-0.12	0.13	-0.12*	-0.13*	0.06	-0.08
PPTV-4	-0.09	-0.09	-0.01	-0.09	-0.07	0.01	-0.05
Age	0.04	0.02	0.24*	-0.01	-0.01	0.11	0.08

Note. ART: Author recognition task. PPVT-4: Peabody Picture Vocabulary Test. $*p < .05$.

Eye-movement Measures

Table 4 and Table 5 show a summary of means for global and local eye-movement measures, respectively. An overview of the models that were finally selected for each eye-movement measure for Picture and Frequency and for Timing, respectively, can be found in Table 6 and Table 7. Given that these measures were standardized, the estimate values in subsequent analyses show proportion of standard deviation in each case (i.e., standardized beta coefficient). Reference values for Picture, Frequency, and Timing was set at *no picture*, *high frequency*, and *concurrent* *presentation*, respectively.

Table 4. Means and standard deviations of global eye-tracking measures.

	Low frequency *M* (*SD*)	**High Frequency** *M* (*SD*)
No picture (*n* = 40)		
Text gaze duration (ms)	2906 (860)	2802 (865)
Text reading time (ms)	4221 (1738)	3933 (1671)
Text regressions out (p)	.09 (.08)	.08 (.07)
Picture (Concurrent Timing; *n* = 20)		
Text gaze duration (ms)	2563 (731)	2485 (723)
Text reading time (ms)	3690 (1389)	3390 (1457)
Text regressions out (p)	.09 (.08)	.07 (.07)
Picture (Consecutive Timing; *n* = 20)		
Text gaze duration (ms)	2748 (855)	2667 (750)
Text reading time (ms)	3804 (1620)	3619 (1493)
Text regressions out (p)	.01 (.05)	.01 (.05)

Note. ms = milliseconds; p = probability.

Table 5. Means and standard deviations of local eye-tracking measures.

	Low frequency *M* (*SD*)	**High Frequency** *M* (*SD*)
No picture (*n* = 40)		
Target-word gaze duration (ms)	308 (224)	249 (158)
Target-word reading time (ms)	532 (394)	357 (297)
Target-word regressions out (p)	.16 (.37)	.15 (.36)
Regression path duration (ms)	544 (581)	493(548)
Picture (Concurrent Timing; *n* = 20)		
Target-word gaze duration (ms)	253 (160)	226 (145)
Target-word reading time (ms)	483 (338)	335 (214)
Target-word regressions out (p)	.22 (.42)	.13 (.34)
Regression path duration (ms)	489 (459)	389 (446)
Picture (Consecutive Timing; *n* = 20)		
Target-word gaze duration (ms)	278 (169)	250 (149)
Target-word reading time (ms)	491 (392)	406 (281)
Target-word regressions out (p)	.03 (.18)	.03 (.16)
Regression path duration (ms)	475 (441)	430 (361)

Note. ms = milliseconds; p = probability.

Table 6. Overview of the models for Picture and Frequency with best goodness of fit for each eye-movement measure.

Global eye-movement measures	**Fixed effects**	**Random effects**	
		Item	**Participant**
Text gaze duration	Picture + Frequency	Intercept	Intercept + Slope (Picture + Frequency)
Text reading time	Picture + Frequency	Intercept	Intercept + Slope (Picture + Frequency)
Text regressions out	Picture + Frequency + Age	Intercept	Intercept + Slope (Picture)
Local eye-movement measures	**Fixed effects**	**Random effects**	
		Item	**Participant**
Target-word gaze duration	Picture + Frequency + ART score	Intercept + Slope (Picture)	Intercept + Slope (Picture)
Target-word reading time	Picture + Frequency + ART score	Intercept	Intercept + Slope (Frequency)
Target-word regressions out	Picture + Frequency	Intercept	Intercept + Slope (Picture + Frequency)
Regression path duration	Picture + Frequency + ART score	Intercept	Intercept + Slope (Picture)

Table 7. Overview of the models for Timing with best goodness of fit for each eye-movement measure.

Global eye-movement measures	**Fixed effects**	**Random effects**	
		Item	**Participant**
Text gaze duration	Null model	Intercept	Intercept
Text reading time	Null model	Intercept	Intercept
Text regressions out	Timing + Age	Intercept	Intercept
Local eye-movement measures	**Fixed effects**	**Random effects**	
		Item	**Participant**
Target-word gaze duration	Null model	Intercept	Intercept
Target-word reading time	Null model	Intercept + Slope (Frequency)	Intercept
Target-word regressions out	Timing + Age	Intercept	Intercept
Regression path duration	Null model	Intercept	Intercept

Global Eye-movement Measures

As can be seen in Table 8, Text gaze duration showed significant effects of Picture and Frequency, indicating that gaze duration was shorter on sentences with picture (2616 vs. 2854 ms.) and sentences with a high-frequency target-word (2651 vs. 2729 ms.; see Table 4). There were also significant effects of Picture and Frequency on Text reading time. Participants spent less time reading the sentences with picture (3625 vs. 4077 ms.) and when the target-word frequency was high (3647 vs. 3905 ms.; see Table 4). In addition, despite the lack of appropriate statistical power (see *Analyses* above), we found a significant effect of Picture and Frequency on Text regressions out indicating that the probability of regressions at text level was significantly lower also in sentences with picture (.05 vs .09) and with a high-frequency target-word (low: .05 vs high: .06; see Table 4).

