

Raquel Serrano* and Ana Pellicer-Sánchez
**Young L2 learners' online processing
of information in a graded reader during
reading-only and reading-while-listening
conditions: A study of eye-movements**

<https://doi.org/10.1515/applirev-2018-0102>

Abstract: Combining reading with auditory input has been shown to be an effective way of supporting reading fluency and reading comprehension in a second language. Previous research has also shown that reading comprehension can be further supported by pictorial information. However, the studies conducted so far have mainly included adults or adolescents and have been based on post-reading tests that, although informative, do not contribute to our understanding of how learners' processing of the several sources of input in multi-modal texts changes with the presence of auditory input and the effect that potential differences could have on comprehension. The present study used eye-tracking to examine how young learners process the pictorial and textual information in a graded reader under reading only and reading-while-listening conditions. Results showed that readers spent more time processing the text in the reading only condition, while more time was spent processing the images in the reading-while-listening mode. Nevertheless, comprehension scores were similar for the readers in the two conditions. Additionally, our results suggested a significant (negative) relationship between the amount of time learners spent processing the text and comprehension scores in both modes.

Keywords: reading, reading-while-listening, eye-tracking, graded readers, reading comprehension

1 Introduction

Mastering the skill of reading is essential for learners to become competent second language (L2) users. Thus, supporting the reading process and exploring ways of

***Corresponding author: Raquel Serrano**, Department of Modern Languages and English Studies, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, SPAIN, E-mail: raquelserrano@ub.edu

Ana Pellicer-Sánchez, UCL Institute of Education, London, UK,
E-mail: a.pellicer-sanchez@ucl.ac.uk

improving reading fluency and reading comprehension have become major concerns in applied linguistics research. Combining reading with auditory input in reading-while-listening (RWL) conditions has been suggested as one possible approach to assist the reading process. Previous research has shown the benefits of RWL for reading fluency and comprehension (e.g., Rasinski 1990), with studies suggesting an advantage of RWL over reading-only (RO) conditions (e.g., Chang and Millett 2014) and listening-only (LO) conditions (e.g., Chang 2009).

The use of visuals has been suggested as a way of further supporting the reading process and this is indeed what we find in most graded readers and in English as a Foreign Language (EFL) textbooks (Elley and Mangubhai 1983; Marcus et al. 1996). The use of visuals is especially salient in the case of graded readers that are targeted to young learners. Some concerns have been raised that children might spend too much time focusing on the images while ignoring the text in the graded reader (Hill 2013), and this might be more clearly the case in the RWL mode, as learners can have access to the textual content of the book both in written and aural mode. The present study aims to address this issue and explore online processing of text and images in RWL as opposed to RO conditions through the use of eye-tracking. There are only a handful of studies that have used eye-tracking to explore learners' processing of text and pictures during reading (e.g., Chang and Choi 2014; Evans and Saint-Aubin 2005; Johnson and Mayer 2012; Justice et al. 2005; Mason et al. 2013a, Mason et al. 2013b; Mason and Tornatora 2015; Roy-Charland et al. 2007), and, to the authors' knowledge, no previous studies have used eye-tracking to examine L2 learners' processing of input in RO and RWL conditions.

2 Background

2.1 Reading-while-listening for L2 learning

There is ample evidence that L2 learning can be promoted through RWL (both extensive RWL and repeated RWL). Although most of the evidence comes from research with older learners (adolescents and adults), there are also some studies that have targeted children. In the Canadian context, Lightbown and colleagues (e.g., Lightbown 1992; Trofimovich et al. 2009) evaluated the L2 learning outcomes of primary school students who were learning English exclusively through autonomous extensive RWL, and compared this group with another group of students following teacher-led instruction through an audio-lingual approach. According to the results on the initial evaluation of these learners' performance from grade 3 until the end of grade 5, the extensive RWL

program was at least as effective as learning English through teacher-led instruction in several measures of receptive and productive skills (Lightbown 1992), and even in terms of pronunciation accuracy (Trofimovich et al. 2009). Moreover, the students in the extensive RWL program demonstrated better attitudes and motivation towards English learning than students receiving regular instruction. These results were replicated in Spain by Tragant et al. (2016), who also found extensive RWL to be very motivating for the students.

In the case of older learners, research indicates more clearly that RWL can lead to increased L2 learning. Several studies have provided evidence for the potential of extensive RWL for vocabulary acquisition, in terms of receptive knowledge of form-meaning mapping (Webb and Chang 2015), with studies showing an advantage of RWL over RO (Webb and Chang 2012) and LO in both meaning recognition and recall (Brown et al. 2008). Similarly, Chang and Millett (2015) reported that a group of Taiwanese EFL secondary school learners improved their reading rates and comprehension levels more clearly in an extensive RWL program (including 26 graded readers) than in an extensive RO program. RWL has also been shown to enhance listening comprehension skills more than LO (Chang 2009), or both LO and RO (Chang and Millett 2014).

In general, the results of previous studies indicate that simultaneous RWL (bimodal input) tends to promote L2 development of different skills more clearly than receiving unimodal input. Several reasons have been proposed for this. One of them is that extensive RWL might contribute more to reading fluency than RO because learners are obliged to follow the pace determined by the audio, which, in many cases, is likely to be higher than the students' own pace (Chang and Millett 2015). This claim, however, needs to be empirically proven, and one way to do so is through the use of eye-tracking, which is the methodology used in the current study. Another reason proposed for the benefits of RWL might be that reading in this mode is more motivating for L2 learners. Some studies have reported that learners generally enjoy receiving input in a RWL mode (Brown et al. 2008; Chang 2009; Chang and Millett 2014; Lightbown et al. 2002; Tragant et al. 2016), and such positive perception might certainly facilitate L2 learning. Chang and Millett (2015) claim that their participants preferred the RWL mode because the sound effects made the stories more interesting and because it helped concentration much more than RO. The present study aims to further explore young learners' perceptions of RO and RWL, as the majority of previous studies have focused on older L2 learners.

