INTERNATIONAL LITERACY ASSOCIATION
READING RESEARCH QUARTERLY

Online Processing Shows Advantages of Bimodal Listening-While-Reading for Vocabulary Learning: An Eye-Tracking Study

Alessandra Valentini
School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

School of Human Sciences, University of Greenwich, London, UK

Centre for Thinking and Learning, Institute for Lifecourse Development, University of Greenwich, London, UK

Institute for Inclusive Communities and Environment, University of Greenwich, London, UK

Rachel E. Pye

Carmel Houston-Price
School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

Jessie Ricketts
Department of Psychology, Royal Holloway, Univeristy of London, London, UK

Julie A. Kirkby
Department of Psychology, Bournemouth University, Poole, UK

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 0(0)
pp. 1–23 | doi:10.1002/rrq.522

ABSTRACT

Children can learn words incidentally from stories. This kind of learning is enhanced when stories are presented both aurally and in written format, compared to just a written presentation. However, we do not know why this bimodal presentation is beneficial. This study explores two possible explanations: whether the bimodal advantage manifests online during story exposure, or later, at word retrieval. We collected eye-movement data from 34 8-to 9-year-old children exposed to two stories, one presented in written format (reading condition), and the second presented aurally and written at the same time (bimodal condition). Each story included six unfamiliar words (non-words) that were repeated three times, as well as definitions and clues to their meaning. Following exposure, the learning of the new words' meanings was assessed. Results showed that, during story presentation, children spent less time fixating the new words in the bimodal condition, compared to the reading condition, indicating that the bimodal advantage occurs online. Learning was greater in the bimodal condition than the reading condition, which may reflect either an online bimodal advantage during story presentation or an advantage at retrieval. The results also suggest that the bimodal condition was more conducive to learning than the reading condition when children looked at the new words for a shorter amount of time. This is in line with an online advantage of the bimodal condition, as it suggests that less effort is required to learn words in this condition. These results support educational strategies that routinely present new vocabulary in two modalities simultaneously.

Vocabulary knowledge is essential for language comprehension; it supports listening and reading comprehension (Ouellette, 2006; Suggate et al., 2018) and is fundamental for academic achievement (Biemiller, 2003; Schuth et al., 2017). This study explores how children process and learn new vocabulary when reading and listening to stories at the same time, and how this bimodal presentation affects online processing and offline learning differently from written-only presentations of new words.

Theoretical Approaches to Word Learning in Bimodal and Unimodal Presentations

Children and adults acquire much of their vocabulary knowledge incidentally when they are exposed to language while listening (Elley, 1989; Wilkinson & Houston-Price, 2013) or reading (Nagy et al., 1987; Ricketts et al., 2011). However, encountering words in the oral and written modality simultaneously (bimodal presentation) appears to be particularly beneficial for vocabulary acquisition. For example, studies exploring the "orthographic facilitation effect" (see Colenbrander et al., 2019 for a review) showed that new words are learned better when both written and oral forms are provided, compared to when the word is presented only orally, both in children (Ricketts et al., 2009) and adults (Miles et al., 2016). Similarly, studies that explore a "phonological facilitation" effect (i.e., the superiority of bimodal presentation to written presentation) show that children asked to read new words aloud while reading stories learned these words better than those who read silently (Rosenthal & Ehri, 2011). Research that has investigated learning from stories has shown that combined presentation of oral and written texts benefits comprehension (Montali & Lewandowski, 1996) and learning of words' meanings (semantic learning; Valentini et al., 2018) compared to written or oral presentation alone. For example, Valentini et al. created a story containing low-frequency words and asked 8- to 9-year-old children to read (a written presentation), listen to (an oral presentation), or read and listen to the story at the same time (a bimodal presentation). They found that children in the bimodal presentation condition were better at identifying the semantic categories of the new words than children exposed to the story in either single-modality presentation conditions.

Two accounts might be proposed to explain the facilitative effect of bimodal presentation on semantic learning. The first account is linked to the Lexical Quality Hypothesis (LQH: Perfetti & Hart, 2002), which suggests that words with higher quality representations are retrieved more easily from memory. According to the LQH, lexical representations are of higher quality if they include better specified and well-integrated information about their forms (phonology, orthography) and meaning (semantics). Bimodal presentation provides information about both phonological and orthographic forms, whereas unimodal presentation only provides information about one form, depending on whether the word is heard (phonology) or seen (orthography). Compared to unimodal presentation, bimodal presentation enables higher-quality representations because children can readily incorporate well-specified information about both forms in the new lexical representation.

An alternative account of the facilitative effect of bimodal presentation derives from Cognitive Load Theory in multimedia learning (CLT: Mayer, 2014; Mayer et al., 1999; Paas et al., 2003). This account (CLT) postulates that situations that reduce cognitive load are more conducive to learning. For bimodal presentation, the provision of the same information in two different modalities (oral and written) can reduce the cognitive load involved in forming a word's representation the first time it is encountered by removing the need for orthography to phonology conversion during reading (as the oral form is directly provided). These freed resources would allow children to devote more attention to the meaning of the word and surrounding text while they read and listen to it. This account would therefore explain orthographic facilitation effects in terms of the processes that occur during the very first encounter with the new word. Interestingly, this account predicts different patterns depending on reading skills; age can be used as a proxy of skill, especially when comparing adults and children. Specifically, based on the "redundancy principle" (the idea that redundant information impairs learning), the account would posit that, for expert readers, the conversion of orthography to phonology happens automatically. For expert readers, therefore, the additional oral information in dual–modality situations (bimodal conditions) is redundant, and it could impair learning by increasing (rather than reducing) cognitive load. Learning could also be impaired for expert readers in the bimodal condition if the provided phonological form differs from the one created through phonological recoding.

While the two accounts both predict benefits in learning new words in bimodal presentation conditions for children, they differ in the *locus* of the beneficial effects. According to the LQH, we could expect that bimodal presentations will lead to higher-quality representations being formed during the first encounter with a new word. When the word is encountered a second or third time, fewer resources are then required to process its form, freeing resources for the encoding of other lexical information (e.g., meaning, context). Therefore, we can hypothesize that any facilitation due to bimodal presentation should be seen only after the first presentation of a new word, as a "delayed" facilitation effect. Indeed, some evidence suggests this is the case, for example. Ricketts et al. (2009) found better performance in vocabulary training sessions for new words presented in bimodal conditions compared to words presented orally, but only after the first training session. To our knowledge, this has not been tested when comparing reading and bimodal conditions. In contrast, the CLT is in line with a reduction in cognitive demands during the bimodal presentation of a new word, suggesting facilitation from the very first presentation, as freed resources allow greater immediate online semantic processing. It is important to note that the two proposed mechanisms (high-quality representations and reduction in cognitive load) are not mutually exclusive and could both facilitate vocabulary acquisition in bimodal conditions. Thus, facilitation might be both immediate, as predicted by the CLT, and delayed, as predicted by the LQH. The current study uses eye tracking during online word learning in a sample of school-aged children to identify the locus of the bimodal facilitation effect.

Online text Processing While Reading or Reading and Listening to Stories

There is a rich literature on how readers explore a written text (Rayner et al., 2012). Eye-movement research assumes that the time spent fixating a word is indicative of its status in the lexicon and whether it has been successfully encoded. Researchers use different measures to explore the time spent on text, particularly *first fixation duration*, which is the duration of the fixation on a word the first time the eye lands on it; *gaze duration*, defined as the time spent on a word before moving to another part of the text (i.e., the sum of first fixation duration and all subsequent fixations on the word before the eye moves to another area); the same measure is called *first-pass reading time* when referring to multi-word clusters; *re-reading time*, the time spent on a word or part of the text after the eye has left the area for the first time; and *total reading time*, the sum of gaze duration and re-reading time. Movements between different areas of text are called *saccades*, and leftward movements while reading are called *regressions* and indicate the reader is re-examining a part of text previously read. On average, readers move their eyes forward by 8–9 characters with each saccade, and an average fixation is 218 ms. However, reader and text differences influence how the text is explored, with more difficult texts prompting longer fixations, smaller saccades, and more regressions (Rayner et al., 2012).

Compared to the literature on eye movements while reading, research on reading behaviors in bimodal conditions is scarce. Attention to the text varies as a function of reading ability, of which age is often used as a proxy. Therefore, by comparing studies with adults and studies with children, we can identify the effect of reading ability, given the differences in reading abilities between the two age groups, at least at the group level. Studies have shown that adults pay careful attention to text in video or image captions (d'Ydewalle et al., 1991; Rayner et al., 2001), even when the text is redundant to the oral information or not useful (Ross & Kowler, 2013). In contrast, in shared picture-book reading contexts, children do not always attend to the text while listening to stories; unsurprisingly, pre-readers spend very little time looking at the print (Evans & Saint-Aubin, 2005), while older and more able readers spend more time on the text. However, even older children do not spend all the given time attending to the written text when adults read them books with pictures (66% time on text at 9–10 years of age); the same is true of young second language learners (Pellicer-Sánchez et al., 2020; Serrano & Pellicer-Sánchez, 2019). However, younger children will read along more often if the text is appropriate for their reading level (Roy-Charland et al., 2007). In sum, readers tend to pay more attention to the text as they get older and their reading skills develop, and their attention to the text is a product of both reading skill and text difficulty. A recent study exploring adults' attention to the text while reading and listening (Conklin et al., 2020) found a tendency to read ahead of the oral presentation, but this tendency was dependent on vocabulary knowledge in both first and second language learners; the eye movements of participants with lower vocabulary skills lagged behind the oral presentation. Adults also made more and longer fixations in the reading-while-listening condition than in the reading condition.

The current study used eye tracking to examine how children explore the text differently in unimodal and bimodal conditions. We expected different patterns of text exploration in the two modalities, possibly with more and longer fixations in the dual modality, as seen in adults (Conklin et al., 2020).

Online Lexical Processing While Reading or Reading and Listening to Stories

Previous studies have explored readers' attention to written text to investigate how adults attend to new words and extract their meaning while reading (Blythe et al., 2012; Brusnighan et al., 2014; Brusnighan & Folk, 2012; Chaffin et al., 2001; Godfroid et al., 2013; Williams & Morris, 2004). Studies have shown that new words are fixated longer than known words, indicating they require more processing (Brusnighan & Folk, 2012; Chaffin et al., 2001; Godfroid et al., 2013). With further exposures, new words become easier to process, with reading time decreasing at each encounter for both adults and children (Joseph et al., 2014; Joseph & Nation, 2018). Word repetition in bimodal conditions seems to have a similar effect to that found in reading: reading time decreases at each presentation of the new word, particularly first fixation durations (Gerbier et al., 2018).

Recent studies have found that 10- to 12-year-old attend to the text longer in reading than in reading-while-listening conditions when pictures are presented alongside the text (Pellicer-Sánchez et al., 2020; Serrano & Pellicer-Sánchez, 2019). Surprisingly, spending more time on the text did not improve comprehension in either condition; indeed, longer time on the text was indicative of processing difficulties in these studies and negatively associated with comprehension in both conditions. However, it is possible that this negative relationship might be specific to studies that present pictures alongside the text and that greater attention to the text might support comprehension when pictures are not presented. In fact, Lowell and Morris (2017) showed that fixation time on novel words and their preceding context in a reading-only condition positively predicted word learning in adults.