In the case of the models including Timing as fixed effect, we found an effect of this factor on Text regressions out (see Table 9), indicating that the probability of regressions out was higher when the picture was concurrent with the sentence (.08 vs .02, see Table 4). Lastly, we found no significant effect of Timing on Text gaze duration and on Text reading time.

Table 8. Results of models for Picture and Frequency on each eye-movement measure.

Model	**Estimate**	***SE***	**95% CI**	***t***	***p***
Text gaze duration					
Intercept	0.20	0.13	[-0.06, 0.45]	1.50	.14
Picture	-0.28	0.05	[-0.38, -0.18]	-5.72	< .001
Frequency	-0.11	0.03	[-0.18, -0.04]	-3.22	< .01
Text reading time					
Intercept	0.22	0.12	[-0.01, 0.46]	1.86	.07
Picture	-0.27	0.07	[-0.40, -0.14]	-4.13	< .001
Frequency	-0.16	0.04	[-0.25, -0.08]	-3.95	< .001
Text regressions out					
Intercept	0.30	0.07	[0.15, 0.44]	1.89	.06
Participants’ age	0.12	0.05	[0.02, 0.23]	2.28	.03
Picture	-0.50	0.10	[-0.69, 0-.31]	-5.17	< .001
Frequency	-0.11	0.05	[-0.20, -0.01]	-2.17	.04

Target-word gaze duration					
Intercept	0.17	0.08	[0.03, 0.32]	2.29	.02
ART score	-0.11	0.05	[-0.20, -0.03]	-2.52	.02
Picture	-0.15	0.05	[-0.25, -0.04]	-2.71	.01
Frequency	-0.24	0.04	[-0.32, -0.16]	-6.17	< .001
Target-word reading time					
Intercept	0.21	0.08	[0.05, 0.37]	2.62	.01
ART score	-0.11	0.05	[-0.20, -0.02]	-2.42	.02
Picture	-0.12	0.04	[-0.19, -0.05]	-3.29	< .01
Frequency	-0.36	0.06	[-0.48, -0.24]	-6.07	< .001
Target-word regressions out					
Intercept	0.10	0.06	[-0.004, 0.22]	1.89	.06
Picture	-0.14	0.08	[-0.30, 0.01]	-1.81	.08
Frequency	-0.09	0.04	[-0.17, -0.01]	-2.16	.03
Target-word regression path duration					
Intercept	0.14	0.08	[-0.01, 0.29]	1.85	.07
ART score	-0.08	0.03	[-0.15, -0.01]	-2.35	.02
Picture	-0.14	0.05	[-0.23, -0.05]	-3.01	< .01
Frequency	-0.13	0.04	[-0.20, -0.05]	-3.39	< .001

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low. More detailed results for each model can be found in the Appendix.

Table 9. Results of models for Timing on each eye-movement measure.

Model	Estimate	*SE*	95% CI	*t*	*p*
Text gaze duration					
Intercept	-0.11	0.15	[-0.41, 0.17]	-0.76	.45
Timing	0.23	0.19	[-0.14, 0.61]	1.22	.23
Text reading time					
Intercept	-0.05	0.13	[-0.31, 0.22]	-0.36	.72
Timing	0.11	0.16	[-0.21, 0.44]	0.69	.50
Text regressions out					
Intercept	0.41	0.10	[0.21, 0.61]	4.10	< .001
Age	0.10	0.07	[-0.04, 0.24]	1.39	.17
Timing	-0.83	0.15	[-1.12, -.55]	-5.72	< .001
Target-word gaze duration					
Intercept	-0.05	0.09	[-0.22, 0.12]	-0.60	.55
ART score	-0.11	0.06	[-0.22, 0.003]	-1.91	.06
Timing	0.08	0.11	[-0.14, 0.30]	0.73	.47
Target-word reading time					
Intercept	-0.04	0.09	[-0.21, 0.14]	-0.39	.70
ART score	-0.13	0.06	[-0.24, -0.02]	-2.23	.03
Timing	0.04	0.11	[-0.19, 0.26]	.32	.75
Target-word regressions out					
Intercept	0.24	0.08	[0.09, 0.39]	3.14	< .01
Timing	-0.49	0.11	[-0.70, -0.27]	-4.49	< .001

Target-word regression path duration					
Intercept	-0.01	0.09	[-0.18, 0.16]	-0.09	.93
Timing	0.03	0.09	[-0.15, 0.21]	0.37	.71

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low. More detailed results for each model can be found in the Appendix.

