Reading Research Quarterly

INTERNATIONAL LITERACY ASSOCIATION

ARTICLE OPEN ACCESS

Using Eye Movements From a "Read-Only" Task to Predict Text Comprehension

Diane C. Mézière[1] | Lili Yu[2] | Titus von der Malsburg[3] | Erik D. Reichle[2] | Genevieve McArthur[4]

[1]Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland | [2]School of Psychological Sciences, Macquarie University, Sydney, Australia | [3]Institute of Linguistics, University of Stuttgart, Stuttgart, Germany | [4]Australian Centre for the Advancement of Literacy, Australian Catholic University, Sydney, Australia

Correspondence: Diane C. Mézière (diane.meziere@utu.fi)

Received: 31 March 2025 | **Revised:** 1 April 2025 | **Accepted:** 9 May 2025

Funding: The study was funded by an International Macquarie University Research Excellence Scholarship (iMQRES) and two Australian Research Council grants (DP190100719 and DP200100311).

ABSTRACT

Recent research on the use of eye movements to predict performance on reading comprehension tasks suggests that while eye movements may be used to measure comprehension, the relationship between eye-movement behavior and comprehension is influenced by differences in task demands between comprehension measures. In this study, we examined the usefulness of eye movements collected during reading with no additional task demands (a "read-only" condition) to predict comprehension ability as measured by a recall task. We collected eye-movement behavior from adult native speakers of English ($N=62$, 46 females, mean age 26 years) while they read nine passages of fictional text in two conditions: a read-only condition with no additional task, and a recall condition where participants were asked to recall the story after reading it. We ran Bayesian logistic regression models to predict performance on the recall tasks from standard eye-tracking measures collected during the two reading conditions (read-only and recall). Eye-tracking measures collected in the read-only and recall conditions were both useful in predicting reading comprehension as measured by recall scores. Additionally, the relationship between eye-movement behavior and recall performance was similar for both reading conditions. In both cases, a combination of early and late measures was the best predictors of performance on the recall task. These findings demonstrate the usefulness of eye movements collected during reading with no additional task as predictors of reading comprehension ability.

Reading comprehension is one of the most complex cognitive tasks that we engage in daily life. Variance models of comprehension suggest that successful text comprehension necessitates good word-reading or decoding skills, as well as good higher-level language comprehension such as the ability to make inferences (Gough and Tunmer 1986; Kim 2016, 2020a, 2020b). Similarly, process models of text comprehension assume that for successful comprehension, the reader must first identify and process the meaning of individual words (i.e., lexical processing), integrate this meaning at the sentence level (e.g., syntactic processing) to ultimately build a mental representation of the meaning of the text (i.e., a situational model: Kintsch 1994; McNamara and Magliano 2009). However, while theories of text comprehension typically all argue that successful reading comprehension relies on good word reading and higher-level language comprehension skills (e.g., SVR, DIER, construction-intregration model), measures of these sub-skills and processes as well as measures of reading comprehension tend to lack validity. Indeed, recent research on reading comprehension assessments has raised doubts about their validity (e.g., see Mézière, Yu, Reichle, et al. 2023). Different assessments employ different combinations of reading materials (e.g., single sentences and long paragraphs of text), tasks (e.g., reading aloud vs. silently), and comprehension measures (e.g., open-ended questions and cloze procedures), which may well explain why their scores are only moderately correlated (Cutting and Scarborough 2006; Keenan et al. 2008; Nation and Snowling 1997; Mézière, Yu, Reichle, et al. 2023). These discrepancies would also explain why scores on reading comprehension assessments relate differently to scores on word-reading, oral language, and working memory assessments (Andreassen and Bråten 2010; Best et al. 2008; Colenbrander et al. 2017; Cutting and Scarborough 2006; Francis et al. 2005; García and Cain 2014; Keenan et al. 2008; Kendeou et al. 2012; Nation and Snowling 1997; Spear-Swerling 2004).

The questionable validity of reading comprehension measures is problematic for the assessment of reading comprehension for both research and clinical purposes (Colenbrander et al. 2017; Collins et al. 2018; Keenan and Meenan 2014). Researchers have suggested a number of ways to overcome this problem, such as comparing performance on multiple assessments that vary in materials and comprehension measures (e.g., Cutting and Scarborough 2006; Keenan et al. 2008), carefully selecting assessments that best match the reading tasks of interest (Mézière, Yu, Reichle, et al. 2023), and perhaps using new techniques that allow for more ecologically valid measurement of reading comprehension, such as eye tracking (Berzak et al. 2018; Copeland et al. 2014, 2016; Mézière, Yu, McArthur, et al. 2023; Mézière, Yu, Reichle, et al. 2023).

Eye-tracking has been widely used in reading research and has been successfully used to investigate the cognitive processes that support reading comprehension in experimental settings (for a review, see Rayner 2009). However, the relationship between eye movements and reading comprehension during what more closely resembles everyday reading (e.g., with more naturalistic texts and goals) is still under investigated. In the current study, we therefore explore whether eye movements can predict reading comprehension scores on a reading task that closely resembles "everyday reading". Specifically, we examine the possibility of developing a measure of reading comprehension based on eye-movement behavior by including a "read-only" condition with no additional explicit comprehension tasks (e.g., comprehension questions) and investigating the predictive relationship between eye movements during this "read-only" condition and text comprehension ability as measured by a recall task.

1 | Using Eye Movements to Investigate Reading Comprehension

Eye-tracking has been widely used in reading research as it provides a noninvasive and online measure of the cognitive processes that support text comprehension (Rayner 1998; Rayner et al. 2006; Rayner 2009; Clifton Jr et al. 2007; for a primer on the method, see Schotter and Dillon 2025). Eye-tracking typically provides two metrics of eye-movement behavior: *fixations* (i.e., brief pauses during which the eye is mostly immobile) and *saccades* (i.e., rapid ballistic movements that move the eye from one fixation to another). In addition, while most saccades move the eye forward in the text, around 10%–15% of them are *regressions* that move the eye backwards. These eye-tracking metrics and other measures derived from them (e.g., gaze durations) have been shown to reflect the cognitive processes that support reading comprehension (Rayner and Reingold 2015), including word identification (e.g., lexical processing: Rayner et al. 2011; Schilling et al. 1998), sentence processing (e.g., syntactic complexity: Staub 2010; syntactic ambiguity: Sturt 2007), and higher-level text processing (e.g., making inferences: Cunnings et al. 2014; Kreiner et al. 2008; Sturt 2003). In addition, eye movements have been shown to reflect differences in cognitive processing and cognitive skills that support successful reading comprehension, including overall reading abilities (e.g., children vs. adult readers: Blythe and Joseph 2012; Mancheva et al. 2015; Reichle et al. 2013), word reading skills (e.g., children with or without dyslexia: Hyönä and Olson 1995; Jones et al. 2007), and higher-level language comprehension skills (e.g., noticing a mismatch during anaphora resolutions: Eilers et al. 2018). Eye movements therefore provide a noninvasive and online measure of the cognitive processes that support text comprehension, and they reflect individual differences in the efficiency of these processes across readers with varying reading comprehension abilities.

Eye-movement measures derived from fixations and saccades can further be divided into so-called global and local or word-level measures. Global measures are typically calculated at the level of a sentence or passage of text and include measures such as mean fixation duration (i.e., the average duration of all fixations made during reading) and mean saccade length (i.e., the average length of all saccades made during reading). These measures are informative about reading behavior as a whole and reflect the influence of linguistic properties of the text (e.g., more difficult texts lead to longer average fixations: Rayner et al. 2006) as well as reading skills (e.g., children vs. adults: Blythe and Joseph 2012; Mancheva et al. 2015; Reichle et al. 2013) on reading comprehension processes. However, they are less informative about which cognitive process(es) are affected by text characteristics or reading skills. For example, while these measures show that children tend to make longer fixations compared to adult/skilled readers, they do not indicate whether this difference is due to children requiring more time for lexical or high-level processes (e.g., slower sentence integration). To more directly investigate the specific cognitive processes involved in reading, eye-tracking researchers compute local measures, which are typically calculated on individual words. These local measures are further divided into *early* and *late* measures. Early measures (also called *first-pass measures*) such as *first-fixation duration* (i.e., the duration of the first fixation on a word) and *gaze duration* (i.e., the sum of all fixations made on a word) are calculated using only first-pass fixations and are reflective of early processes such as lexical processing. For example, one of the most robust findings in eye-movement research is the *word-frequency effect*, showing that common or frequent words receive shorter first fixations compared to less frequent words (e.g., Schilling et al. 1998), with those longer fixations being interpreted as reflecting lexical processing difficulty. Indeed, studies have shown that readers with poorer word identification skills such as children with dyslexia exhibit larger word-frequency effects (Jones et al. 2007). On the other hand, late measures are typically used to investigate higher-level processes as they tend to reflect later processes of reading such as sentence processing (e.g., syntactic integration: Clifton Jr et al. 2007; Vasishth et al. 2013) or higher-level text comprehension processes such as anaphora processing (e.g., pronoun resolution: Sturt 2003). Typical late measures include *go-past time* (i.e., the total amount of time from when a word is first fixated until the word is exited to the right) and *total-reading times* (i.e., the sum of all fixations on a word including any regressions). For example, research shows that readers tend to spend more time re-reading ambiguous sentences (e.g., garden-path sentences) and more specifically, the sources of ambiguity and disambiguating regions (Sturt 2007). These longer re-reading times are interpreted as reflecting difficulties processing the syntactic and/or semantic information in a sentence or text. In addition, studies have shown that late measures also reflect individual differences in the processes that support reading comprehension. For example, Wonnacott et al. (2016) found that when children encountered garden-path sentences, those who spent additional time re-reading the critical ambiguous segments performed better on a related comprehension task that targeted the ambiguity of the sentence. Hence, late measures of re-reading behavior have been shown to reflect higher-level reading processes, as well as individual differences in those processes.

In sum, eye-tracking provides researchers with a variety of measures that reflect both global reading behavior as well as word-, sentence-, and text-level processes that support reading comprehension. In addition, these measures have been shown to reflect individual differences in processing among readers of varying ability and can thus be used to investigate individual differences in reading comprehension processes and skills.