The current study used eye tracking to explore how children attend to new words differently in unimodal and bimodal conditions. This comparison allowed us to distinguish between an encoding facilitation effect of the bimodal condition, indicated by shorter looking times at the first repetition of a word, shorter gaze durations, or retrieval facilitation effects that would manifest as shorter reading of the second or third presentation or shorter re-reading times.

Vocabulary Learning While Reading or Reading and Listening to Stories

Eye-tracking studies of word learning while reading have explored how adults attend to the words and surrounding context in which new words are embedded, showing that readers spend more time reading the context when presented with new words compared to when known words are presented (Brusnighan & Folk, 2012; Chaffin et al., 2001), especially if the context is informative (Chaffin et al., 2001). They also make more regressions out of the surrounding context for new words than for known words (Williams & Morris, 2004). These findings suggest that readers are trying to link their representation of the new word to contextual information.

To explore whether reading patterns are connected to learning, it is necessary to measure both eye movements while reading and offline vocabulary learning, yet very few studies have included both types of measures (see review by Pellicer-Sánchez & Siyanova-Chanturia, 2018). The few studies that have done so find an association between time spent reading sentences and the learning of new words within them (e.g., Brusnighan & Folk, 2012). Total reading times are also typically longer for learned words compared to unlearned words (Godfroid et al., 2013, 2018; Mohamed, 2018; Pellicer-Sánchez, 2016). The measure of reading time used and the measure of learning adopted by different studies may influence the nature of the associations found. One study found shorter gaze duration but longer re-reading time for learned words compared to unlearned new words when learning was assessed using a synonym test (Williams & Morris, 2004), while Mohamed (2018) reported positive associations between gaze duration and the ability to produce the meaning of the new word, and between total reading time and both meaning recognition and meaning production, and Pellicer-Sánchez (2016) found a positive association between total reading time and meaning production but no association between any reading time measure and meaning recognition. Lowell and Morris (2017) found positive effects of longer first-pass time and longer re-reading time on a meaning recognition task, although the positive effect of re-reading was particularly noticeable when words were presented in a less constraining context. Interestingly, they also found that higher re-reading time in the informative context preceding the new words had a positive effect on learning. Despite the discrepancies in these findings, overall, these studies suggest that readers who spend more time attending to new words learn them better and that later fixation time measures (especially total reading time) might be more reliable predictors of learning than gaze duration. As these studies were conducted with adult readers, some of whom were second language learners (Godfroid et al., 2013, 2018; Mohamed, 2018; Pellicer-Sánchez, 2016), it remains to be seen whether children who are learning to read in their first language show similar patterns.

It is presently unknown whether and how reading time is related to learning in multimodal conditions, given the lack of research in this area. Studies from the field of multimedia research show that participants who attend to subtitles efficiently and spend more time reading them perform better in comprehension tasks about the subtitles than participants who do not (Kruger & Steyn, 2014). Very few studies have examined how looking times relate to vocabulary learning in multimodal conditions, however. Montero Perez et al. (2015) assessed adult second language learners' ability to acquire new words from videos with captions when participants were made aware (intentional group) or not aware (incidental group) of a subsequent vocabulary test. In this study, longer gaze duration was positively associated with word recognition, while longer re-reading times were negatively related to learning in the incidental group but positively related to learning in the intentional group. The authors proposed that longer re-reading reflects processing difficulties in multimodal presentations in incidental conditions. However, this interpretation contrasts with the results of studies from the reading literature that find positive effects on word learning of longer later reading measures, particularly total reading times (Godfroid et al., 2013, 2018; c.f. Williams & Morris, 2004). This might suggest opposite effects of longer re-reading or total reading times in the two conditions: a positive effect in the reading condition and a negative effect in the bimodal condition. Interestingly, the difference between effects for the incidental and intentional groups might suggest that looking times have different effects depending on the approach participants take to the task.

Some hypotheses regarding the relationship between looking patterns and vocabulary learning in multimodal conditions can be drawn from research into shared storybook reading in pre-school children (Evans & Saint-Aubin, 2013) and younger readers (Duckett, 2003), which supports the idea that providing information in more than one modality facilitates comprehension. For example, children who looked at the relevant parts of pictures (i.e., those parts that provided information about the meaning of words) while a word was being spoken, or soon after, were more likely to learn it (Evans & Saint-Aubin, 2013). This indicates that children can use the link between oral and visual presentation modalities to learn new words. Other studies show that synchronization of the oral and written presentations in bimodal conditions can enhance semantic learning, though results are not always consistent (Gerbier et al., 2015, 2018). Gerbier and colleagues presented short paragraphs containing pseudo-words repeated four times to 10- to 12-year-old (2015) and 8- to 11-year-old (2018). Paragraphs were presented either in a conventional bimodal presentation (non-synchronous presentation), where children read and listened to the text at the same time, or in a synchronous presentation, where they were instructed to follow the reading closely and the spoken word was highlighted, karaoke-style, in the text. The difference between these two presentation modalities could be likened to the difference between children paying attention to the text while the text is read (synchronous presentation) or not doing so (non-synchronous presentation). The older children showed enhanced category learning of the pseudo-words in the synchronous condition, while the younger children showed a disadvantage in this condition. The authors attributed the difference to the slower reading pace of the younger children. This suggests that synchronous presentation, or paying close attention to words while they are spoken, might be a positive strategy for older children, while younger ones might benefit from bimodal presentations that allow them to attend to the written text more sporadically. This could also suggest that reading along closely might be a good strategy for learning when the text is at the right difficulty level for the participant, while it might not be a good strategy for slower or younger readers, in line with the results of research on reading picture books (Roy-Charland et al., 2007). Given the use of age-appropriate texts in the present study, we expected children to learn words better when they followed the oral presentation more closely, in line with the results for synchronous presentations in older readers.

The Present Study

The current study uses an eye-tracking paradigm to investigate how children allocate their attention when encountering new words while reading versus reading and listening to stories at the same time. The study bridges the gap between two research areas: the literature on children's vocabulary acquisition in different presentation modalities, which highlights a positive effect of bimodal presentation (Valentini et al., 2018), and the literature on eye movements, mostly with adult participants, which shows that the time spent on new vocabulary items is related to word learning in both reading-only (Godfroid et al., 2013) and multimodal conditions (Montero Perez et al., 2015). We investigate both the online processes children use to acquire new words when exposed to stories in two different modalities and the products of this process in terms of how well children learn the link between novel word forms and their meanings. The study is novel in its exploration of children's allocation of attention to new vocabulary items in bimodal and unimodal conditions and in its attempt to distinguish between two theoretical approaches that might explain the bimodal advantage (i.e., the LQH (Perfetti & Hart, 2002) and the CLT (Mayer, 2014; Mayer et al., 1999; Paas et al., 2003)).

Thirty-four children in Year 4 of UK primary school (8- to 9-year-old) were exposed to word-like pronounceable non-words within two stories in two conditions: a reading-only condition, where they were presented with stories in the written modality, and a bimodal condition, where they listened to and read stories simultaneously. Stories were divided into passages presented on a computer screen, each containing one new word repeated three times. A brief definition was included in the text at the first mention of each new word, while in-text clues accompanied the second and third mentions of the word. Eye-movement data were collected while the children were exposed to the stories, and an offline category recognition task and a definition production task were used to assess their learning of the words' meanings following story exposure.

The study addressed the following primary research questions:

1. Do children explore the text differently in the two presentation modalities? To answer this question, we examined eye movements in the text during bimodal (oral and written) versus unimodal (written-only) presentations. Based on previous literature, we expected children of this age to spend most of the time reading along with the story when presented with text at their reading level (Roy-Charland et al., 2007).
   In terms of specific predictions, we expected children to spend more time fixating the text in the reading condition than in the bimodal condition (Pellicer-Sánchez et al., 2020; Serrano & Pellicer-Sánchez, 2019). Previous studies have found more and longer fixations in bimodal conditions in adults (Conklin et al., 2020), but fewer and shorter fixations during comparable conditions (synchronous presentations) in children (Gerbier et al., 2018). Given the age of our participants, we expected to find fewer fixations in the bimodal condition compared to the reading condition, as suggested by the results of Gerbier et al. (2018); however, these fixations might be longer in the bimodal condition, in line with the idea that bimodal presentation facilitates reading by increasing children's reading span.
2. Do children pay attention to the new written words and in-text definitions or clues differently in the two conditions? For this second question, we compared the two conditions similarly to Research Question 1, but restricted our focus to the specific areas of

interest within the passages (i.e., new words, definitions, and clues). Specifically, we compared eye-movement measures to the target words between the two conditions, distinguishing between the first, second, and third times they appeared in the text. We also compared eye-movement measures to the definitions and clues in the two conditions.

In the analysis for target non-words, we explored the effect of presentation of the target on eye-movement measures, comparing reading times at the first, second, and third presentation of the new words, expecting a reduction in reading time across presentations (Joseph & Nation, 2018). A steeper decrease across presentations in the bimodal condition than in the reading condition (that could be highlighted by a significant interaction between presentation of the target and condition) would suggest faster integration into the lexicon, which would be in line with the delayed facilitation account predicted on the basis of the LQH (Perfetti & Hart, 2002).

In this analysis, we also explored the effect of condition, comparing reading times in the two conditions and their interaction with the presentation of the target (first, second, and third presentations). In this analysis, shorter reading times in the bimodal condition compared to the reading condition at the first presentation of the word would support the online facilitation account (as per the CLT, Mayer, 2014; Mayer et al., 1999; Paas et al., 2003). Shorter first-pass reading times (gaze duration) in the bimodal condition for all presentations of the word would also support the idea of online facilitation. In both instances, a difference between conditions at the first presentation of a word would be due to encoding facilitation (online effects) rather than retrieval, as no information can be retrieved before the word has been presented the first time. However, if the results show shorter reading times in the bimodal condition only at the second or third presentation of the items and on second pass reading measures (i.e., re-reading and total reading times rather than gaze duration), this might suggest facilitation occurring at a later stage, namely retrieval (as predicted by the LQH).

Differences between exploration of the definition and clues in the two conditions were similarly examined to establish whether definitions or clues were encoded faster in the combined condition.

3. Do children learn new words better in the bimodal condition? We predicted that we would find enhanced semantic learning following bimodal presentation in comparison to reading-only presentation, in line with previous findings (Valentini et al., 2018).
4. Do eye movements to the new words and their definitions predict word learning? And does this relationship differ between presentation conditions? To answer these questions, we explored the data as described for Question 2, but in relation to the vocabulary learning task. We aimed to elucidate whether eye movements in the areas of interest (new words and their definitions) predict vocabulary learning in the two conditions. (Given that contextual clues were varied in terms of length and frequency and not as well matched between stories as definitions, the interpretation of any relationships between looking at clues and word learning was more difficult. Therefore, we do not include attention to clues among our main analyses; the relevant analyses are reported in Appendix C).

   To explore these effects, we included condition and reading time measures, as well as their interaction, as predictors of word learning. In terms of specific effects, we expected that longer reading times, especially total reading time and re-reading time, would be associated with better word learning in the reading condition, as found in previous research (Godfroid et al., 2013, 2018; Mohamed, 2018; Williams & Morris, 2004). In contrast, based on findings from multimedia learning studies (Montero Perez et al., 2015), we expected longer gaze duration and shorter re-reading time to predict word learning in the bimodal condition.
5. Finally, we explored a secondary research question in the bimodal condition: Does looking at the words or their definitions while the word or definition is spoken improve word learning? It was hypothesized that looking at the specific areas of interest in the text at the same time the oral text was heard would predict word learning in the bimodal condition. Specifically, we hypothesized that children might learn the meaning of the target non-words more easily if they looked at the word while hearing either the word itself (coincident time) or its definition or clues (cross-coincident time). Similarly, children were expected to learn word meanings better if they were reading the definition or clues while hearing the word (cross-coincident time). This hypothesis was based on the idea that hearing and reading a word or its definition or clues at the same time might result in a higher-quality representation of the word in memory.