Local Eye-movement Measures

Models for the effects of Picture and Frequency on the eye-movement measures at target-word level are depicted in Table 8. Target-word gaze duration showed a significant effect of Picture, with shorter gaze duration in sentences with picture (252 vs. 279 ms), and an effect of Frequency, showing shorter gaze duration when the target-word frequency was high (242 vs. 280 ms; see Table 5). It is noteworthy that this effect was found despite the insufficient statistical power of our sample size for this eye-movement measure. Moreover, Picture and Frequency also significantly affected Target-word reading time and Regression path duration. The results indicated that, in sentences with picture, reading time of the target-word was shorter than in no-picture sentences (429 vs 445 ms) and regression path duration was also shorter (446 vs 519 ms). Similarly, when target-word frequency was high (vs low), reading time of the target word was shorter (366 vs 502 ms) and regression path duration was shorter (437 vs 503 ms; see Table 5). Lastly, Target-word regressions out was significantly affected only by Frequency, as the probability of regressions on the target-word was lower when target-word frequency was high (.10 vs .14). Although this latter measure was also lower when the picture was present (.11 vs .15), the difference did not reach statistical significance (see Table 8).

Regarding the models for Timing, they also revealed a significant effect of this factor on Target-word regressions out (see Table 9), indicating that the target word had higher probability of regressions out when the picture was concurrent with the sentence than when picture and sentence appeared consecutively (.18 vs .03, see Table 5). As found for global eye-tracking measures, there was no significant effect of Timing on Target-word gaze duration, Target-word reading time, and on Regression path duration. More detailed information including statistical data for random effects and model fit for each of the LMMs reported above can be found in the Appendix.

Discussion

The primary aim of the study was to examine the effect of visual support and how varying the timing of visual support affected sentence processing. We used eye tracking to investigate online sentence processing, that is, reading comprehension processes as they occur in real time through fixation durations and regressions. We had expected that presenting visual support prior to the sentence would have a facilitating effect on reading and comprehension. In the remainder of the present section, we first go through the results of the global text eye-movement measures, and then, the local target-word eye-movement measures.

Effects of Frequency, Picture, and Timing on global eye-movement measures

As expected, all the global text measures (i.e., gaze duration, reading time, and regressions on the entire sentence) were affected by the presence of a picture supporting the sentences, regardless of the presentation timing (i.e., consecutive or concurrent with the sentence). In such cases, participants spent shorter time in reading the sentence and the probability of regressions was lower. A similar pattern was found for target-word frequency. When frequency was high, the participants also spent shorter times in reading the sentence and their probability of regressions was also lower than when frequency was low.

It is noteworthy that, contrary to our prior work (Rivero-Contreras et al., 2021, 2023), whose findings showed that the effects of the supporting picture on global measures were not consistently significant either in university students with dyslexia and with typical development, the effects of picture on gaze duration and total reading time at text level in the present study were significant and larger than the effects of target-word frequency (standardized estimates: -0.28 and -.027 vs -0.11 and -0.16, respectively). However, these results are consistent with the idea that the effect of picture is expected to be larger than the one of frequency, as it has the potential to affect the processing of the sentence in its entirety, whereas the effect of frequency would be necessarily limited to a smaller portion of the sentence (i.e., one, two, or at most three words). A possible explanation for the differences between present findings and those in Rivero-Contreras et al. (2021, 2023) is that the samples in these latter studies included adult readers with dyslexia and with low levels of education.

Regarding the effect of Timing, we found a large effect (standardized estimate: 0.83; Cohen, 1988) showing that when the pictures were presented before the sentences, the probabilities of regressions at text level was lower. We provide two possible complementary explanations to this finding. On the one hand, it could be the case that presenting a picture before the sentence facilitates the construction of a situation model or the event mental representation (Kintsch, 1998; McRae et al., 2021). Visual information would provide a context for understanding written information, thus fostering the activation of students' prior knowledge which, in turn, would speed up the generation of inferences facilitating comprehension. This would explain why participants' probability of regressions was lower on sentences that were presented after the participants visualized the picture.

On the other hand, it could be also possible that the lower probability of regressions when the picture is presented before the sentence was at least partially due to participants adapting their reading strategy to the experimental materials. Being able to look at the picture concurrent with the sentence could have led the participants to make an increased number of regressions, as they could made gaze transitions between the text and the picture. Although regressions on the sentences that occurred immediately after picture-to-text transitions were excluded from the analyses, it is possible that these transitions led the participants to make more subsequent regressions within the sentences. However, we see these two possible explanations as complementary rather than exclusive. If this is the case, it would explain the larger size of the effect, as it would be the consequence of summing both the effect of activating prior knowledge before reading the sentence, and the influence of our experimental materials.