The outcomes of RO and RWL have been shown to be affected by a variety of factors. Success in reading comprehension seems to be determined by L2 proficiency and L1 reading ability (Carrell 1991), motivation to read, processing strategies (Grabe 2009), and different types of background knowledge

(Alderson and Urquhart 1985; Carrell 1983), among other variables. One factor that has attracted particular attention from researchers is vocabulary knowledge. Many studies have reported the strong relationship between vocabulary knowledge and reading comprehension (Anderson and Freebody 1983; Jeon and Yamashita 2014; Koda 2005; Qian 1999). This factor has also been reported to affect other outcomes of RWL programs. Webb and Chang (2015), for example, found that RWL to 20 graded readers led to significantly larger vocabulary gains for learners with better previous vocabulary. The present study will also explore further how previous vocabulary knowledge affects reading behavior in RWL and RO conditions.

2.2 The role of visuals in reading

When, in addition to aural and written input, the materials also include images (as is the case of graded readers for children), the allocation of students' attention is divided between the three input sources. The presence of images together with textual content has been shown to be beneficial for language learning, as it helps learners decode unknown words (Center et al. 1999). Additionally, images contribute to the improvement of incidental vocabulary learning (e.g. Bisson et al. 2015), and have a positive effect on reading comprehension (e.g. Elley and Mangubhai 1983; Omaggio 1979). Tang (1992) not only found that visuals led to higher comprehension scores but that students also perceived this facilitation effect. Overall, visuals seem to provide additional contextual information and strengthen the links between the verbal and imagery systems, facilitating, among other areas, word decoding, vocabulary learning and reading comprehension. A recent study by Rodgers (2018) examined the potential of imagery in documentary and narrative television for vocabulary learning. Results showed that the way imagery in television co-occurs with words in the soundtrack supports vocabulary learning, with documentary television having a higher percentage of co-occurrence of aural words and supporting images. However, it has also been claimed that visuals in graded readers might detract young learners' attention from the text, making them spend more time focusing on the images (Hill 2013). As argued earlier, this might be more clearly the case in RWL conditions, where the verbal input is provided through both written and aural modes. As Kiss and Weninger (2016) claim, despite the important role of visuals for language learning, very little is known about how learners engage with visuals. The present study addressed these concerns by examining learners' processing of text and pictures in graded readers and how it changed with the presence of auditory input.

2.3 Eye-tracking and input modalities

Eye-tracking technology allows researchers to examine participants' eye movements while simultaneously processing written verbal stimuli (reading), non-verbal stimuli (images), and auditory stimuli (listening). Eye-tracking measures the movements of the eyes to parts of a text or visual stimuli (saccades), stops of the eyes (fixations), and movements back to previous sections of the text while reading (regressions). It records the number and duration of fixations, the number and length of saccades, and regressions made within a particular area of interest (AOI), allowing for a very rich record of eye-movement behavior. It can therefore be considered a very close approximation of natural reading or viewing behavior. Eye-tracking allows researchers to identify what participants are considering in the visual and written stimuli, under the assumption that what is being fixated is what is considered. In the processing of texts, eye-tracking measures allow us to identify parts where participants show more processing effort, manifested in more and longer fixations (Pickering et al. 2004).

In the L1 context, research has shown that the combination of written and auditory verbal input leads to differences in how readers process words. Rayner (1998, 2009) explains that in oral reading fixation durations and number of fixations on a given word are longer than in reading only conditions. However, as Rayner (2009) highlights, most eye-tracking reading research has focused on reading only conditions. No previous studies have examined differences in L2 learners' eye-movement behavior when auditory input is also present.

Particularly relevant for the present investigation are those studies that have used eye-tracking to examine readers' processing of different sources of input in multimodal materials. A few studies have examined learners' eye movements to subtitled videos, with studies showing that learners read subtitles regardless of the subtitle condition (e.g., Bisson et al. 2014) and that the distribution of attention across different redundant sources of information results in the partial processing of subtitles (Kruger and Steyn 2014). Previous research has also suggested a relationship between subtitle reading and reading comprehension (e.g. Kruger and Steyn 2014). In the L1 context, a few studies have used eye-tracking to examine children's attention to print and static illustrations in storybooks at different stages of reading development (e.g., Evans and Saint-Aubin 2005; Justice et al. 2005; Roy-Charland et al. 2007), and in the context of content learning (e.g., Johnson and Mayer 2012; Mason et al. 2013b). However, only one previous study (to the authors' knowledge) has used eye-tracking to examine L2 learners' processing of text and static pictures and its relationship to reading comprehension. Chang and Choi (2014) looked at the effect of seductive sentences (i.e. emotionally interesting text segments) and attention-grabbing

pictures on advanced L2 learners' recall and reading comprehension. Korean EFL learners of English read a text containing these seductive details while their eye movements were recorded. The results of the study showed that increased attention to seductive sentences was a major determinant of poor performance in the two tests. There was a significant, positive relationship between relative gaze duration for the seductive sentences and recall and reading comprehension. However, relative gaze to the seductive pictures or the baseline text (i.e. segments of the text without seductive details) were not significant predictors of recall and reading comprehension. This study provided useful insights about learners' engagement with the print and pictures in a RO condition and its link with comprehension. To the authors' knowledge, no previous studies have used eye-tracking to examine L2 learners' reading behavior in RO and RWL conditions and its relation with comprehension.