Given this relationship between eye-movement measures and reading comprehension, eye movements have also been used to predict performance on reading comprehension tests that use different response formats. One goal of such studies is to examine the possibility of developing a reading comprehension measure based on eye-movement behavior. In an early study, Copeland et al. (2014) presented text (tutorial slides) and two comprehension questions (one cloze, one multiple-choice) to 39 university students in one of four conditions that manipulated *when* the questions were presented relative to the text. In condition A, participants were given the text and then the questions and the text concurrently. In condition B, participants were given the text and questions concurrently. In condition C, participants were given the text and then the questions without access to the text. Finally, in condition D, participants were given the questions and then the text and then the questions again without the text. Using artificial neural networks, the authors found that reading comprehension performance was successfully predicted by global eye-movement measures (e.g., average fixation duration, number of fixations, regression rate, and forward saccade length accounted for 79%–89% accuracy in predicting whether a reader could answer a comprehension question correctly) when readers could read the questions before the text (conditions A, B, and D). It was most difficult to predict comprehension when participants were given the questions after reading the text (condition C resulted in 49% prediction accuracy). This demonstration indicates that global eye-tracking measures can be used to predict performance on both cloze and multiple-choice questions, although there was no explicit attempt to compare the eye movements across the two response formats (see also Copeland et al. 2016; Copeland and Gedeon 2013; Martínez-Gómez and Aizawa 2014).

Inhoff et al. (2018) investigated the relationship between two latent eye-tracking variables and reading comprehension measured with two response formats: yes/no and multiple-choice questions. The two latent variables were an "acquisition" variable (a composite of early eye-tracking measures) and a "correction" variable (regression-related measures). Fifty adults were given (1) a set of 70 sentences, 10 of which were followed by a yes/no question; and (2) a short biographical passage of text followed by multiple-choice questions. These two measures of comprehension were grouped into a single measure of comprehension accuracy, with no attempt made to distinguish between the two measures. A regression analysis suggested that participants who had higher acquisition and correction scores (i.e., made longer fixations and more regressions) tended to have higher comprehension accuracy, whereas participants who read in a more "linear" left-to-right fashion (e.g., made few regressions) tended to have lower accuracy. However, a path analysis showed that, although the acquisition variable was correlated with the correction variable, only the correction variable was useful in predicting comprehension. This suggests that eye movements related to re-reading may be more useful for predicting comprehension than eye movements related to the initial reading of a text. In addition, the results suggest that corrective behavior may be associated with better text comprehension.

More recently, Southwell et al. (2020) tested if global eye-movement measures (e.g., average fixation duration and number of fixations) or corrective eye movements (i.e., proportion of fixations preceded by a regression) predict comprehension as measured using multiple-choice questions across three large datasets. For each dataset, 104–147 adults read long passages of text silently while their eye movements were tracked. The participants also answered multiple-choice questions after each page of text for the first dataset and then immediately after reading the whole passage for all datasets. Regression models indicated that the relationships between global eye movements and multiple-choice scores were similar across datasets, with more and shorter fixations associated with higher scores (see also D'Mello et al. 2020). In contrast, corrective eye movements were not significantly associated with multiple-choice scores, which contrasts with the findings of Inhoff et al. (2018). This might be explained by differences in methodology since Southwell and colleagues used individual variables as predictors instead of latent variables.

Finally, Mézière, Yu, McArthur, et al. (2023); Mézière, Yu, Reichle, et al. (2023) used eye movements to directly compare the cognitive processes recruited by three popular reading comprehension assessments: (1) the York Assessment of Reading for Comprehension (*YARC*; Snowling et al. 2009); (2) the Gray Oral Reading Test (*GORT-5*; Wiederholt and Bryant 2012); and (3) the sentence comprehension subtest of the Wide Range Achievement Test (*WRAT-4*; Wilkinson and Robertson 2006). Each used different types of stimuli (i.e., passages: YARC and GORT; single sentences: WRAT); access to stimuli (i.e., text removed before questions: GORT; text available during questions: YARC and WRAT); reading modality (i.e., silent reading: YARC and WRAT; reading aloud: GORT); and response formats (i.e., open-ended questions: YARC and GORT; cloze items: WRAT). These design differences across the three tests are summarized in Table 1. Linear regression models and leave-one-out cross-validation analyses revealed that scores on all

TABLE 1 | List of passages for the recall task.

Set	Title	Author	Length	FRES
1	The Arrest of Arsène Lupin	M. Leblanc	1063	68.5
	A Handful of Clay	H. van Dyke	970	78.2
	The Tremendous Adventure of Major Brown	G. K. Chesterton	823	71.4
2	The Awful Reason for the Vicar's Visit	G. K. Chesterton	919	68.7
	A Strange Story	O. Henry	871	80.7
	The Story of O-Tei	L. Hearn	1078	74
3	The Unpresentable Appearance of Colonel Crane	G. K. Chesterton	827	65.5
	The Story of the Late Mr. Elvesham	H. G. Wells	1032	77.6
	The Disintegration Machine	A. Conan Doyle	994	73.7

Note: FRES is the Flesch Reading Ease score (Flesch 1951).

three comprehension measures were successfully predicted by a different combinations of early and late eye-movement measures: YARC scores were best predicted by reading speed, skipping rate, gaze duration, and go-past time; GORT scores by average fixation duration, saccade length, first-fixation duration, and total reading time; and WRAT scores by reading speed, skipping rate, and regression rate. In addition, YARC scores were most strongly associated with early eye-tracking measures, GORT scores with global and early measures, and WRAT scores with reading speed.

In sum, findings from the previous eye-tracking studies suggest that eye movements can (or can be combined to) predict reading comprehension. Specifically, global eye-movement behavior (e.g., fixations durations) and corrective eye movements (e.g., making regressions) are associated with performance on reading comprehension assessments. However, the findings also suggest that the relationship between eye-movement measures and comprehension accuracy varies across tasks and response formats. That is, the best eye-movement predictors of comprehension accuracy varied depending on which comprehension test was used, to the extent that no single eye-movement measure could be identified as a useful predictor of comprehension across the different comprehension tasks. This may be in part due to readers adapting their reading behavior to the secondary tasks associated with typical pen-and-paper comprehension assessments, such as having to answer questions after reading. Importantly, the tracking of eye movements during reading does not require any secondary comprehension task (e.g., answering questions). Hence, it may be possible to develop a measure of reading comprehension that minimizes such effects of typical secondary tasks used in comprehension assessments (e.g., questions) by creating a measure based on eye movements during reading with no such additional task. Yet no study to date has examined the predictive relationship between eye movements collected during reading without any secondary tasks and text comprehension. Hence, the primary aim of this study was to investigate the possibility of using eye movements collected during a "read-only" condition with no secondary task to predict text comprehension ability as measured by a separate recall task.

2 | Recall as a Measure of Text Comprehension

Another notable gap in the reading comprehension literature is eye-movement studies that use recall as a response format for comprehension assessments. Recall, also referred to as "retell" or "summary" (Reed and Vaughn 2012), simply asks participants to summarize the content of a text they have just read (see Reed 2011 and Reed and Vaughn 2012, for reviews). Like most reading comprehension measures, recall can be used with a variety of texts (e.g., narrative and expository) that can be read silently or aloud, with verbal or written summaries of the text then being provided and quantitatively or qualitatively scored (Cao and Kim 2021; Reed 2011; Reed and Vaughn 2012; Reed and Petscher 2012). Unlike most reading comprehension measures, recall resembles most closely the type of reading done by adult readers, namely reading for comprehension with the possible future task of relating the contents of the text to others. For example, for most adults, the activity of reading a novel, news article, or social media post does not then entail answering a question or providing a missing word in a sentence, but can often entail providing a quick summary to someone (e.g., relating the main points of a news article to a friend or colleague). We therefore use recall as our measure of comprehension as this measure is most similar to reading tasks that resemble everyday reading done by adult readers most, in comparison to other commonly used comprehension tasks such as answering comprehension questions or "fill-in-the-blank" tasks.

A number of studies have examined the strength of the relationships between scores on reading comprehension tests that use recall versus open-ended questions (r = 0.43, Collins et al. 2019; r = 0.41, Keenan et al. 2008), cloze procedures (r = 0.48, Keenan et al. 2008; rs = 0.32–0.37, Kendeou et al. 2012), multiple-choice questions (r = 0.37, Collins et al. 2019; r = 0.31, Keenan et al. 2008; r = 0.41, Loyd and Steele 1986), and picture-selection (r = 0.45, Keenan et al. 2008; see Cao and Kim 2021 for a review). The fact that these relationships are moderate in strength indicates that like most reading comprehension tasks—including those that use exactly the same response format—recall tasks tax reading comprehension in a different way than other task formats (Cao and Kim 2021; Collins et al. 2019; Cutting and Scarborough 2006; Keenan et al. 2008; Kendeou et al. 2012; Loyd and Steele 1986; Mézière, Yu, Reichle, et al. 2023; Nation and Snowling 1997; Reed and Vaughn 2012). However, in contrast to other measures that use highly structured response formats, recall may minimize the effects of secondary task demands on reading comprehension, making it a good candidate for investigating the relationship between eye movements and comprehension (e.g., Mézière, Yu, McArthur, et al. 2023; Mézière, Yu, Reichle, et al. 2023). Thus, a second aim of this study was to investigate the relationship between eye movements and reading comprehension as measured by a recall task, and to test the assumption that reading for recall may have minimal influence on eye movements compared to reading with no additional task.

3 | Using Eye Movements to Investigate Recall

To our knowledge, the only study that has directly investigated whether eye movements can predict performance in recalling naturalistic passages of text was reported in an unpublished PhD thesis (Fenton 1982). He found that progressive and regressive eye movements across six passages of text explained an additional 15%–86% of the variance in recall scores over and above grade (3rd vs. 5th) or reading skills (high vs. low). Unfortunately, Fenton does not report the eye-movement measures used in his models or the nature of the relationship between the measures and recall performance (e.g., whether longer fixations times are associated with better recall).

A few other studies have examined eye-movement behavior when reading for recall of more formal expository texts (e.g., Hyönä et al. 2002; Leon et al. 2019; Vauras et al. 1992). For example, using a cluster analysis, Hyönä et al. (2002) identified four types of readers within a sample of 48 university students; (1) 19 "fast linear readers" who read quickly from left to right with few regressions; (2) 11 "slow linear readers" who did the same as (1) more slowly; (3) three "nonselective readers" who re-read the text more often than (1) and (2); and (4) seven "topic structure processors" who read slowly and re-read topic headings. With regards to recall, slow linear readers tended to provide the poorest summaries, while the topic structure processors provided the best summaries. This suggests that reading patterns characterized by efficient eye-movement behavior and targeted corrective eye movements tend to be associated with better text recall.