Methods

Participants

Thirty-four children aged 8 or 9 years participated in the study (M_{age} = 8.93 years; SD = .29 years; 15 boys). The sample size was in line with previous research using eye tracking with children (Gerbier et al., 2015, 2018; Pellicer-Sánchez et al., 2020). Participants were recruited from three primary schools in the South of England. Informed parental consent was received for all participants. All children had normal or corrected-to-normal vision, and teachers confirmed the absence of special educational needs and neurological disabilities. All children were monolingual native English speakers, and their performance in standardized tasks of non-verbal abilities (Colored Progressive Matrices: CPM; Rust, 2008), vocabulary knowledge (British Picture Vocabulary Scale: BPVS—3; Dunn et al., 2009) and word and non-word reading (Set A of the Test of Word Reading Efficiency: TOWRE—Second edition; Torgesen et al., 1999) were within the normal range (see Appendix A).

Materials and Procedure

Non-word Presentation

Design

Two stories were used, each including six word-like, pronounceable non-words. Story presentation modality was manipulated within subjects so that all children were presented one story in the reading condition and the other in the bimodal condition. The order of condition (reading first vs. bimodal first), story ("*The Pirate Story*" first vs. "*The Knight Story*" first), and list of target non-words included (List A vs. List B) were counterbalanced.

Non-words

Twelve non-words were chosen from existing datasets of non-words that specify a correct pronunciation: sets B, C, and D of the TOWRE—Second edition (Torgesen et al., 1999), the Diagnostic Test of Word Reading Processes (DTWRP—Forum for Research in Literacy and Language, 2012), the Wechsler Individual Achievement Test—Second UK Edition (WIAT-II—Wechsler, 2005), and Chaffin (1997). Words were split into two lists, matched on length, bigram frequency (Medler & Binder, 2005), and phonotactic probability (Vitevitch & Luce, 2004; all $ps > .10$). Lists were also matched for pronunciation accuracy, word-likeness ratings, and ease of pronunciation by a pilot sample of 13 adults (all $ps > .40$). Items in the two lists were paired, with each pair associated with a category (animal, building, clothing, food, job, and object).

Stories

Two stories and eight passages were written for this study. Each story began with two introductory passages (each approximately 50 words in length), followed by six passages that each introduced one target non-word, repeated three times, accompanied by clues to its meaning (101–133 words in length). The order in which new word categories were presented was the same across the stories. The stories were similar in length (821 & 848, words respectively) and had a Flesch reading ease and Flesch–Kinkaid Grade Level appropriate for the age of the children (Flesch reading ease: $M_{Knight} = 84.14$; $M_{Pirate} = 82.93$; Flesch–Kinkaid Grade Level: $M_{Knight} = 4.61$; $M_{Pirate} = 4.29$). Passages in the two stories did not differ on any of these measures (all $ps > .30$). Stories can be found at: https://osf.io/mqsgr/?view_only=1d6d4bce382a473b9bf8f0fceeb11fb6.

The meanings of the non-words were provided by a definition the first time the word was mentioned and by clues to the word's meaning on its second and third presentations. Definitions were four words in length and provided the word's sub-category and further specified information. For example, for the category *animal* in the knight story, the definition *"dragon that eats sheep"* comprises both the sub-category *"dragon*," and the specific characteristic *"that eats sheep"*. Predictability and plausibility of the definitions were assessed by asking 24 adult English speakers to supply the last word of each definition given the first three (predictability) and ratings of internal plausibility on a 5-point Likert scale obtained from 15 further adults. Definitions were matched across stories for length, plausibility, and predictability, as well as for word frequency and the number of orthographic and phonological neighbors for each word in the definition (Masterson et al., 2003; all $ps > .10$). Clues accompanied the second and third mentions of each word and gave information regarding part of the definition; for example, for *"dragon that eats sheep*," the first clue was "*gigantic creature*," and the second was "*meat-eating*". Definitions and clue positions were at similar distances and positions relative to the non-words in each passage. The first time non-words were presented in a passage, they were always preceded by an adjective to minimize the probability of skipping the previous word and control for preview benefit (see Appendix B).

Recordings of the stories for the bimodal condition were read by a female native English speaker. Pilot data were used to calculate a reading speed that would match children's silent reading speed in the reading condition to ensure exposure time would be matched across conditions and reduce the likelihood of reading speed affecting children's performance (see Gerbier et al., 2018). When comparing exposure time in the bimodal condition and reading time in the reading condition for the participants in the study, we found no significant difference in exposure time between the two conditions (Mean narration = 46 s, SD = .62 s, mean reading = 49 s, SD = 11.56 s, $T = 231.00$, $p = .256$).

Eye-Tracking Set Up

Children's right eye movements were recorded during story reading by an Eyelink 1000+ eye-tracker with a refresh rate of 1000 Hz. The viewing was binocular, but only the right eye was recorded. The eye-tracker was interfaced with a computer that controlled stimulus display and data storage and a second computer screen where the passages were presented (screen resolution: 1920 by 1080, refresh rate: 59 Hz, length: 33.8 cm, height: 26.6 cm). Participants viewed the screen with their heads positioned on a deep chin rest and a forehead rest to minimize movements, positioned 60 cm from the screen. A 9-point calibration procedure was used, which was accepted when the average calibration error was less than 0.3° of visual angle; recalibration was performed when necessary.

For the listening condition, HP 530 headset headphones were used (frequency range: 20 Hz–20,000 Hz; sensitivity: 105 dB S.P.L at 1 KHz; rated power: 100 mW).

Procedure

Tasks were administered in two sessions on different days. During one session, participants completed the standardized tasks described in the participants section. In the other session, participants were exposed to the two stories, one in the reading condition and the other in the combined condition. The first story was presented, followed by a task to assess learning of the link between phonology and orthography of the presented items (in this phono-orthographic task, children heard each non-word from the story twice while the word was presented on screen, and they were asked to judge whether the given pronunciation was correct. The results of this task are reported in Valentini (2018), as they are not pertinent to the hypotheses of interest in this paper), and then two tasks to assess learning of the words' meanings. Tasks were presented in a fixed order to minimize the impact of earlier tasks on later tasks. The second story and tasks related to this were then presented. The sessions were completed in a quiet room within the child's school and lasted around 1 hour.

Story Presentation Procedure

Stories were presented on a computer screen while participants' eye movements were recorded. Children were told that they would see a story on the screen and were asked to either read the story at their own pace (reading condition) or to listen to the story via headphones and read along at the same time (bimodal condition). Children wore headphones in both conditions, making presentation modalities as similar as possible. Children were told that the stories contained some new words to reduce possible head movements due to surprise. Children were also told that there would be questions at the end of each passage and at the end of the story, and to try their best to understand the story so that they could answer them. Once calibration was successful, participants were instructed to look at a gray square at the top left of the screen before each trial to ensure standardization of the initial gaze location. When a stable fixation was detected, the square was replaced with the start of the paragraph.

In the bimodal condition, the presentation of each passage on screen ended automatically 500 ms after the end of the oral presentation to ensure children could not re-read the passage. In the reading condition, children were asked to read the passage once and to press a button to indicate that they had finished reading it; their eye movements were monitored closely to ensure that, once children reached the end of a paragraph, they did not start to read it a second time.

At the end of each passage, the computer displayed a comprehension question that required participants to answer either YES or NO by pressing buttons on a response device. These questions served to assess basic comprehension of the passages and maintain children's attention to the story. Performance in this task was used to confirm basic story comprehension in both conditions (see Results).

After the presentation of the first story, children's learning about the new words in the story was assessed.

Vocabulary learning tasks

In the *category recognition task*, the experimenter presented each target non-word in turn, in random order; words were simultaneously spoken and presented in written form on a card. A bimodal presentation at testing ensured comparability with the results of Valentini et al. (2018). Children then heard and saw a list of eight possible categories on paper and were asked to choose the one associated with the target non-word. The list of categories included the six categories associated with the target non-words in the story plus two additional categories (plant and vehicle). Each item was given a score of 0 or 1, and the number of correct categories was summed for each participant (maximum score of 6 for each condition). Next children completed a *definition production task* designed to elicit the production of all the information children remembered regarding each non-word. Children were asked whether they remembered the meaning of each word in turn and were invited to say "everything they remembered". If children were unable to produce a full definition, they were given up to two prompts. The first prompt was the correct category of the word; for example, for "dragon that eats sheep," children were told "X was an animal in the story. Do you remember something more about it? What animal was X in the story?" If the child still failed to produce the entire definition for the item, the first part of the definition was provided; for example, for "dragon that eats sheep," children were told that the item was a dragon and asked if they remembered anything further regarding the dragon. This task was scored on a 0–4 scale, with children scoring a 4 when able to produce a complete definition without prompt, 3 if able to produce part of the definition without prompt, 2 if able to produce the entire definition after the category prompt, 1 if they produced only part of the definition after the category prompt or produced the second part of the definition after the second prompt, or 0 if they failed to produce any part of the definition, even after prompts. We computed three different scores for this task rather than relying on an overall score in order to fully explore the complexity of the tasks. As well as a mean overall score for each child, which captures general performance, we also included the total number of full definitions produced per child per condition (i.e., the number of non-words for which a score of 4 was obtained) to explore the production/recall of items as a potentially purer measure of learning than recognition measures. We also measured the total number of non-words for which at least partial information was produced (i.e., where a score of at least 1 out of 4 was obtained) to explore whether items in the process of being learned, but not yet fully available, were impacted by the bimodal and unimodal conditions differently.

Eye-movement Measures

Fixations shorter than 80 ms were excluded (1.9% of the data), since such short fixations are unlikely to reflect meaningful processing (see Inhoff & Radach, 1998). No cut-off was applied to long fixations (fixations longer than 1200 ms formed 0.08% of fixations in the reading condition and 0.27% in the bimodal condition). Eye-movement data were analyzed in two ways. First, reading behavior was compared between the two conditions across all passages (except the first introductory passage, which we used as a practice trial). Second, specific comparisons were made between eye movements and areas of interest surrounding target words and definitions. The measures used in each set of analyses are shown in Table 1.

We identified six areas of interest in each passage, corresponding to the three repetitions of the target non-word, the area including the definition and clues (see Figure 1). As described in Table 1, we measured gaze duration (or first-pass reading time), re-reading time, and total reading time for each area of interest. Gaze duration (or first-pass reading time when considering multi-word clusters) refers to the sum of initial fixations within an interest area prior to the eyes moving outside the area, while re-reading time is the sum of all the subsequent fixations in the interest area. Total reading time is the total time spent on the area of interest (the sum of gaze duration, or first-pass reading time, and re-reading time).

