UC Irvine
UC Irvine Previously Published Works

Title

The Relations Between Children's Comprehension Monitoring and Their Reading Comprehension and Vocabulary Knowledge: An Eye-Movement Study.

Permalink

https://escholarship.org/uc/item/5pk9d8d9

Journal

Reading and writing, 33(3)

ISSN

0922-4777

Authors

Zargar, Elham
Adams, Ashley
Connor, Carol McDonald

Publication Date

2020-03-01

DOI

10.1007/s11145-019-09966-3

Peer reviewed

Check for updates

The relations between children's comprehension monitoring and their reading comprehension and vocabulary knowledge: an eye-movement study

Elham Zargar[1] · Ashley Marie Adams[1] · Carol McDonald Connor[1]

Abstract

Poor reading comprehension may be due to having ineffective comprehension monitoring, the metacognitive process of evaluating and regulating comprehension. When comprehension breaks down due to an inconsistency either at the word-level (e.g., due to an unfamiliar word) or at the sentence-level (e.g., due to contradictory information), readers may identify the misunderstanding and take steps to regulate their comprehension. In the current study, we utilized two eye-movement tasks (one newly developed) to examine comprehension monitoring in third through fifth grade students ($n = 123$), when confronted with word- and sentence-level inconsistencies, by measuring the amount of time they read (gaze duration) and reread the target inconsistent words. We investigated how this skill may be associated with individual differences in age, reading comprehension ability, and vocabulary knowledge. The results showed that generally, all students detected the word-level inconsistencies, indicated by longer gaze durations, and attempted to regulate their comprehension after detecting both word- and sentence-level inconsistencies, as indicated by more time spent rereading. Students with stronger reading comprehension (when controlling for their vocabulary), and stronger vocabulary knowledge (when controlling for their reading comprehension) were more likely to attempt regulating their comprehension. In general, the difference between the control words and the inconsistent words was smaller for third graders and larger for fourth and fifth graders, which we argue indicates greater levels of comprehension monitoring—specifically employing repair strategies. With eye-tracking technology becoming more accessible, these tasks may be useful in assessing children's reading processes to better understand at which level of comprehension monitoring they may be struggling, which in return will allow us to develop more individualized instruction for all readers.

✉ Elham Zargar
elham.zargar@uci.edu

Extended author information available on the last page of the article

🕿 Springer

Reading for understanding goes beyond the fundamental aspects of reading, such as word decoding and reading fluency (Ehri, 2014). According to a cognitive view, the processes underlying reading comprehension may be structured into two categories of lower level and higher level processes (Kendeou, van den Broek, Helder, & Karlsson, 2014). As suggested by this view, the lower level processes consist of creating meaningful units of language from written codes, which depend on word decoding (Ehri, 2014), reading fluency (Fuchs, Fuchs, Hosp, & Jenkins, 2001), and vocabulary knowledge (Nagy, Herman, & Anderson, 1985; Quinn, Wagner, Petscher, & Lopez, 2015). The higher level processes (component skills of comprehension) consist of inference making, executive functioning and attention-allocation abilities, such as comprehension monitoring (Cain, Oakhill, Barnes, & Bryant, 2001; Connor et al., 2015; Oakhill, Hartt, & Samols, 2005). These skills enable the reader to focus on relevant aspects of the text, while making simple inferences, drawing conclusions, and making judgments and connecting parts of the text. Both lower and higher level processes develop early and before formal instruction of reading, as these same skills are needed for oral language comprehension (Kim, 2017; Storch & Whitehurst, 2002). However, research findings indicate that these skills may predict reading comprehension abilities at a later age, independently (Kendeou, van den Broek, White, & Lynch, 2009). This suggests the importance of studying the different aspects of comprehension processes, to better understand why students might succeed or fail at reading comprehension.

Previous studies have found that many students struggling with reading comprehension have ineffective comprehension monitoring skills, the ability to evaluate and regulate one's own comprehension (e.g., Cain, Oakhill, & Bryant, 2004; Rapp & van den Broek, 2005). However, assessing this construct using written assessments in young children can be difficult due to the metacognitive nature of comprehension monitoring, and because this skill may not be fully conscious or completely developed in children (Kinnunen & Vauras, 2010; Rayner, Chace, Slattery, & Ashby, 2006). Moreover, using traditional methods to measure this skill, such as asking students to read and to think aloud, may interfere with automatic and natural reading and comprehension monitoring (Kinnunen & Vauras, 2010). Therefore, eye-movement methods allow for examining students' comprehension monitoring, without interfering with this partially unconscious process. To this end, using eye-movement technology, this study aims to examine students' variability in comprehension monitoring and how this skill may be related to reading comprehension and vocabulary knowledge during middle childhood as metacognition and other higher order processes are becoming more developed (Del Giudice, 2014).

Comprehension monitoring

Comprehension monitoring is generally strongly related to effective reading comprehension and is found to explain unique variance in reading comprehension ability, when controlling for word reading and vocabulary skills (Cain et al., 2004). Although cognitive and educational psychologists may refer to the process of comprehension monitoring using different terms such as meta-memory of text,

calibration of comprehension, meta-comprehension, or self-regulated comprehension (Hacker, 1998), this metacognitive skill is commonly viewed as processing that involves evaluation and regulation of comprehension while reading. The ongoing evaluation process informs the reader whether or not comprehension is occurring when reading. Once an inconsistency or misunderstanding is noticed, the reader may take steps to resolve the problem to establish consistency and regulate comprehension. More recently, researchers have similarly defined comprehension monitoring as the conscious and unconscious strategies used to (1) evaluate comprehension and identify inconsistencies that might occur during text reading and (2) regulate comprehension or repair the misunderstandings and facilitate reading (Cain et al., 2004; Connor et al., 2015). Therefore, it is crucial not to refer to comprehension monitoring as a unitary skill at which a reader is either effective or ineffective. A student may be capable of evaluating his/her comprehension, but he/she may be poor at regulating it (Baker, 1984).

Inconsistencies in text that may lead to comprehension breakdowns and necessitate the reader to employ repair strategies to regulate comprehension may be caused by different types of obstacles, such as scrambled or contradictory sentences, or information that conflict with the reader's existing knowledge such as unfamiliar words (Cain et al., 2004). Therefore, comprehension monitoring may take place at different levels of the linguistic structure including the word- and sentence-level (Nagy, 2007). For example, comprehension monitoring at the word-level is provoked when the reader becomes aware of a breakdown in comprehension after encountering an unfamiliar word, and at the sentence-level when the reader is confronted with implausibility in the global context of the text, or when the structure of the sentence is not fully understood. Baker (1984, 1985) suggested that the different criteria utilized for comprehension can be categorized into three types, each operating at different levels of text processing. These include lexical standards for monitoring individual words, syntactic standards for monitoring the syntax, and semantic standards that are used for monitoring the overall semantic representations, logical consistency, and construction of meaning of the text. Semantic standards are utilized for both external inconsistencies, where prior knowledge is violated, or when confronted with internal inconsistencies, where the text is inconsistent or presents contradictory information (Kinnunen & Vauras, 2010). These standards demand different cognitive processes and are likely different in their ease of application. Therefore, it is important to distinguish between them and not overgeneralize failure to use one standard as having ineffective comprehension monitoring.

Since comprehension monitoring is a metacognitive task, and therefore developmentally sensitive, this skill may differ for students of different ages (Gombert, 1992; Connor et al., 2015; Kinnunen & Vauras, 2010; Oakhill & Cain, 2012). Thus, we aimed to investigate how age might play a role in individual differences in comprehension monitoring. The current study uses two recently developed eye-tracking tasks to examine comprehension monitoring in third through fifth grade students when they are confronted with either a word- or sentence-level inconsistency, and how this metacognitive skill relates to reading comprehension ability and vocabulary knowledge as measured by standardized assessments.

Eye-movement methodology

There is a close relation between the amount of time one spends viewing a linguistic unit and the mental effort needed to process it while reading. Therefore, eye-movement methods have been widely used to examine moment-to-moment information processing in reading (Rayner, 1998). Newer eye-movement methods are used to examine comprehension monitoring without further taxing young children's reduced metacognitive skills, as they provide precise analyses at the word-level, while permitting children to move their heads freely (Garrett, Mazzocco, & Baker, 2006; Kinnunen & Vauras, 2010). There are three oculomotor measures, reflecting different stages of word processing, that have been consistently used in reading research. These are *initial fixation duration*, *gaze duration*, and *rereading time* for a target word. Initial fixation duration is when the eye first views the word and it represents orthographic and pre-lexical or early lexical processing. Gaze duration is the summed duration of all eye fixations before the eye leaves the word and it reflects later stages of word reading including lexical access. Rereading time is the summed duration of all fixations made after leaving the word for the first time or the time spent rereading previously attended words. Rereading time is known to be reflective of the post-lexical integration of meaning at the sentence level (Radach & Kennedy, 2004; 2013).

Gaze duration and rereading time are found to be sensitive to inconsistencies in text, such that a more proficient reader spends a longer time viewing and rereading an inconsistency (Connor et al., 2015). Therefore, the two aspects of comprehension monitoring, detecting an inconsistency (or comprehension evaluation) and attempting to repair comprehension breakdowns (comprehension regulation), are examined by measuring the amount of time one spends reading (gaze duration) and rereading a target inconsistent word, respectively. A target inconsistent word purposely causes an error or inconsistency in the text and provokes the need for monitoring at different levels of text processing. For example, a target word may cause a word-level inconsistency when it contradicts with the readers' vocabulary knowledge, or a sentence-level inconsistency when it provides logically inconsistent, contextually implausible, or contradictory information (Kinnunen & Vauras, 1995, 2010). Thus, a longer gaze duration and rereading time for target inconsistent words compared to control words might be diagnostic of the two aspects of comprehension monitoring (e.g., Connor et al., 2015). A longer gaze duration for the target word indicates that the reader detects an error and slows down, whereas a greater rereading time suggests an attempt to resolve the inconsistency and regulate comprehension (e.g., by using repair strategies), which goes beyond simply detecting them.

To measure word-level comprehension monitoring in this study, we developed the Word vs. Non-word task, where a non-word (or pseudoword) is embedded within an otherwise simple English sentence. For measuring sentence-level monitoring, we utilized the Plausible vs. Implausible task previously developed by Connor and colleagues (2015), where a contextually implausible word is embedded in the second sentence of a two-sentence passage. By measuring gaze duration and rereading time for the target words, this study examines the students' ability to evaluate

and regulate their comprehension, when confronted with word- and sentence-level inconsistencies.

Vocabulary knowledge

Language and cognitive skills and processes are organized in a hierarchical structure, such that lower level skills such as vocabulary knowledge contribute to higher order processes of reading comprehension, including comprehension monitoring (Kim, 2016). In fact, research shows that vocabulary knowledge is found to be highly associated with reading comprehension ability (e.g., Carroll, 1993; Oakhill & Cain, 2012), and may explain variance in reading comprehension beyond traditional predictors of reading comprehension such as word recognition and listening comprehension (Tunmer & Chapman, 2012).