Effects of Frequency, Picture, and Timing on local eye-movement measures

The local target-word measures were also affected by the presence visual support, the target-word frequency, and the timing of visual support, showing the same direction that those found for the entire sentences. When visualizing the supporting picture, the students' spent less time in reading the target-word, which was reflected both in shorter gazed duration and reading time. However, in this case, the effect of visual support on the probability of regressions in the target word did not reach statistical significance, and the effect of frequency was larger than the effect of picture both on gaze duration (standardized estimates: -0.24 vs. -0.16) and on reading time (-

0.36 vs. -0.12). Given that frequency is a factor inherent to the target-word, contrary to what was found among the global eye-movement measures, its influence was more salient on the target than on the time spent reading the whole sentence. These findings corroborate previous evidence indicating that visual support facilitates lexical access and comprehension of unfamiliar words (Huettig et al., 2011; Montag et al., 2015; Qu et al., 2016). Altogether, our results demonstrate that the use of pictures supporting textual information facilitates text processing, especially when the text contains low-frequency vocabulary.

In addition, we also found a main effect of the timing of picture presentation on participants' probability of regressions on the target-word. Similar to what was found in the global measures, presenting the picture before the sentence decreased participants' regressions. As suggested above, this result could indicate that the students behaved differently when the picture was concurrent with the text, which supports Ferreira et al. (2013), who indicate that previewing visual information (consecutive presentation) allows readers to generate certain expectations or predictions in relation to the linguistic information in the text. However, as also argued, we cannot exclude the possibility that the presence of a concurrent picture increased the number of regressions as students made more gaze transitions between the picture and the text, which could increase rereading.

Before turning to the limitations of the current study, we should point out that, in general, most of the effects on the local measures found in the present study replicate findings in prior works (Rivero-Contreras et al., 2021, 2023). For the local measures, there is a clear and robust effect of word frequency, and picture has positive effects on reading time, gaze duration and regression path durations. In addition, we also found a robust effect of the timing of presentation of visual support on reading behaviour, both at text and target-word level. Although we do not have a clear explanation for this effect, we suggest that, in our study, visual support preceding the reading of the text (i.e., consecutive presentation) at least partially facilitated subsequent processing of textual information.

Limitations

This study has yielded some interesting and novel findings, but it is not exempt of some limitations. Firstly, as already reported, the sample was rather small to detect possible effects of Picture, Frequency, and Timing on some of the eye-movement measures similar to those found in our previous research work. Therefore, we cannot discard that the timing of the visual support also affects reading time or gaze duration at target-word level due to the lack of sufficient statistical power.

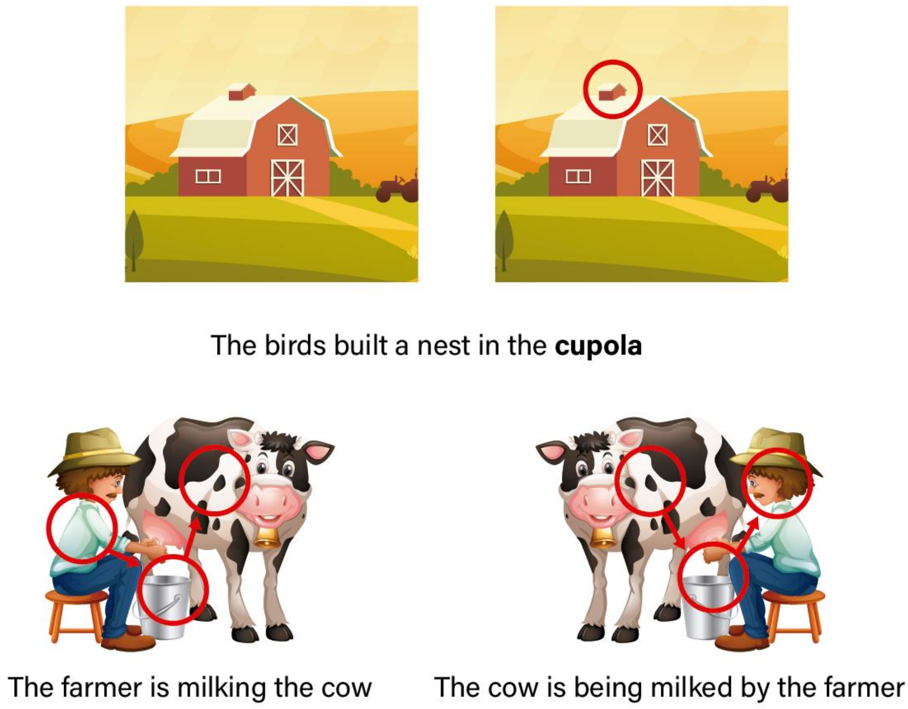
Secondly, the participants consisted of a skilled and homogenous group of college students. Further research could replicate our study in a larger sample and, more important, with a higher heterogeneity in terms of reading skills. Thirdly, regarding the effects of supporting pictures, we used coloured illustrations that provided many contextual elements and mirrored the content of the sentence. Whereas this type of images has been shown to be highly beneficial for reading (Carney & Levin, 2002) we did not explore the possible differences of other types of visual support such as visual scenes, infographics, or symbols. Thus, future studies can examine how other types of visual support and the timing of that visual support benefits language comprehension.