3 Aims and research questions

Overall, the studies reviewed in the previous section have shown the benefits of RWL for the acquisition of a variety of linguistic skills and point to the contribution of RWL for improving reading fluency and reading comprehension. However, little is known about how the processing of the text and pictures in multimodal reading conditions changes when auditory input is also present, and the effect that these potential processing differences have on reading comprehension. In addition, despite their widespread use in the classroom, no previous studies have used eye-tracking to investigate how learners engage with the different input sources in the type of multimodal materials used in the present study, i.e. an illustrated graded reader.

In order to address these gaps, the present study investigates young learners' processing of the text and images while RO and RWL to an illustrated graded reader and the relationship between processing and reading comprehension. Our aim is to describe how the presence of auditory input in the RWL condition (as compared to RO) leads to changes in the online processing of the text and pictures, which could lead to potential differences in reading comprehension. The role of previous vocabulary knowledge on this process is also examined. Finally, we are interested in analyzing learners' perception of reading in the two modes in order to examine whether our participants were as enthusiastic about the RWL mode as previous studies have reported (e.g., Chang and Millett 2015; Tragant et al. 2016). The following research questions were addressed:

1. Does the presence of auditory input in RWL conditions lead to changes in the online processing of text and pictures in multimodal reading materials?
2. How is young learners' online reading behavior related to their reading comprehension?
3. How does previous vocabulary knowledge relate to reading behavior and comprehension?
4. Finally, what is young learners' perception of these two multimodal learning modes?

4 Methodology

4.1 Participants

The participants of this study were a group of 36 Catalan/Spanish bilingual children (age 10–11), who were learning English as a foreign language in two different schools in Barcelona. They were all in grade 5 and there were 21 girls and 15 boys. Twenty three students attended School A (from three different classes) and 13 attended School B (from two different classes). The students in the two schools were selected on the following bases: they had to be judged by their teacher as having average English skills considering the whole class (i.e., they should not be among the strongest or weakest students); their reading skills in Catalan should also be average and they should not have a long record of English extra-curricular classes. This information was obtained through background questionnaires from the students and interviews with the teachers.

Because the data collection process required some alterations to the school routine, we considered that it was more appropriate to include two schools rather than one, as the data collection period would be shorter for each school if two schools were involved. We made sure the two schools were comparable in terms of students' profiles: both are located in an affluent part of the city and thus attract middle to upper-middle class families. Additionally, the two schools have a good reputation for the English proficiency that their pupils demonstrate at the end of primary school (age 11–12). The proficiency level of the students in the two schools included in this study, as assessed by their receptive vocabulary knowledge, is not significantly different, according to the Peabody Picture Vocabulary Test (PPVT): Mean School A = 92.17; SD = 28.88; Mean School B = 87.23; SD = 27.84; $t(34) = .499$, $p = .621$.

4.2 Materials

Graded Reader. In order to analyze participants' reading behavior, we chose a graded reader that was deemed appropriate for our target students, after consulting with their English teachers. The title of the book was *Uncle Jack and the Bakonzi tree*, by Jane Cadwallader. It had 300 headwords and was aimed at learners with level A1.1 (Basic User) according to the Common European Framework of Reference (Council of Europe 2001). As expected from materials addressed to children, the book included many colorful images (typically occupying at least two thirds of the page) and three to six lines of text in each page. The book had 26 pages, and for the purpose of this study, it was divided in two parts according to the plot of the story (with Part 1 including pages 2–11 and Part 2 including pages 12–26). The graded reader had a CD with the audio rendition of the story, which was used for the RWL condition.

Comprehension questions. A battery of 20 short open questions referring to key aspects of the story was originally designed in order to examine reading comprehension and relate it to reading mode. These questions were administered in the learners' L1. Since the main aim of the study was to examine online reading behavior and its relation to reading comprehension, the answers to the questions required the comprehension of the text/audio and not simply having processed the images. In order to ensure that that was the case, we asked a group of grade-5 learners ($n = 9$), who were not part of our study, to tell us what they thought the story in *Uncle Jack and the Bakonzi tree* was about by only looking at the pictures (we covered the text). The comprehension questions that were included in the test did not address any of the aspects the learners were able to infer by just looking at the images.

The 20 comprehension questions were initially piloted with two groups of grade-5 learners from a different school ($n = 49$). The results of the pilot study indicated that the comprehension questions were too difficult ($M = 5.94/20$). After close examination of the students' performance in each question, we discarded those questions for which more than $2/3$ of the students provided incorrect answers, which led to the elimination of 10 out of the initial 20 questions. The final test, thus, included 10 comprehension questions and had high internal consistency (Cronbach's $\alpha = .874$).

The comprehension questions were analyzed in terms of accuracy, and one point was awarded for completely accurate responses, 0.5 points for answers that were partially correct (e.g., when the answer was "Winnie and the monkeys", just saying "Winnie" or "the monkeys"), and 0 points for incorrect answers. Two raters evaluated all the answers from all the learners and inter-rater reliability was high (Cronbach's $\alpha = .998$).

Peabody Picture Vocabulary Test (PPVT). The purpose of using the PPVT (Dunn and Dunn 2007) was twofold. First, it was used as a measure of English proficiency to ensure that the learners in the two schools were comparable. The PPVT is appropriate to learners of all ages and proficiency levels so we considered it a good instrument for our target population. The test includes 228 target items in order of difficulty. The target words are organized in 19 sets of 12 words each and the test stops when the participant makes eight or more errors in a set. The participants' task is to identify the word that they hear from the researcher and pick the right picture out of four different pictures that appear on each page.