Researchers have also examined how vocabulary knowledge may be associated with higher level processes. For example, previous studies have investigated the relation between vocabulary knowledge and inference making (e.g., Cain & Oakhill, 2014), which has been found to be bidirectional (Oakhill, Cain, & McCarthy, 2015). However, the relation between vocabulary knowledge and comprehension monitoring might be different. In a longitudinal study examining the precursors of reading ability, Cain and Oakhill (2012) found that comprehension monitoring in third grade predicted vocabulary knowledge in sixth grade, whereas vocabulary knowledge did not predict comprehension monitoring. Moreover, Oakhill and Cain (2012) found that vocabulary knowledge and comprehension monitoring (at the sentence-level) may not be associated for seven to eight-year-old children, whereas these two skills were found to be correlated when the children were eight to nine and ten to eleven years of age. Although this study did not differentiate between the two aspects of comprehension monitoring (evaluation and regulation) and used a written assessment to measure this skill; nonetheless, it supports the link between vocabulary knowledge and comprehension monitoring. To the best of our knowledge, the nature of the relations between vocabulary knowledge and comprehension monitoring has not been extensively investigated in younger students using eye-movement techniques. To this end, one of our aims in this study is to examine how vocabulary knowledge may be associated with individual differences in comprehension monitoring, above and beyond reading comprehension abilities.

Individual differences in comprehension monitoring and reading achievement

A study examining online comprehension monitoring in beginning readers found that students as early as in second grade, were sensitive to sentence-level inconsistencies (i.e., an implausible word in a sentence), such that they spent longer fixating on and rereading the inconsistent compared to consistent words (Kim, Vorstius, & Radach, 2018). Comprehension monitoring may be developmentally sensitive, such that students of different ages may differ in their ability to evaluate and regulate their

comprehension, perhaps due to later development of metacognition and metalinguistic awareness (Gombert, 1992; Kinnunen & Vauras, 2010). A recent longitudinal study found that both aspects of comprehension monitoring improved in the span of eight months for fifth grade students, but only for those with stronger literacy and academic language skills (Connor et al., 2015). Moreover, it is proposed that the evaluation aspect of comprehension may improve with age due to the development of information processing abilities (Baker, 1984).

Other individual differences are also found to be related to comprehension monitoring performance, and uniquely to each aspect of this skill. Regarding comprehension evaluation, previous studies have suggested contradictory claims. Rubman and Waters (2000) discussed that detecting sentence-level inconsistencies requires the ability to construct a coherent representation of the text. Similarly, Ehrlich (1996) discovered that less skilled comprehenders were found to be weaker in self-evaluating their comprehension and detecting inconsistencies in text, as they overestimated their understanding due to being unaware of their deficiencies. However, contradictory to these claims, Cain and colleagues (2004) discussed that comprehension evaluation or error detection can be done through comparison of statements and may not require the reader to engage in higher-level processes of comprehension, and therefore may not be dependent on the reader's literacy skills. Similarly, Connor and colleagues (2015) found that, on average, all the fifth-grade students in their study reacted to sentence-level inconsistencies in text by slowing down their reading regardless of their literacy or academic language skills. In other words, the findings of this study suggested that students with stronger literacy skills were not any more likely to identify inconsistencies and slow down their reading, compared to those with weaker literacy skills. The results of these studies indicate that detecting inconsistencies in text is likely to be automatic and independent of the reader's higher-level literacy skills, especially for older students. For this reason, this study aims to further investigate how students' reading comprehension ability and vocabulary knowledge may play a role in detecting word-level and sentence-level inconsistencies in text while considering students' grade level.

After detection of an error, regulating comprehension requires knowledge of repair strategies. According to previous research, comprehension regulation is suggested to be dependent on the readers' individual differences in language achievement. In a study with fifth grade students, Connor and colleagues (2015), elucidated that students' academic language predicts the likelihood that they will repair their misunderstandings. In other words, those students with stronger academic language were found to spend more time rereading and attempting to repair breakdowns in comprehension. However, the same study suggested that students' literacy (a construct of reading comprehension, vocabulary, and word reading) did not predict greater rereading time indicating that these skill sets may not be related to individual differences in comprehension monitoring. In contrast, other researchers have found that comprehension monitoring may explain unique variance in reading comprehension ability beyond student' word reading and vocabulary skills (Cain et al., 2004). Therefore, further research is necessary to investigate how comprehension regulation may be associated with reading comprehension and vocabulary knowledge.

Current study

The current study examines the relation between comprehension evaluation and regulation and students' reading comprehension and vocabulary knowledge during middle childhood—third, fourth and fifth grade. We studied how students' individual differences in reading comprehension and vocabulary knowledge may be related to their comprehension monitoring skills, when confronted with either word- or sentence-level inconsistencies in text, during this crucial time in reading comprehension development. During middle childhood, children have generally achieved reasonably fluent word-reading skills and have the cognitive resources available to focus on comprehending (Perfetti, 1985; Perfetti & Lesgold, 1979). Using eye-tracking technology, students' gaze duration and rereading time for the target words were measured to examine their comprehension evaluation and regulation, respectively. Furthermore, since comprehension monitoring processes are found to be developmentally sensitive, this study examined how students in third, fourth, and fifth grade may differ in their abilities to evaluate and regulate their comprehension.

The following are the specific research questions guiding the current study: (1) Are students sensitive to the word- and sentence-level inconsistencies, such that they spend a longer time reading and rereading the target inconsistent words compared to the control words? (2) Is students' reading comprehension ability associated with comprehension evaluation or regulation when confronted with word- or sentence-level inconsistencies? (3) Is students' vocabulary knowledge associated with comprehension evaluation or regulation when confronted with word- or sentence-level inconsistencies? (4) When controlling for students' reading comprehension ability and vocabulary knowledge, do students in third through fifth grade differ in their comprehension monitoring skills? We conjectured that all students would be generally sensitive to the target inconsistent words and would view these words longer than control words (i.e., have longer gaze durations). However, we hypothesized that students who either (a) had stronger reading comprehension ability or vocabulary knowledge and/or (b) were older would be more likely to regulate their comprehension and thus have longer rereading times for the target inconsistent words.

Methods

Participants

Of the potential 129 children eligible for this study, three parents declined to provide consent for their child to participate. Of the remaining students for whom we had consent, two left the district, and one was unable to complete the tasks primarily due to weak decoding skills. This provided a total sample of 123 students. The 123 students (M age = 9.80 years, SD = 0.9, range = 8.17–12.17 years) who

participated in this study attended a charter elementary school located in Arizona. The participants were from two third-grade classrooms (n = 48), two fourth-grade classrooms (n = 42), and two fifth-grade classes (n = 33). This sample consisted of 58% White, 15% African American, and 15% Hispanic students, with the remaining belonging to other ethnicities. Fifty-nine percent of the participants were girls. Overall, students scored at approximately the 44th percentile on the standardized reading comprehension, and at the 52nd percentile on the vocabulary knowledge assessment.

Measures and procedures

Students were assessed on their comprehension monitoring and reading achievement in winter of the 2015–2016 academic year. The procedures for each measure are described below.

Comprehension monitoring assessment

Comprehension evaluation and regulation were examined using eye-movement tasks by measuring gaze duration and rereading time for target words, respectively.

Procedures Students were assessed using the eye-movement tasks individually. Before starting the tasks, students were instructed to read the sentences on the computer screen for understanding and answer the occasional reading comprehension questions to the best of their ability. The experimenter explained that the tasks were self-paced and students were to proceed to the next item by making a mouse click. Participants were also asked to keep their physical movements to a minimum to assure that eye-tracker recorded their eye-movements. The participants were seated 66–72 cm from the screen. Before starting students performed the calibration process, which required following a moving dot on the display monitor with their eyes. Throughout the assessment, the experimenter assured that each child was engaged and carefully reading; otherwise the child was encouraged to continue as instructed.

Eye-movement tasks materials The eye movement trials consisted of 2 different tasks—20 items each—for a total of 40 items. The first task, Word vs. Non-word, is researcher-developed and aims to assess comprehension monitoring at the word-level. The second task, Plausible vs. Implausible, is a replication from a previous eye-movement study conducted with fifth grade students (Connor et al., 2015) to assess sentence-level monitoring. Two alternate "mirror image" forms of the tasks were developed, such that the corresponding control item of a target inconsistent word appeared in the other form. Both forms consisted of 10 control and 10 target items from each task. Sixty-seven participants were randomly assessed using Form A and the remaining 56 were given Form B. The list of stimuli is in the "Appendix".

Word vs. Non-word task This task was composed of 20 simple sentences (18 declarative, one exclamatory, and one interrogative). The control version of the

sentence (or the *word* version of the sentence) made a simple statement (e.g., Rosita climbed the *mountain* in the morning), including a target word expected to be common knowledge for students. A foil version of the sentence contained a *non-word* in the target position (e.g., Rosita climbed the *floggorn* in the morning). Words and non-words were matched exactly on number of letters ($M = 5.95$ letters, *range* = 3–9 letters) and morphemes ($M = 1.25$ morphemes, range = 1–2 morphemes). When developing the non-words, we assured that the letter patterns and the orthographic structures of letter-strings in the non-words were not different than the control words, as this could influence the students' word processing. This was done by examining the bigram frequency for each target word, the frequency with which adjacent pairs of letters (bigrams) occur in text (Rice & Robinson, 1975). The mean bigram frequency of words and non-words were calculated using WordGen software and were not significantly different $t(38) = 1.72$, $p = 0.093$ (Duyck, Desmet, Verbeke, & Brysbaert, 2004). In 10 sentences target words served as nouns, in 4 sentences they were adjectives, in 5 sentences they were verbs, and in one sentence it was an interjection (i.e., *zop!*). Items were counterbalanced in terms of position within the sentence. The purpose of this task was to control for any prior vocabulary knowledge. Regardless of their literacy skills, the non-words acted as unfamiliar words for all participants, causing a word-level inconsistency.

Plausible vs. Implausible task This task is composed of 20 pairs of simple declarative sentences, with each stimulus consisting of 2 sentences. The first sentence sets up a scenario involving an event or action (e.g., Every day Rover barked at the passing animals on the street.), which is then explained further in the second sentence. The second sentence of each item contained either a plausible (e.g., He was the most alert *puppy* in the neighborhood) or an implausible target word (e.g., He was the most alert *kitten* in the neighborhood). This task was originally developed and used in a previous study, where all stimuli were matched on word length, number of syllables, and morphological complexity (see Connor et al., 2015 for further details).

Check-point questions Throughout the eye-movement assessment, 20 items were randomly followed by a short and simple multiple-choice comprehension question. For example, the item "My cat is big and orange", was followed by the question "What color is my cat?" with four different choices. The purpose of this section was to assure that students were reading for understanding and not clicking through the assessment mindlessly. These were inserted in between stimuli in a random order for the two forms.