4 | Aims

Previous studies have used eye-movement behavior to predict performance on various reading comprehension tests that use a range of response formats (open-ended questions, multiple-choice questions, and cloze tasks). Yet, to test whether it would be possible to develop a measure of reading comprehension based solely on eye-movement behavior during reading, it is necessary to understand the relationship between eye movements and reading comprehension with eye movements collected during reading with no overt comprehension task (e.g., questions). However, no study to date has investigated the relationship between eye movements while reading with no overt comprehension task (e.g., questions) and text comprehension accuracy. In this study, we measured eye-movement behavior in a "read-only" condition (silent reading with no additional task) and two recall conditions (both entailing silent reading followed a recall task but using different texts). We examined the relationship between eye movement behavior in the "read-only" condition and recall scores from the first recall condition to test whether eye movements collected while reading with no additional tasks could predict comprehension accuracy (Aim 1). We also investigated the relationship between eye-movement behavior and recall performance within the second recall condition to examine the predictive relationship between eye-movements and comprehension accuracy as measured by recall (Aim 2). Finally, we tested whether eye-movement behavior differed between the read-only condition and the first recall condition to test whether readers adapted their reading behavior to having the additional recall task (Aim 3).

Given the lack of existing relevant studies, it is difficult to predict the outcomes, and hence this study is exploratory. As such, regardless of direction or strength, the results will provide much-needed insight into whether eye movements can predict an index of reading comprehension that closely resembles everyday reading—reading for recall.

5 | Method

5.1 | Ethics

This project was approved by the [redacted for anonymity] University Human Research ethics committee and conforms with the guidelines of the National Statement on Ethical Conduct in Human Research (2007). All participants signed an informed consent form prior to starting the experiment.

5.2 | Participants

Sixty-two native speakers of English took part in our experiment (46 females, mean age 26 years, range 18–51 years). Of those, 13 were birth multilingual speakers who had both English and at least one other language as their native language[1]. No participant indicated a history of reading or language difficulties. Participants were compensated $60 AUD for their participation or given course credit.

5.3 | Design

The design of the experiment and corresponding analyses are presented in Figure 1. The design of the experiment was set up so that the materials (consisting of nine short stories) were divided into three sets, with three stories per set and with each set corresponding to one of three conditions: (1) *read-only*; (2) *recall-1*; and (3) *recall-2*. The order of the conditions was kept consistent across participants such that the "reading-only" condition is not "contaminated" by the later presented recall task. All participants read all nine stories. The stories in the read-only and recall-1 conditions were counterbalanced (set 1 and set 2), but all participants read the same stories in the recall-2 condition (set 3). The inclusion of the recall-2 condition allowed for the relationship between eye movements collected during reading and recall performance to be investigated when both the stimuli and task were consistent across all participants. The order of the stories within each set was randomized. Eye movements were collected while participants read all nine stories, and recall scores were collected only in the recall-1 and recall-2 conditions.

As indicated in Figure 1, the eye-movement data and recall scores were used in two analyses, corresponding to our two aims. In the first analysis (Aims 1), the eye movements from the

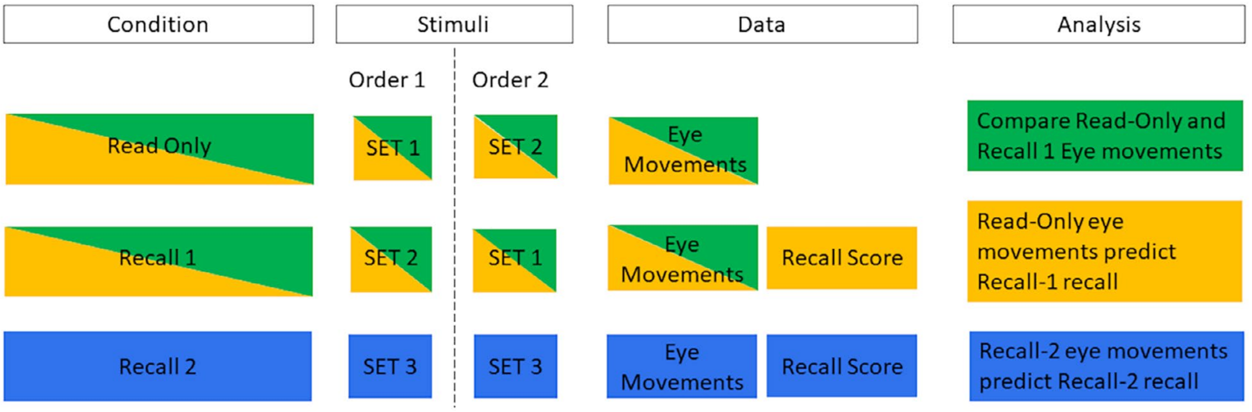

FIGURE 1 | Visualization of experiment design. It shows the design of the experiment, as well as how the eye movement and recall data were used in the different analyses. The first two datasets were both used in the first two analyses, and the third dataset was only used in the final analysis. (The analyses are color-coded to indicate what stimuli and conditions were used.)

read-only condition were used to predict the recall scores from the recall-1 condition (analysis shown in yellow in Figure 1). In the second analysis (Aims 1 and 2), eye movements from the recall-2 condition were used to predict recall scores of the same condition (analysis shown in blue in Figure 1). In the third analysis (Aim 2), the eye movements from the read-only and recall-1 conditions were compared to each other (analysis shown in green in Figure 1). The stimuli, procedure, and analyses are further described in the following sections.

5.4 | Stimuli

Stimuli were nine passages of text. The passages were adapted by the first author from fictional short stories available in the public domain (see Table 1 for titles and authors). The passages ranged from 823 words to 1078 words in length ($M=953$ words) and varied in difficulty from 65.5 to 78.2 ($M=73.1$; Flesch Reading Ease scores; FRES; Flesch 1951) indicating good readability. The FRES is a measure of text readability based on the length of both words and sentences in the text, with higher FRES indicating higher readability. The nine passages were separated into three sets of three passages that minimized differences between length and difficulty between sets (see Table 1). The order of the passages in each set was randomized between participants.

The passages were presented in Courier New, font size 18, in black color on a gray background (RGB: 204, 204, 204) on an AOC G2770PF Monitor with a screen resolution of 1920×1080 pixels and a refresh rate of 144 Hz. The passages were spread across 6–8 screen pages each, and participants could move forward using the space bar. They were not able to go back.

5.5 | Procedure

5.5.1 | Conditions

For the read-only condition, participants were instructed "to read the passages normally, at your own pace, making sure you are reading for comprehension." For both recall conditions (recall-1 and recall-2), participants were also instructed "to read the story normally, at your own pace, making sure they you are reading for comprehension," and that they would be "asked to recall each passage when you are done reading it." At the end of the passage, the experimenter further instructed them to: "Tell me the story in as much detail as possible, in your own words." At the end of their recall, participants were then given a second prompt: "Is there anything else that you can remember from the story?" (Reed and Petscher 2012).

5.5.2 | Eye Movements

Eye movements were recorded with an Eye Link 1000 (SR Research, Toronto Ontario, Canada) eye-tracker. Participants were seated about 95 cm from the display screen, such that each letter subtended approximately 0.24° of the visual angle on the screen. A headrest and a chinrest were provided in order to minimize head movements. The experiment started with a 9-point calibration, which was repeated at the start of each story. The maximum calibration error allowed was 0.50°. Each story also started with a drift-correction point located at the beginning of the text.

5.6 | Data Processing

5.6.1 | Recall Task

Participants' recalls were transcribed before scoring. Scoring was based on propositional text bases (Bovair and Kieras 1985; Kintsch and Van Dijk 1978; Kintsch 1994). Propositions consisted of both information explicitly stated in the text, as well as information only implied in the text and which require making inferences. A text base was created for each passage, with text bases ranging from 163 to 250 propositions. Because the passages in this experiment were long, the text bases were structured into "events" in order to make scoring easier. Participant's verbal recall protocols were then compared to the passage text bases and scored based on how many correct propositions they contained. The percentage of correctly recalled propositions was then calculated from the raw scores to make scores on all passages comparable. Two stories from each set were scored by three of the authors prior to finalizing the scoring system, to resolve any ambiguities or disagreements. Following this, all recalls were scored by the first author. Recall scores for stories which participants did not read in their entirety (e.g., skipped a page) were not calculated and were treated as missing data (2 stories in total).

5.6.2 | Eye Movements

The eye-movement data was preprocessed in Data Viewer (SR Research, Toronto, Ontario, Canada). One participant's data was discarded due to an eye-tracker malfunction. We further excluded participants and items based on visual inspection of the eye-movement data (e.g., when calibration was poor). This led to the exclusion of another three participants. This procedure resulted in the loss of 8.5% of all trials. Fixations shorter than 80 ms or longer than 800 ms were excluded from the analysis, which are the most commonly used cut-off criteria for excluding outlier fixation durations in eye-tracking studies of reading (Schotter and Dillon 2025; Eskenazi 2024). Additionally, forward saccades longer than the perceptual span (20 characters; Rayner 2009) were excluded from the analysis (5.6% of all forward saccades) as they typically represent outliers (e.g., long sweep saccades to look at the bottom of the page/page number).

5.7 | Data Analysis

5.7.1 | Predicting Comprehension From Eye Movements in the "Read-Only" Condition

In our first analysis, we used eye movements from the read-only condition to predict recall performance on the recall-1 condition (yellow in Figure 1). We fitted a series of logistic regression models within the Bayesian framework using the "brms" package (Bürkner 2017, 2018) in *R* (R Core Team 2020). We included nine predictors including both global measures: (1) reading speed (i.e., words per minute); (2) average fixation duration (i.e., average duration of all fixations in a text); (3) average forward saccade length (i.e., length of all rightward eye movements in the text, in character space); and local measures: (4) first-pass skipping rate (i.e., proportion of words skipped during first-pass); (5) first-fixation duration (i.e., duration of the first fixation on a word); (6) gaze duration (i.e., sum of all first-pass fixations on a word), (7) regression rate (i.e., proportion of regressions in the text); (8) go-past time (i.e., the sum of all fixations made on a word up to when the word is exited to the right, including time spent during regressions); and (9) total reading time (i.e., the sum of all fixations on a word). All measures were aggregated per participant, and centered prior to analysis to make the model estimates comparable to each other.

In the models, recall scores were treated as a binary variable based on the propositions data (coded as: recalled = 1, not recalled = 0) and then entered into the model as the total number of correctly recalled propositions across the set of stories (i.e., the three stories in the sets were treated as one set of trials), compared to the total number of propositions in the set. The models therefore predicted the odds of participants recalling individual propositions correctly, not the percentage of correctly recalled propositions. The output of the models should therefore be interpreted as showing log-odds of recalling individual propositions, and not as a change in the percentage of propositions correctly recalled (see Section 6).