Results

The first set of analyses explored differences in the pattern of eye movements in passages presented in bimodal versus unimodal conditions (Research Question 1). Subsequent analyses explored differences between conditions in looking times on the specific areas of interest (the three repetitions of the target non-words, definitions, and clues; Research Question 2). We used mixed-effects models to explore whether looking times on words, definitions, and clues were predicted by conditions. Next, we use paired-sample *t*-tests (or appropriate non-parametric tests) to explore the effect of condition on children's word learning (Research Question 3). Finally, we used mixed-effects models to identify reading strategies more likely to be associated with word learning in the two conditions. Using category recognition scores as the dependent variable, we looked at the predictive value of eye-movement measures, condition, and the interaction between these (Research Question 4). To explore Research Question 5, we analyzed eye movements during specific interest periods and computed the total time participants spent on the relevant non-words, definitions, and clues while hearing them (coincident time), along with the time participants spent on the non-words while their definitions or clues were spoken or vice versa (cross-coincident time). Mixed-effects models were used to explore whether either coincident or cross-coincident time predicted learning in the bimodal condition.

Analyses for research questions 1 and 3 were conducted on IBM SPSS Statistics (Version 25). Analyses for research questions 2, 4, and 5 were conducted in R version 4.1.1 (R Core Team, 2021), and scripts are available at: https://osf.io/mqsgr/?view_only=1d6d4bce382a473b9bf8f0fceeb11fb6.

Comparison between Conditions: Overall Reading Times (Research Question 1)

Table 2 reports the mean number of fixations, mean fixation duration (ms), and mean saccade amplitude (degrees of visual angle) per condition. Paired-sample *t*-tests were used to compare the two conditions, or, wherever normality assumptions were not met, Wilcoxon signed-rank tests were used. Children made significantly fewer fixations in the bimodal compared to the reading condition, but their fixations were significantly longer (see Table 2). The mean distance between two fixations (saccade amplitude) was also significantly longer in the bimodal condition than in the reading condition. Furthermore, children made significantly more downward and upward movements but fewer leftward and rightward movements on the text in the bimodal condition than in the reading condition. Overall, children's approach to the written text was qualitatively different when the text was read versus both spoken and read.

TABLE 1
Measures Used in Analyses of Eye-Movement Data and How They Were Computed

Analyses and passages used	Measure name	Main measure	Computation
Comparison between conditions: Overall reading times (passages 2–8; Research Question 1)	Mean number of fixations	Number of fixations of each passage	Averaged for each child by condition
	Total number of fixations—with directionality	Number of fixations moving in a particular direction with respect to the previous one (i.e., rightward, leftward, upward, or downward)	Count of total number of fixations in each direction for each child by condition
	Mean fixation duration	Duration of each fixation	Averaged for each child, by passage, by condition
	Mean saccade amplitudes	Spatial distance between two fixation points	Averaged for each child, by passage, by condition
Comparison between conditions: Fixation on areas of interest (passages 3–8)—Research Questions 2, 4, and 5	Gaze duration/first-pass reading time	For each area of interest (word & definition): sum of initial fixations prior to the eyes moving outside the area	Averaged for each child by condition
	Re-reading time	For each area of interest (word & definition): sum of all the subsequent fixations after the eyes have moved outside the area for the first time	Averaged for each child by condition
	Total reading time	For each area of interest (word & definition): sum of gaze duration (or first-pass reading time) and re-reading time	Averaged for each child by condition

Comparisons between Conditions: Fixations on Target Items and Definitions (Research Question 2)

Figure 2 presents eye-movement measures for the six areas of interest: the three presentations of the target non-word, the definition, and clues. Children spent more time looking at the target non-words in the reading condition than in the bimodal condition, as shown by gaze duration and total reading time measures. This difference was particularly noticeable in the first presentation of the non-word. Furthermore, time spent on the non-words seemed to diminish with exposure, with the first presentation of the target being fixated significantly longer than the second and third presentations across all three measures. There was no difference between conditions in terms of time spent reading definitions.Three linear mixed-effects models were carried out with gaze duration, re-reading time, and total reading time on non-words (all centered around the mean) as the dependent variables, respectively, and condition (reading vs. bimodal) and presentation of the target (first vs. second vs. third) as the independent factors, as well as the interaction between these. Items and Participants were entered as random factors. Analyses were conducted using

FIGURE 1
Example of a Story Passage introducing the new word "Cynthor" (Chaffin, 1997), highlighting the Areas of Interest within the Passage. Example comprehension question asked following the Passage: "Were people happy and safe when Fred reached the city?"

Fred continued his journey. He came out of the forest, and reached a city. No one was in sight. An enormous cynthor, a dragon that eats sheep, was roaming in the sky. Fred discovered that the people of the city were too frightened to come out of their houses.
"I'll try to free you from it," offered Fred.
"The cynthor is a gigantic creature, but I am not afraid."
Fred set off into the fields and engaged it in battle. People looked at the battle from a safe distance. They couldn't see much, just a lot of flames, but after a few days Fred returned. He was alive.
Nobody saw the meat-eating cynthor ever again.
Once the king heard of Fred's brave adventures, he allowed him to join his knights.

TABLE 2
Eye Movements to Passages in the Two Conditions

	Reading condition		Bimodal condition		Comparison between conditions	
	Mean (SD)	Range	Mean (SD)	Range	*T*	*p*
Number of fixations						
Total (per passage)	183 (39.05)	110–259	162 (12.15)	138–186	131.00	.004*
Rightward (per story)[a]	769 (142.34)	468–1099	668 (74.60)	531–816	3.90	<.001*
Leftward (per story)	359 (103.51)	217–590	323 (55.96)	225–435	177.00	.039*
Upward (per story)	13 (10.24)	4–52	18 (9.66)	6–46	90.00	<.001*
Downward (per story)	12 (9.09)	2–43	20 (8.65)	6–35	75.00	<.001*
Fixation duration[a]	240 (23.44)	196–298	256 (20.72)	222–300	−6.14	<.001*
Saccade amplitude[a]	3.30 (.48)	2.35–4.37	3.46 (.35)	2.88–4.27	−2.63	.013*

[a]*t*-test was computed since the differences between measures were normally distributed. Wilcoxon signed-rank tests were used in all other cases.
*Significant at $p < .05$.

generalized linear mixed models for binomial data (Jaeger, 2008), using the function "glmer" from the package "lme4" (Bates et al., 2014), computed with the software R (R Core Team, 2021). The analysis for gaze duration highlighted a main effect of condition (χ^2 (1) = 8.23, $p = .004$); specifically, gaze durations on target words were shorter in the bimodal than the reading condition. There was also a main effect of presentation of the target (χ^2 (2) = 31.44, $p < .001$), with a decrease in gaze duration from first to second ($p_{reading} < .001$; $p_{bimodal} = .036$) and first to third ($p_{reading} < .001$; $p_{bimodal} = .024$) presentations of the word but no difference between second and third presentations ($p_{reading} = .769$; $p_{bimodal} = .988$), in both conditions. The interaction between condition and word presentation was also significant (χ^2 (2) = 7.01, $p < .001$). There was a reduction in gaze duration from first to subsequent presentations and no difference between second and third presentations in both conditions; however, the difference between the two conditions was significant at the first presentation of the target word but not at subsequent presentations ($p_{target\ 1} < .001$; $p_{target\ 2} = .681$; $p_{target\ 3} = .897$). This suggests that the benefit of the bimodal condition was evident primarily in the first presentation of new words.

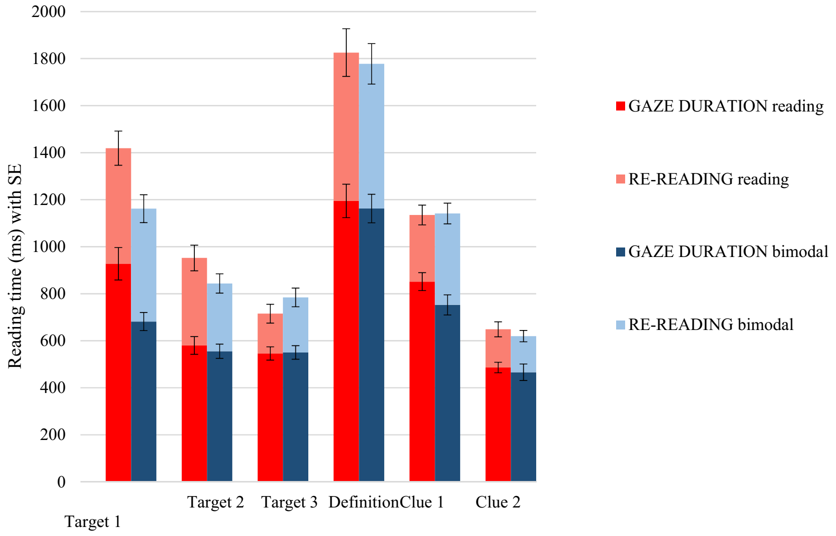
The analysis for re-reading time highlighted a main effect of presentation of the target (χ^2 (2) = 22.84, $p = .004$), with re-reading time decreasing from first to second to third presentations (target 1 *vs.* target 2: $p < .001$; target 1 vs. target 3: $p < .001$; target 2 *vs.* target 3: $p = .006$), but no effect of condition (χ^2 (1) = .05, $p = .816$), and no interaction between condition and presentation of the target (χ^2 (2) = 1.59, $p = .205$).

The analysis of total reading time also highlighted a significant main effect of presentation of the target (χ^2 (2) = 76.58, $p < .001$) and a main effect of condition (χ^2 (1) = 7.44, $p = .006$). The analysis also found an interaction between the presentation of the target and the condition (χ^2 (1) = 6.99, $p < .001$). In line with the results for gaze duration, the interaction reflected a significant difference between conditions in total reading time at the first presentation of the target ($p < .001$), that still reached significance at the second presentation of the target ($p = .049$) but not at the third presentation ($p = .213$). The interaction also highlighted a marked decrease in total reading time across presentations in the reading condition (all $ps < .002$), while in the bimodal condition, total reading time reduced from the first presentation to subsequent presentations (both $ps < .001$), but there was no decrease in total reading time between the second and third presentations ($p = .718$).

Three sets of analyses (for definitions, clue 1, and clue 2) were carried out using linear mixed-effects models with gaze duration, re-reading time, and total reading time (all centered around the mean) as the dependent variables, respectively, and condition (reading vs. bimodal) as the independent factors. Items and Participants were entered as random factors. There were no differences between conditions in the time spent looking at definitions on any measure ($Estimate_{first\text{-}pass} = -.03$, $p = .719$; $Estimate_{re\text{-}reading\ time} = -.01$, $p = .880$; $Estimate_{total\ reading\ time} = -.02$, $p = .815$). Gaze duration on the first in-text clue was longer in the reading condition than in the bimodal condition, ($Estimate = -.19$, $p = .038$). No other differences were highlighted in eye-movement measures for either clue 1 or clue 2 (clue 1: $Estimate_{re\text{-}reading\ time} = .13$, $p = .200$; $Estimate_{total\ reading\ time} = -.08$, $p = .405$; clue 2: $Estimate_{first\text{-}pass} = -.05$, $p = .630$; $Estimate_{re\text{-}reading\ time} = .06$, $p = .523$; $Estimate_{total\ reading\ time} = -.02$, $p = .833$).

FIGURE 2

Means of Gaze Duration (or First-Pass) and Re-Reading Times (in ms) with SE, on the three presentations of the Target Non-words, the Definitions, and the Clues, by condition. Total Reading Time is represented by the sum of Gaze Duration and Re-Reading measures in each case



Performance on Story Comprehension and Vocabulary Learning Tasks (Research Question 3)

After each paragraph of each story, children were asked to answer a yes/no comprehension question. Accuracy was significantly better than chance (4 out of 8) in both conditions (bimodal condition: Median = 6.00; $W = 508.50$, $p < .001$; reading condition: *Median* = 6.00; $W = 373.00$, $p < .001$), with no difference between conditions ($T = 146.50$, $p = .455$).