Lastly, we did not analyse participants' eye-gaze on the pictures. This was due to several reasons (e.g., image presentation was time limited in one condition and self-paced in the other). Therefore, we could not explore what parts of the pictures could be more beneficial for textual processing. In the consecutive condition, participants inspected the image before the text (unless they looked away from the screen). In contrast, in the concurrent condition participants were free to view the picture (or not), and there were no constraints on when they viewed the image versus when they read the sentence (before, after, or intermittently while reading the text).

Applied Implications and Future Directions

Regarding practical implications of our findings, it seems that visual support can positively impact skilled adult readers' textual processing. In fact, the current study has shown the strongest effects we have observed to date for visual support. One clear extension of this work is how readers deal with unfamiliar words and how visual support might help access or direct readers to the meaning of an unknown word. One example is "cupola", which is an example of a "difficult" low-frequency word in some vocabulary texts. In Figure 3, it will be important to test what readers actually look at (if given the upper left picture) after reading the word "cupola" in the sentence (shown below the sentence). Second, for word learning, if participants read "cupola" and then the indicated object is subsequently highlighted in an image (e.g., right upper picture), are participants then more likely to retain that knowledge (i.e., to what extent does visual support enhance word learning).

Figure 3. Examples images and texts.



Another relevant future direction concerns the images that are used as visual support. Two possible research questions stand out as particularly critical to us. The first is to better understand how images are incrementally processed (i.e., the most common scan paths associated with an image), and whether images can be manipulated to encourage particular scan paths, which would then map onto the text in a meaningful way (see Figure 3, bottom panels). The expectation would be that "good" visual support is processed in accordance with the grammatical roles and thematic relations contained in the sentence (e.g., farmer is first fixated and is the subject/topic/agent of the sentence). Perhaps an experiment with the mirror reversed image may facilitate processing of the same sentence in passive voice (e.g., the image shown in lower right panel). Given that algebraic equations can prime syntax (Scheepers et al., 2011), it only makes sense that images would too. The second question pertains saliency (Henderson, 2017; Itti & Koch, 2000). Here the key issue is how visual attention can be directed/capture to direct the viewer to particular regions of an image, which should then link up with particular regions of text.

Conclusions

The current study focused largely on visual support to facilitate reading, and particularly, whether the timing of visual support impacts reading. Our results suggested that visual support benefitted the overall reading of sentences. Moreover, we did find a main effect of timing on text and target regressions out. Based on this, we conclude that visual support is generally beneficial, and is so even for highly skilled readers. In addition, the presentation of the picture before the text appears to contribute to the activation of participants' prior knowledge and/or adaptation of their reading strategies. Finally, we argued that more research is needed with respect to visual support, and we have highlighted how it can benefit by drawing upon different strands of psychological research. There is an old saying *"a picture is worth a thousand words",* what this research shows is that a picture (as visual support) is worth a good handful of words.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgments

This research was supported by the V Plan Propio de Investigación y Transferencia of the University of Seville (PPI2016-IV.5). The work of Miriam Rivero-Contreras is supported by a research grant from the University of Seville (PP2019-12336). The work of Pablo Delgado is supported by the grant Ayudas a los Agentes del Sistema Andaluz del Conocimiento para la Contratación de Personal Investigador Doctor (PAIDI

DOCTOR 21; Regional Government of Andalusia, Spain).

Data availability statement

The data that support the findings of this study and the syntaxes used to perform the statistical analyses are available at: https://osf.io/s28f5/?view_only=b5ce705485b04684bd4074cd61f7ee0a

References

Acheson, D.J., Wells, J.B. & MacDonald, M.C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods, 40*(1), 278-289. https://doi.org/10.3758/BRM.40.1.278

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264. https://doi.org/10.1016/S0010-0277(99)00059-1

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye Movements of Highly Skilled and Average Readers: Differential Effects of Frequency and Predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086. https://doi.org/10.1080/02724980443000476

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2021). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-27.1*. https://cran.r-project.org/web/packages/lme4/index.html

Bischl, B., Lang, M., Bossek, J., Horn, D., Richter, J., & Surmann, D. (2017). *BBmisc: Miscellaneous Helper Functions for B. Bischl. R package version 1.11.* https://cran.r-project.org/web/packages/BBmisc/index.html

Brookshire, J., Scharff, L. F. V., & Moses, L. E. (2002). The Influence of Illustrations on Children's Book Preferences and Comprehension. *Reading Psychology*, *23*(4), 323–339. https://doi.org/10.1080/713775287

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review, 14*, 5–26. https://doi.org/10.1023/A:1013176309260

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

Dowdall, N., Melendez-Torres, G.J., Murray, L., Gardner, F., Hartford, L., & Cooper, P.J. (2020). Shared picture book reading interventions for child language development: a systematic review and meta-analysis. *Child Development*, *91*(2), e383-e399. https://doi.org/10.1111/cdev.13225

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test (4th ed.)*. Pearson.