To identify the best set of predictors, we ran a model for every possible combination of our nine predictors (512 models in total), with the number of predictors ranging from none (i.e., the null model) to nine (i.e., the full model). Within each set of 512 models, individual models were then evaluated and compared using *leave-one-out* cross-validation (*LOO*, Gelman et al. 2014; Vehtari et al. 2017). The LOO estimates a model's ability to predict new data by running the model as many times as there are data points in the dataset, leaving out a separate data point each time. Through this process, the *estimated log predictive density* (*elpd*) is estimated, which is a measure of how good a model is at predicting new data. The elpd-LOO can then be used to compare models. As the elpd-LOO is only a measure of how good a model is at predicting new data, we also calculated the Bayes R^2 for each model indicating how much variance the models explain in our data. However, the LOO is more conservative and more robust against overfitting; hence we base our inferences on the elpd-LOO. We then looked at the output of the "top" 10 models according to the elpd-LOO, as well as the full model, a model with only the eye-movement measures (i.e., all measures except reading speed), and a model with reading speed alone. This method is identical to previous work using Bayesian regression and the LOO to investigate the predictive relationship between eye movements and comprehension outcomes (Mézière, Yu, McArthur, et al. 2023; Mézière, Yu, Reichle, et al. 2023).

5.7.2 | Predicting Recall Scores From Eye Movements in the Recall Condition

In our second analysis, we used data from the final set of stories (i.e., set 3, recall-2 condition) which were read by all participants to predict recall scores from eye movements collected on the same stories (blue in Figure 1). This analysis was conducted identically to analysis 1 described above. To inform Aim 1, we also compared the results to those of the first analysis to examine possible differences in the relationship between eye movements and recall performance across datasets.

5.7.3 | Comparing Eye Movements During the Read-Only and Recall Conditions

In the third analysis, we used eye movements collected while participants read the first two sets of stories and compared the read-only condition to the recall-1 condition because the stories in these two conditions were counter-balanced. We investigated differences in the same nine variables (reading speed, etc.) by running linear regression models, one for each eye-tracking measure as the outcome variable, with condition (recall vs. no recall) as the predictor. All outcome variables were aggregated per story and participant prior to analysis. All reading time variables and saccade length were treated as log-normal (i.e., normally distributed residuals on the log-scale). Skipping was treated as a binary variable (skipped = 1, fixated = 0) and regression number was treated as normal. We ran a regression models within the Bayesian framework for each measure as the outcome variable and made inferences based on the 95% credible interval.

To investigate the effect of task, the condition (recall vs. no recall) was contrast coded as −0.5 (read-only) and 0.5 (recall) such that the intercept of the models could be interpreted as the mean across conditions. Because the order of the two conditions (read-only and recall-1) was kept constant across participants, we controlled for the potential effect of time by adding story number as a predictor. The story number was logged and scaled prior to analysis. Finally, we included random intercepts and slopes for participants and items (i.e., individual stories).

6 | Results

6.1 | Recall Task Scores

Participants were able to recall around 28% of propositions on average across the nine stories, which is comparable to previous findings (e.g., Reed and Petscher 2012; Reed and Vaughn 2012). The average percentage of recalled propositions for each story is shown in Table 2 and illustrated in Figure 2. We interpreted the quality of recalls based on the mean and standard deviation of recall scores, such that most participants (n = 44) performed within one standard deviation of the mean (18%–38% of propositions recalled on average), 10 participants had low performance (< 18% of propositions recalled on average), and the remaining eight performed very well (> 38% of propositions recalled on average). Participants

TABLE 2 | Summary statistics of the recall scores.

Set	Story	Mean (SD)	Range	Length	FRES	Propositions
1	Lupin	22.5 (12)	3.6–40.8	1063	68.5	250
	Clay	23.5 (11)	4.6–44.6	970	78.2	175
	Major Brown	28.2 (16)	1.7–63.6	823	71.4	176
2	Vicars	28.5 (12)	14.8–63.9	919	68.7	169
	Strange Story	16.5 (8)	6.5–43.5	871	80.7	186
	O-Tei	22.5 (8)	8.2–38.6	1078	74	220
3	Cabbage	38.7 (12)	16.6–68.7	827	65.5	163
	Elvesham	27.1 (11)	5.6–60.5	1032	77.6	233
	Disintegration Machine	29.2 (12)	6.6–62.8	994	73.7	226

Note: FRES is the Flesch Reading Ease score (Flesch 1951).

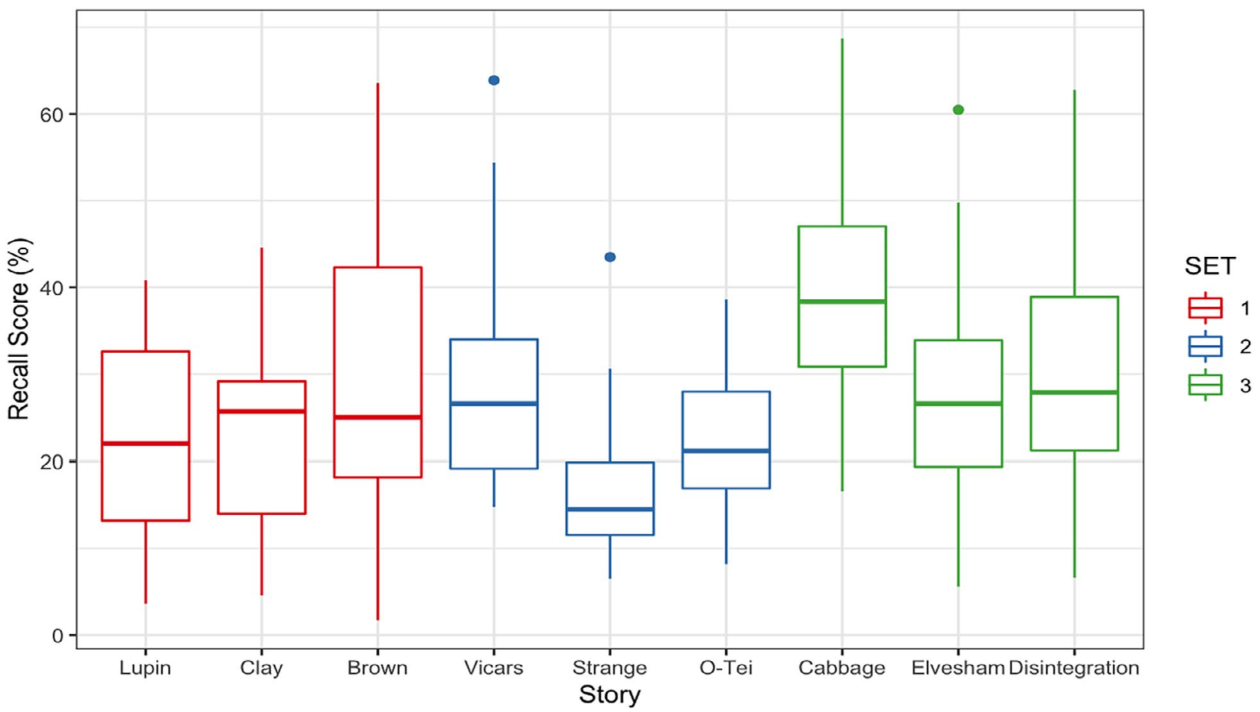

FIGURE 2 | Distribution of recall score per story. It shows participants' average recall scores for each story.

generally performed better for set 3 ($M = 31.8\%$) than in the other two sets. This seems to be in part driven by the fact that one story in that set ("Cabbage") appeared much easier than the others. Recall performance for sets 1 and 2 was similar ($M = 24.7\%$ and 22.5%, respectively), although performance in one story in set 2 ("Strange Story") was much lower than for any other story. The scores between the sets were highly correlated ($rs = 0.87–0.90$; all $ps < 0.05$), as illustrated in Figure 3. This indicates that, although some stories were more difficult than others, participants' performance was fairly consistent, relative to the story's difficulty, such that participants who performed well tended to perform well across stories, and vice versa.

6.2 | Aim 1: Predicting Comprehension From Eye Movements in the "Read-Only" Condition

In this analysis, we predicted recall-1 scores from eye-movement data collected while participants were not given an additional task (i.e., the "read-only" condition). The output of the best 10 models, the model with only reading speed, the model with only the eye-movement measures, and the full model are shown in Table 3. The best predictors of recall were reading speed, average fixation duration, average forward saccade length, regression rate, and total reading time, closely followed by gaze duration. The results indicate that faster reading speed and shorter total reading times were associated with worse performance on the recall task, and other predictors suggested that indicators of efficient eye-movement behavior (e.g., shorter average fixations, fewer regressions, and longer saccades) were predictive of better performance. While reading speed was the strongest predictor, the model including eye-movement measures significantly improved predictions over and above the model with reading speed alone. Indeed, the full model explained 36% of the variance and the eye-movements only model explained 34% of the variance while the speed-only model only explained only 10% of the variance.

6.3 | Aim 2: Predicting Recall Scores From Eye Movements in the Recall Condition

In this second analysis, we predicted recall scores from the recall-2 condition from eye movements collected during recall-2. The output of the best 10 models, the model with only reading speed, the model with only eye-movement measures, and the full model are shown in Table 4. The output of the models shows that the best predictors of recall scores were reading speed, forward saccade length, first-fixation duration, and regression rate, closely followed by total reading time. Similarly to the first analysis, the models show that faster reading speed and shorter total reading times were predictive of worse performance on the recall task, while the other predictors suggest that more efficient eye-movement behavior (i.e., shorter first-fixations, longer saccades, and fewer regressions) were predictive of better performance. Although reading speed was the strongest predictor, the elpd-LOO values show that models with both reading speed and eye-tracking measures make significantly better predictions. The full model explained 29% of the variance in our data, and the eye-movement only model explained 16% of the variance, whereas the speed-only model explained only 8%.

The full model for this analysis using eye movements from the recall-2 condition has a smaller R^2 than the model in the reading-only analysis (29% vs. 36%). However, there are multiple possible explanations for this phenomenon. For example, such models will have a high R^2 if performance in the outcome variable is close to 0 (high rate of failure) or 1 (high rate of success). Hence, differences in R^2 between the two sets of models could be due to differences in scores between the sets of stories investigated, as the average recall performance is closer to 0 in sets 1 or 2 (i.e., recall-1 used in this analysis; average recall = 23.6) than in set 3 (i.e., recall-2; average recall = 31.8). We tested this possibility by running the same analysis using eye movements from recall-1 and the read-only condition to predict recall-2 performance and found that the R^2 for the full models were nearly identical (27% and 26%, respectively). As the results of these analyses were nearly identical to the ones reported here, we do not describe them further (but see Appendix S1 for details).