Apart from ensuring comparability among our learners, the PPVT also offers information about learners' receptive vocabulary, which was used in the present study to explore the relationship between lexical knowledge and reading behavior. The score that we used for the analyses was the raw score, which is calculated by subtracting the number of errors from the number of the ceiling item (standard scores were not used as they are based on native speaker norms).

Oral interview. In order to find out about students' perceptions about the graded reader and their preferences for reading mode, we performed an oral interview, which included the following open questions in the students' L1: (1) Did you like the book? (2) Did you think the story was difficult? (3) Which mode did you prefer, RO or RWL? and (4) In which mode was it easier for you to understand the content?

4.3 Procedure

The learners were tested individually in a quiet room. After the explanation of the reading procedure, the eye-tracker was set up and calibrated using a 5-point calibration. Participants' eye-movements were recorded using Tobii T 120 (Tobii, www.tobii.com). This is a remote, desktop eye-tracker, with the camera and infrared light integrated in the monitor. It has a sampling rate of 120 Hz, which is considered adequate for the examination of fixations to larger regions of interest (Conklin and Pellicer-Sánchez 2016). It has a typical accuracy of 0.5° (measured in ideal conditions) and 0.2° resolution. The visual and auditory stimuli were displayed on the screen using Tobii Pro Studio (version 3.4.2).

We adopted a counterbalanced design, where participants in Condition 1 ($n = 21$) received written and aural input for Part 1 (RWL mode) and they only read Part 2, while the other group (Condition 2, $n = 15$) read Part 1 and simultaneously read and listened to Part 2. All the participants used headphones for reading the whole book (technically they were only necessary in the RWL mode,

but they were also helpful to isolate the students from any potential outside noise). In the RWL part the pace was marked by the audio (pages automatically turned when the audio for the page finished), while in the RO mode, the students followed their own pace and clicked the mouse to move from one page to the next. Another 5-point calibration was performed between the two parts. The whole reading process took on average 10 minutes, after which, the learners answered orally a set of 10 comprehension questions (5 for each of the two parts in which the book was divided). Additionally, the researcher asked each student about their perception of reading in the two modes. Finally, the participants completed the PPVT with a research assistant in a different room. The duration of the PPVT depended on participants' proficiency and it ranged from 10 to 24 minutes.

4.4 Analysis

Online measures. Three late eye-movement measures were analyzed: *dwell time* (i.e. the sum of all fixation durations made within an AOI), *number of fixations* (i.e. the total number of fixations made within an AOI), and *average fixation duration* (i.e. the mean of the duration of each individual fixation within an AOI). In relation to the processing of the text, longer reading times and more fixations are considered a reflection of greater processing difficulty. These late measures reflect not only lexical factors, but are also influenced by the contextual and discourse-level features of the text being read (Conklin et al. 2018). In the processing of the image, longer processing time and higher number of fixations are usually related to salient or appealing aspects of the image (Conklin et al. 2018).

The text and image areas of each page of the illustrated graded reader were selected as AOIs. The AOIs were manually drawn. The text AOIs were drawn making sure that the space around the text was the same in all text AOIs. The position and size of the image and text AOIs differed across pages. In the RWL condition AOIs were time-locked to the auditory stimuli and were therefore only active during the duration of the audio. Image and text AOIs were both active at the same time. Those AOIs corresponding to parts of the auditory stimuli that were particularly salient or different, e.g. including a song or other noises related to the events of the story, were discarded. After these deletions, a total number of 24 pages were included in the analyses (10 in part one and 14 in part two). Fixations to the text and image AOIs were analyzed.

Data from two participants in the RO mode were discarded because of low data quality. The percentage of eye-tracking samples correctly identified for the

remaining participants was above 78% (Max = 100%). Single fixations shorter than 70 ms were discarded. There were no significant differences between the data lost in both conditions, i.e. RO (5303 data points; 15.8% of the data) and RWL (4483 data points; 16.3% of the data). These percentages of data loss are higher than those normally reported by eye-tracking reading studies, which is expected given that the study was conducted with younger participants in remote mode, and considering the properties of the equipment used.

The use of an unmodified graded reader in this study had the advantage of adding ecological validity to the study but this meant having to perform a more careful analysis of eye-movement data. First, *dwelt time*, the *number of fixations*, and the *average fixation duration* were calculated for each text and image AOIs and averaged for each participant. Analyses with these raw mean values were conducted first. However, because of the different sizes across text and image AOIs, bigger areas would attract more fixations and this would affect *dwelt time* and *number of fixations* of the AOIs. Eye-movement data was normalized accounting for these size differences. *Dwelt time* and the *number of fixations* calculated per AOIs were divided by the size of the AOI in megapixels (width x depth in pixels/1,000,000).

In order to further confirm the pattern of results, and following the procedure used in previous studies looking at subtitles (e.g. Bisson et al. 2014), data of the text area was further normalized dividing the fixation measures by the number of words in each AOI. Analyses of *dwelt time* and *number of fixations* in text AOIs were then conducted with these normalized values. This second stage of normalization would confirm the patterns found with the previous normalization procedure and would make results more comparable with previous studies. Analyses of eye-movement data were therefore conducted in three steps, i.e. with the non-normalized data, with data normalized by the size of the text and image AOIs, and finally data of the text AOIs normalized by the number of words in each AOI.