Eitel, A. & Scheiter, K. (2015). Picture or text first? Explaining sequence effects when learning with pictures and text. E*ducational Psychology Review, 27,* 153–180. https://doi.org/10.1007/s10648-014-9264-4

Eitel, A., Scheiter, K., & Schüler, A. (2013). How Inspecting a Picture Affects Processing of Text in Multimedia Learning. *Applied Cognitive Psychology*, *27*(4), 451–461. https://doi.org/10.1002/acp.2922

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language, 69*(3), 165-182. https://doi.org/10.1016/j.jml.2013.06.001

Fletcher, K., & Reese, E. (2005). Picture book reading with young children: a conceptual framework. *Developmental Review*, *25*(1), 64–103. https://doi.org/10.1016/j.dr.2004.08.009

Guo, D., Zhang, S., Wright, K. L., & McTigue, E. M. (2020). Do You Get the Picture? A Meta-Analysis of the Effect of Graphics on Reading Comprehension. *AERA Open*, *6*(1), 1-20. https://doi.org/10.1177/2332858420901696

Henderson, J. M. (2017). Gaze control as prediction. *Trends in Cognitive Sciences, 21*(1), 15-23. https://psycnet.apa.org/doi/10.1016/j.tics.2016.11.003

Hibbing, A. N., & Rankin-Erickson, J. L. (2003). A Picture Is Worth a Thousand Words: Using Visual Images to Improve Comprehension for Middle School Struggling Readers. *The Reading Teacher*, *56*(8), 758–770. https://www.jstor.org/stable/20205292

Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye Movement Measures to Study Global Text Processing. In J. Hyönä, R. Radach, & H. B. T. Deubel (Eds.), *The Mind's Eye* (pp. 313–334). North-Holland. https://doi.org/https://doi.org/10.1016/B978-044451020-4/50018-9

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *21*(6), 1430–1440. https://doi.org/10.1037/0278-7393.21.6.1430

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431–439. https://doi.org/10.3758/BF03208203

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2018). Mixedpower: a library for estimating simulation-based power for mixed models in R. https://doi.org/10.5281/zenodo.1341047

Kumle, L., Võ, M.LH., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Researcg Method*, *53*, 2528–2543. https://doi.org/10.3758/s13428-021-01546-0

Leech, G., Rayson, P. and Wilson, A. (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. Longman. http://ucrel.lancs.ac.uk/bncfreq/

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1), 43-52. https://doi.org/10.1207/S15326985EP3801_6

McRae, K., Brown, K. S., & Elman, J. L. (2021). Prediction-Based Learning and Processing of Event Knowledge. *Topics in Cognitive Science, 13*(1), 206-223. https://doi.org/10.1111/tops.12482

Mol, S. E., & Bus, A. G. (2011). To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*(2), 267. https://doi.org/10.1037/a0021890

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The Words Children Hear: Picture books and the statistics for language learning. *Psychological Science*, *26*(9), 1489–1496. https://doi.org/10.1177/0956797615594361

Moore, M., & Gordon, P. C. (2015). Reading ability and print exposure: item response theory analysis of the author recognition test. *Behavior research methods*, *47*(4), 1095–1109. https://doi.org/10.3758/s13428-014-0534-3

Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology*, *60*(3), 329–346. https://doi.org/10.1111/j.2044-8295.1969.tb01204.x

Qu, Q., Zhang, Q., & Damian, M. F. (2016). Tracking the time course of lexical access in orthographic production: An event-related potential study of word frequency effects in written picture naming. *Brain and Language*, *159*, 118–126. https://doi.org/10.1016/j.bandl.2016.06.008

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, *58*(5), 734–747. https://doi.org/10.3758/BF03213106

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.R-project.org/

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1188–1200. https://doi.org/10.1037/0096-1523.22.5.1188

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *Behavioural and Brain Sciences*, *26*(4), 445-526. https://doi.org/10.1017/s0140525x03000104

Rivero-Contreras, M., Engelhardt, P. E, & Saldaña, D. (2021). An experimental eye-tracking study of text adaptation for readers with dyslexia: effects of visual support and word frequency. *Annals of Dyslexia*, *71*, 170–187. https://doi.org/10.1007/s11881-021-00217-1

Rivero-Contreras, M., Engelhardt, P. E, & Saldaña, D. (2023). Do easy-to-read adaptations really facilitate sentence processing for adults with a lower level of education? An experimental eye-tracking study. *Learning and Instruction, 84,* 101731. https://doi.org/10.1016/j.learninstruc.2022.101731

Scheepers, C., Sturt, P., Martin, C. J., Myachykov, A., Teevan, K., & Viskupova, I. (2011). Structural priming across cognitive domains: From simple arithmetic to relative-clause attachment. *Psychological Science, 22*(10), 1319-1326. https://doi.org/10.1177/0956797611416997

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1280–1293. https://doi.org/10.1037/a0016896

Strouse, G. A., Nyhout, A., & Ganea, P. A. (2018). The role of book features in young children's transfer of information from picture books to real-world contexts. *Frontiers in Psychology*, *9*(50), 1-14. https://doi.org/10.3389/fpsyg.2018.00050