In the category recognition task, children recognized the correct category for an average of 1 word in the reading condition (M = 1.32, SD = 1.41) and 2 words in the bimodal condition (M = 2.09, SD = 1.31) out of a maximum of six. Chance performance was set at .75 (the probability of selecting the correct answer from eight alternatives on six trials). Performance was significantly better than chance in the bimodal condition ($W = 559.00$, $p < .001$), but only approached significance in the reading condition ($W = 409.00$, $p = .054$). Children performed significantly better at category recognition in the bimodal condition ($T = 78.50$, $p = .022$); the same result was obtained in a *by item* analysis ($T = 58.50$, $p = .022$). To check that order of story presentation did not affect performance, we compared performance in the first story presented ($M = 1.53$, $SD = 1.28$) compared to the second ($M = 1.91$, $SD = 1.50$) and found no significant order effect ($T = 181.50$, $p = .361$). When entered into a repeated measures ANOVA alongside condition, order of presentation did not interact with condition ($F(1,32) = 3.50$, $p = .071$) in determining performance in the category recognition task. The order of story presentation was therefore not considered in any further analyses.

Table 3 reports the results of the definition production task for the two conditions, in terms of the total number of full definitions produced, the total number of words for which children produced at least one correct feature when given prompts (*n*. partial definitions), and the mean overall score for each condition. The same analyses were conducted *by item*, with the same results (all *ps* > .200). This task was very difficult for the children, and means for all the measures were very low, with half or more of the words receiving a score of 0 for every child. Scores were higher for the number of partial definitions (children were able to produce some information for 2 or 3 words out of 6), showing that extensive help was needed for children to produce any definitions. Given the very low scores observed in this task and the lack of differences between conditions, this measure was not used further in analyses.

Relationship between Looking Times to Target Words and Definitions and Word Learning (Indexed by Category Recognition) (Research Question 4)

Mixed-effects models were used to explore the relationship between eye-movement patterns and vocabulary learning. Analyses were conducted using generalized linear mixed models for binomial data (Jaeger, 2008), using the function "glmer" from the package "lme4" (Bates et al., 2014), computed with the software R (R Core Team, 2021). Category recognition scores were used as the dependent variable; these were coded dichotomously (1 or 0) for each

TABLE 3
Scores for the Definition Production Task

	Reading condition		Bimodal condition		Difference between conditions	
	M (SD)	Range	M (SD)	Range	*T*	*p*
n. correct full definitions (max = 6)	.53 (.96)	0–3	.82 (1.17)	0–4	58.00	.123
n. partial definitions (max = 6)	2.12 (1.75)	0–6	2.59 (1.88)	0–6	132.00	.161
Mean overall score[a] (max = 4)	.78 (.84)	0–2.67	.98 (.91)	0–3.33	1.53	.136

[a]Paired-sample *t*-test is reported in place of Wilcoxon signed-rank test—the distribution of the differences is normal.

item. Eye-movement measures, condition (bimodal vs. reading), and the interaction between these were included as predictors of learning. The eye-movements measures included were gaze duration, re-reading time, and total reading time for each repetition of the target non-words (Table 4) or definitions (Table 5). Eye-tracking measures were centered around the mean for ease of comparison.

All models included random intercept terms for both participants and items. For the models that considered the words' interest areas (Table 4), we computed maximum models that included the hypothesized effects: condition (bimodal vs. reading), a single eye-movement measure, and word repetition (Targets 1, 2, and 3). A three-way interaction between word repetition, condition, and the eye-movement measure was also included, as were all the two-way interactions. The models were simplified by eliminating non-significant interactions. Final models were compared to an "empty" model that only included the random intercept terms (using pairwise Likelihood Ratio Test comparisons; Barr et al., 2013). For the models that considered the interest area of the definition, the full model included the hypothesized effects: condition (bimodal vs. reading), a single eye-movement measure, and the interaction between condition and the eye-movement measure. In the models of time spent looking at the target non-words, each child provided three data points, one for each repetition of the word (Target 1, 2, and 3), for each of the six target non-words in each condition. For the models that considered time spent on the definition, each child provided six data points per condition, one for each definition.

All three models that considered looking times at the target non-words (Table 4) confirmed the significant effect of condition: children learned more words in the bimodal than the reading condition. Gaze duration for the non-words predicted category learning, while re-reading time

TABLE 4
Generalized Linear Mixed Models of Accuracy in the Category Recognition Task as predicted by Gaze Duration, Re-Reading Time, and Total Reading Time on the three repetitions of the Target Non-words

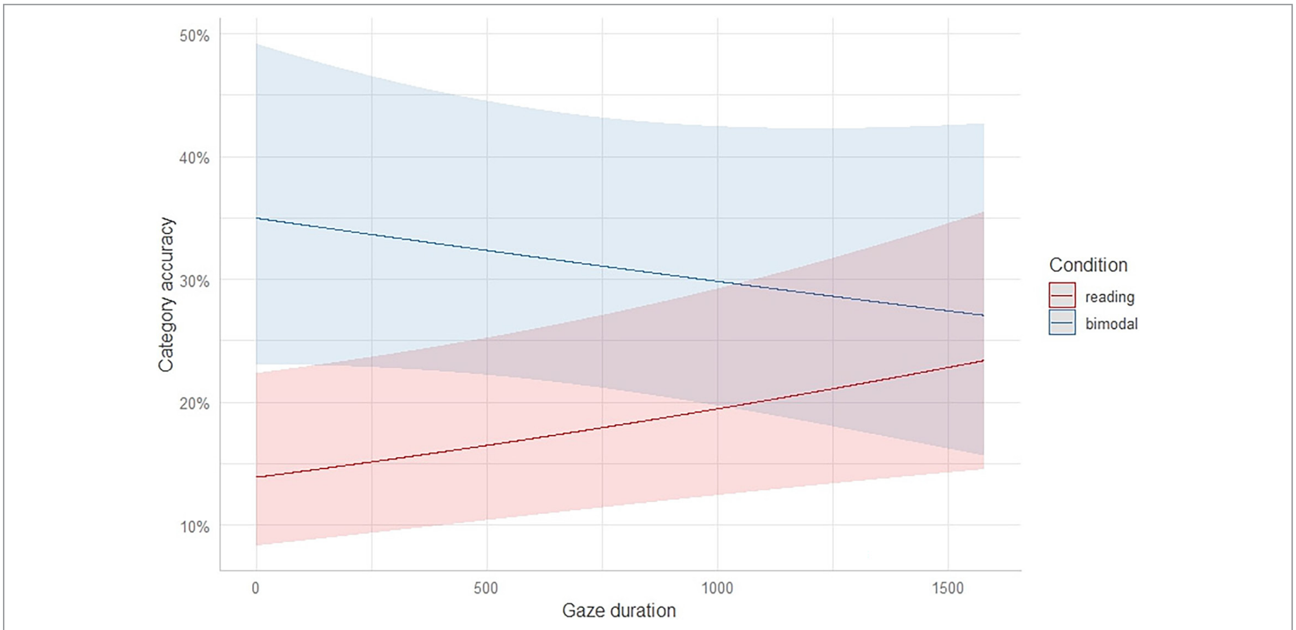
	Model 1: Gaze duration		Model 2: re-reading time		Model 3: Total reading time	
Fixed effects	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
(intercept)	−1.66	<.001*	−1.55	<.001*	−1.58	<.001*
Condition	.80	<.001*	.77	<.001*	.78	<.001*
Word repetition	.05	.885	.01	.903	.05	.555
Gaze duration	.23	.009*	–	–	–	–
Re-reading time	–	–	−.01	.981	–	–
Total reading time	–	–	–	–	.10	.181
Gaze duration* Condition	−.34	.026*	–	–	–	–
Random effects	**Var**	**SD**	**Var**	**SD**	**Var**	**SD**
Subject	.74	.86	.74	.86	.73	.85
Item	.46	.68	.45	.67	.45	.67
Final model vs. Empty model	$x^2(4) = 35.92$, $p < .001$*		$x^2(3) = 28.07$, $p < .001$*		$x^2(3) = 29.80$, $p < .001$*	

TABLE 5
Generalized Linear Mixed Models of Accuracy in the Category Recognition Task with predictors of Gaze Duration, Re-Reading Time, and Total Reading Time to Definitions

	Model 1: Gaze duration		Model 2: re-reading time		Model 3: Total reading time	
Factors	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
(intercept)	−1.42	<.001*	−1.42	<.001*	−1.42	<.001*
Condition	.71	.003*	.70	.003*	.71	.003*
First-pass reading time	.08	.515				
Re-reading time			.02	.814		
Total reading time					.06	.605
Random effects	**Var**	**SD**	**Var**	**SD**	**Var**	**SD**
Subject	.40	.63	.39	.62	.41	.64
Item	.25	.50	.25	.50	.25	.50
Fixed factor model vs. Empty model	$x^2(2)=9.32$, $p=.009$*		$x^2(2)=8.92$, $p=.012$*		$x^2(2)=9.47$, $p=.009$*	

and total reading time did not. However, the interaction between condition and gaze duration was also significant, suggesting that first-pass reading time predicted learning differently in the two conditions. Figure 3 suggests that word learning was better at longer gaze durations in the reading condition, while word learning was better at shorter gaze durations in the bimodal condition. At shorter gaze durations, learning was better in the bimodal condition, while at longer gaze durations, learning was similar in the two conditions. The likelihood of learning a word increased with longer gaze duration in the reading condition but decreased with gaze duration in the bimodal condition. To explore the interaction, we first computed separate models for each condition and then computed separate models by splitting data at the mean for gaze duration. Separate models for each condition highlighted no effect of gaze duration in either condition (reading: *Estimate* = .16, $p=.083$; bimodal: *Estimate* = −.05, $p=.743$). When the gaze duration data were split at the mean to create two datasets, one including all data at or below the

FIGURE 3
Predicted Probability of Category Learning by Gaze Duration and Condition. Note that the Figure presents Gaze Duration in its original metric (ms) for clarity



mean, and one including all data above the mean, the difference in accuracy on the category recognition task between conditions was significant in the analysis including the shorter gaze duration data (*Estimate* = 1.09, $p < .001$), but not in the analysis with the longer gaze duration data (*Estimate* = .26, $p = .264$). This suggests that the advantage for the bimodal condition was particularly evident when children looked more briefly at the new words.

All three models that considered looking at definitions (Table 5) highlighted a significant effect of condition: children performed better in the bimodal than the reading condition. However, no measure of looking predicted category learning. The same was true for looking times at both in-text clues (Tables in Appendix C).

Looking Times During Specific Interest Periods in the Bimodal Condition (Research Question 5)

In the bimodal condition, children heard the words and definitions spoken while reading them. It was hypothesized that looking times toward the specific areas of interest in the text at the same time that the relevant oral text was heard would predict word learning. Table 6 reports generalized linear mixed-effects models that explore whether accuracy in the category recognition task in the bimodal condition was predicted by the time spent reading the corresponding areas of text (non-words, definitions, or clues) while these were heard (coincident time) or by the time spent reading non-words while the relevant definitions or clues were spoken, or vice versa (cross-coincident time). Results showed that neither coincident nor cross-coincident time predicted word learning.