Statistical analyses. In order to answer the first research question (RQ1), ten Generalized Linear Mixed Models (GLMMs) were run using SPSS v25. These analyses were adopted because the data were not normally distributed for the measures of attention to the images, and also because the means for the text and the image areas were so different (as well as the standard deviations). One model was conducted for each of the measures examined (i.e. *dwelt time*, *dwelt time* normalized by AOI size, *number of fixations*, *number of fixations* normalized by AOI size, *average fixation duration*) and each of the input sources (i.e. text, and image). Additionally, two other GLMMs were run for *dwelt time* and *number of fixations* normalized by number of words in text AOIs. These models included a repeated measures

Diagonal structure, plus a random intercept for Subject (except for non-normalized analyses of *dwelt time* and *number of fixations* on the text, in which this parameter led to a non-positive definiteness of the Hessian matrix). Due to the temporal/count nature of the original dependent variables, which are typically skewed towards positive values, they were analyzed following a Gamma distribution (Log link). The relationship between processing behavior and reading comprehension was examined by means of correlations (RQ2). In order to examine the role of participants' receptive vocabulary knowledge in their reading behavior in the two modes (RQ3), correlations were calculated between the scores of the PPVT and the online reading measures. Finally, descriptive statistics were performed to examine participants' perception of reading in the two modes (RQ4).

5 Results

5.1 Relationship between eye-movement behavior and reading mode

Table 1 presents the descriptive statistics of the three eye-tracking measures (*dwelt time*, *number of fixations*, and *average fixation duration*) in the two modes for the two input sources. This table includes the non-normalized data first in ms and then the normalized data when applicable.

Table 1: Descriptive statistics of eye-tracking measures.

Measure	RWL Mean (SD)	RO Mean (SD)
Dwell time text (ms)	10,049 (2,930)	14,239 (4,000)
Dwell time text (norm.)	148,986 (35,256)	186,535 (55,420)
Dwell time images (ms)	1,918 (1,365)	1,216 (869)
Dwell time images (norm.)	10,474 (7,332)	3,523 (2,702)
Number of fixations text	36 (9)	48 (11)
Number of fixations text (norm.)	519 (107)	636 (143)
Number of fixations images	9 (5)	6 (4)
Number of fixations images (norm.)	44 (28)	17 (13)
Average fix. dur. text (ms)	284 (43)	285 (50)
Average fix. dur. images (ms)	215 (37)	186 (36)

Note: *norm* = normalized

It can be seen that learners did process both sources of input. Importantly, for all measures and regardless of reading mode, learners spent more time on the text than on the images. The presence of pictorial support did not fully detract them from reading the text. The time devoted to processing the text was generally higher in the RO mode than in the RWL, while the opposite was true for the images. These results were the same for the normalized and the non-normalized data.

Table 2 shows the results of the ten GLMMs, in which “ β (RO)” stands for the fixed coefficient assigned to the RO mode taking RWL as the reference value; i.e., a positive β value indicates that RO > RWL, and vice versa. As can be seen from the results, when text was the input source, all measures except for *average fixation duration* showed significantly more processing time in the RO than in the RWL mode. Conversely, when image was the input source, all the measures indicate that images were significantly more attended to in the RWL mode than in the RO mode.

Table 2: Results of the GLMMs.

Input Source	Measurement	F	df1	df2	Sig.	β (RO)	SE	t
Text	Dwell time	26.349	1	68	.000	.348	.0678	5.133
	Dwell time (norm.)	19.483	1	68	.000	.228	.0516	4.414
	Number of fixations	21.119	1	68	.000	.264	.0574	4.595
	Number of fix. (norm.)	26.080	1	68	.000	.203	.0398	5.107
	Average fixation duration	.044	1	68	.834	.004	.0195	.211
Image	Dwell time	9.004	1	68	.004	-.488	.1628	-3.001
	Dwell time (norm.)	48.340	1	68	.000	-1.125	.1618	-6.953
	Number of fixations	8.873	1	68	.004	-.393	.1319	-2.979
	Number of fix. (norm.)	45.285	1	68	.000	-.969	.1439	-6.729
	Average fixation duration	19.109	1	68	.000	-.144	.0330	-4.371

Normalized by number of words. In order to explore the effect of reading mode on *dwell time* and *number of fixations* to the text normalized by the number of words in the AOIs, two additional GLMMs were run using the same structure as described in the previous section (i.e., a repeated measures Diagonal structure, plus a random intercept for Subject). The descriptive statistics appear in Table 3:

The results of the GLMMs (Table 4) confirmed the previous results in that significantly more processing time took place in the RO than in the RWL mode.

Table 3: Descriptive statistics of eye-tracking measures in the text areas (data normalized by number of words).

Measure	RWL Mean (SD)	RO Mean (SD)
Dwell time (norm. words)	385 (72)	442 (123)
Number of fixations (norm. words)	1.3 (0.2)	1.5 (0.3)

Table 4: Results of the GLMMs (data normalized by number of words).

Measure	F	df1	df2	Sig.	β (RO)	SE	t
Dwell time (norm. words)	8.531	1	68	.005	.140	.0480	2.921
Number of fix. (norm. words)	10.748	1	68	.002	.137	.0418	3.278

5.2 Eye-movement behavior and reading comprehension in RWL and RO

The mean results of the reading comprehension test (there were 5 questions for each mode) were similar in the two reading modes: $M = 2.23$, $SD = 1.62$ (RWL), and $M = 2.15$, $SD = 1.58$ (RO). The scores ranged from 0–5 in both conditions, and there were no significant differences between reading modes: $t(34) = .422$, $p = .676$. In order to examine whether processing of text was related to reading comprehension in the two reading modes, Pearson correlations were calculated between the three eye-tracking measures and the comprehension scores in the two modes (i.e. comprehension scores for those questions relating to the part of the story presented in RO or RWL mode). For these correlations the *average fixation duration*, and the size-normalized values for *dwell time* and *number of fixations* were used.