6.4 | Aim 3: Comparing Eye Movements During the Read-Only and Recall Conditions

In the final analysis, we examined whether participants adapted their eye-movement behavior during reading when they had to perform a recall task compared to reading with no additional task. The output of the models for each variable is shown in Table 5. The results from all models show a similar pattern: participants

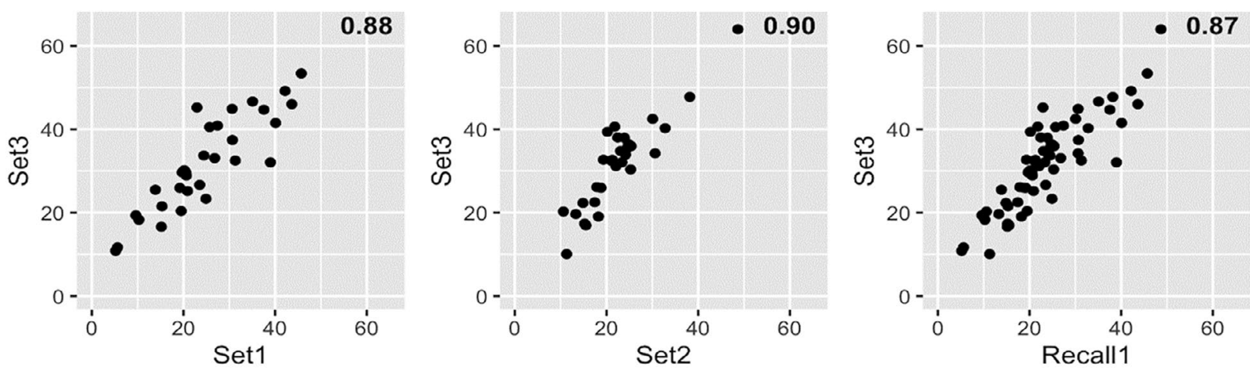

FIGURE 3 | Correlations of recall scores. It shows scatterplots and correlation coefficients of recall scores between sets of stories and between the recall-1 (i.e., sets 1 and 2 combined) and recall-2 (i.e., set 3) conditions. Each data point represents one participant's score on each set of text, with Recall 1 including both Set 1 and Set 2.

TABLE 3 | Output of the best 10 models for the read-only condition.

Predictors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Speed only	EM only	Full model
Intercept	−1.23	−1.23	−1.23	−1.23	−1.23	−1.23	−1.23	−1.23	−1.23	−1.23	−1.22	−1.23	−1.23
Speed (wpm)	−0.27	−0.36	−0.23	−0.32	−0.39	−0.25	−0.32	−0.32	−0.26	−0.35	−0.17		−0.29
Average fixation duration	−0.39	−0.23	−0.39	−0.59		−0.51	−0.23		−0.38			−0.95	−0.69
Saccade length	0.36	0.36	0.36	0.33	0.30	0.36	0.36	0.38	0.37	0.30		0.16	0.27
Skipping									−0.03			0.20	0.21
First-fixation duration				0.44		0.14		−0.39				0.87	0.59
Gaze duration		−0.25		−0.32	−0.46		−0.24			−0.46		−0.31	−0.41
Regressions	−0.32	−0.38	−0.28	−0.33	−0.34	−0.30	−0.34	−0.35	−0.31	−0.30		−0.11	−0.26
Go-past time			−0.12				−0.12			−0.11		−0.51	−0.48
Total time	0.51	0.55	0.63	0.56	0.50	0.51	0.66	0.49	0.52	0.60		1.05	0.97
R^2 Bayes	0.32	0.33	0.32	0.34	0.32	0.32	0.34	0.30	0.32	0.32	0.10	0.34	0.36
ELDP_LOO	−877.8	−881.3	−885.8	−885.9	−886.3	−886.7	−888.5	−890.0	−891.3	−892.0	−1010.5	−905.5	−906.8

Note: Models 1–10 are ordered based on their ELPD LOO (descending). EM model = eye-movement measures only model. Green = 95% credibility interval does not include 0; yellow = 90% credibility interval does not include 0; white = 90% credibility interval includes 0; blank = predictor not included in the model.

TABLE 4 | Output of the best 10 models for the recall-2 condition.

Predictors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Speed model	EM model	Full model
Intercept	−0.84	−0.85	−0.84	−0.85	−0.85	−0.85	−0.85	−0.85	−0.84	−0.85	−0.84	−0.84	−0.85
Speed (wpm)	−0.53	−0.56	−0.51	−0.49	−0.54	−0.50	−0.52	−0.54	−0.54	−0.57	−0.15		−0.55
Average fixation duration		0.44			0.41					0.44		0.07	0.45
Saccade length	0.25	0.25	0.28	0.24	0.28	0.28	0.27	0.24	0.26	0.25		0.14	0.23
Skipping				0.13			0.03	0.03				0.08	0.14
First-fixation duration	−0.27	−0.73	−0.33	−0.37	−0.76	−0.36	−0.34	−0.28	−0.29	−0.71		−0.53	−0.73
Gaze duration						0.03				−0.02		0.30	−0.14
Regressions	−0.17	−0.22	−0.29	−0.28	−0.33	−0.29	−0.31	−0.20	−0.24	−0.22		−0.07	−0.31
Go-past time				−0.30					0.07			−0.36	−0.40
Total time			0.16	0.44	0.15	0.16	0.16					0.58	0.50
R^2 Bayes	0.25	0.26	0.27	0.28	0.27	0.27	0.27	0.25	0.26	0.26	0.08	0.16	0.29
ELPD-LOO	−998.0	−999.3	−999.4	−1001.6	−1002.5	−1002.6	−1002.7	−1003.2	−1003.7	−1006.2	−1115.8	−1142.4	−1010.9

Note: Models 1–10 are ordered based on their ELPD LOO (descending). EM model = eye-movement measures only model. Green = 95% credibility interval does not include 0; yellow = 90% credibility interval does not include 0; white = 90% credibility interval includes 0; blank = predictor not included in the model.

adapted their reading behavior to the task by slowing down (i.e., slower reading speed and longer fixations) but seemed to get faster over time. The effect of the task was found for all variables except first-pass skipping rate, as participants had slower reading speed, longer fixations on all fixation measures, shorter saccades, and made more regressions while reading for recall compared to the read-only condition. The effect of time (i.e., story number) was in the opposite direction and was present for global (speed, average fixation duration, and saccade length) and late measures (regression rate, go-past time, and total reading time) but not for early measures. These effects of task and time are illustrated in Figure 4.

TABLE 5 | Model estimates of the effect of task on eye movements.

	Model estimates	
Outcome variable	**Task**	**Time**
Speed[1] (words per minute)	−0.15	0.05
Average fixation duration[1]	0.03	−0.01
Forward saccade length[1]	−0.02	0.02
First fixation duration[1]	0.03	−0.01
Gaze duration[1]	0.03	−0.01
First-pass skipping[2]	0.09	−0.02
Regressions[3]	0.05	−0.01
Go-past time[1]	0.07	−0.02
Total reading time[1]	0.09	−0.02

Note: The model estimates indicate the effect of having the recall task on each variable compared to the average of the two conditions (recall and no recall) due to contrast coding. A separate model was run for each eye-tracking variable and reading speed. Green = 95% interval does not include 0; yellow = 90% interval does not include 0; white = credible interval includes 0.
[1]Variable on log scale.
[2]Log-odds.
[3]Normal scale.

7 | Discussion

The first aim of this study was to investigate whether eye movements collected while reading with no additional task could be used to predict comprehension accuracy as measured by a recall task. To do this, we used eye-movement data collected while participants read texts with no secondary task demands (i.e., the read-only condition) to predict participants' performance on a separate recall task. In addition, we compared this relationship to the predictive relationship between recall scores and eye movements collected when reading for recall. The second aim of this study was to investigate the predictive relationship between eye movements and recall performance. To do this, we used eye movements collected while participants read for recall to predict their performance on the recall task. Additionally, we directly compared eye-movement behavior during the read-only condition to eye-movement behavior while reading for recall. We expected that the two patterns would be similar, although participants may adapt to the additional demands of the recall task to some extent.

7.1 | Predicting Comprehension From Eye Movements in the "Read-Only" Condition

The primary aim of this study was to investigate the relationship between eye movements collected during a read-only condition and comprehension ability measured by a recall

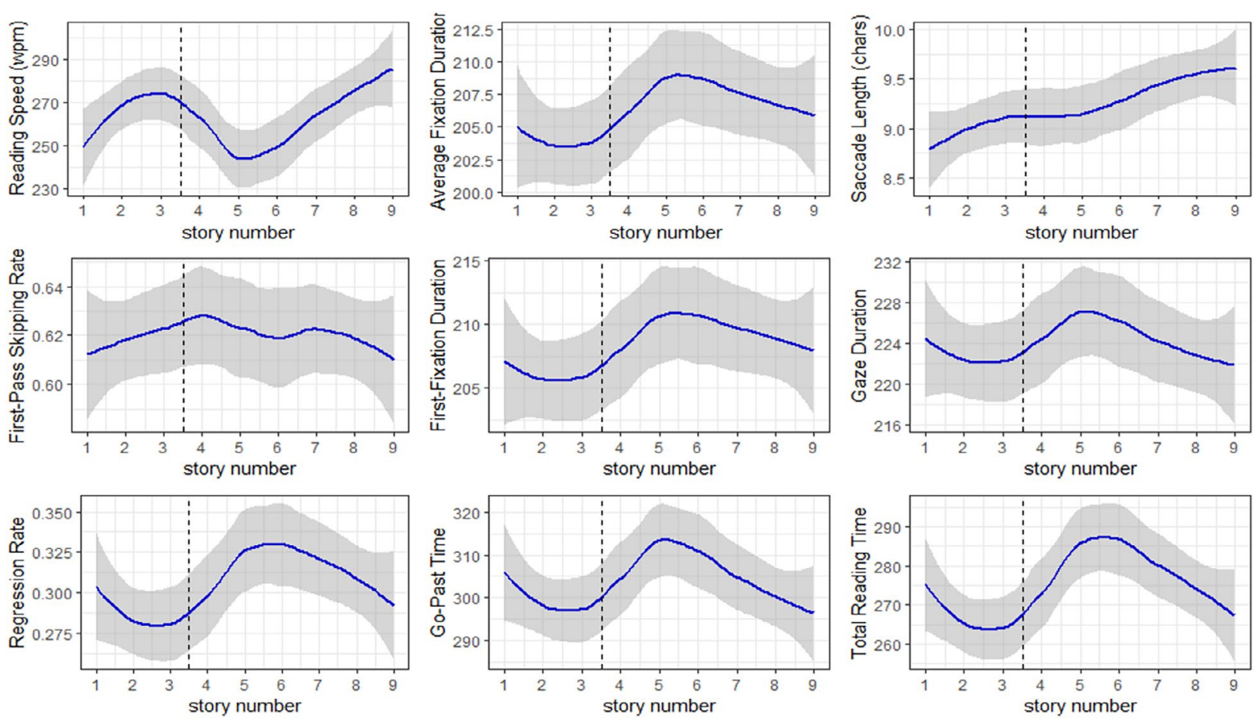

FIGURE 4 | Eye-movement behavior over time. It shows the average of nine eye-movement measures over time (i.e., story number) across the duration of the experiment. The vertical dotted lines indicate the change from the read-only condition to the recall conditions.

task. The results from the logistic regression models run with the eye movements collected while reading with no secondary task demands show that indicators of efficient eye-movement behavior (e.g., shorter average fixation durations) and indicators of slower or more careful reading strategies (e.g., slower reading speed and longer total reading times) during reading in the read-only condition were both predictive of better performance on the recall task. The results suggest that eye movements collected during reading with no additional task can be used to predict comprehension accuracy of an unrelated text. In addition, the results suggest that eye movements collected while reading without any overt comprehension task could potentially be used to develop an alternative to standard measures of reading comprehension. Indeed, the full model in this analysis explained 36% of the variance in the data, which is similar to the variance explained by models using eye movements collected while participants read text with different comprehension tasks (29%–46%: Mézière, Yu, Reichle, et al. 2023; 31%–41%: Mézière, Yu, McArthur, et al. 2023; 29%: recall task in this study). This in turn suggests that eye movements can be useful predictors of comprehension regardless of whether they are collected during a comprehension task or during reading with no overt comprehension task.