TABLE 6
Generalized Linear Mixed Models for Accuracy in the Category Recognition Task in the Bimodal Condition considering Total Coincident and Cross-Coincident Time spent on the three repetitions of the Target Non-words, Definitions, and Clues

	Model 1: Coincident time		Model 2: Cross-coincident time	
Factors	Estimate	*p*	Estimate	*p*
(intercept)	−1.00	.012*	−.90	.007*
Total coincident time	−.02	.785		
Total cross-coincident time			.13	.305
Random effects	Var	SD	Var	SD
Subject	1.78	1.33	1.14	1.07
Item	1.14	1.07	.75	.87
Fixed factor model vs. Empty model	$x^2(1) = .07$, $p = .788$		$x^2(1) = 1.02$, $p = .312$	

Discussion

In this study, children were exposed to new word forms (non-words) in stories presented in two conditions: when children were reading the story on their own (reading condition), and when children listened to the story while reading it (bimodal condition). Children were exposed to each story once, and each story contained six target non-words repeated three times. Children's eye movements while reading the stories were analyzed in terms of overall reading times plus gaze duration (or first-pass reading time), re-reading, and total reading times for specific areas of interest (the words, their definitions, and clues). The learning of words' categories and the ability to produce the words' definitions were assessed as indices of their semantic learning.

Comparison between Conditions: Overall Reading Times (Research Question 1)

As expected, participants explored the text differently in the presence and absence of oral narration. Children made fewer but longer fixations and made longer saccades in the bimodal condition than in the reading condition. The difference in the pattern of children's eye movements between the bimodal and reading conditions in our study resembles Gerbier et al. (2018) findings for synchronous and non-synchronous presentations of audio and written stories: in the synchronous condition, children showed longer fixations and longer saccades, in line with the bimodal condition of the current study. Our findings differ, however, from those reported for adults: like adult readers in Conklin et al. (2020) study, children made longer fixations in the bimodal condition. However, adults also made more fixations in this condition, while children in the present study made fewer fixations in the bimodal condition. Adults also tended to read ahead of the oral text rather than reading along. This difference might reflect the different utility of the oral presentation for adults and children; children might find the oral narration a helpful support for the task of reading, which allows them to move their eyes further in the text, while adults might find it redundant, and thus unhelpful, if it engages cognitive resources without benefit. Adults' shorter saccade amplitude in this condition might reflect an attempt to avoid moving too far ahead in the text compared to the narration. While we do not have data on whether children read ahead of the text in the bimodal condition, we have explored whether looking at words, definitions, and clues while these were spoken had an impact on learning in this condition, and this seems not to be the case; thus, whether or not children attended the words while these were spoken did not have an effect on their learning.

With respect to the direction of the eye movements on the text, the majority of children's fixations in both conditions were horizontal movements (mostly left to right), a pattern similar to typical adult reading of English text. However, children moved their eyes upwards and downwards significantly more often in the bimodal than in the reading condition. Similar patterns have been reported previously. In a study of shared story-book reading, for example, Roy-Charland et al. (2007) found that children in grades 3 and 4 made more than 70% horizontal or "reading-like" saccades and 20–30% "non-reading-like" ones.

Several explanations of these differences in eye-movement patterns in the two conditions suggest themselves. One possibility is that, in bimodal conditions, attentional resources are freed to explore the text. Hearing the oral narration may make the written text redundant, allowing participants to choose where to place their attention and to skim through the text, leading to more vertical eye movements. This idea is supported by the results of studies exploring looking during multimodal presentations of text and pictures: participants attended to the text more closely in the reading condition while using their freed resources to look more often at the pictures in the multimodal condition (Pellicer-Sánchez et al., 2020; Serrano & Pellicer-Sánchez, 2019). Children might also choose to attend more closely to the oral presentation than the written text if they find listening less taxing than reading, especially while reading skills are still developing. This hypothesis is supported by previous findings that younger children read along more closely in shared story reading situations if stories are at the child's reading level than if texts are more difficult (Roy-Charland et al., 2007). Both accounts would suggest that the presence of the oral narration frees children from the task of reading to some extent, either because the text is redundant to the information presented orally or because children find it easier to follow the oral narration than reading the text.

An alternative explanation for children's non-reading-like eye movements in the bimodal condition is that this condition poses a particular challenge for children. Children may find linking two streams of redundant information difficult due to their still developing executive functions (Altemeier et al., 2008), making it more challenging to follow the text closely in this condition and causing their eyes to wander away from the written passages. According to this account, the narration interferes with children's ability to follow the written text.

Yet another account, particularly to explain the longer saccades and fixations in the bimodal condition compared to the reading condition, is that the presence of the oral narration has a facilitative effect on the reading process by widening children's perceptual span. Specifically, as phonological information is provided orally in the bimodal condition, this frees up cognitive resources and potentially allows children to make greater use of parafoveal preview and predictability of upcoming words in the bimodal condition. A similar facilitation effect is seen in the widening of the visual span with experience when the eye movements of experts are compared to those of novices when reading musical scores or interpreting specialized images like x-rays (Gegenfurtner et al., 2011; Truitt et al., 1997). In these studies, experts show longer saccades but shorter fixation times on relevant information, which is thought to result from the reduced cognitive load associated with their specialist expertise. A widening of attention distribution in the current task might therefore indicate that children experience a reduced cognitive load in bimodal conditions.

In summary, children move their eyes differently during bimodal presentation, making fewer but longer fixations and longer saccades, and more vertical movements, which might be interpreted as an indication of either greater challenge or greater facilitation associated with listening while reading. Given that vertical eye movements provide opportunities to seek out information in support of text comprehension and word learning, and in light of children's superior word learning in the bimodal condition, a facilitation account seems more likely. We return to consideration of this issue later in our Discussion when we consider children's differential learning of the new words across conditions and how this relates to attention to the different areas of interest in the text.

Comparisons between Conditions: Target Items, Definitions, and Clues (Research Question 2)

Our second set of hypotheses concerned the difference between conditions in how children attend to the new vocabulary items. Gaze duration (i.e., time spent on a word before moving to another one) and total reading times (total time spent on a word, including re-reading) on target non-words were longer in the reading than in the bimodal condition. This difference was particularly significant the first time words were presented, suggesting that participants experienced a processing facilitation in the bimodal condition at their very first encounter with the new items.

We also explored whether reading times decreased with repeated presentation of the item, as would be expected if the non-words became more familiar and therefore easier to process. Similar to previous studies of adults reading new words in context (Joseph et al., 2014), there was a reduction in reading time from the first to subsequent presentations. Gaze duration decreased only from first to second presentation in both conditions, and more markedly so in the reading condition. Re-reading time decreased from first to second to third presentations in both conditions, while total reading times decreased steadily in the reading condition, but only from first to second presentation in the bimodal condition. This pattern suggests that the target non-words were processed more easily at each encounter in the reading condition, in line with previous findings with adults (Joseph et al., 2014). The lack of difference between second and third presentations in the bimodal condition in gaze duration and total reading times, on the other hand, might be interpreted as evidence of faster integration of the word into the lexicon in this condition.

These results, taken together, suggest that bimodal presentation facilitates both online encoding and increases the quality of the new item's lexical representation for easier later retrieval. Specifically, the shorter gaze durations seen at the very first encounter with new items in the bimodal condition suggest that facilitation happens online, pointing toward a reduction in cognitive load due to the simultaneous oral presentation. This finding is consistent with the predictions of Cognitive Load Theory (CLT; Mayer, 2014; Mayer et al., 1999; Paas et al., 2003). At the same time, the reduction in total reading time at each presentation of the new items in the reading condition versus the plateau after the second presentation in the bimodal condition suggests that the latter condition supports faster integration into the lexicon and facilitation at the retrieval stage, in line with the Lexical Quality Hypothesis (LQH; Perfetti & Hart, 2002).

No difference between conditions was found in reading times for definitions, suggesting similar processing of the information these provided or, perhaps, that eye-movement measures are less sensitive indices when applied to groups of words rather than individual items. Similarly, there were no differences in reading times on clues, except for a longer gaze duration on the first clue in the reading condition. This difference might be interpreted as a faster integration of relevant information into the lexicon for the bimodal condition. It is interesting to note that both the definition and the first clue followed the relevant non-word, and it is possible that the similar syntactic structure might have prompted a deeper analysis of the first clue and therefore prompted a condition effect. A similar effect might have been masked by a novelty effect for definitions, as definitions were the first semantic information provided and thus more likely to be deeply analyzed in both conditions. The positioning of the second clue before the relevant non-word might have masked any such effect. However, it must be noted that clues were not as controlled in terms of length or frequency as definition, so any insight arising from an analysis of clues should be taken with caution.

Word Learning and how this is related to Looking Times to Target Words and Definitions (Research Questions 3, 4, and 5)

In line with previous studies (Valentini et al., 2018), presenting stories both orally and in writing facilitated vocabulary acquisition in terms of learning new words' categories. Children performed similarly in their comprehension of the passages across conditions, in line with previous findings (Pellicer-Sánchez et al., 2020; Serrano & Pellicer-Sánchez, 2019), but learned new words better in the bimodal condition, at least in terms of category learning. Children failed to show a difference between conditions in the definition of the production task, but this lack of effect might have been driven by their low overall performance on the task.

Performance on the category recognition task confirms that presenting words both orally and in writing has a facilitative effect on word learning. The nature of this facilitation effect was the focus of the final set of analyses. It was hypothesized that, if the bimodal presentation of oral and written text facilitates vocabulary acquisition by freeing attentional resources online, as proposed by the CLT (Mayer, 2014), children would need to spend less time on words to learn them in this condition. To address this hypothesis, our analyses explored whether reading times predicted word learning and whether this effect differed between conditions. Only gaze duration on target items predicted category learning, but this effect interacted with the effect of condition. Specifically, children were facilitated by the presence of the oral narration (i.e., learned more words in the bimodal than the reading condition) at shorter gaze durations but not at longer gaze durations. Also, when reading only, children were somewhat more likely to learn the new words if they looked longer at them before looking away (gaze duration), which was not the case in the bimodal condition. However, this effect did not reach significance in the reading condition, tempering this interpretation.

The models therefore suggest that category learning is a function of the interaction between presentation modality and gaze duration on the target word. We interpret this interaction as follows: If longer gaze duration is assumed to reflect processing effort, then it makes sense that learning was greater when children's gaze durations were longer in the reading condition—and worth noting that, at longer gaze durations, learning became comparable to that seen in the bimodal condition. When children were not assisted by the presence of the oral narration, spending more time on the target words had a positive effect on learning. At very short gaze durations, in contrast, children were significantly advantaged in the bimodal condition relative to the unimodal condition, suggesting that less effort is required to acquire semantic information when it is presented in more than one modality simultaneously. However, longer gaze duration was not associated with further learning in the bimodal condition. On the basis that children show longer gaze durations in the reading condition, where learning was poorer, we interpret cases of longer gaze duration in the bimodal condition as reflecting processing difficulties and likely to be associated with lower levels of learning as a result, which is in line with the pattern of results found in this study.

The idea that bimodal presentation supports vocabulary acquisition by freeing attentional resources from the task of reading (CLT) appears to be supported by the data: children learned both more words and spent less initial time on the new items in the bimodal condition. The bimodal condition was also more conducive to learning at shorter gaze durations; this, accompanied by overall shorter gaze durations in this condition, suggests that less effort is required to learn words in this condition. This modality effect might be partially compensated in the reading condition by looking longer at the items. This provided only partial compensation, however, as, overall, children learned more words in the bimodal condition.