Takacs, Z. K., & Bus, A. G. (2018). How pictures in picture storybooks support young children’s story comprehension: An eye-tracking experiment. *Journal of Experimental Child Psychology*, *174*, 1–12. https://doi.org/10.1016/j.jecp.2018.04.013

Wasik, B. A., Hindman, A. H., and Snell, E. K. (2016). Book reading and vocabulary development: a systematic review. *Early Childhood Research Quarterly*, *37*, 39–57. https://doi.org/10.1016/j.ecresq.2016.04.003

Wassenburg, S. I., de Koning, B. B., Bos, L. T., & van der Schoot, M. (2020). Inspecting a picture before reading affects attentional processing but not comprehension. *Educational Psychology*, *40*, 4–21. https://doi.org/10.1080/01443410.2019.1645306

Whaley, C. P. (1978). Word—nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143–154. https://doi.org/10.1016/S0022-5371(78)90110-X

Appendix

Table A1. Model for main effects of Picture and Frequency on Text gaze duration.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.20	0.13	[-0.06, 0.45]	1.50	.14
Picture	-0.28	0.05	[-0.38, -0.18]	-5.72	< .001
Frequency	-0.11	0.03	[-0.18, -0.04]	-3.22	< .01
Random effects					
			Variance	***SD***	**Correlation**
Item (intercept)			0.20	0.45	-
Participant (intercept)			0.53	0.73	-
Participant (Picture)			0.07	0.26	-.68
Participant (Frequency)			0.02	0.14	-.45; .40
Model fit					
Marginal R^2			**Conditional R^2**		
.02			0.62		

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low.

Table A2. Model for main effects of Picture and Frequency on Text reading time.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.22	0.12	[-0.01, 0.46]	1.86	.07
Picture	-0.27	0.07	[-0.40, -0.14]	-4.13	< .001
Frequency	-0.16	0.04	[-0.25, -0.08]	-3.95	< .001
Random effects					
			Variance	***SD***	**Correlation**
Item (intercept)			0.20	0.44	-
Participant (intercept)			0.42	0.65	-
Participant (Picture)			0.14	0.38	-.66
Participant (Frequency)			0.04	0.19	-0.31; .01
Model fit					
Marginal R^2			**Conditional R^2**		
0.03			0.53		

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low.

Table A3. Model for main effects of Picture and Frequency on Text regressions out.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.11	0.06	[-0.004, 0.21]	1.89	.06
Age	0.12	0.05	[0.02, 0.23]	2.28	0.03
Picture	-0.50	0.10	[-0.69, -0.31]	-5.17	<.001
Frequency	-0.11	0.05	[-0.20, -0.01]	-2.17	.04
Random effects					
			Variance	***SD***	**Correlation**
Item (Intercept)			0.02	0.14	-
Participant (Intercept)			0.18	0.42	-
Participant (Picture)			0.32	0.57	-.37
Participant (Frequency)			0.05	0.22	-.49; -.16
Model fit					
Marginal R^2			**Conditional R^2**		
.08			.31		

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low.

Table A4. Model for main effect of Picture and Frequency on Target-word gaze duration.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.17	0.08	[0.06, 0.37]	2.29	.02
ART score	-0.11	0.05	[-0.20, -0.03]	-2.71	.02
Picture	-0.15	0.04	[-0.25, -0.04]	-3.50	.01
Frequency	-0.24	0.04	[-0.32, -0.16]	-5.94	< .001
Random effects					
			Variance	***SD***	**Correlation**
Item (Intercept)			0.12	0.34	-
Item (Picture)			0.03	0.17	-.88
Participant (Intercept)			0.11	0.33	-
Participant (Picture)			0.04	0.19	-.68
Model fit					
Marginal R^2			**Conditional R^2**		
.03			0.20		

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low.

Table A5. Model for main effects of Picture and Frequency on Target-word reading time.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.21	0.08	[0.05, 0.37]	2.62	.01
ART score	-0.11	0.05	[-0.20, -0.02]	-2.42	.02
Picture	-0.12	0.04	[-0.19, -0.05]	-3.29	< .01
Frequency	-0.36	0.06	[-0.48, -0.24]	-6.07	< .001
Random effects					
			Variance	***SD***	**Correlation**
Item (Intercept)			0.12	0.35	-
Participant (Intercept)			0.13	0.37	-
Participant (Frequency)			0.09	0.29	-.73
Model fit					
Marginal R^2			**Conditional R^2**		
.05			.27		

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low.

Table A6. Model for main effects of Picture and Frequency on Target-word regressions out.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.11	0.06	[-0.004, 0.22]	1.89	.06
Picture	-0.14	0.08	[-0.30, 0.01]	-1.81	.08
Frequency	-0.09	0.04	[-0.17, -0.01]	-2.18	.03
Random effects					
			Variance	***SD***	**Correlation**
Item (intercept)			0.01	0.12	-
Participant (intercept)			0.07	0.27	-
Participant (Picture)			0.18	0.42	-.62
Model fit					
Marginal R^2			**Conditional R^2**		
.01			.11		

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low.