Apparatus Text stimuli were displayed one at a time written in black text and Courier New 28-point font, on a light gray background, using a 17-in Tobii T-120 eye-tracker monitor with a display resolution of 1280×768 pixels and a data rate of 120 Hz. Prior to starting the task, a 5-point calibration was performed at a medium calibration

speed. The distance between the participant's eyes and the monitor was 66–72 cm, while the eye-tracker recorded movements from both eyes, with the average binocular tracking enabled. No chin or forehead rest was used and the apparatus enabled participants to behave naturally with freedom of head movements throughout the trials. Data was recoded and analyzed using Tobii Studio software version 3.3.2.1150, and the default Tobii I-VT fixation filter, with a window length of 20 ms, a velocity threshold of 30 degrees/second, a minimum fixation duration of 60 ms (to discard fixations shorter than 60 ms), and a maximum time between fixations of 75 ms with a 0.5 maximum angle between fixations (to merge adjacent fixations).

Reading comprehension and vocabulary knowledge

Students' reading comprehension and vocabulary knowledge were assessed using the Gates-MacGinitie Reading Test, Fourth Edition (MacGinitie, MacGinitie, Maria, & Dreyer, 2002). The grade-appropriate assessments were group administered in the students' classrooms. The reading comprehension subtest requires students to read various passages and to answer multiple choice comprehension questions pertaining to the passage. The reading vocabulary subtest includes sentences with target words and students are asked to identify the correct definition of the target word among four choices. For our analyses, we used the Extended Scaled Scores (ESS), with a mean of 500 and a standard deviation of 15. The published reliability for this test is 0.96.

Analytic strategy

To answer our first research question, students' sensitivity to word- and sentence-level inconsistencies, we conducted t-tests. For the remaining research questions, due to the nesting of item within student, Hierarchical Linear Modeling (HLM) was used to control for the nested nature of the data. This method allowed for correctly estimating the standard errors by considering the shared variance among the items nested in students. A separate set of HLM models were utilized to examine the relation between each aspect of comprehension monitoring and reading comprehension and vocabulary. The 2-level HLM models consisted of item-level measures for each of the tasks including item type, and either gaze duration or rereading time (in milliseconds) as the outcome variable in the level-1 data. Child-level variables, reading comprehension, vocabulary scores, and grade level (as a measure of students' age) were entered at level 2. For all models, residuals were tested to be normally distributed with means of zero. All models were analyzed with restricted maximum likelihood and fixed effects for each model are reported with robust standard errors.

Reading comprehension models

Y_{mj} represents the outcome variable, either gaze duration or rereading time (milliseconds) for item m for child j, as a function of grade level, comprehension score, vocabulary knowledge score, item type ($0 =$ control, $1 =$ target word), and the interaction between reading comprehension and item type (γ_{11}), as well as the residual (see Eq. 1). Therefore, the coefficient γ_{11}, or the interaction term of reading comprehension and item type will answer our research question as to how reading comprehension is associated with the amount of time spent reading or rereading the target words in each task, when controlling for vocabulary and grade level.

$$Y_{mj} = \gamma_{00} + \gamma_{01} * (Grade\ level)_j + \gamma_{02} * (Comprehension)_j + \gamma_{03} * (Vocabulary)_j$$
$$+ \gamma_{10} * (item\ type)_{mj} + \gamma_{11}(Comprehension)_j(Item\ type)_{mj} + u_{0j} + e_{mj}$$
$$(1)$$

Vocabulary knowledge models

Similar to the previous model, the coefficient γ_{11}, or the interaction term of vocabulary knowledge and item type will answer our research question as to how vocabulary is associated with the amount of time spent reading or rereading the target words, when controlling for comprehension and grade level (see Eq. 2).

$$Y_{mj} = \gamma_{00} + \gamma_{01} * (Grade\ level)_j + \gamma_{02} * (Comprehension)_j + \gamma_{03} * (Vocabulary)_j$$
$$+ \gamma_{10} * (item\ type)_{mj} + \gamma_{11}(Vocabulary)_j(Item\ type)_{mj} + u_{0j} + e_{mj}$$
$$(2)$$

Grade level models

To examine how students in third through fifth grade may differ in their comprehension monitoring skills, after controlling for their vocabulary and reading comprehension, a dummy variable for each of the 3 grade levels was made and entered at the child level. Fourth grade was first left out to be the reference group for comparison. Here, the coefficient γ_{11} will tell us how different third graders may be in processing the target words compared to fourth graders, and γ_{12} will allow us to compare fifth graders to fourth graders (see Eq. 3). In order to further examine grade level differences, third grade was then set to be the reference group in another HLM model.

$$Y_{mj} = \gamma_{00} + \gamma_{01} * (Comprehension)_j + \gamma_{02} * (Vocabulary)_j + \gamma_{03} * (Third\ grade)_j$$
$$+ \gamma_{04} * (Fifth\ grade)_j + \gamma_{10} * (Item\ type)_{mj} + \gamma_{11} * (Third\ grade)_j * (Item\ type)_{mj}$$
$$+ \gamma_{12} * (Fifth\ grade)_j * (Item\ type)_{mj} + u_{0j} + e_{mj}$$
$$(3)$$

Data analysis

For each of the two tasks, a total of 2460 data points for gaze duration and rereading time was collected from the participants. If the student did not view a target word or the eye-movement behavior was not recorded, the gaze duration or rereading time was registered as zero milliseconds. Following the methodology used by Kim,

Table 1 Descriptive statistics for gaze duration and rereading time (in milliseconds) for each of the eye-movement tasks, and reading achievement assessment scores

Task	N	Mean	SD	Range	Skewness
Gaze duration (ms)					
Word vs. Non-word task					
Word	1225	220.77	260.147	0–1949	2.662
Non-word	1198	328.76	367.824	0–1999	1.818
Total	2423	274.17	322.464	0–1999	2.202
Plausible vs. Implausible task					
Plausible	1222	197.45	247.092	0–1999	2.494
Implausible	1225	201.29	241.214	0–1849	2.302
Total	2447	199.37	244.124	0–1999	2.400
Rereading time (ms)					
Word vs. Non-word task					
Word	1218	117.72	282.022	0–1799	3.147
Non-word	1173	297.07	466.425	0–1999	1.615
Total	2391	205.71	393.986	0–1999	2.204
Plausible vs. Implausible task					
Plausible	1223	61.76	172.225	0–1682	4.103
Implausible	1220	95.11	247.879	0–1982	3.694
Total	2443	78.41	213.992	0–1982	4.015
Assessment	N	Mean	SD	Range	
Reading achievement					
Reading comprehension	123	475.22	46.207	320.00–596.00	− 0.217
Reading vocabulary	123	484.07	41.254	354.00–593.00	− 0.521

Vorstius, and Radach (2018), gaze durations or rereading times longer than 2000 ms were considered outliers and were excluded from data analyses. For the Word vs. Non-word task, approximately 1.50% of the gaze duration data, and approximately 2.80% rereading time data were excluded. For the Plausible vs. Implausible task, approximately 0.53% of the gaze duration and 0.69% of the rereading time data were excluded. To this end, for the Word vs. Non-word task, analyses for gaze duration was based a total of 2423 cases and 2391 cases for rereading time. For the Plausible vs. Implausible task, analyses for gaze duration was based on a total of 2447 cases and 2443 cases for rereading time. See Table 1 for descriptive statistics.

Results

The associations between comprehension evaluation, comprehension regulation, reading comprehension, and vocabulary knowledge were analyzed using HLM. Descriptive statistics are provided in Table 1, and correlations between gaze duration and rereading time in each of the two eye-movement tasks are provided in

Table 2 Correlations Between Gaze Duration and Rereading Time for Each Eye-Movement Task

Variables	1	2	3	4	5	6	7	8
1. Gaze duration for control words	–							
2. Gaze duration for non-words	.183**	–						
3. Rereading time for control words	.144**	.063*	–					
4. Rereading time for non-words	.062*	.144**	.097**	–				
5. Gaze duration for plausible words	.164**	.183**	.099**	.077*	–			
6. Gaze duration for implausible words	.207**	.185**	0.042	0.057	.216**	–		
7. Rereading time for plausible words	.115**	.078**	0.051	.081**	.220**	.101**	–	
8. Rereading time for implausible words	.112**	.103**	0.053	.159**	.165**	.178**	.137**	–

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Table 2. Students' reading comprehension and vocabulary knowledge scores were found to be significantly correlated ($r = .75$, $p < .001$). Students in our sample demonstrated a wide range of reading skills. Students achieved Gates-MacGinitie Reading Comprehension and Reading Vocabulary percentile rank scores between the 1st and 99th (with the expected mean of 50th percentile). On average, students scored at approximately the 44th percentile on the reading comprehension subtest, and at the 52nd percentile on the vocabulary assessment.

Due to the complexity of this data, we would like to reiterate our hypotheses before going into the results. We expected main effects of gaze duration and rereading time such that readers read and reread non-words and implausible words longer than the control words, respectively. We also expected to find that students with stronger reading comprehension and vocabulary would reread inconsistent target words in both tasks longer compared to their peers with lower comprehension and vocabulary skills. Finally, we expected to see that older students reread inconsistent target words longer than younger students.