7.2 | Predicting Recall Scores From Eye Movements in the Recall Condition

The second aim of this study was to investigate whether eye-tracking measures could accurately predict reading comprehension ability as measured by story recall. Results from logistic regression models show that eye-tracking measures could predict participants' performance in the story recall task and explained 29%–36% of the variance in the data. Specifically, the results show that efficient eye-movement behavior (i.e., shorter first-pass fixations, longer saccades, and fewer regressions) along with slower reading speed and longer total reading times were associated with better performance in the recall task. This is consistent with previous findings suggesting that efficient eye-movement behavior during reading tends to be associated with better performance on comprehension tasks (Kim et al. 2019; Parshina et al. 2022). These results also provide further evidence that eye-tracking measures can be used to predict reading comprehension. Specifically, our results show that recall scores were best predicted by a combination of early and late measures, which is in line with previous findings suggesting that reading comprehension scores are typically best predicted by measures associated with both lower (e.g., lexical processing) and higher-level processing (e.g., sentence integration; Inhoff et al. 2018; Mézière, Yu, McArthur, et al. 2023; Mézière, Yu, Reichle, et al. 2023), suggesting that both types of processing are necessary for reading comprehension. The results also suggest that recall scores were associated with later eye-tracking measures, namely regression rate and total reading times, which is in line with Keenan et al.'s (2008) finding that recall captures higher-level processes of reading to a greater extent than lower-level (e.g., lexical) processing.

However, as illustrated in Table 6, the relationship between eye-tracking measures and comprehension as measured by recall varies from the pattern of results found in previous studies (Mézière, Yu, McArthur, et al. 2023; Mézière, Yu, Reichle, et al. 2023). These earlier studies investigated the predictive relationship between eye movements and three reading comprehension assessments that varied in terms of how they measured comprehension, with a key finding being that the relationships between eye movements and the three test scores were mediated by differences in task demands between comprehension assessments. The results from the current study also differ from the pattern of results reported by Mézière, Yu, McArthur, et al. (2023); Mézière, Yu, Reichle, et al. (2023). This is true both in terms of the best predictors of comprehension and the direction of the relationship between said predictors and comprehension scores. This is not unexpected, however, because such differences were previously found between reading comprehension measures. It is likely that participants adapted their reading behavior to the specific task demands of the recall task, and/or that the different comprehension tasks do not rely on the same cognitive processes to the same extent, leading to differences in the relationship between eye movements and comprehension between studies, similar to differences found between comprehension measures in Mézière, Yu, McArthur, et al. (2023);

TABLE 6 | Summary of best predictors for the five comprehension measures.

	Measure	YARC	GORT	WRAT	Read-Only	Recall-2
Global	Speed (words per minute)	↑		↑	↓	↓
	Average fixation duration		↓		↓	
	Saccades length		↑		↑	↑
Early	First-pass skipping	↓		↓		
	First fixation duration		↑			
	Gaze duration	↓				
Late	Regression rate			↑	↓	↓
	Go-past time	↑				
	Total reading time		↓		↑	↑

Note: Table 6 is adapted from Mézière, Yu, Reichle, et al. (2023) and summarizes the results from linear models predicting comprehension scores from eye-tracking measures for five comprehension measures. Read-Only: eye movements without recall task predicted recall performance. Recall-2: eye movements with recall task predicted recall performance.

Mézière, Yu, Reichle, et al. (2023). Interestingly, although the recall task may be thought to have more in common with the YARC (silent reading of long texts followed by questions and recall), the pattern of results is actually most similar to that of the GORT (reading aloud of long texts). Indeed, with the exception of reading speed and regression rate, the best predictors of GORT scores were also the best predictors of performance on story recall (saccade length, first-fixation duration/average fixation duration, and total reading time), although the direction of the relationship was not always identical (e.g., longer total reading times were associated with lower GORT scores but better recalls). One possible explanation for this discrepancy is that, although our reading task is most similar to the YARC (i.e., reading long passages silently), our comprehension task is most similar to the GORT (i.e., no access to the text when answering questions or during recall). Hence, it may be that some aspects of the comprehension task (e.g., access to text) have more influence on the relationship between eye-tracking measures and text comprehension than others (e.g., reading modality). This is speculative, however, and further research is needed to investigate the extent to which various aspects of the test design (e.g., reading modality and access to the text) affect reading behavior and its relationship to reading comprehension accuracy.

Although the relationship between efficient eye-movement behavior and good comprehension is consistent with previous findings, faster reading speed is not typically found to be associated with lower comprehension scores (Fuchs et al. 2019; Price et al. 2016). Indeed, faster reading speed was associated with poorer recall scores even in the single-predictor model in which only reading speed was included as a predictor, which is in contrast with previous findings suggesting that as a single predictor reading speed is positively associated with comprehension scores (e.g., Mézière, Yu, Reichle, et al. 2023). Nevertheless, some studies do suggest that high reading speed (e.g., skimming or "linear reading") can negatively impact comprehension (Hyönä et al. 2002; Strukelj and Niehorster 2018; see also Rayner et al. 2016, for a discussion on "speed reading" and comprehension). Hence, in our data, reading speed may be an indication of participants' depth of processing of the text (e.g., skimming vs. thorough reading), and reflect participants' attention or motivation to do the task, as well as their reading ability. Another possible explanation is that the predictive relationship between reading speed and comprehension is partially moderated by task demands. Indeed, as discussed above, the relationship between eye-movement behavior and comprehension seems to be moderated by the varying task demands of reading comprehension measures, such that differences were previously found in the direction of the relationship between comprehension and eye-tracking measures. In Mézière, Yu, Reichle, et al. (2023), faster reading speed was predictive of better comprehension scores on the WRAT and YARC, but of poorer performance on the GORT. Although the recall task and the GORT differ in several aspects of their design (i.e., reading modality and comprehension task), they put the highest demands on readers' memory as they do not allow readers to return to the text after reading. Hence, it may be that for such tasks, thoroughness (i.e., how long people spent reading the text) and efficiency (i.e., how quickly they can read the text) are both important predictors of performance, as they require readers to remember the text in detail after reading it. Although simple regression models do not allow for differences between strategies to be clearly identified (e.g., skimming vs. regular reading vs. thorough reading; but see Parshina et al. 2022 for an example of how scanpath analysis can be used to investigate reading strategies), it is likely that eye-tracking measures and reading speed are indicative of both reading comprehension ability and aspects of motivation, interest, and/or attention, which can influence reading behavior (Catrysse et al. 2018). This is also in line with the results of Mézière, Yu, McArthur, et al. (2023); Mézière, Yu, Reichle, et al. (2023), which suggest that while reading speed is a robust correlate of comprehension, it is not necessarily a good predictor of comprehension ability, and that statistical models with reading speed and eye-tracking measures together make more accurate predictions. This finding was replicated here, as models with both eye-tracking measures and reading speed as predictors explained more variance and made more accurate predictions than models with reading speed as the sole predictor.

7.3 | Comparing Eye Movements During the Read-Only and Recall Conditions

Lastly, we examined the possible differences in eye-movement behavior between reading for recall and reading without an additional task. Results from the multiple-regression models discussed in the previous sections show that the relationship between recall and eye movements in the read-only and recall conditions is highly similar, such that the differences in the two patterns of results were minimal. Indeed, the only difference between the two patterns of results is that, in the read-only condition, average fixation duration was a good predictor of recall instead of first-fixation duration. This is possibly due to the fact that these eye movements were not directly associated with the recall task. However, a similar pattern of results was found when comparing the best predictors of recall-2 scores from the read-only and recall-1 eye movements, such that first-fixation duration was a useful predictor of recall only from eye movements collected during recall-1, while average fixation duration was a useful predictor in both conditions (see Appendix S1 for details). Another possible factor is that readers adapted to the additional demands of having to recall the text after reading and hence read it more carefully, placing higher demands on (and predictive power on measures associated with) lexical processing compared to the read-only condition. Overall, this similarity suggests that, at least when the materials, reading tasks, and comprehension measures are kept consistent, the relationship between eye movements collected during reading with no secondary task demands and comprehension is similar (at least to some degree) that of eye movements collected with reading for recall and comprehension. However, the current experiment does not allow us to generalize this finding to instances where the materials, reading task, and comprehension measures are not kept consistent. Indeed, although different passages were used in the two analyses, the reading task (silent reading of long passages) and comprehension measure (story recall) were kept consistent. An interesting line of inquiry for further research would be to investigate whether eye movements during a read-only condition are equally predictive of comprehension accuracy across comprehension measures, and whether these predictions are influenced by the way comprehension is measured (e.g., recall vs. questions).

Nevertheless, results from the regression models show that participants' eye movements did differ when reading for recall compared to the read-only task. Specifically, our results show that participants generally slowed down when reading for recall as opposed to when reading with no additional task. This effect was found both for early and late measures, suggesting that participants not only spent more time re-reading the text, but also read more carefully during first pass. This suggests that participants generally read the text more carefully and thoroughly when they anticipated having to recall it than when reading in the read-only condition. This is in line with studies showing that readers adapt their reading strategies to the task (Bax and Chan 2019; Kaakinen and Hyönä 2010; O'Reilly et al. 2018; Radach et al. 2008; Schotter et al. 2014) and suggests that, although recall may closely resemble everyday reading with no overt comprehension task, it is not identical to reading with no additional task. The fact that an effect of time was found is also noteworthy, as it shows that participant's performance on the task (i.e., reading) changed over time, in a nonlinear fashion. Specifically, participants read faster over time, and spent less time re-reading the text. This is also in line with recent methodological papers suggesting that the "human factor" (e.g., fatigue, learning effect, and variations in attention) needs to be taken into account in psycholinguistic studies, as participants' task performance can vary over time (Baayen et al. 2017).