The results of the correlations indicated that, in the RWL mode, there was a negative correlation between reading comprehension and *dwell time* and *average fixation duration*. In the RO mode, the scores in the three online measures negatively correlated with the reading comprehension scores (see Table 5).

None of the correlations between the online processing of images and reading comprehension scores were significant. This result was expected, as the comprehension questions tapped issues that could only be inferred from processing the text.

Table 5: Correlations between eye-tracking measures for the text AOIs and reading comprehension.

	RWL Comprehension	RO Comprehension
RWL Total dwell time	-.472**	
RWL Number of fix.	n.s.	
RWL Average fix. dur.	-.471**	
RO Total dwell time		-.444***
RO Number of fix.		-.383*
RO Average fix. dur.		-.520**

*** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$, n.s. = non-significant

5.3 The role of previous vocabulary knowledge on reading performance in RWL and RO

Correlations were calculated between previous vocabulary knowledge, as assessed by learners’ performance on the PPVT, and their reading behavior on the one hand, and their reading comprehension on the other, for the two reading modes under study. The scores of the PPVT correlated significantly (and negatively) with only one online reading measure; namely, *average fixation duration* to the text in the RO mode ($r = -.374$, $p = .030$). Additionally, there was a positive significant correlation between vocabulary knowledge and reading comprehension in the RWL ($r = .336$, $p = .049$) and also in the RO mode ($r = .446$, $p = .013$). Because of this correlation, and the potential role of vocabulary knowledge in reading comprehension, the relationship between eye-movement behavior and reading comprehension presented in the previous section was reexamined through a partial correlation, including the scores of the PPVT as a covariate and a similar pattern of results emerged.

5.4 Students’ perceptions

We asked the students four open questions in their L1 about their perceptions on the difficulty of the book and their preference for reading mode. Regarding the first question, “Did you like this book?”, only 2.9% said that they did not like it much, 42.9% said that it was OK, 31.4% said they liked it; and finally, 22.9% said that they liked it a lot. While it might be true that there might have been cases of acquiescence bias, it can be said that in general the students’ perception of the book was positive. When asked about the difficulty in understanding the

contents of the book, 28.6% of the learners considered the book difficult; 62.9% said it was OK, and only 8.6% said it was easy. In general, then, most of the students considered the book appropriate in terms of difficulty.

Regarding preference for reading mode, most learners preferred RWL (67.6%), only 11.8% preferred RO, and 20.6% did not have any preference. Finally, we asked the students whether it was easier to understand the story in one of the modes. The answers to this question were similar to the previous one: most students found RWL helped comprehension more than RO (64.7%), 20.6% thought the opposite, while 14.7% thought the two modes were comparable in this sense.

6 Discussion and conclusion

The present study used eye-tracking to examine the processing behavior of a group of children in RO and RWL modes, while relating online processing to offline comprehension measures. Moreover, the role of previous vocabulary knowledge and students' perceptions was also analyzed.

Concerning the first research question, which examined learners' processing of the two input sources under examination (text and images) in the two reading modes, results showed that learners did process both sources of input in this type of multimedia materials. Contrary to what has been previously suggested, learners did process the text in the presence of pictures and, in fact, the text areas attracted most of the learners' attention regardless of reading mode. This finding goes against some concerns that have been raised regarding the potential distracting effect that images could have on attention to the text in graded readers (Hill 2013). Importantly, the presence of audio caused differences in the way young learners processed the text and pictures in these conditions. Regarding the text, although no differences were found in *average fixation duration* between RO and RWL, the results of the GLMMs for *dwelt time* and *number of fixations* suggest that learners spent more time and fixated more on the text in the RO mode than in the RWL mode. The results were the same for the three different types of analyses performed (non-normalized data, data normalized by size, data normalized by number of words), indicating a robust effect. The differences in the processing of the textual AOIs were expected as in the RO mode learners could read for as long as they wanted before turning the page, whereas in RWL the speed of the auditory input would determine learners' reading speed. With respect to the processing of the image areas, significant differences were also found in the three eye-tracking measures; crucially, the

presence of the audio allowed readers to pay more attention to the features of the visual input.

The lack of difference between the *average fixation duration* on the text in RO and RWL contrasts with Rayner's (1998) claims for adult L1 readers: fixation durations tend to be longer and number of fixations higher in reading aloud and RWL than in silent reading. Rayner (1998) explains that in the case of reading aloud and RWL readers make an effort not to get too far ahead of the voice, resulting in longer fixations. This suggests that RWL slows down the normal reading pace of adult L1 readers, and they would read faster in a RO condition. The findings of the present study suggest that this is not the case for young L2 learners.

Although average fixation durations to the text were similar in RWL and RO conditions, overall, there was more reading (longer dwell time and higher number of fixations) in the RO than the RWL mode. The learners under the RO mode spent more time reading the text, which is a reflection of the fact that the pace in the RWL was faster than the reading pace of the average student included in this study, which led to shorter dwell time and fewer fixations in RWL.

The fact that the learners in the RWL mode spent significantly more time focusing on the images than in the RO mode indicates that they could "move away" from the text because they had the audio support and could then check the images more often. This finding suggests that the presence of audio allows for a better integration of the text and images in this type of multimodal materials. Similarly, thanks to the audio support, the RWL mode probably helped learners segment the information in the text more easily (Webb and Chang 2012). In fact, Han and Chen (2010) suggest that a common explanation for reading dysfluency is that some learners might experience difficulties in segmenting the written information into meaningful constituents, because of the absence of prosodic cues in the reading mode (Schreiber 1980). The faster reading found in RWL might be an indication of the positive effect that having access to prosodic cues (via the oral rendition of the text) could have on reading fluency, and might also help to explain the advantages shown by the RWL condition in previous research. Also, RWL possibly helps the recognition of some words that might be more familiar to the learners in the spoken form (Webb and Chang 2012) and that is why less time is necessary to process the written information and thus the learners could divide more easily their attention between the text and the images than in the RO mode. Interestingly, the delivery of the verbal input in auditory mode seems to have freed some of the cognitive demands imposed by the presence of multimodal input in the RWL condition, which allowed the students to spend more time processing the images.