Research question 1: sensitivity to word- and sentence-level inconsistencies

The analysis for the Word vs. Non-word task demonstrated that on average, both gaze duration and rereading time for non-words were significantly longer compared to the control words, as hypothesized. Gaze duration for non-words was 107.99 ms longer; t (2421) = 8.36, $p < .001$, and rereading time was 179.35 ms longer, t (2389) = 11.43, $p < .001$, compared to that of the control words (see Fig. 1 top). Contrary to our expectations, for the Plausible vs. Implausible task, gaze duration was not significantly different for implausible words compared to plausible words, t (2445) = 0.39, $p = .520$, whereas, rereading time for implausible words was significantly longer by 33.35 ms as expected, t (2441) = 3.86, $p < .001$, compared to that of control plausible words, on average (see Fig. 1 bottom).
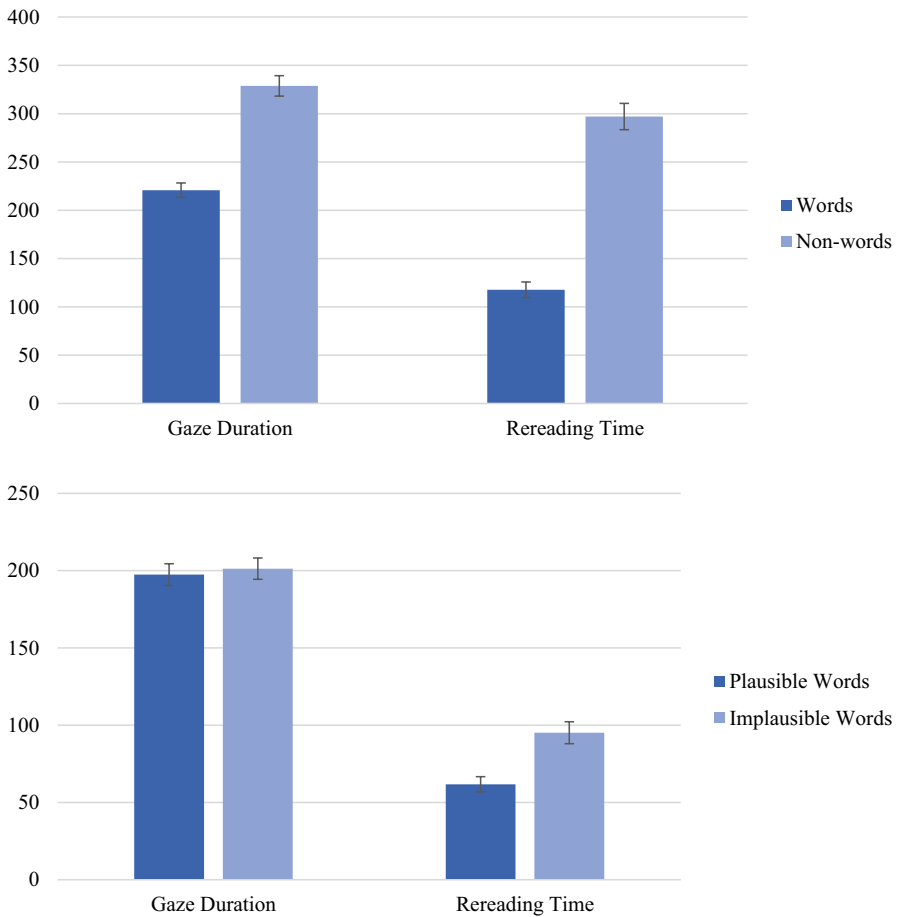
Fig. 1 Fitted means for gaze duration and rereading time by task and item type, controlling for grade level, reading comprehension, and vocabulary. In general, in the Word vs. Non-word Task, gaze duration and rereading time for target non-words were significantly longer than the control words (see top figure), whereas in the Plausible vs. Implausible Task, only rereading time for implausible words were significantly longer than that of the plausible words (see bottom figure). Error bars represent approximate standard errors

Research question 2: reading comprehension models

The nature of the relation between reading comprehension ability and comprehension evaluation and regulation, when confronted with a word- or sentence- level inconsistency, controlling for students' vocabulary knowledge and grade level was examined. We present the results by task.

Word vs. Non-word task

The analysis for this task showed no significant interaction between reading comprehension and gaze duration of non-words, $\gamma = .18$, $p = .50$, such that students did not differ in their gaze duration for non-words according to their reading comprehension skills, as expected and in line with the findings of Connor et al. (2015). However, as we expected, the results showed a significant interaction between reading comprehension and rereading time for non-words, $\gamma = .92$, $p = .02$ (see Tables 3 and 4). This demonstrates that students with stronger reading comprehension skills spent more time rereading the non-words compared to their peers with lower reading comprehension skills (See Fig. 2 top). There was no significant association between reading comprehension ability and reading or rereading of the control words, $\gamma = -0.17$, $p = .67$ and $\gamma = .10$, $p = .77$, respectively.

Plausible vs. Implausible task

Analysis for students' gaze duration for this task revealed that students did not spend significantly different amounts of time processing the implausible words, compared to the plausible words, contrary to our expectation, $\gamma = .53$, $p = .13$ (see Table 3). However, as we expected, analysis for rereading time revealed a significant relation between reading comprehension and rereading the implausible words, $\gamma = .57$, $p = .001$, such that stronger comprehenders generally spent more time rereading the implausible words, compared to their peers (see Table 4 and Fig. 2 bottom). There was no significant relation between reading comprehension and reading or rereading the control plausible words, $\gamma = -0.45$, $p = .21$, and $\gamma = .02$, $p = .91$, respectively (see Tables 3 and 4).

Research question 3: vocabulary models

Word vs. Non-word task

There was a significant interaction between vocabulary knowledge and both gaze duration and rereading time of non-words, $\gamma = .57$, $p = .04$, and $\gamma = 1.28$, $p = .001$ (see Tables 5 and 6 and Figs. 3 and 4 top). This means that students with stronger vocabulary knowledge generally spent more time reading and rereading the non-words, compared to their peers with lower vocabulary. There was no significant relation between vocabulary knowledge and reading or rereading the control words, $\gamma = -0.43$, $p = .39$, and $\gamma = -0.69$, $p = .09$, respectively (see Tables 5 and 6).

Plausible vs. Implausible task

While our results showed that overall students' gaze duration did not differ for plausible and implausible words, there was a significant interaction between vocabulary knowledge and gaze duration for plausible words, $\gamma = -0.97$, $p = .03$,

Table 3 Hierarchical linear modeling results for gaze duration (milliseconds) as a function of students' grade level, reading comprehension and vocabulary, and reading comprehension by item type interaction

Fixed effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p value
Word vs. Non-word task					
For word targets, ψ_0					
Fitted mean, γ_{00}	236.763	66.048	3.585	119	<0.001
Grade level, γ_{01}	−4.113	16.910	−0.243	119	0.808
Reading comprehension, γ_{02}	−0.171	0.406	−0.422	119	0.674
Vocabulary, γ_{03}	−0.149	0.551	−0.270	119	0.788
For Non-word targets, ψ_1					
Item type, γ_{10}	110.090	12.701	8.668	2298	<0.001
RC × item type, γ_{11}	0.178	0.261	0.681	2298	0.496
Plausible vs. Implausible task					
For plausible targets, ψ_0					
Fitted mean, γ_{00}	211.993	50.671	4.184	119	<0.001
Grade level, γ_{01}	−3.657	13.094	−0.279	119	0.780
Reading comprehension, γ_{02}	−0.445	0.356	−1.249	119	0.214
Vocabulary, γ_{03}	−0.501	0.408	−1.227	119	0.222
For Implausible targets, ψ_1					
Item type, γ_{10}	3.668	9.953	0.368	2322	0.713
RC × item type, γ_{11}	0.534	0.354	1.509	2322	0.131

Random effect	Standard deviation	Variance component	$d.f.$	χ^2	p value
Word vs. Non-word task					
Intercept, u_0	130.334	16,986.865	119	584.294	<0.001
Level-1, e	291.220	84,808.928			
Plausible vs. Implausible Task					
Intercept, u_0	94.087	8852.319	119	537.399	<0.001
Level-1, e	223.923	50,141.404			

RC stands for reading comprehension

$Gaze\ Duration_{mj} = \gamma_{00} + \gamma_{01} * (Grade\ level)_j + \gamma_{02} * (Comprehension)_j + \gamma_{03} * (Vocabulary)_j$

$\qquad + \gamma_{10} * (item\ type)_{mj} + \gamma_{11}(Reading\ Comprehension)_j(Item\ type)_{mj} + u_{0j} + e_{mj}$

as well as for implausible words, $\gamma=.93$, $p=.003$ (see Table 5 and Fig. 3 bottom). More specifically, students with stronger vocabulary knowledge were found to have shorter gaze durations for plausible words, and longer gaze durations for implausible words, compared to their peers. Similarly, the results for rereading time revealed that vocabulary knowledge was reversely associated with rereading the plausible words and positively related to rereading the implausible words, $\gamma=-0.50$, $p=.02$, and $\gamma=.53$, $p=.01$, respectively (see Table 6 and Fig. 4 bottom). In other words, as we expected, those with stronger vocabulary knowledge generally had lower rereading time for the plausible words, whereas they spent
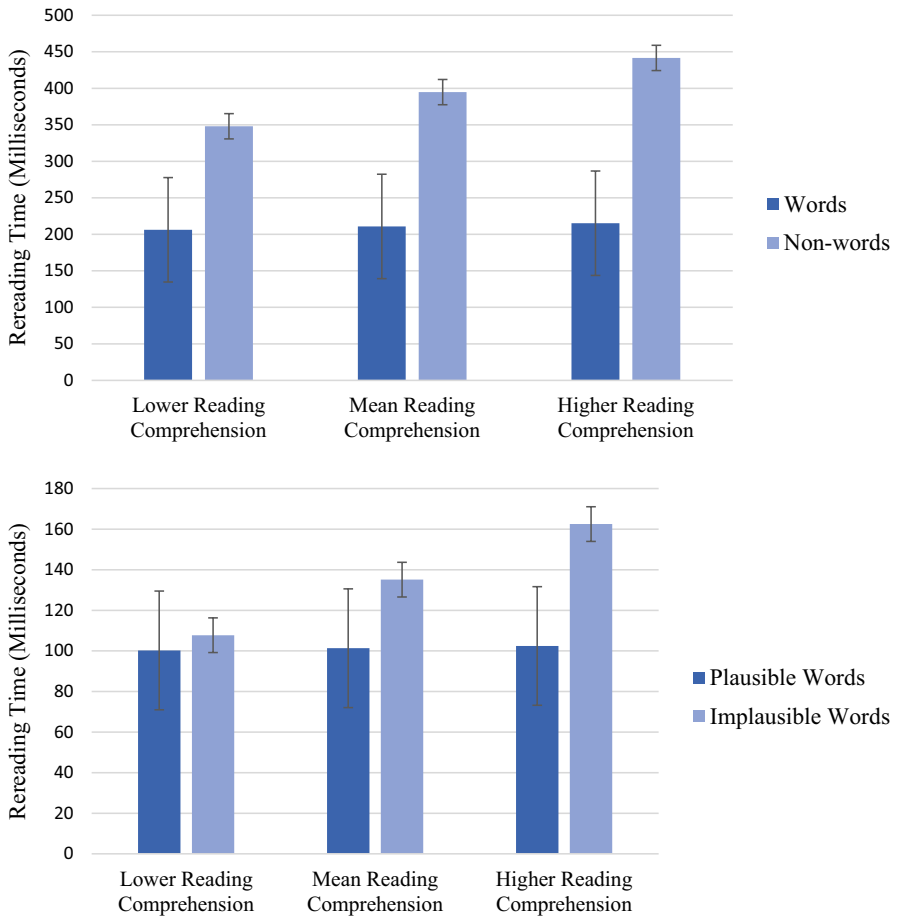
Fig. 2 Rereading time for words vs. non-words (top) and plausible vs. implausible words (bottom) modeled for students whose reading comprehension falls one standard deviation below the mean (lower reading comprehension); at the mean (mean reading comprehension); or one standard deviation above the mean (higher reading comprehension). Error bars represent approximate standard errors (see Table 4 for exact standard errors)

more time rereading the implausible words, when controlling for students' reading comprehension and grade level.

Research question 4: grade level

The nature of relation between students' grade level in school with their ability to evaluate and regulate comprehension, controlling for their reading comprehension and vocabulary, was examined in each eye-movement task. We present the results below.