8 | Conclusion

The results presented in this article provide further evidence for the usefulness of eye movements as predictors of performance on reading comprehension tasks. Specifically, the results provide support for the validity and usefulness of eye movements collected during a text comprehension task that closely resembles everyday reading (i.e., reading with no overt task) as measures of reading comprehension processes. However, more research is necessary to better understand this relationship, which might inform further work to develop an online, ecologically valid measure of reading comprehension based on eye-movement behavior during reading.

In addition, our results suggest that while eye-movement behavior during reading for recall may be similar to eye movements while reading without any overt comprehension task, the results also show that readers adapt their eye movements and strategies to some extent between the two reading activities (recall vs. no recall). More research is necessary in order to better understand differences in task demands between different response formats (e.g., open-ended questions and multiple-choice questions) and reading with no overt task (i.e., "everyday" reading).

Acknowledgments

Open access publishing facilitated by Turun yliopisto, as part of the Wiley - FinELib agreement.

Ethics Statement

The study was approved by the Macquarie University Human Research ethics committee and conforms with the guidelines of the National Statement on Ethical Conduct in Human Research.

Consent

All participants signed an informed consent form before participating in the experiment.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data from this study is not available.

Endnotes

[1] One reviewer pointed out that the inclusion of bi-/multilinguals may have impacted the results. However, we did not find any difference in the results of these participants compared to the monolinguals in our sample. This lack of difference is also in line with previous work in which the inclusion of second-language learners in the sample did not impact the results in terms of the predictive relationship between eye movements and reading comprehension scores (see Mézière, Yu, McArthur, et al. 2023; Mézière, Yu, Reichle, et al. 2023).

References

Andreassen, R., and I. Bråten. 2010. "Examining the Prediction of Reading Comprehension on Different Multiple-Choice Tests." *Journal of Research in Reading* 33, no. 3: 263–283. https://doi.org/10.1111/j.1467-9817.2009.01413.x.

Baayen, H., S. Vasishth, R. Kliegl, and D. Bates. 2017. "The Cave of Shadows: Addressing the Human Factor With Generalized Additive Mixed Models." *Journal of Memory and Language* 94: 206–234. https://doi.org/10.1016/j.jml.2016.11.006.

Bax, S., and S. Chan. 2019. "Using Eye-Tracking Research to Investigate Language Test Validity and Design." *System* 83: 64–78. https://doi.org/10.1016/j.system.2019.01.007.

Berzak, Y., B. Katz, and R. Levy. 2018. "Assessing Language Proficiency From Eye Movements in Reading. *arXiv*." https://doi.org/10.48550/arXiv.1804.07329.

Best, R. M., R. G. Floyd, and D. S. Mcnamara. 2008. "Differential Competencies Contributing to Children's Comprehension of Narrative and Expository Texts." *Reading Psychology* 29, no. 2: 137–164. https://doi.org/10.1080/02702710801963951.

Blythe, H. I., and H. S. S. L. Joseph. 2012. "Children's Eye Movements During Reading." In *The Oxford Handbook of Eye Movements*, edited by S. P. Liversedge, I. Gilchrist, and S. Everling, 644–662. Oxford University Press.

Bovair, S., and D. E. Kieras. 1985. "A Guide to Propositional Analysis for Research on Technical Prose." In *Understanding Expository Text*, 315–362. Routledge.

Bürkner, P. 2017. "brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80, no. 1: 1–28.

Bürkner, P. 2018. "Advanced Bayesian Multilevel Modeling With the R Package brms." *R Journal* 10, no. 1: 395–411.

Cao, Y., and Y. S. G. Kim. 2021. "Is Retell a Valid Measure of Reading Comprehension?" *Educational Research Review* 32: 100375. https://doi.org/10.1016/j.edurev.2020.100375.

Catrysse, L., D. Gijbels, and V. Donche. 2018. "It Is Not Only About the Depth of Processing: What if Eye Am Not Interested in the Text?" *Learning and Instruction* 58: 284–294. https://doi.org/10.1016/j.learninstruc.2018.07.009.

Clifton, C., Jr., A. Staub, and K. Rayner. 2007. "Eye Movements in Reading Words and Sentences." In *Eye Movements*, 341–371. Elsevier.

Colenbrander, D., L. Nickels, and S. Kohnen. 2017. "Similar but Different: Differences in Comprehension Diagnosis on the Neale Analysis of Reading Ability and the York Assessment of Reading for Comprehension." *Journal of Research in Reading* 40, no. 4: 403–419.

Collins, A. A., D. L. Compton, E. R. Lindström, and J. K. Gilbert. 2019. "Performance Variations Across Reading Comprehension Assessments: Examining the Unique Contributions of Text, Activity, and Reader." *Reading and Writing* 33, no. 3: 605–634. https://doi.org/10.1007/s11145-019-09972-5.

Collins, A. A., E. R. Lindström, and D. L. Compton. 2018. "Comparing Students With and Without Reading Difficulties on Reading Comprehension Assessments: A Meta-Analysis." *Journal of Learning Disabilities* 51, no. 2: 108–123. https://doi.org/10.1177/0022219417704636.

Copeland, L., and T. Gedeon. 2013. "Measuring Reading Comprehension Using Eye Movements." In *4th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2013—Proceedings*, 791–796. IEEE. https://doi.org/10.1109/CogInfoCom.2013.6719207.

Copeland, L., T. Gedeon, and S. Caldwell. 2016. "Effects of Text Difficulty and Readers on Predicting Reading Comprehension From Eye Movements." In *6th IEEE Conference on Cognitive Infocommunications, CogInfoCom 2015—Proceedings*, 407–412. IEEE. https://doi.org/10.1109/CogInfoCom.2015.7390628.

Copeland, L., T. Gedeon, and S. Mendis. 2014. "Predicting Reading Comprehension Scores From Eye Movements Using Artificial Neural Networks and Fuzzy Output Error." *Artificial Intelligence Research* 3, no. 3: 35–48. https://doi.org/10.5430/air.v3n3p35.

Cunnings, I., C. Patterson, and C. Felser. 2014. "Variable Binding and Coreference in Sentence Comprehension: Evidence from Eye Movements." *Journal of Memory and Language* 71, no. 1: 39–56.

Cutting, L. E., and H. S. Scarborough. 2006. "Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills Can Depend on How Comprehension Is Measured." *Scientific Studies of Reading* 10, no. 3: 277–299. https://doi.org/10.1207/s1532799xssr1003_5.

D'Mello, S. K., R. Southwell, and J. Gregg. 2020. "Machine-Learned Computational Models Can Enhance the Study of Text and Discourse: A Case Study Using Eye Tracking to Model Reading Comprehension." *Discourse Processes* 57, no. 5–6: 420–440. https://doi.org/10.1080/0163853X.2020.1739600.

Eilers, S., S. P. Tiffin-Richards, and S. Schroeder. 2018. "Individual Differences in Children's Pronoun Processing During Reading: Detection of Incongruence Is Associated With Higher Reading Fluency and More Regressions." *Journal of Experimental Child Psychology* 173: 250–267. https://doi.org/10.1016/j.jecp.2018.04.005.

Eskenazi, M. A. 2024. "Best Practices for Cleaning Eye Movement Data in Reading Research." *Behavior Research Methods* 56, no. 3: 2083–2093.

Fenton, T. 1982. *The Role of Eye Movement Analysis in the Study of Cognitive Processes in Reading* (Publication No. 8223249) [Doctoral Dissertation, Harvard University]. ProQuest Dissertations & Theses Global.

Flesch, R. F. 1951. *How to Test Readability*. Harper.

Francis, D. J., J. M. Fletcher, H. W. Catts, and J. B. Tomblin. 2005. *Dimensions Affecting the Assessment of Reading Comprehension*, 369–394. Children's Reading Comprehension and Assessment. https://doi.org/10.4324/9781410612762.

Fuchs, L. S., D. Fuchs, M. K. Hosp, and J. R. Jenkins. 2019. "Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis." In *The Role of Fluency in Reading Competence, Assessment, and Instruction*, vol. 8438, 239–256. Routledge. https://doi.org/10.4324/9781410608246-3.

García, J. R., and K. Cain. 2014. "Decoding and Reading Comprehension: A Meta-Analysis to Identify Which Reader and Assessment Characteristics Influence the Strength of the Relationship in English." *Review of Educational Research* 84, no. 1: 74–111. https://doi.org/10.3102/0034654313499616.

Gelman, A., J. Hwang, and A. Vehtari. 2014. "Understanding Predictive Information Criteria for Bayesian Models." *Statistics and Computing* 24, no. 6: 997–1016.

Gough, P. B., and W. E. Tunmer. 1986. "Decoding, Reading, and Reading Disability." *Remedial and Special Education* 7, no. 1: 6–10. https://doi.org/10.1177/074193258600700104.

Hyönä, J., and R. K. Olson. 1995. "Eye Fixation Patterns Among Dyslexic and Normal Readers: Effects of Word Length and Word Frequency." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, no. 6: 1430–1440.

Hyönä, J., R. F. Lorch, and J. K. Kaakinen. 2002. "Individual Differences in Reading to Summarize Expository Text: Evidence From Eye Fixation Patterns." *Journal of Educational Psychology* 94, no. 1: 44–55. https://doi.org/10.1037/0022-0663.94.1.44.

Inhoff, A. W., J. Gregg, and R. Radach. 2018. "Eye Movement Programming and Reading Accuracy." *Quarterly Journal of Experimental Psychology* 71, no. 1: 3–10. https://doi.org/10.1080/17470218.2016.1226907.

Jones, M. W., M. L. Kelly, and M. Corley. 2007. "Adult Dyslexic Readers Do Not Demonstrate Regularity Effects in Sentence Processing: Evidence from Eye-Movements." *Reading and Writing* 20, no. 9: 933–943.

Kaakinen, J. K., and J. Hyönä. 2010. "Task Effects on Eye Movements During Reading." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36, no. 6: 1561–1566. https://doi.org/10.1037/a0020693.

Keenan, J. M., and C. E. Meenan. 2014. "Test Differences in Diagnosing Reading Comprehension Deficits." *Journal of Learning Disabilities* 47, no. 2: 125–135. https://doi.org/10.1177/0022219412439326.

Keenan, J. M., R. S. Betjemann, and R. K. Olson. 2008. "Reading Comprehension Tests Vary in the Skills They Assess: Differential Dependence on Decoding and Oral Comprehension." *Scientific Studies of Reading* 12, no. 3: 281–300. https://doi.org/10.1080/10888430802132279.

Kendeou, P., T. C. Papadopoulos, and G. Spanoudis. 2012. "Processing Demands of Reading Comprehension Tests in Young Readers." *Learning and Instruction* 22, no. 5: 354–367. https://doi.org/10.1016/j.learninstruc.2012.02.001.