Importantly, despite the reported differences in online processing of the information included in the graded reader caused by the presence of auditory input, these were not reflected in reading comprehension (question addressed in RQ2). On the one hand, readers under the RWL mode did not show worse reading comprehension, despite the fact that they spent more time processing the images and less time processing the text than in RO mode. This finding indicates that looking at the images included in graded readers did not have a detrimental effect in comprehending the text.

On the other hand, regardless of mode, longer and more fixations to the text were negatively correlated with reading comprehension. Indeed, spending more time processing the text was not necessarily associated with better comprehension. Longer reading time might instead be an indication of processing difficulties and increased processing effort that is later reflected in comprehension difficulties. Many studies have shown negative correlations between reading fluency and reading comprehension. For instance, Han and Chen (2010) claim that “Poor readers often spend a great deal of their cognitive resources on decoding and have little left for comprehension. Conversely, good readers decode words quickly and accurately, thus conserving more resources for comprehension” (p. 243). This is partially in line with Chang and Choi (2014), who found that more attention to the seductive text also led to worse comprehension. It seems that text that requires extra cognitive effort and therefore more processing time (because of lower proficiency in our case and because of emotionally salient content in the case of Chang and Choi 2014), leads to lower comprehension scores.

The results of the present study in relation to RQ3 indicated that previous vocabulary knowledge, as assessed by the PPVT, was negatively correlated with *average fixation duration* in the RO mode. This result suggests that those learners with high vocabulary scores had shorter fixations; or, what is the same, read more quickly. No correlations were found between previous vocabulary knowledge and any of the eye-tracking measures in the RWL mode, which might be related to the fact that all learners (regardless of their vocabulary knowledge and reading speed) were following the pace marked by the audio. This suggests that previous vocabulary knowledge might be a better indicator of reading speed in RO conditions.

Finally, our results concerning students’ perceptions (RQ4) are consistent with those reported in the literature for RWL vs. RO. As in other studies (Brown et al. 2008; Chang 2009; Chang and Millett 2014, Chang and Millett 2015), students preferred the RWL mode and found it easier to understand information when they were simultaneously reading and listening as opposed to when they were only reading.

These findings have shed light onto our understanding of young learners’ reading behaviour in the presence of pictures and auditory input. Adding audio

to reading materials pushes young learners to read faster and allows for a better integration of text and pictures in multimodal materials. Because the aim of our study was to examine how processing the text affected reading comprehension, we only included comprehension questions referring to information in the text exclusively. In the future, it would be interesting, however, to examine the role of processing images in understanding the whole story told in a graded reader (and compare it to the processing of text). The design should include questions about aspects of the story that could be inferred from the images alone, text alone, and both text and images and analyze how processing of the two input sources affects comprehension in the two modes. There are several studies that claim that pictorial information help L2 reading comprehension (e.g. Elley and Mangubhai 1983; Tang 1992), and this is definitely a topic worth exploring further. Future studies should also be conducted with learners of different age and proficiency to explore how the processing patterns found here change with different populations.

To conclude, the fact that RWL pushes learners to read faster than they would do in a RO situation, possibly contributes to the development of reading fluency. Importantly, RWL allows them to focus on both text and images in a way that the RO mode cannot do. Our findings thus suggest that RWL conditions allow younger learners to attend to the different sources of input in multimodal conditions without negative consequences for reading comprehension. This, together with the overwhelming preference for RWL among learners, supports its use with young EFL learners as a way to promote learners' engagement and practice with the English language.

Acknowledgements: We would like to thank Andrej Krasnansky and Radha Chandy for their help in this study.

Funding: This work was supported by the Ministerio de Ciencia e Innovación, (Funder Id: <http://dx.doi.org/10.13039/501100004837>, Grant Number: FFI2016-80576-P).

References

- Alderson, J. C. & A. H. Urquhart. 1985. The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing* 2. 192–204.
- Anderson, R. C. & P. Freebody. 1983. Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutson (ed.), *Advances in reading/language research: A research annual*, 231–256. Greenwich, CT: JAI Press.