Table 4 Hierarchical linear modeling results for rereading time (milliseconds) as a function of students' grade level, reading comprehension and vocabulary, and reading comprehension by item type interaction

Fixed effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p value
Word vs. Non-word task					
For word targets, ψ_0					
Fitted mean, γ_{00}	210.721	71.522	2.946	119	0.004
Grade level, γ_{01}	−23.744	18.176	−1.306	119	0.194
Reading comprehension, γ_{02}	0.097	0.331	0.292	119	0.771
Vocabulary, γ_{03}	−0.062	0.422	−0.146	119	0.884
For non-word targets, ψ_1					
Item type, γ_{10}	184.042	17.311	10.632	2266	<0.001
RC×item type, γ_{11}	0.915	0.398	2.299	2266	0.022
Plausible vs. Implausible task					
For Plausible targets, ψ_0					
Fitted mean, γ_{00}	101.346	29.247	3.465	119	<0.001
Grade level, γ_{01}	−10.162	7.049	−1.442	119	0.152
Reading comprehension, γ_{02}	0.024	0.204	0.115	119	0.908
Vocabulary, γ_{03}	−0.236	0.213	−1.108	119	0.270
For implausible targets, ψ_1					
Item type, γ_{10}	33.793	8.544	3.955	2318	<0.001
RC×item type, γ_{11}	0.569	0.174	3.270	2318	0.001

Random effect	Standard deviation	Variance component	d.f.	χ^2	p value
Word vs. Non-word task					
Intercept, u_0	133.563	17,839.055	119	430.312	<0.001
Level-1, e	360.234	129,768.293			
Plausible vs. Implausible task					
Intercept, u_0	58.875	3466.294	119	311.791	<0.001
Level-1, e	205.009	42,028.487			

RC stands for reading comprehension

$$Rereading\ Time_{mj} = \gamma_{00} + \gamma_{01} * (Grade\ level)_j + \gamma_{02} * (Comprehension)_j + \gamma_{03} * (Vocabulary)_j$$
$$+\gamma_{10} * (item\ type)_{mj} + \gamma_{11}(Reading\ Comprehension)_j(Item\ type)_{mj} + u_{0j} + e_{mj}$$

Word vs. Non-word task

There was no significant interaction between students' grade level and gaze duration for non-words vs. real words, suggesting that the third through fifth grade students did not significantly differ in the extent to which they viewed the non-words compared to the real words. In all cases, gaze duration was longer for non-words than words.

When we examined differences in rereading time for words vs. non-words, we found a significant effect of grade (see Table 7 and Fig. 5, top). Overall, the

Table 5 Hierarchical linear modeling results for gaze duration (milliseconds) as a function of students' grade level, reading comprehension and vocabulary, and vocabulary by item type interaction

Fixed effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p value
Word vs. Non-word task					
For Word targets, ψ_0					
Fitted mean, γ_{00}	236.738	66.047	3.584	119	<0.001
Grade level, γ_{01}	−4.100	16.910	−0.242	119	0.809
Reading comprehension, γ_{02}	−0.086	0.423	−0.203	119	0.840
Vocabulary, γ_{03}	−0.427	0.495	−0.864	119	0.390
For Non-word targets, ψ_1					
Item type, γ_{10}	109.899	12.547	8.759	2298	<0.001
Vocabulary × item type, γ_{11}	0.569	0.281	2.026	2298	0.043
Plausible vs. Implausible task					
For plausible targets, ψ_0					
Fitted mean, γ_{00}	212.011	50.754	4.177	119	<0.001
Grade level, γ_{01}	−3.631	13.092	−0.277	119	0.782
Reading comprehension, γ_{02}	−0.178	0.314	−0.568	119	0.571
Vocabulary, γ_{03}	−0.965	0.432	−2.235	119	0.027
For implausible targets, ψ_1					
Item type, γ_{10}	3.475	9.618	0.361	2322	0.718
Vocabulary × item type, γ_{11}	0.926	0.307	3.018	2322	0.003

Random effect	Standard deviation	Variance component	$d.f.$	χ^2	p value
Word vs. Non-word task					
Intercept, u_0	130.344	16,989.585	119	585.035	<0.001
Level-1, e	291.007	84,685.159			
Plausible vs. Implausible task					
Intercept, u_0	94.113	8857.322	119	539.457	<0.001
Level-1, e	223.439	49,925.016			

$$Gaze\,Duration_{mj} = \gamma_{00} + \gamma_{01} * (Grade\,level)_j + \gamma_{02} * (Comprehension)_j + \gamma_{03} * (Vocabulary)_j$$
$$+\gamma_{10} * (item\,type)_{mj} + \gamma_{11}(Vocabulary)_j(Item\,type)_{mj} + u_{0j} + e_{mj}$$

difference between words and non-words was smaller for third graders than for fourth and fifth graders—that is, for third graders, rereading time was longer for words and shorter for non-words than it was for fourth and fifth graders (see Table 7 and Fig. 5 top). Fourth and fifth graders both demonstrated a large and significant difference between rereading times for words and non-words.

Table 6 Hierarchical linear modeling results for rereading time (milliseconds) as a function of students' grade level, reading comprehension and vocabulary, and vocabulary by item type interaction

Fixed effect	Coefficient	Standard error	t-ratio	Approx. $d.f.$	p value
Word vs. Non-word task					
For word targets, ψ_0					
Fitted mean, γ_{00}	210.614	71.628	2.940	119	0.004
Grade level, γ_{01}	− 23.678	18.175	− 1.303	119	0.195
Reading compre-hension, γ_{02}	0.543	0.376	1.443	119	0.152
Vocabulary, γ_{03}	− 0.692	0.402	− 1.720	119	0.088
For non-word targets, ψ_1					
Item type, γ_{10}	183.762	17.050	10.778	2266	<0.001
Vocabulary×item type, γ_{11}	1.282	0.402	3.186	2266	0.001
Plausible vs. Implausible task					
For plausible targets, ψ_0					
Fitted mean, γ_{00}	101.417	29.207	3.472	119	<0.001
Grade level, γ_{01}	− 10.182	7.048	− 1.445	119	0.151
Reading compre-hension, γ_{02}	0.308	0.222	1.386	119	0.168
Vocabulary, γ_{03}	− 0.499	0.213	− 2.344	119	0.021
For implausible targets, ψ_1					
Item type, γ_{10}	33.778	8.646	3.907	2318	<0.001
Vocabulary×item type, γ_{11}	0.527	0.193	2.732	2318	0.006

Random effect	Standard deviation	Variance component	$d.f.$	χ^2	p value
Word vs. Non-word task					
Intercept, u_0	133.482	17,817.532	119	430.613	<0.001
Level-1, e	359.884	129,516.512			
Plausible vs. Implausible task					
Intercept, u_0	58.770	3453.932	119	310.854	<0.001
Level-1, e	205.155	42,088.562			

$$Rereading\ Time_{mj} = \gamma_{00} + \gamma_{01} * (Grade\ level)_j + \gamma_{02} * (Comprehension)_j + \gamma_{03} * (Vocabulary)_j$$
$$+ \gamma_{10} * (item\ type)_{mj} + \gamma_{11}(Vocabulary)_j(Item\ type)_{mj} + u_{0j} + e_{mj}$$

Plausible vs. Implausible task

As with the Word vs. Non-word task, students did not differ by grade in the difference between plausible vs. implausible words for gaze duration, whereas they differed by grade in their rereading times differences for plausible words vs. implausible words (see Table 8 and Fig. 5 bottom). For third graders, there was no significant difference between plausible vs. implausible rereading times. There was a significant difference between plausible and implausible words for both fourth and fifth graders, which was significantly greater than the non-significant
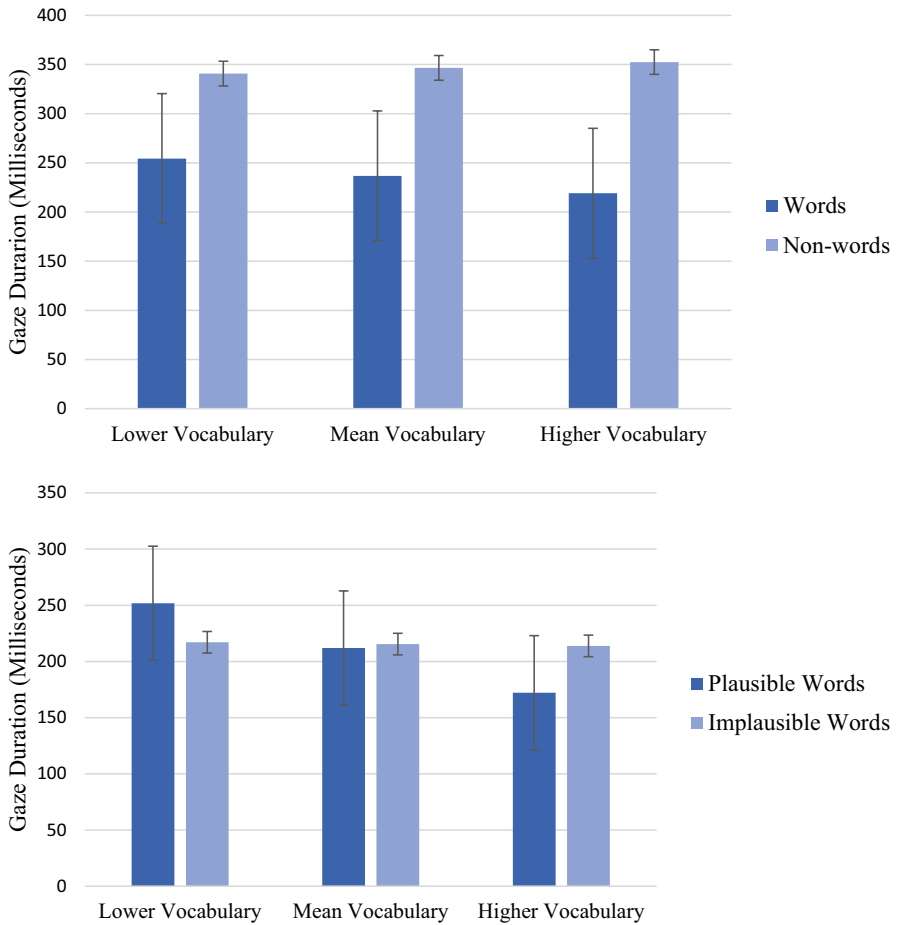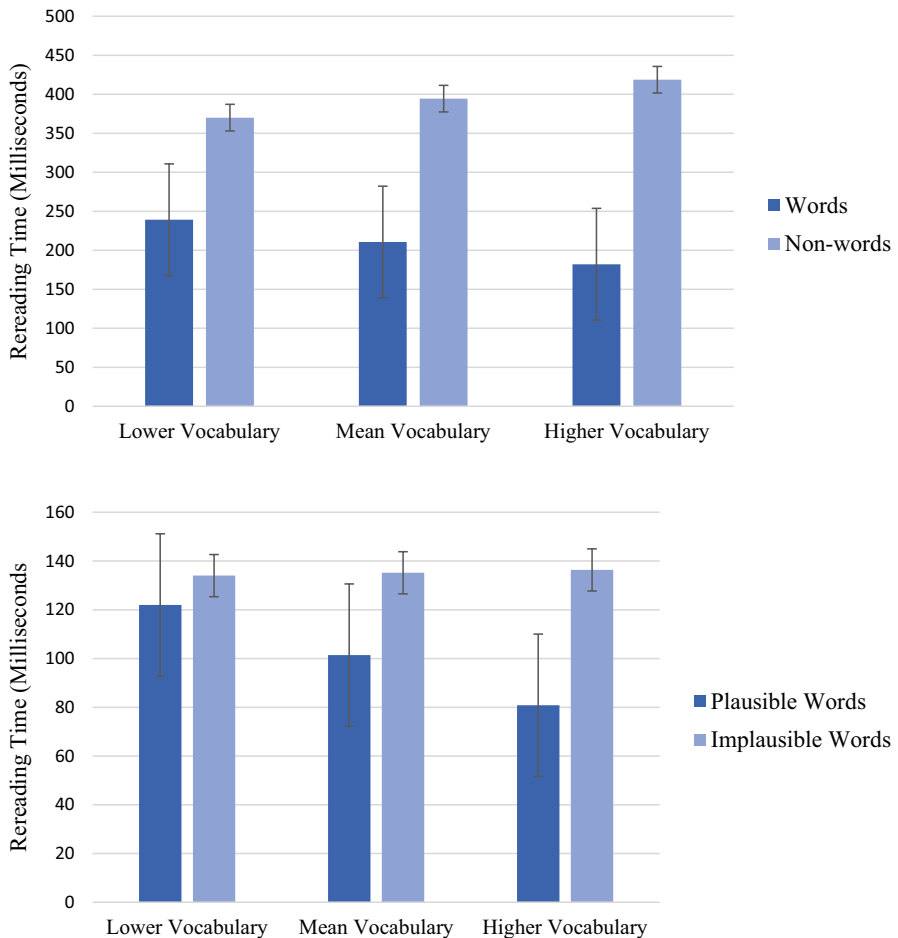
Fig. 3 Gaze duration for words vs. non-words (top) and plausible vs. implausible words (bottom) modeled for students whose vocabulary knowledge falls one standard deviation below the mean (lower vocabulary); at the mean (mean vocabulary); or one standard deviation above the mean (higher vocabulary). Error bars represent approximate standard errors (see Table 5 for exact standard errors)