Kim, Y. S. G., Y. Petscher, and C. Vorstius. 2019. "Unpacking Eye Movements During Oral and Silent Reading and Their Relations to Reading Proficiency in Beginning Readers." *Contemporary Educational Psychology* 58: 102–120.

Kim, Y.-S. 2016. "Direct and Mediated Effects of Language and Cognitive Skills on Comprehension of Oral Narrative Texts (Listening Comprehension) for Children." *Journal of Experimental Child Psychology* 141: 101–120. https://doi.org/10.1016/j.jecp.2015.08.003.

Kim, Y.-S. 2020a. "Hierarchical and Dynamic Relations of Language and Cognitive Skills to Reading Comprehension: Testing the Direct and Indirect Effects Model of Reading (DIER)." *Journal of Educational Psychology* 112, no. 4: 667–684. https://doi.org/10.1037/edu0000407.supp.

Kim, Y.-S. 2020b. "Toward Integrative Reading Science: The Direct and Indirect Effects Model of Reading." *Journal of Learning Disabilities* 53, no. 6: 469–491. https://doi.org/10.1177/0022219420908239.

Kintsch, W. 1994. "Text Comprehension, Memory, and Learning." *American Psychologist* 49, no. 4: 294–303.

Kintsch, W., and T. A. Van Dijk. 1978. "Toward a Model of Text Comprehension and Production." *Psychological Review* 85, no. 5: 363–394.

Kreiner, H., P. Sturt, and S. Garrod. 2008. "Processing Definitional and Stereotypical Gender in Reference Resolution: Evidence from Eye-Movements." *Journal of Memory and Language* 58, no. 2: 239–261.

Leon, J. A., J. D. Moreno, I. Escudero, and J. K. Kaakinen. 2019. "Selective Attention to Question-Relevant Text Information Precedes High-Quality Summaries: Evidence From Eye Movements." *Journal of Eye Movement Research* 12, no. 1: 1–16. https://doi.org/10.16910/jemr.12.1.6.

Loyd, B. H., and J. L. Steele. 1986. "Assessment of Reading Comprehension: A Comparison of Constructs." *Reading Psychology* 7, no. 1: 1–10. https://doi.org/10.1080/0270271860070102.

Mancheva, L., E. D. Reichle, B. Lemaire, S. Valdois, J. Ecalle, and A. Guérin-Dugué. 2015. "An Analysis of Reading Skill Development Using EZ Reader." *Journal of Cognitive Psychology* 27, no. 5: 657–676.

Martínez-Gómez, P., and A. Aizawa. 2014. "Recognition of Understanding Level and Language Skill Using Measurements of Reading Behavior. In Proceedings of the 19th International Conference on Intelligent User Interfaces. 95-104."

McNamara, D. S., and J. Magliano. 2009. "Chapter 9 Toward a Comprehensive Model of Comprehension." In *Psychology of Learning and Motivation—Advances in Research and Theory*, vol. 51. Elsevier Inc. https://doi.org/10.1016/S0079-7421(09)51009-2.

Mézière, D., L. Yu, E. D. Reichle, T. Von Der Malsburg, and G. McArthur. 2023. "Using Eye-Tracking Measures to Predict Reading Comprehension." *Reading Research Quarterly* 58, no. 3: 425–449. https://doi.org/10.1002/rrq.498.

Mézière, D., L. Yu, G. McArthur, E. D. Reichle, and T. Von Der Malsburg. 2023. "Scanpath Regularity as an Index of Reading Comprehension." *Scientific Studies of Reading* 28, no. 1: 79–100. https://doi.org/10.1080/10888438.2023.2232063.

Nation, K., and M. Snowling. 1997. "Assessing Reading Difficulties: The Validity and Utility of Current Measures of Reading Skill." *British Journal of Educational Psychology* 67, no. 3: 359–370. https://doi.org/10.1111/j.2044-8279.1997.tb01250.x.

O'Reilly, T., D. G. Feng, D. J. Sabatini, D. Z. Wang, and D. J. Gorin. 2018. "How Do People Read the Passages During a Reading Comprehension Test? The Effect of Reading Purpose on Text Processing Behavior." *Educational Assessment* 23, no. 4: 277–295. https://doi.org/10.1080/10627197.2018.1513787.

Parshina, O., I. A. Sekerina, A. Lopukhina, and T. von Der Malsburg. 2022. "Monolingual and Bilingual Reading Processes in Russian: An Exploratory Scanpath Analysis." *Reading Research Quarterly* 57, no. 2: 469–492.

Price, K. W., E. B. Meisinger, M. M. Louwerse, and S. D'Mello. 2016. "The Contributions of Oral and Silent Reading Fluency to Reading Comprehension." *Reading Psychology* 37, no. 2: 167–201. https://doi.org/10.1080/02702711.2015.1025118.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Radach, R., L. Huestegge, and R. Reilly. 2008. "The Role of Global Top-Down Factors in Local Eye-Movement Control in Reading." *Psychological Research* 72, no. 6: 675–688. https://doi.org/10.1007/s00426-008-0173-3.

Rayner, K. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124, no. 3: 372.

Rayner, K. 2009. "Eye Movements and Attention in Reading, Scene Perception, and Visual Search." *Quarterly Journal of Experimental Psychology* 62, no. 8: 1457–1506. https://doi.org/10.1080/17470210902816461.

Rayner, K., E. R. Schotter, M. E. J. Masson, M. C. Potter, and R. Treiman. 2016. "So Much to Read, So Little Time: How Do We Read, and Can Speed Reading Help?" *Psychological Science in the Public Interest* Supplement 17, no. 1: 4–34. https://doi.org/10.1177/1529100615623267.

Rayner, K., K. H. Chace, T. J. Slatterly, and J. Ashby. 2006. "Eye Movements as Reflections of Comprehension Processes in Reading." *Scientific Studies of Reading* 10, no. 3: 241–255. https://doi.org/10.1207/s1532799xssr1003.

Rayner, K., and E. M. Reingold. 2015. "Evidence for Direct Cognitive Control of Fixation Durations During Reading." *Current Opinion in Behavioral Sciences* 1: 107–112.

Rayner, K., T. J. Slattery, D. Drieghe, and S. P. Liversedge. 2011. "Eye Movements and Word Skipping During Reading: Effects of Word Length and Predictability." *Journal of Experimental Psychology: Human Perception and Performance* 37, no. 2: 514–528. https://doi.org/10.1037/a0020990.

Reed, D. K. 2011. "A Review of the Psychometric Properties of Retell Instruments." *Educational Assessment* 16, no. 3: 123–144. https://doi.org/10.1080/10627197.2011.604238.

Reed, D. K., and S. Vaughn. 2012. "Retell as an Indicator of Reading Comprehension." *Scientific Studies of Reading* 16, no. 3: 187–217. https://doi.org/10.1080/10888438.2010.538780.

Reed, D. K., and Y. Petscher. 2012. "The Influence of Testing Prompt and Condition on Middle School Students' Retell Performance." *Reading Psychology* 33, no. 6: 562–585. https://doi.org/10.1080/02702711.2011.557333.

Reichle, E. D., S. P. Liversedge, D. Drieghe, et al. 2013. "Using EZ Reader to Examine the Concurrent Development of Eye-Movement Control and Reading Skill." *Developmental Review* 33, no. 2: 110–149.

Schilling, H. E. H., K. Rayner, and J. I. Chumbley. 1998. "Comparing Naming, Lexical Decision, and Eye Fixation Times: Word Frequency Effects and Individual Differences." *Memory and Cognition* 26, no. 6: 1270–1281. https://doi.org/10.3758/BF03201199.

Schotter, E. R., and B. Dillon. 2025. "A Beginner's Guide to Eye Tracking for Psycholinguistic Studies of Reading." *Behavior Research Methods* 57, no. 2: 68.

Schotter, E. R., K. Bicknell, I. Howard, R. Levy, and K. Rayner. 2014. "Task Effects Reveal Cognitive Flexibility Responding to Frequency and Predictability: Evidence From Eye Movements in Reading and Proofreading." *Cognition* 131, no. 1: 1–27.

Snowling, M. J., S. E. Stothard, P. Clarke, et al. 2009. *YARC York Assessment of Reading for Comprehension Passage Reading*. GL Publishers.

Southwell, R., J. Gregg, R. Bixler, and S. K. D'Mello. 2020. "What Eye Movements Reveal About Later Comprehension of Long Connected Texts." *Cognitive Science* 44, no. 10: e12905. https://doi.org/10.1111/cogs.12905.

Spear-Swerling, L. 2004. "Fourth Graders' Performance on a State-Mandated Assessment Involving Two Different Measures of Reading Comprehension." *Reading Psychology* 25, no. 2: 121–148. https://doi.org/10.1080/02702710490435727.

Staub, A. 2010. "Eye Movements and Processing Difficulty in Object Relative Clauses." *Cognition* 116, no. 1: 71–86.

Strukelj, A., and D. C. Niehorster. 2018. "One Page of Text: Eye Movements During Regular and Thorough Reading, Skimming, and Spell Checking." *Journal of Eye Movement Research* 11, no. 1: 1–22. https://doi.org/10.16910/jemr.11.1.1.

Sturt, P. 2007. "Semantic Re-Interpretation and Garden Path Recovery." *Cognition* 105, no. 2: 477–488.

Sturt, P. 2003. "The Time-Course of the Application of Binding Constraints in Reference Resolution." *Journal of Memory and*

Language 48, no. 3: 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3.

Vasishth, S., T. von der Malsburg, and F. Engelmann. 2013. "What Eye Movements Can Tell Us About Sentence Comprehension." *Wiley Interdisciplinary Reviews: Cognitive Science* 4, no. 2: 125–134.

Vauras, M., J. Hyönä, and P. Niemi. 1992. "Comprehending Coherent and Incoherent Texts: Evidence From Eye Movement Patterns and Recall Performance." *Journal of Research in Reading* 15, no. 1: 39–54. https://doi.org/10.1111/j.1467-9817.1992.tb00020.x.

Vehtari, A., A. Gelman, and J. Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC." *Statistics and Computing* 27, no. 5: 1413–1432.

Wiederholt, J. L., and B. R. Bryant. 2012. *Gray Oral Reading Tests—Fifth Edition (GORT-5).* Pro-Ed.

Wilkinson, G. S., and G. J. Robertson. 2006. *Wide Range Achievement Test (WRAT4).* Psychological Assessment Resources.

Wonnacott, E., H. S. S. L. Joseph, J. S. Adelman, and K. Nation. 2016. "Is Children's Reading "Good Enough"? Links Between Online Processing and Comprehension as Children Read Syntactically Ambiguous Sentences." *Quarterly Journal of Experimental Psychology* 69, no. 5: 855–879.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.