- Bisson, M.-J., W. Van Heuven, K. Conklin & R. Tunney. 2014. Processing of native and foreign language subtitles in films: An eye-tracking study. *Applied Psycholinguistics* 35. 399–418.
- Bisson, M.-J., W. Van Heuven, K. Conklin & R. Tunney. 2015. The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology* 68. 1306–1326.
- Brown, R., R. Waring & S. Donkaewbua. 2008. Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language* 20. 136–163.
- Cadwallader, J. 2009. *Uncle Jack and the Bakonzi Tree*. Cheltenham: ELI Publishing.
- Carrell, P. 1991. Second language reading: Reading ability or language proficiency? *Applied Linguistics* 12(2). 159–179.
- Carrell, P. L. 1983. Three components of background knowledge in reading comprehension. *Language Learning* 33. 183–203.
- Center, Y., L. Freeman, G. Robertson & L. Outhred. 1999. The effect of visual imagery training on the reading and listening comprehension of low listening comprehenders in year 2. *Journal of Research in Reading* 22. 241–256.
- Chang, A. 2009. Gains to L2 listeners from reading while listening only in comprehending short stories. *System* 37. 652–663.
- Chang, A. & S. Millett. 2014. The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal* 68. 31–40.
- Chang, A. & S. Millett. 2015. Improving reading rates and comprehension through audio-assisted extensive reading for beginner learners. *System* 52. 91–102.
- Chang, Y. & S. Choi. 2014. Effects of seductive details evidenced by gaze duration. *Neurobiology of Learning and Memory* 109. 131–138.
- Conklin, K. & A. Pellicer-Sánchez. 2016. Using eye-tracking in applied linguistics and second language research. *Second Language Research*. Epub ahead of print doi:10.1177/0267658316637401
- Conklin, K., A. Pellicer-Sánchez & G. Carrol. 2018. *Eye-tracking: A guide for applied linguistics research*. Cambridge: CUP.
- Council of Europe. 2001. *Common European framework of reference for languages*. Cambridge: CUP.
- Dunn, L. M. & D. M. Dunn. 2007. *Peabody picture vocabulary test*, Fourth edn (PPVT™-4). Minneapolis, MN: Pearson Education.
- Elley, W. B. & F. Mangubhai. 1983. The impact of reading on second language learning. *Reading Research Quarterly* 19. 53–67.
- Evans, M. A. & J. Saint-Aubin. 2005. What children are looking at during shared storybook reading: Evidence from eye movement monitoring. *Psychological Science* 16. 913–920.
- Grabe, W. 2009. *Reading in a second language: Moving from theory to practice*. Cambridge: Cambridge University Press.
- Han, Z. & C. A. Chen. 2010. Repeated-reading-based instructional strategy and vocabulary acquisition: A case study of a heritage speaker of Chinese. *Reading in a Foreign Language* 22. 242–262.
- Hill, D. 2013. Graded readers. *ELT Journal* 67(1). 85–125.
- Jeon, E. H. & J. Yamashita. 2014. L2 reading comprehension and its correlates: A meta-analysis. *Language Learning* 64. 160–212.
- Johnson, C. I. & R. E. Mayer. 2012. An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*. 18. 178–191.

- Justice, L. M., L. Skibbe, A. Canning & C. Langkford. 2005. Pre-schoolers, print and storybooks: An observational study using eye movement analysis. *Journal of Research in Reading* 28. 229–243.
- Kiss, T. & C. Weninger. 2016. Culture learning in the EFL classroom: The role of visuals. *ELT Journal*. Advance online publication: doi:10.1093/elt/ccw072
- Koda, K. 2005. *Insights into second language reading: A cross-linguistic approach*. NY: Cambridge University Press.
- Kruger, J. L. & F. Steyn. 2014. Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly* 49. 105–120.
- Lightbown, P.M. 1992. Can they do it themselves? A comprehension-based ESL course for young children. In R. Courchene, J. St John, C. Therien & J. I. Glidden (eds.), *Comprehension-based second language teaching*, 353–370. Ottawa: University of Ottawa Press.
- Lightbown, P.M., R.H. Halter, J. White & M. Horst. 2002. Comprehension-based learning: the limits of “do it yourself.”. *The Canadian Modern Language Review* 58. 427–464.
- Marcus, N., M. Cooper & J. Sweller. 1996. Understanding instructions. *Journal of Educational Psychology* 88. 49–63.
- Mason, L., P. Pluchino, M. C. Tornatora & N. Ariasi. 2013a. An eye-tracking study of learning science text with concrete and abstract illustrations. *Journal of Experimental Education* 81. 356–384.
- Mason, L. & M. C. Tornatora. 2015. Integrative processing of verbal and graphical information during re-reading predicts learning from illustrated text: An eye-movement study. *Reading and Writing* 28. 851–872.
- Mason, L., M. C. Tornatora & P. Pluchino. 2013b. Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers and Education* 60. 95–109.
- Omaggio, A. C. 1979. Pictures and second language comprehension: Do they help? *Foreign Language Annals* 12(1979). 107.
- Pickering, M., S. Frisson, B. McElree & M. Traxler. 2004. Eye movements and semantic composition. In M. Carreriras & C. Clifton (eds.), *On-line study of sentence comprehension: Eye tracking, ERPs and beyond*, 33–50. New York: Psychology Press.
- Qian, D. 1999. Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review* 56. 282–308.
- Rasinski, T. V. 1990. Effects of repeated reading and listening-while-reading on reading fluency. *The Journal of Educational Research* 83. 147–151.
- Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124. 372–422.
- Rayner, K. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* 62(8). 1457–1506. doi:10.1080/17470210902816461.
- Rodgers, M.P.H. 2018. The images in television programs and the potential for learning unknown words: The relationship between on-screen imagery and vocabulary. *ITL - International Journal of Applied Linguistics* 169(1). 191–211. doi:10.1075/itl.00012.rod.
- Roy-Charland, A., J. Saint-Aubin & M. A. Evans. 2007. Eye movements in shared book reading with children from kindergarten to grade 4. *Reading and Writing* 20. 909–931.
- Schreiber, P. A. 1980. On the acquisition of reading fluency. *Journal of Reading Behavior* 12. 177–186.

- Tang, G. 1992. The effect of graphic representation of knowledge structures on ESL reading comprehension. *Studies in Second Language Acquisition* 14. 177–195.
- Tragant, E., C. Muñoz & N. Spada. 2016. Maximizing young learners' input: An intervention program. *The Canadian Modern Language Review* 72. 234–257.
- Trofimovich, P., P.M. Lightbown, R. Halter & H. Song. 2009. Comprehension based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition* 31. 609–639.
- Webb, S. & A. Chang. 2012. Vocabulary learning through assisted and unassisted repeated reading. *The Canadian Modern Language Review* 68. 267–290.
- Webb, S. & A. Chang. 2015. Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research* 19. 667–686.