difference observed for third graders. Fifth graders spent less time rereading plausible words than did fourth graders, creating a larger difference between plausible and implausible rereading times for the oldest children in our sample.

Discussion

This study examined how students in third through fifth grade may differ in their comprehension monitoring skills at the word- and sentence-level, and how these skills may be associated with their general reading comprehension, vocabulary knowledge, and age/grade level. Our hypotheses were supported but only to some

Fig. 4 Rereading time for words vs. non-words (top) and plausible vs. implausible words (bottom) modeled for students whose vocabulary knowledge falls one standard deviation below the mean (lower vocabulary); at the mean (mean vocabulary); or one standard deviation above the mean (higher vocabulary). Error bars represent approximate standard errors (see Table 6 for exact standard errors)

extent. We hypothesized that gaze duration, the amount of time spent reading a word for the first time, would be longer for inconsistent target words, as compared to control words, when students were presented with word-level inconsistencies (as tested by the Word vs. Non-word task) and with sentence-level inconsistencies (as tested by the Plausible vs. Implausible task). This would indicate an inconsistency has been identified. Furthermore, we hypothesized that students would attempt to repair the breakdown in comprehension caused by these inconsistencies by spending more time rereading the target inconsistent words, as compared to the control words, in both tasks. Finally, we hypothesized that longer rereading time, but not necessarily longer gaze durations, would be positively associated with higher baseline reading

Table 7 Hierarchical linear modeling results for rereading time (milliseconds) for word vs. non-word task as a function of students' reading comprehension and vocabulary, and grade level as dummy variables

Fixed effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p value
Fourth grade as the reference group					
For word targets, ψ_0					
Fitted mean, γ_{00}	107.085	17.992	5.952	118	<0.001
Comprehension, γ_{01}	0.514	0.373	1.378	118	0.171
Vocabulary, γ_{02}	−0.006	0.436	−0.015	118	0.988
Third grade, γ_{03}	50.370	25.404	1.983	118	0.050
Fifth grade, γ_{04}	−30.583	31.989	−0.956	118	0.341
For non-word targets, ψ_1					
Item type, γ_{10}	223.399	34.243	6.524	2265	<0.001
Third grade, γ_{11}	−84.458	41.856	−2.018	2265	0.044
Fifth grade, γ_{12}	−23.398	47.558	−0.492	2265	0.623
Third grade as the reference group					
For word targets, ψ_0					
Fitted mean, γ_{00}	157.455	19.154	8.221	118	<0.001
Comprehension, γ_{01}	0.514	0.373	1.378	118	0.171
Vocabulary, γ_{02}	−0.006	0.436	−0.015	118	0.988
Fourth grade, γ_{03}	−50.370	25.404	−1.983	118	0.050
Fifth grade, γ_{04}	−80.952	33.309	−2.430	118	0.017
For non-word targets, ψ_1					
Item type, γ_{10}	138.941	24.064	5.774	2265	<0.001
Fourth grade, γ_{11}	84.458	41.856	2.018	2265	0.044
Fifth grade, γ_{12}	61.060	40.870	1.494	2265	0.135

Random effect	Standard deviation	Variance component	d.f.	χ^2	p value
Fourth grade as the reference group					
Intercept, u_0	134.002	17,956.488	118	428.328	<0.001
Level-1, e	360.455	129,927.496			
Third grade as the reference group					
Intercept, u_0	134.002	17,956.488	118	428.328	<0.001
Level-1, e	360.455	129,927.496			

Equation for Fourth grade as the reference group

$$Rereading\ Time_{mj} = \gamma_{00} + \gamma_{01} * (Comprehension)_j + \gamma_{02} * (Vocabulary)_j + \gamma_{03} * (Third\ grade)_j$$
$$+ \gamma_{04} * (Fifth\ grade)_j + \gamma_{10} * (Item\ type)_{mj} + \gamma_{11} * (Third\ grade)_j * (Item\ type)_{mj}$$
$$+ \gamma_{12} * (Fifth\ grade)_j * (Item\ type)_{mj} + u_{0j} + e_{mj}$$

comprehension and vocabulary scores, and with grade level. That is, students with stronger reading comprehension and vocabulary skills and older students would be more likely to attempt to repair breakdowns in comprehension. Overall, our results found evidence of comprehension monitoring among third through fifth graders.
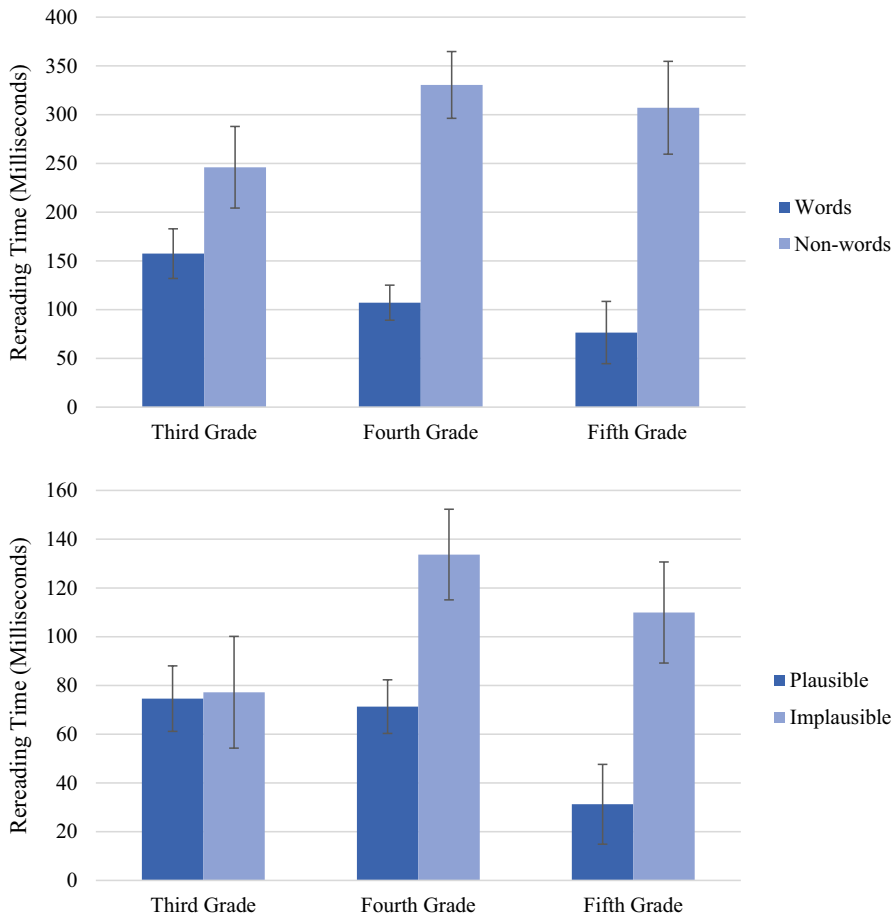
Fig. 5 Rereading time for the Word vs. Non-word Task (top) and Plausible vs. Implausible Task (bottom) by grade level when controlling for comprehension and vocabulary skills. In the Word vs. Non-word task, fourth graders spent significantly longer rereading the non-words compared to third graders. In the Plausible vs. Implausible task, fourth and fifth graders spent significantly longer reading the implausible words compared the third graders. Error bars represent approximate standard errors (see Tables 7 and 8 for exact standard errors)

However, this varied by students' age/grade level and their reading comprehension and vocabulary skills. Again, we examined the two aspects of comprehension monitoring: detecting an incongruity using gaze duration, and working to repair the misunderstanding using rereading time, at both the word-level (Word vs. Non-word) and the sentence-level (Plausible vs. Implausible).