In summary, participants spent less time looking at new items in the bimodal condition but showed greater learning in this condition compared to when reading alone. In our study, unlike previous similar studies (Pellicer-Sánchez et al., 2020; Serrano & Pellicer-Sánchez, 2019), we directly measured total exposure time in both conditions and ensured matched exposure time between conditions. The comparable total exposure time, paired with shorter looking time at the new items in the bimodal condition, suggests that children had spare resources and time to allocate to other parts of the text in this condition. We hypothesized that they would use their freed resources (and time) to explore the definitions of the non-words or the in-text clues, but we found no difference between conditions in time spent on definitions and only an effect in the opposite direction for gaze duration on the first clue (i.e., longer gaze duration in the reading condition). The models also failed to find any relationship between time spent on definitions and clues and vocabulary learning. It therefore remains unclear how children used their freed resources in the bimodal condition and how this supported vocabulary acquisition. It is possible that participants used their freed cognitive resources to connect the new words to the text more generally, supporting their understanding of the story as a whole, but we found no evidence of better story comprehension in the bimodal condition to support this claim. The positive effect of looking at definitions for word learning might also be too subtle to be detected with the present design (i.e., comparison across conditions of different passages). In fact, previous research comparing more controlled sentences found that readers spend more time reading the context for new words than known words (Brusnighan & Folk, 2012; Chaffin et al., 2001). Future research might compare the time course of word learning by comparing the time spent on the same definition for the same item over multiple presentations.

To analyze how children explored the text in the bimodal condition, we explored whether the time spent looking at the word, definitions, or clues while hearing them (coincident time) or the time spent on words while hearing definitions or clues or vice versa (cross-coincident time) affected learning. We hypothesized that the bimodal condition might improve learning by allowing children to connect a word with its semantic features in ways impossible to achieve during a single-modality presentation. Our results, however, did not highlight any significant cross-modality effect on learning the new words' meanings, so we can conclude that children did not seem to use this specific strategy to maximize learning in the bimodal condition. The results of coincident time lend support to the CLT account over the LQH account: hearing and looking at a word at the same time should produce stronger lexical representation in memory, according to the LQH; however, this did not affect learning. Children did not need to attend to the words while they were spoken to learn them better in the bimodal condition, suggesting that this condition freed resources from the task of reading itself rather than improving performance by strengthening word representations.

Our results show that, in the reading condition, children learn words' meanings better if they look at the words for longer. Contrary to studies of adults' learning of new words in their second language, which found effects of both gaze duration and total reading time (Godfroid et al., 2013, 2018; Mohamed, 2018), only gaze duration predicted learning in our study. This suggests that it was initial effort when attending to new items that assisted children's semantic encoding, especially in the reading condition. This result is striking when one considers that it is the time spent on a word at first-pass, even before moving to parts of the text that provide information about the word's meaning, that determines whether words are learned. This is in line with the results of Lowell and Morris (2014), who found first-pass time to be greater for new words than known words. The authors suggested that the longer time spent on new words might allow the encoding of the new orthographic form in memory, supporting the next step in word learning, the linking of the new form with its meaning. Initial attention to new word forms appears to be primary, suggesting that a crucial aspect of word learning is noticing that a word form is unfamiliar in the first place. The aforementioned differences in findings between children and adults might then be due to differences in how these two groups explore new texts. Compared to adults, children are more likely to encounter words they have never seen or heard before while reading, and this might prompt them to pay more attention to the form of a new word at first-pass rather than its meaning. This reasoning is in line with the delayed effects of implausibility (a word-meaning effect) found in the eye movements of children compared to adults (Joseph et al., 2008). In the current study, new words were fixated multiple times and for long periods. Children may have been attempting to encode the new forms in memory, and this may have taken precedence over determining the word's meaning. Heightened initial processing time on learned words might therefore reflect the child's efforts to encode the word's form. Adults, in contrast, might adopt a different strategy, building a more general representation of the text by exploring words' meanings and returning to previously read words if they find them to be important for understanding the text, leading to stronger effects of total reading time and re-reading time on adult word learning. If children focused to a greater extent on the process of decoding, it would make sense that they would spend more initial time on the words and that their learning would be driven more strongly by gaze duration. On the other hand, if adults focus more on text comprehension, they might spend more re-reading and total time on words they consider important, and these times might predict their learning more strongly.

In conclusion, bimodal presentation seems to support vocabulary acquisition online by freeing attentional resources, as predicted by the CLT: children spent less initial time on new items in the bimodal condition, especially at first presentation of the item, but still learned these better. The online facilitation provided by bimodal presentation also seems to lead to higher quality representations: total reading time plateaued after the second presentation in the bimodal, but not in the reading condition, suggesting faster integration in the lexicon in the bimodal condition. This higher-quality representation also supports subsequent retrieval; the offline results, in fact, support the LQH. The process to create this better representation seems supported by a facilitation in online processing; thus, both the CLT and the LQH have a role in explaining our results.

Limitations

This study contributes to the existing literature by elucidating the process of bimodal facilitation for vocabulary acquisition. However, we acknowledge some limitations. First, with regards to methodology, we used passages that together made full stories to engage children's attention and reproduce a realistic situation in which words were presented in context. Although we controlled for a large number of potentially confounding differences between the stimuli used in each condition, it was not possible to control for all the variables that might impact eye movements in the areas of interest. For example, target non-words were preceded by words that were different in length and frequency, the positions of words within the text differed between stories, and sentence structures were not controlled. We counterbalanced item lists and stories between conditions to avoid a confounding effect of story variability on the difference between conditions, but we acknowledge that this variability might have had a more general effect on the time children spent on each item, independent of condition. Second, in relation to eye-movement methodology in general, we must also acknowledge that eye movements in the text were recorded rather than manipulated. Thus, while our data suggest a relationship between time spent on new words and the subsequent ability to recognize their category, this relationship might not be causal. Future studies might use different methodologies, such as word-by-word presentations, to determine the causal link between looking time and learning, although these would, of course, be less generalizable to real-life reading conditions. Third, we must acknowledge a potential effect of presentation modality at testing, where children were presented with words both orally and written, regardless of condition. This testing procedure could have provided an advantage for items presented in bimodal condition, as it aligns more closely in a bimodal presentation than a reading-only presentation. Nevertheless, we consider this presentation modality at testing the best choice, ensuring comparability with previous studies (Valentini et al., 2018). Previous results also suggest that children create phonological representations of new words while reading (Valentini et al., 2018), and analyses carried out on the present sample on the phono-orthographic task showed equal performance in the two conditions, suggesting that children learned the phonological form of the new words even in the reading condition (Valentini, 2018). This would suggest that a bimodal testing modality would not present a big disadvantage for items presented in the reading-only condition. A further consideration is that, while performance in the category recognition task was higher than chance in the bimodal condition and not at the floor in either condition, children's learning was nonetheless quite low. Presenting material multiple times, as in Valentini et al. (2018), would allow us to explore effects in vocabulary production tasks, as well as avoid possible floor effects in all tasks due to lack of learning, and it would allow exploration of differences between conditions over a longer exposure period. Similarly, while the sample size is in line with previous studies (Gerbier et al., 2015, 2018), increasing the number of participants would improve the reliability and generalizability of the results.

Conclusion

In conclusion, the results of this study support the hypothesis that the process of encoding new words is less effortful in bimodal presentation conditions. Children spent less time fixating words overall in this condition and did not need to spend as much time on words that they learned in this condition as they did in the reading condition. The final product of word learning was better in the bimodal condition, supporting the idea that, in bimodal conditions, children create lexical representations of better quality (Perfetti & Hart, 2002). This study is the first to show that this facilitation happens online, from the very first encounter with a new word, suggesting that bimodal presentation frees attentional resources online during text processing, even before a representation of the new word has been formed (Mayer et al., 1999). Further research is needed to understand how these freed resources are used to facilitate semantic learning, as this was not related to the processing of words' definitions or clues in the current study, and to develop educational approaches that embed multimodal learning techniques in the classroom to better support children's learning of vocabulary.

Funding Information

The study was conducted as part of Alessandra Valentini's PhD, funded by a Research Studentship in Social Sciences from the University of Reading. Jessie Ricketts was supported by the Economic and Social Research Council (Grant ES/K008064/1) while the research was conducted.

Conflict Of Interest Statement

The authors have no conflict of interest to declare.

Data Availability Statement

Participants' data cannot be widely shared in Open Access format due to specific features of the consent originally requested. Data is available upon request from the leading author. The script used for data analysis in R can be found at: https://osf.io/mqsgr/?view_only=1d6d4bce382a473b9bf8f0fceeb11fb6.

Ethics Approval Statement

The study was given a favorable ethical opinion for conduct by the School Research Ethics Committee of the University of Reading. Permission to reproduce material from other sources: all materials used were written specifically for the present research. All references to non-words extracted from other sources are acknowledged in the paper. Non-words derived from published tests are not included.

REFERENCES

Altemeier, L. E., Abbott, R. D., & Berninger, V. W. (2008). Executive functions for reading and writing in typical literacy development and dyslexia. *Journal of Clinical and Experimental Neuropsychology*, *30*(5), 588–606. https://doi.org/10.1080/13803390701562818

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7, http://CRAN.R-project.org/package=lme4

Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology*, *24*(3–4), 323–335. https://doi.org/10.1080/02702710390227297

Blythe, H. I., Liang, F., Zang, C., Wang, J., Yan, G., Bai, X., & Liversedge, S. P. (2012). Inserting spaces into Chinese text helps readers to learn new words: An eye movement study. *Journal of Memory and Language*, *67*(2), 241–254. https://doi.org/10.1016/j.jml.2012.05.004

Brusnighan, S. M., & Folk, J. R. (2012). Combining contextual and morphemic cues is beneficial during incidental vocabulary acquisition: Semantic transparency in novel compound word processing. *Reading Research Quarterly*, *47*(2), 172–190. https://doi.org/10.1002/rrq.015

Brusnighan, S. M., Morris, R. K., Folk, J. R., & Lowell, R. (2014). The role of phonology in incidental vocabulary acquisition during silent reading. *Journal of Cognitive Psychology*, *26*(8), 871–892. https://doi.org/10.1080/20445911.2014.965713

Chaffin, R. (1997). Associations to unfamiliar words: Learning the meanings of new words. *Memory & Cognition*, *25*(2), 203–226. https://doi.org/10.3758/bf03201113

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 225–235. https://doi.org/10.1037/0278-7393.27.1.225

Colenbrander, D., Miles, K. P., & Ricketts, J. (2019). To see or not to see: How does seeing spellings support vocabulary learning? *Language, speech, and hearing services in schools*, *50*(4), 609–628. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0135

Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaitė-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*, *36*(3), 257–276. https://doi.org/10.1177/0267658320921496

Duckett, P. (2003). Envisioning story: The eye movements of beginning readers. *Literacy Teaching and Learning*, *7*, 77–89.