Table A7. Model for main effects of Picture and Frequency on Target-word regression path duration.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.14	0.08	[-0.01, 0.29]	1.85	.07
ART score	-0.08	0.03	[-0.15, -0.01]	-2.35	.02
Picture	-0.14	0.05	[-0.23, -0.05]	-3.01	< .01
Frequency	-0.13	0.04	[-0.20, -0.05]	-3.39	< .001
Random effects					
			Variance	***SD***	**Correlation**
Item (Intercept)			0.19	0.43	-
Participant (Intercept)			0.07	0.26	-
Participant (Picture)			0.03	0.17	-.80
Model fit					
Marginal R^2			**Conditional R^2**		
.02			.25		

Note. Reference value for *Picture:* No picture. Reference value for *Frequency:* Low.

Table A8. Model for main effect of Timing on Text gaze duration.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	-0.11	0.15	[-0.41, 0.18]	-0.76	.45
Timing	0.23	0.19	[-0.14, 0.61]	1.22	.23
Random effects					
			Variance	***SD***	**Correlation**
Item (intercept)			0.25	0.47	-
Participant (intercept)			0.36	0.60	-
Model fit					
Marginal R^2			**Conditional R^2**		
.01			.58		

Note. Reference value for *Timing:* Concurrent. Reference value for *Frequency:* Low

Table A10. Model for main effect of Timing on Text reading time.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	-0.05	0.13	[-0.31, 0.22]	-0.35	.72
Timing	0.11	0.17	[-0.21, 0.44]	0.69	.50
Random effects					
			Variance	***SD***	**Correlation**
Item (intercept)			0.25	0.50	-
Participant (intercept)			0.26	0.51	-
Model fit					
Marginal R^2			**Conditional R^2**		
.003			.50		

Note. Reference value for *Timing:* Concurrent. Reference value for *Frequency:* Low

Table A11. Model for main effect of Timing on Text regressions out.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.41	0.10	[0.21, 0.61]	4.10	< .001
Age	0.10	0.07	[-0.04, 0.24]	1.39	0.17
Timing	-0.83	0.15	[-1.12, -.55]	-5.72	< .001
Random effects					
			Variance	***SD***	**Correlation**
Item (intercept)			0.01	0.09	-
Participant (intercept)			0.16	0.40	-
Model fit					
Marginal R^2			**Conditional R^2**		
.21			.38		

Note. Reference value for *Timing:* Concurrent*.* Reference value for *Frequency:* Low

Table A11. Model for main effect of Timing on Target-word gaze duration.

			Fixed effects		
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	-0.05	0.09	[-0.22, 0.12]	-0.59	.55
ART score	-0.11	0.06	[-0.22, 0.003]	-1.91	.06
Timing	0.08	0.11	[-0.14, 0.30]	0.73	.47
			Random effects		
			Variance	***SD***	**Correlation**
Item (Intercept)			0.07	0.26	-
Participant (Intercept)			0.09	0.29	-
			Model fit		
Marginal R^2			**Conditional R^2**		
.02			0.17		

Note. Reference value for *Timing:* Concurrent. Reference value for *Frequency:* Low.

Table A12. Model for main effect of Timing on Target-word reading time.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.04	0.09	[-0.22, 0.14]	-0.39	.70
ART score	0.13	0.06	[-0.24, 0.26]	-2.23	.03
Timing	0.04	0.11	[-0.19, 0.36]	0.32	.75
Random effects					
			Variance	***SD***	**Correlation**
Item (Intercept)			0.13	0.37	-
Participant (Intercept)			0.09	0.30	-
Model fit					
Marginal R^2			**Conditional R^2**		
.02			.24		

Note. Reference value for *Timing:* Concurrent. Reference value for *Frequency:* Low.

Table A13. Model for main effect of Timing on Target-word regressions out.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	0.24	0.08	[0.09, 0.39]	3.14	< .01
Timing	-0.49	0.11	[-0.70, -0.27]	-4.49	< .001
Random effects					
			Variance	***SD***	**Correlation**
Item (Intercept)			0.01	0.09	-
Participant (Intercept)			0.09	0.29	-
Model fit					
Marginal R^2			**Conditional R^2**		
.06			.15		

Note. Reference value for *Timing:* Concurrent. Reference value for *Frequency:* Low.

Table A14. Model for main effect of Timing on Target-word regression path duration.

Fixed effects					
	Estimate	***SE***	**95% CI**	***t***	***p***
Intercept	-0.01	0.08	[-0.18, 0.16]	-0.09	.93
Timing	0.03	0.09	[-0.15, 0.21]	0.37	.71
Random effects					
			Variance	***SD***	**Correlation**
Item (Intercept)			0.19	0.43	-
Participant (Intercept)			0.05	0.23	-
Model fit					
Marginal R^2			**Conditional R^2**		
.0002			.24		

Note. Reference value for *Timing:* Concurrent. Reference value for *Frequency:* Low.