Table 8 Hierarchical linear modeling results for rereading time (milliseconds) for plausible vs. implausible task as a function of students' reading comprehension and vocabulary, and grade level as dummy variables

Fixed effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p value
Fourth grade as the reference group					
For plausible targets, ψ_0					
Fitted mean, γ_{00}	71.325	11.013	6.477	118	<0.001
Comprehension, γ_{01}	0.240	0.201	1.191	118	0.236
Vocabulary, γ_{02}	−0.125	0.220	−0.567	118	0.572
Third grade, γ_{03}	3.269	13.455	0.243	118	0.808
Fifth grade, γ_{04}	−40.083	16.381	−2.447	118	0.016
For implausible targets, ψ_1					
Item type, γ_{10}	62.357	18.596	3.353	2317	<0.001
Third grade, γ_{11}	−56.478	22.908	−2.465	2317	0.014
Fifth grade, γ_{12}	−23.778	20.730	−1.147	2317	0.252
Third grade as the reference group					
For plausible targets, ψ_0					
Fitted mean, γ_{00}	74.594	10.230	7.292	118	<0.001
Comprehension, γ_{01}	0.240	0.201	1.191	118	0.236
Vocabulary, γ_{02}	−0.125	0.220	−0.567	118	0.572
Fourth grade, γ_{03}	−3.269	13.455	−0.243	118	0.808
Fifth grade, γ_{04}	−43.352	15.145	−2.863	118	0.005
For implausible targets, ψ_1					
Item type, γ_{10}	5.879	13.391	0.439	2317	0.661
Fourth grade, γ_{11}	56.478	22.908	2.465	2317	0.014
Fifth grade, γ_{12}	32.700	16.223	2.016	2317	0.044

Random effect	Standard deviation	Variance component	d.f.	χ2	p value
Fourth grade as the reference group					
Intercept, u_0	56.3852	3179.221	118	293.594	<0.001
Level-1, e	205.108	42,069.221			
Third grade as the reference group					
Intercept, u_0	56.385	3179.221	118	293.594	<0.001
Level-1, e	205.108	42,069.221			

Equation for Fourth grade as the reference group

$$Rereading\ Time_{mj} = \gamma_{00} + \gamma_{01} * (Comprehension)_j + \gamma_{02} * (Vocabulary)_j + \gamma_{03} * (Third\ grade)_j$$
$$+ \gamma_{04} * (Fifth\ grade)_j + \gamma_{10} * (Item\ type)_{mj} + \gamma_{11} * (Third\ grade)_j * (Item\ type)_{mj}$$
$$+ \gamma_{12} * (Fifth\ grade)_j * (Item\ type)_{mj} + u_{0j} + e_{mj}$$

Word-level comprehension monitoring

When examining gaze duration and rereading time for the Word vs. Non-word task, the results suggested that, on average, all students were found to be sensitive to the word-level inconsistencies, such that they spent a longer time reading the non-words than the control real words. Additionally, we found evidence of attempts to repair comprehension inasmuch as rereading times were longer for the non-words compared to the real words (see Fig. 1, top). Considering that non-words were unfamiliar to the students, it is not surprising that they were easier to detect and students were more likely to take a longer time processing them compared to the control words. It is also encouraging that they worked to repair their comprehension breakdown caused by these inconsistencies by rereading them. Whereas reading comprehension was not significantly associated with gaze duration for non-words, the effect did vary depending on students' vocabulary knowledge even after controlling for grade level and reading comprehension.

All students, regardless of their reading comprehension skills, read non-words longer than words. This further supports the findings of Connor et al. (2015) that detection of inconsistencies in text does not depend on reading comprehension skills, but may be more automatic than previously hypothesized (Cain et al., 2004; Connor, 2013; Oakhill et al., 2005). However, we did find that children's ability to detect an inconsistency at the word-level did depend on their vocabulary skills. This finding suggests that children with better vocabulary knowledge may have a better sense of what they do and do not know (i.e., word knowledge calibration) and are therefore better at detecting unfamiliar words. Similarly, and in line with our expectations, we found that stronger vocabulary knowledge and stronger reading comprehension was related to a higher likelihood of repairing misunderstandings (i.e., rereading) or regulating comprehension after a breakdown was caused by a word-level inconsistency. Students with stronger vocabulary knowledge, when controlling for their comprehension, and vice versa, generally spent longer rereading non-words compared to real words. These findings lead us to conclude that stronger literacy skills are critical for regulating comprehension, using repair strategies such as rereading.

Importantly, although varying by grade, reading comprehension, and vocabulary, all the students in our sample generally detected an inconsistency when confronted with a non-word, slowed down their reading, and spent more time rereading to repair their misunderstanding. This indicates that the newly developed Word vs. Non-word Task successfully introduced a word-level inconsistency, caused a breakdown in students' comprehension, which in return necessitated the readers to regulate their comprehension. Thus, this task may be an effective method to assess word-level comprehension monitoring for third through fifth graders.

Sentence-level comprehension monitoring

Our hypotheses for sentence-level comprehension monitoring were supported only partially. Students did not react to the sentence-level inconsistencies (implausible words) with a longer gaze duration compared to plausible words (see Fig. 1,

bottom), contrary to our expectations and the previous study using this eye-movement task (Connor et al., 2015). Connor and colleagues (2015) found that the fifth-grade students, regardless of their literacy skills, generally had a longer gaze duration for the implausible words than for plausible words. However, we did replicate the findings for rereading time, in that rereading time was longer for implausible vs. plausible words. When we examined the effect of students' grade level (see Fig. 5 bottom), we found that this was really only the case for fourth and fifth graders, who did spend more time rereading implausible words compared to plausible words with the greatest difference in fourth grade. There was no significant difference in rereading time for plausible and implausible words for third graders. Plus, overall, students with stronger vocabulary skills, controlling for grade and reading comprehension, spent longer times rereading implausible vs. plausible words than their peers with weaker vocabulary skills (see Fig. 4 bottom). This does replicate the Connor et al. (2015) findings. Unfortunately, our sample size was too small to adequately power a potential three-way interaction effect of task by grade by vocabulary; however, the idea is worth pursuing. Overall, these findings suggest that the Plausible vs. Implausible task may be useful for assessing comprehension monitoring but only for fourth and fifth graders.

General discussion

Taking the results altogether (see Table 9), we found evidence of comprehension monitoring in all three grades for the word-level task. However, for the sentence-level task, we found evidence of comprehension monitoring only for fourth and fifth graders. For both tasks, rereading time (comprehension regulation) varied by students' reading comprehension and vocabulary skills, such that students with stronger literacy skills engaged in longer rereading of inconsistent target words. However, for gaze duration (comprehension evaluation), we found that this depended only on vocabulary knowledge and not on reading comprehension. These findings demonstrate, for both word- and sentence-level comprehension monitoring, that comprehension evaluation depends more strongly on vocabulary skills while comprehension regulation depends on both vocabulary and comprehension skill.

It has been proposed that comprehension monitoring reflects the readers' goals, such as the intended level of comprehension as well as ability to utilize different levels of criteria for effective comprehension (Kinnunen & Vauras, 2010). We conjecture that students with stronger reading comprehension, vocabulary knowledge, or who are older may be reading with the goal of understanding the gist of the text and therefore are more aware and sensitive to breakdowns in their comprehension at the sentence-level. The results of the Plausible vs. Implausible task support this hypothesis; however, we also found evidence for this sensitivity at the word-level in the Word vs. Non-Word task. Moreover, it may be that students with stronger reading comprehension, vocabulary knowledge, or who are older have a stronger ability to utilize higher levels of comprehension regulation and repair strategies. However, it is important to note that this link may be bidirectional, such that those students with higher level comprehension monitoring

Table 9 Summary of Results

	Word *vs.* Non-word	Plausible *vs.* Implausible
Gaze duration	Longer for non-words than for control words	No significant differences between plausible and implausible words
Rereading time	Longer for non-words than for control words	Longer for implausible than for plausible words
Relation with reading comprehension	No significant relation with gaze duration, but Stronger comprehenders had longer rereading times for non-words	No significant relation with gaze duration, but stronger comprehenders had longer rereading times for implausible words
Relation with vocabulary	Those with stronger vocabulary had longer gaze durations and rereading times for non-words	Those with stronger vocabulary had longer gaze durations and rereading times for implausible words
Relation with grade	No significant relation with gaze duration. Fourth graders had longer rereading times for non-words than third graders	No significant relation with gaze duration. Fourth and fifth graders had longer rereading times for implausible words than third graders

skills (regulation at the sentence-level) are more likely to develop better reading comprehension and vocabulary skills.

When we examined how students' vocabulary knowledge may be linked to the amount of time spent rereading the target words, we observed similar patterns across the two tasks as we hypothesized. We expected students with stronger vocabulary knowledge to be more likely to regulate their comprehension when confronted with both word- and sentence-level inconsistencies. This trend was observed for both the non-words and the implausible words. More specifically, when controlling for students' reading comprehension, those with stronger vocabulary knowledge demonstrated a higher likelihood to attempt repairing their misunderstandings when the reading obstacle was a non-word or an implausible word. This suggests that stronger vocabulary knowledge may be associated with greater ability to regulate comprehension when confronted with either a word-level or sentence-level error. This could be due to students with stronger vocabulary knowledge having a better sense of what they do not know and attempting to regulate their comprehension (i.e., word knowledge calibration; Connor et al., 2019). This is similar to the findings of a previous study with older adults. Kavé and Halamish (2015) found that older adults (70–84 years of age) are more likely to have stronger vocabulary knowledge and are also more proficient in judging their own knowledge compared to young or middle-aged adults.

Whereas we found no effect of age/grade level for gaze duration for either task, there was a significant interaction effect on rereading for both tasks. In general, the difference between the control word (i.e., word or plausible word) and the inconsistent word (non-word or implausible word) was smaller for third graders and larger for fourth and fifth graders, which we argue indicates greater levels of comprehension monitoring—specifically employing repair strategies. Our results demonstrated that fourth and fifth graders engaged in more rereading when faced with sentence-level inconsistencies compared to third graders, which could indicate a developmental trend of stronger skills in the use of repair strategies. Similarly, only on the sentence-level comprehension monitoring task was there a difference between fourth and fifth graders with fifth graders spending less rereading time on plausible words than fourth graders; rereading time for implausible words did not differ for fourth and fifth graders. These age/grade level differences are consistent with developmental theories that metacognitive skills are becoming more fully developed during middle childhood (Del Giudice, 2014). During middle childhood, metacognitive skills are developing and this would suggest that comprehension monitoring, also a metacognitive skill is strengthening as students enter fourth and fifth grade. These results also may be indicating that sentence-level comprehension monitoring is more difficult than is the word-level comprehension monitoring. However, replication studies are needed to support this claim.

Limitations and future directions

Although the newly developed eye-movement task was validated, such that students viewed and reread the target non-words longer, this task may be improved.

It would be interesting to test whether providing more context rather than only one sentence, might increase the differences between gaze duration and rereading time for the words vs. non-words. Although this study did not allow for any conclusions regarding a causal relationship between vocabulary knowledge and regulating comprehension, we believe that this relationship may be bidirectional, similar to the link between vocabulary knowledge and inference making (Oakhill, Cain, & McCarthy, 2015). On one hand, students with greater ability to regulate their comprehension by using repair strategies, such as the use of context clues or other word learning strategies, tend to develop stronger vocabulary knowledge by learning new words from the text (Fukkink & de Glopper, 1998; McNamara, 2007; Nagy, McClure, & Mir, 1997). On the other hand, students with better vocabulary knowledge may have a higher metalinguistic awareness and are more likely to identify unfamiliar words or contradictory information in text and may engage in using repair strategies (Nagy, 2007). Therefore, future longitudinal studies are needed in order to examine the direction in which vocabulary knowledge and comprehension monitoring skills may be linked. This bidirectionality may also hold true for the relation between reading comprehension skills and comprehension monitoring. Additionally, this study demonstrated grade level effects, such that the fourth graders in our sample engaged in more rereading than third graders when confronted with a word-level inconsistency, while fourth and fifth graders engaged in more rereading when confronted with sentence-level inconsistencies, compared to their younger peers in third grade. However, this study used a cohort design. Longitudinal studies are needed to fully examine the effect of age on comprehension monitoring, and whether it improves as students mature and gain more experience with reading. Finally, our sample size was too small to examine three-way interactions. Thus, we could not fully disentangle the effects of reading comprehension, vocabulary knowledge, and grade level. Replicating this study with a larger longitudinal sample would address this issue.