Dunn, L. M., Dunn, D. M., & NFER. (2009). *British picture vocabulary scale* (3rd ed.). GL Assessment Ltd.

d'Ydewalle, G., Praet, C., Verfaillie, K., & Rensbergen, J. V. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research*, *18*(5), 650–666. https://doi.org/10.1177/009365091018005005

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*(2), 174–187. https://doi.org/10.2307/747863

Evans, M. A., & Saint-Aubin, J. (2005). What children are looking at during shared storybook reading: Evidence from eye movement monitoring. *Psychological Science*, *16*, 913–920. https://doi.org/10.1111/j.1467-9280.2005.01636.x

Evans, M. A., & Saint-Aubin, J. (2013). Vocabulary acquisition without adult explanations in repeated shared book reading: An eye movement study. *Journal of Educational Psychology*, *105*(3), 596–608. https://doi.org/10.1037/a0032465

Forum for Research in Literacy and Language. (2012). *Diagnostic test of word Reading processes (DTWRP)*. GL Assessment.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Gerbier, E., Bailly, G., & Bosse, M. L. (2015, September). Using karaoke to enhance reading while listening: Impact on word memorization and eye movements. *Speech and Language Technology for Education (SLaTE)*, 59–64.

Gerbier, E., Bailly, G., & Bosse, M. L. (2018). Audio–visual synchronization in reading while listening to texts: Effects on visual behavior and verbal learning. *Computer Speech & Language*, *47*, 74–92. https://doi.org/10.1016/j.csl.2017.07.003

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental l2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3), 483–517. https://doi.org/10.1017/s0272263113000119

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in Reading and scene perception* (pp. 29–53). Elsevier Science Ltd.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Joseph, H., & Nation, K. (2018). Examining incidental word learning during reading in children: The role of context. *Journal of Experimental Child Psychology*, *166*, 190–211. https://doi.org/10.1016/j.jecp.2017.08.010

Joseph, H. S., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, *61*(5), 708–723.

Joseph, H. S., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*(1), 238–248. https://doi.org/10.1016/j.cognition.2014.06.015

Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, *49*(1), 105–120. https://doi.org/10.1002/rrq.59

Lowell, R., & Morris, R. K. (2014). Word length effects on novel words: Evidence from eye movements. *Attention, Perception, & Psychophysics*, *76*, 179–189. https://doi.org/10.3758/s13414-013-0556-4

Lowell, R., & Morris, R. K. (2017). Impact of contextual constraint on vocabulary acquisition in reading. *Journal of Cognitive Psychology*, *29*(5), 551–569. https://doi.org/10.1080/20445911.2017.1299155

Masterson, J., Stuart, M., Dixon, M., Lovejoy, D., & Lovejoy, S. (2003). The children's printed word database. Retrieved from: https://www1.essex.ac.uk/psychology/cpwd/

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). Cambridge University Press.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology*, *91*(4), 638–643. https://doi.org/10.1037/0022-0663.91.4.638

Medler, D. A., & Binder, J. R. (2005). *MCWord: An on-line orthographic database of the English language*. Retrieved from www.neuro.mcw.edu/mcword/

Miles, K. P., Ehri, L. C., & Lauterbach, M. D. (2016). Mnemonic value of orthography for vocabulary learning in monolinguals and language minority English-speaking college students. *Journal of College Reading and Learning*, *46*(2), 99–112. https://doi.org/10.1080/10790195.2015.1125818

Mohamed, A. A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*(2), 269–293. https://doi.org/10.1017/S0272263117000092

Montali, J., & Lewandowski, L. (1996). Bimodal reading: Benefits of a talking computer for average and less skilled readers. *Journal of Learning Disabilities*, *29*(3), 271–279. https://doi.org/10.1177/002221949602900305

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, *99*(2), 308–328. https://doi.org/10.1111/modl.12215

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*(2), 237–270. https://doi.org/10.2307/1162893

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/s15326985ep3801_1

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, *38*(1), 97–130.

Pellicer-Sánchez, A., & Siyanova-Chanturia, A. (2018). Eye movements in vocabulary research. *ITL-International Journal of Applied Linguistics*, *169*(1), 5–29.

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, A. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, *42*(3), 577–598.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). John Benjamins.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing http://www.R-project.org/

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). *Psychology of Reading*. Psychology Press.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, *7*(3), 219–226. https://doi.org/10.1037/1076-898X.7.3.219

Ricketts, J., Bishop, D. V., & Nation, K. (2009). Orthographic facilitation in oral vocabulary acquisition. *The Quarterly Journal of Experimental Psychology*, *62*(10), 1948–1966. https://doi.org/10.1080/17470210802696104

Ricketts, J., Bishop, D. V., Pimperton, H., & Nation, K. (2011). The role of self-teaching in learning orthographic and semantic aspects of new words. *Scientific Studies of Reading*, *15*(1), 47–70. https://doi.org/10.1080/10888438.2011.536129

Rosenthal, J., & Ehri, L. C. (2011). Pronouncing new words aloud during the silent reading of text enhances fifth graders' memory for vocabulary words and their spellings. *Reading and Writing*, *24*(8), 921–950. https://doi.org/10.1007/s11145-010-9239-x

Ross, N. M., & Kowler, E. (2013). Eye movements while viewing narrated, captioned, and silent videos. *Journal of Vision*, *13*(4), 1–19. https://doi.org/10.1167/13.4.1

Roy-Charland, A., Saint-Aubin, J., & Evans, M. A. (2007). Eye movements in shared book reading with children from kindergarten to grade 4. *Reading and Writing*, *20*(9), 909–931. https://doi.org/10.1007/s11145-007-9059-9

Rust, J. (2008). *Coloured progressive matrices and Chrichton vocabulary scale manual*. Pearson.

Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction*, *49*, 157–165. https://doi.org/10.1016/j.learninstruc.2017.01.005

Serrano, R., & Pellicer-Sánchez, A. (2019). Young L2 learners' online processing of information in a graded reader during reading-only and reading-while-listening conditions: A study of eye-movements. *Applied Linguistics Review*, 1(ahead-of-print), *13*, 70. https://doi.org/10.1515/applirev-2018-0102

Suggate, S., Schaughency, E., McAnally, H., & Reese, E. (2018). From infancy to adolescence: The longitudinal links between vocabulary, early literacy skills, oral narrative, and reading comprehension. *Cognitive Development*, *47*, 82–95. https://doi.org/10.1016/j.cogdev.2018.04.005

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word Reading efficiency*. Pro-ed.
Truitt, F. E., Clifton, C., Pollatsek, A., & Rayner, K. (1997). The perceptual span and the eye-hand span in sight reading music. *Visual Cognition*, *4*(2), 143–161. https://doi.org/10.1080/713756756
Valentini, A. (2018). *How do reading and listening to stories facilitate vocabulary acquisition?* [Doctoral thesis, University of Reading].
Valentini, A., Ricketts, J., Pye, R. E., & Houston-Price, C. (2018). Listening while reading promotes word learning from stories. *Journal of Experimental Child Psychology*, *167*, 10–31. https://doi.org/10.1016/j.jecp.2017.09.022
Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 481–487. https://doi.org/10.3758/s13428-017-0872-z
Wechsler, D. (2005). *Wechsler individual achievement. (WIAT II)*. The Psychological Corp.
Wilkinson, K. S., & Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess...: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics*, *34*(3), 591–613. https://doi.org/10.1017/s0142716411000889
Williams, R., & Morris, R. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*(1–2), 312–339. https://doi.org/10.1080/09541440340000196

Submitted January 11, 2023
Final revision received August 24, 2023
Accepted September 25, 2023

ALESSANDRA VALENTINI, Lecturer, School of Human Sciences, University of Greenwich, Greenwich, London, UK; email: a.valentini@greenwich.ac.uk

RACHEL E. PYE, Associate Professor, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK; email: rachel.pye@reading.ac.uk

CARMEL HOUSTON-PRICE, Professor and Head of School, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK; email: carmel.houston-price@reading.ac.uk

JESSIE RICKETTS, Professor, Department of Psychology, Royal Holloway, Univeristy of London, London, UK; email: jessie.ricketts@rhul.ac.uk

JULIE A. KIRKBY, Principal Academic in Psychology, Department of Psychology, Bournemouth University, Poole, UK; email: jkirkby@bournemouth.ac.uk

APPENDIX A

Standardized Scores on the Background Measures Based on Published Norms (M = 100, SD = 15)

	Mean (SD)	Range
TOWRE sight word efficiency	104.94 (10.98)	85–134
TOWRE phonemic decoding efficiency	108.47 (13.70)	77–135
BPVS—3	95.06 (14.09)	72–126
CPM	99.26 (16.43)	70–135

Colored Progressive Matrices: BPVS—3, British Picture Vocabulary Scale; CPM, Colored Progressive Matrices; TOWRE, Test of Word Reading Efficiency.

APPENDIX B

Definitions Used in the Stories and their Features

Category	Story	Definition	Plausibility	Predictability proportion	Length
Animal	Knight	Dragon that eats sheep	2.73	0.08	22
Animal	Pirate	Elephant that pulls carriages	2.87	0.13	29
Building	Knight	Tower with no windows	1.67	0.25	21
Building	Pirate	Grave for many people	1.67	0.08	21
Clothing	Knight	Shirt made of chains	2.87	0.00	20
Clothing	Pirate	Dress worn by men	2.27	0.00	17
Food	Knight	Soup eaten by farmers	1.40	0.00	21
Food	Pirate	Potato wrapped in ham	1.53	0.04	21

APPENDIX B (Continued)

Category	Story	Definition	Plausibility	Predictability proportion	Length
Job	Knight	Someone who colors leather	2.47	0.00	27
Job	Pirate	Someone who sells furs	1.13	0.00	22
Object	Knight	Spear made of gold	1.67	0.00	18
Object	Pirate	Sofa used during meals	2.20	0.00	22

Plausibility = Mean plausibility computed from the adult sample (each judging plausibility on a scale from 1—very implausible to 5—very plausible); Predictability proportion = proportion of adults correcting predicting the final word of the definition from the previous ones; Length = length of the definition in characters.

APPENDIX C

Models for Looking Time at Clue 1 and Clue 2

Generalized linear mixed models of accuracy in the category recognition task with predictors of gaze duration, re-reading time, and total reading time to clue 1.

	Model 1: Gaze duration		Model 2: re-reading time		Model 3: Total reading time	
Factors	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
(intercept)	−1.40	<.001*	−1.41	<.001*	−1.42	<.001*
Condition	.69	.004*	.70	.004*	.72	.003*
Gaze duration	−.09	.502	—	—	—	—
Re-reading time	—	—	.13	.257	—	—
Total reading time	—	—	—	—	.08	.513
Random effects	**Var**	**SD**	**Var**	**SD**	**Var**	**SD**
Subject	.37	.60	.38	.62	.40	.63
Item	.25	.50	.26	.51	.26	.51
Fixed factor model vs. Empty model	$x^2(2) = 9.43$, $p = .009$		$x^2(2) = 10.23$, $p = .006$		$x^2(2) = 9.25$, $p = .010$	

Generalized linear mixed models of accuracy in the category recognition task with predictors of gaze duration, re-reading time, and total reading time to clue 2

	Model 1: Gaze duration		Model 2: re-reading time		Model 3: Total reading time	
Factors	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
(intercept)	−1.38	<.001*	−1.38	<.001*	−1.38	<.001*
Condition	.71	.004*	.71	.004*	.71	.004*
Gaze duration	.03	.807	—	—	—	—
Re-reading time	—	—	−.04	.723	—	—
Total reading time	—	—	—	—	−.01	.964
Random effects	**Var**	**SD**	**Var**	**SD**	**Var**	**SD**
Subject	.36	.60	.36	.60	.36	.60
Item	.19	.43	.19	.44	.19	.43
Fixed factor model vs. Empty model	$x^2(2) = 8.56$, $p = .013$*		$x^2(2) = 8.63$, $p = .013$*		$x^2(2) = 8.51$, $p = .014$*	

* Significant at $p < .05$.