Implications

Understanding the level at which a student may be struggling with reading comprehension will allow for better individualization of literacy instruction. As discussed earlier, a student may be able to evaluate his/her comprehension, while struggling with regulating his/her comprehension. Other times, students may be struggling with monitoring their comprehension at the sentence-level, while they do not find regulating their comprehension at lower levels difficult. Recognizing what skillset might be needed for having stronger comprehension monitoring and reading comprehension is important for developing literacy instruction for students, especially in early elementary years. Therefore, valid and reliable assessments need to be more accessible to educators. With the new eye-movement technology and the number of different eye-movement tasks developed, such as the one we have developed, assessing students' reading and comprehension monitoring abilities has become easier and more reliable. In the present study, we introduced different eye-movement tasks that may allow educators to precisely examine comprehension monitoring and eye-movement behavior in young children. Moreover, this will allow researchers and

educators to develop educational interventions that target aspects of reading comprehension with which a student may be struggling. For example, after examining and considering the important role of vocabulary knowledge, comprehension monitoring, and utilizing different repair strategies, we have developed an E-Book aiming to improve these skills in elementary school students (Connor et al., 2019). Therefore, developing personalized instructional interventions, with the use of technology or in the classroom, can be informed through assessing students' individual differences and abilities and eye-tracking methodology is a promising technology.

Appendix

Stimuli used in Word vs. Non-word eye-movement task

1. David wanted to *explore/spettle* the wild island.
2. The bug *crawled/thecked* up Daisy's arm.
3. Rosita climbed the *mountain/floggorn* in the morning.
4. The loud *noise/ploof* scared Ethan.
5. The sun was shining and it was a *beautiful/smotterup* day.
6. The cheetah was fast, but the lion was *stronger/roolaps*.
7. At his birthday, Dan *laughed/shalled* at the clown.
8. *Wow/Zop*, this cake is really tasty!
9. Mickey did not enjoy the apple because it was not *ripe/bope*.
10. The ballerina danced gracefully on the *stage/quode*.
11. The *dentist/cranter* said Jimmy needed to floss.
12. Electricity allows for the *toaster/hearner* to work.
13. The soccer team *won/jub* the playoffs.
14. My father is a judge and goes to *court/sneet*.
15. *Rubber/Trople* tires are the best for a car.
16. *School/Phleen* is where Timmy learned science.
17. *Water/Klopa* is very important to drink on a hot day.
18. Lindsey had to *take/chup* her dog on a walk.
19. Did you know tomatoes are a *fruit/grutt*?
20. When it rains, my mom uses an *umbrella/nepronto* to stay dry.

Stimuli used in Plausible vs. Implausible eye-movement task

1. For the wedding, Linda wore her best outfit. The colorful *dress/plant* was one of her favorites.
2. Jeremy carefully sawed through a large oak tree in his yard. The steel *blade/bolt* became dull after only a few minutes.
3. Last week Kyle flew to visit his family in another city. The large *plane/truck* was spacious and quickly transported them.
4. Charles and his friends tossed the new toy outside in the yard. They were upset when the *ball/book* went over the fence.
5. Marcus washed dishes every night to earn his allowance. The new *sponge/broom* was great for getting rid of the grit.
6. Before school Jack printed his report for history class. He became angry when the *paper/pencil* got stuck.
7. Every day Rover barked at the passing animals on the street. He was the most alert *puppy/kitten* in the neighborhood.
8. In the evening, Nicholas pounded two boards together in his garage. His new *hammer/handsaw* was a really useful tool.
9. Tyler cautiously sipped his fresh sweetened tea. He dropped the *glass/plate* because it was so hot.
10. Last night Bobby slept very well for many hours. His new boots were cozy and comfortable. His new *boots/sheets* were cozy and comfortable.
11. Tim and Landon rowed along the river in the park. Tim lost his *paddle/saddle* in the middle of the river.
12. Justin threw the ball to his friend during the game. He liked *football/jogging* more than any other sport.
13. Yesterday evening Jenny sewed patches onto her jeans. Sadly, she lost the *needle/wrench* and could not finish.
14. When she got mad Sarah stomped on the floor. It really hurt when her *foot/hand* hit the ground with such force.
15. Today Sean arrested an unruly criminal as everyone watched. As an experienced *officer/minister* he quickly took control of the situation.
16. Amanda sat outside and read about a man named Arthur. She loved the *novel/movie* about ancient times.
17. To plant a tree Shannon dug a large hole in the ground. Her sturdy *shovel/blower* helped make the job easy.
18. Michelle poured a fresh cup of coffee for her friend. She dropped the *kettle/shoes* and got very annoyed.
19. Today Janet rode with her friends to the mall. The yellow *taxi/kayak* quickly got them to their favorite places.
20. They were all happy as they sailed along the coast. The swift *boat/bike* raced near the beautiful beach.

References

Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluation comprehension: Effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology, 38,* 289–311.

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating comprehension. In D. Forrest-Pressley, G. McKinnon, & T. Waller (Eds.), *Metacognition, cognition and human performance* (pp. 155–205). New York, NY: Academic.

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique [Topics in Cognitive Psychology], 114,* 647–662.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference making ability, and the relation to knowledge. *Memory & Cognition, 29,* 850–859.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* 31–42.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, NY: Cambridge University Press.

Connor, C. M. (2013). Intervening to support reading comprehension development with diverse learners. In B. Miller & L. E. Cutting (Eds.), *Unraveling the behavioral, neurobiological and genetic components of reading comprehension: The Dyslexia Foundation and NICHD* (pp. 222–232). Baltimore, MD: Brookes.

Connor, C. M., Day, S. L., Zargar, E., Wood, T. S., Taylor, K. S., Jones, M. R., et al. (2019). Building word knowledge, learning strategies, and metacognition with the Word-Knowledge e-Book. *Computers & Education, 128,* 284–311.

Connor, C. M., Radach, R., Vorstius, C., Day, S. L., McLean, L., & Morrison, F. J. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading, 19,* 114–134. https://doi.org/10.1080/10888438.2014.943905.

Del Giudice, M. (2014). Middle childhood: An evolutionary-developmental synthesis. *Child Development Perspectives, 8*(4), 193–200. https://doi.org/10.1111/cdep.12084.

Duyck, W., Desmet, T., Verbeke, L., & Brysbaert, M. (2004). WordGen: A tool for word selection and non-word generation in Dutch, German, English, and French. *Behavior Research Methods, Instruments & Computers, 36*(3), 488–499.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18,* 5–12.

Ehrlich, M. F. (1996). Metacognitive monitoring in the processing of anaphoric devices in skilled and less skilled comprehenders. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and remediation* (pp. 221–249). Mahwah. NJ: Lawrence Erlbaum Associates Inc.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239–256.

Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of Educational Research, 68*(4), 450–469.

Garrett, A. J., Mazzocco, M. M. M., & Baker, L. (2006). Development of the metacognitive skills of prediction and evaluation in children with or without math disability. *Learning Disabilities Research & Practice, 21,* 77–88. https://doi.org/10.1111/j.1540-5826.2006.00208.x.

Gombert, J. E. (1992). *Metalinguistic development*. Chicago, IL: University of Chicago Press.

Hacker, D. J. (1998). Self-regulated comprehension during normal reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 165–191). Mahwah, NJ: Erlbaum.

Kavé, G., & Halamish, V. (2015). Doubly blessed: Older adults know more vocabulary and know better what they know. *Psychology and Aging, 30*(1), 68–73.

Kendeou, P., Van Den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice, 29*(1), 10–16.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101,* 765–778.

Kim, Y.-S. G. (2016). Direct and mediated effects of language and cognitive skills on comprehension or 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. G. (2017). Why the simple view of reading is not simplistic: Unpacking the simple view of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading, 21,* 310–333. https://doi.org/10.1080/10888438.2017.1291643.

Kim, Y. S. G., Vorstius, C., & Radach, R. (2018). Does online comprehension monitoring make a unique contribution to reading comprehension in beginning readers? Evidence from eye movements. *Scientific Studies of Reading, 22,* 1–17.

Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning & Instruction, 5,* 143–165.

Kinnunen, R., & Vauras, M. (2010). Tracking online metacognition: Monitoring and regulating comprehension in reading. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 209–258). New York, NY: Springer.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates-Macginitie reading tests, fourth edition technical report for forms S & T*. Itasca, IL: Riverside Publishing.

McNamara, D. S. (Ed.). (2007). *Reading comprehension strategies: Theory, interventions, and technologies*. Mahwah, NJ: Erlbaum.

Nagy, W. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. Wagner, A. Muse, & K. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52–77). New York: Guilford Press.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233–253.

Nagy, W. E., McClure, E. F., & Mir, M. (1997). Linguistic transfer and the use of context by Spanish-English bilinguals. *Applied Psycholinguistics, 18*(4), 431–452.

Oakhill, J., & Cain, K. (2012). The precursors of reading comprehension and word reading in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16,* 91–121. https://doi.org/10.1080/10888438.2010.529219.

Oakhill, J., Cain, K., & McCarthy, D. (2015). Inference processing in children: The contributions of depth and breadth of vocabulary knowledge. In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 140–159). New York, NY: Cambridge University Press.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing: An Interdisciplinary Journal, 18*(7–9), 657–686.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A., & Lesgold, A. M. (1979). Coding and comprehension in skilled reading and implications for reading instruction. *Theory and Practice of Early Reading, 1,* 57–84.

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development, 86,* 159–175.

Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for the future. *European Journal of Cognitive Psychology, 16,* 3–26.

Radach, R., & Kennedy, A. (2013). Eye movements in reading: Some theoretical context. *Quarterly Journal of Experimental Psychology, 66,* 4429–4452.

Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension: An integrative view of reading. *Current Directions in Psychological Science, 14,* 276–279.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10,* 241–255.

Rice, G. A., & Robinson, D. O. (1975). The role of bigram frequency in the perception of words and nonwords. *Memory & Cognition, 3*(5), 513–518.

Rubman, C. N., & Waters, H. S. (2000). A, B seeing: The role of constructive processes in children's comprehension monitoring. *Journal of Educational Psychology, 92,* 503–514.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities, 45*(5), 453–466.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Elham Zargar[1] · Ashley Marie Adams[1] · Carol McDonald Connor[1]

1   School of Education, University of California, Irvine, 3200 Education Bldg., Irvine, CA 92697-5500, USA