Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

The effects of question previewing on response accuracy and text processing: An eye-movement study

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Moss, Corrin	Mount Holyoke College	
Ardoin, Scott P.		0000-0003-2176-4221
Mellott, Joshua		0000-0003-4504-1984
Binder, Katherine S.		0000-0003-0711-5751

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 2023

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Journal of School Psychology.

DOI or URL to published work (if available)

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Institute of Education Scie through **[Grant number]** R305A1703 to **Institution]** University of G. The opinions expressed are those of the authors and do not represent views of the **[Office name]** or the U.S. Department of Education.

2) Title page with author details

The Effects of Question Previewing on Response Accuracy and Text Processing: An Eye-Movement Study

Corrin Moss[1]

Scott P. Ardoin[2]

Joshua A. Mellott[3]

Katherine S. Binder[1]

[1]Department of Psychology, Mount Holyoke College

[2]Department of Educational Psychology, University of Georgia

[3] Kennedy Krieger Institute and Johns Hopkins University School of Medicine

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A17036 to the University of Georgia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Neither the Institute nor the U.S. Department of Education contributed to the study design; the collection, analysis, or interpretation of data; or to the writing and submission of the manuscript.

Declarations of interest: none

Correspondence regarding this article should be directed to Katherine S. Binder, kbinder@mtholyoke.edu.

Key words: reading comprehension; eye tracking; reading strategies; individual differences; working memory capacity; reading achievement

Abstract

The current study investigated the impact of manipulating reading strategy, reading the questions first (QF) or the passage first (PF), during a reading comprehension test, and we explored how reading strategy was related to student characteristics. Participants' eye movements were monitored as they read 12 passages and answered multiple-choice questions. We examined differences in (a) response accuracy, (b) average total time on words in the text, (c) total task reading time, and (d) time reading text relevant to questions as a function of PF and QF strategies. Analyses were conducted to examine whether findings varied as a function of student characteristics (reading achievement and working memory capacity) and grade level (third-, fifth-, and eighth grade). There were several interesting findings from our study: (1) there was a limited effect of strategy use on response accuracy, with only eighth graders demonstrating better accuracy in the QF condition, and (2) there were several demonstrations of PF leading to more efficient test-taking processes: (a) longer average total reading times on words in the passage in the PF condition which could be associated with creating a better mental model of the text, (b) PF was often associated with less total-task time, and (c) PF was associated with more successful search strategies. Implications for providing teachers and students with strategies are discussed.

The Effects of Question Previewing on Response Accuracy and Text Processing: An Eye-Movement Study

A large body of research indicates that providing readers with goal-directed reading instructions allows them to better process text sections relevant to the defined goal (Kaakinen et al., 2002; Lewis & Mensink, 2012; McCrudden & Schraw, 2007). Accordingly, some argue that question first (QF) strategies—reading questions before the text—help to define readers' text goals, allowing them to strategically focus their attention while reading the passage during reading comprehension tests (McCrudden et al., 2005). Therefore, QF strategies should lead to higher accuracy scores on multiple-choice reading comprehension tests and longer reading times on relevant text areas than a passage first (PF) strategy (reading the passage before the questions). However, findings from existing studies are inconsistent on the impact of these two strategies on reading comprehension test outcomes (e.g., Guo et al., 2016; Reed et al., 2019; Yeari et al., 2021). Additionally, researchers have devoted less attention to examining the impact of these two strategies on the reading comprehension test outcomes of school-aged students, as most related research has investigated the impact of these strategies on the performance of college-aged readers (e.g., Cerdán et al., 2009; Daneman & Hannon, 2001; Farr et al., 1990). Considering school psychologists regularly administer and interpret the results from multiple-choice reading comprehension tests when making educational decisions for students and frequently provide students with test-taking strategies, it is vital for them to understand how test-taking strategies might impact test outcomes. Thus, the current study examined the efficacy of PF and QF strategies used by children. Specifically, using eye-tracking technology, we examined differences in (a) response accuracy, (b) average total reading time reading on words in the texts, (c) total-task time, and (d) time reading text relevant to questions as a function of PF and QF strategies across third-, fifth-, and eighth-grade students.

Although educators and researchers often look to response accuracy to understand reading comprehension, that measure only tells us about the product of comprehension. To gain a more comprehensive picture, we must strive to understand how readers process the information in real-time during reading. Despite there being several ways to accomplish this goal, recent advances in eye-tracking technologies allow us to examine these processes in children. In a typical eye-tracking experiment, a child is seated in front of a computer monitor that displays reading material. An eye-tracking camera is positioned in front of the monitor, and it records every eye movement that the child makes while reading. Even though it feels like our eyes smoothly sweep across the page, we actually make a series of discrete movements while reading or viewing a scene. The "jumps" are called saccades, and fixations are periods during which our eyes are relatively still. It is during the fixations that we extract visual information and begin to process it. There are a number of eye movement measures a researcher can extract from a reading episode, with some of these measures reflecting lower-level word processing (i.e., word recognition). Other measures are thought to reflect higher-level processing such as text integration and inference making. A major advantage of using eye movements to understand reading is that a researcher can examine a constellation of eye movement measures to gain an understanding of how processing unfolds over time (Foster et al., 2018; Rayner et al., 2013). Likewise, measuring eye movements can allow us to evaluate whether different test-taking strategies result in students engaging in reading behavior consistent with mindless reading and searching the text for answers versus reading for comprehension (Nguyen et al., 2014; Zawoyski et al., 2023)

Eye-movement studies are based on the assumption that eye movements can reveal the underlying cognitive processing associated with reading (Rayner, 2009). There is variability in the duration, number, and direction of fixations and saccades, and researchers have linked this variability to word-level or text difficulty and/or the skill of the reader. For example, low frequency words have longer fixations associated with them compared to words with higher frequency, even when word length is controlled (e.g., Rayner & Duffy, 1986). Ashby et al. (2013) revealed relationships between eye movement measures and other indicators of reading ability. They found that inefficient phonemic processing was associated with silent reading fluency, as measured by eye movements. In addition, Valle et al. (2013) demonstrated changes in the durations of eye movements of 2nd-grade readers who were exposed to a repeated reading intervention. There is compelling evidence that eye movements can be used to gain a deeper understanding of the processes readers use when trying to comprehend material.

Test-Taking Strategy

Completing a reading comprehension test is a task-oriented activity, with qualitative (Farr et al., 1990) and eye-tracking research (Authors, submitted) suggesting that students are not necessarily consistent in the strategies that they use across texts. In a study conducted by Authors (submitted), students in third, fifth, and eighth grade were provided with passages and multiple-choice questions concurrently and allowed to use the test-taking strategy of their choice. Differences were observed in the strategies chosen by students as a function of grade level, with older students being more likely than younger students to choose a QF strategy, and younger students being more likely to read passages in their entirety (PF strategy) before leaving the text to read and answer questions. It was speculated that the change in the use of strategies might indicate that across the school years, there is a greater likelihood that students (a) were encouraged to use a QF strategy, (b) developed a belief that the QF strategy required less effort, or (c) developed a belief that the QF strategy resulted in greater response success.

Although teachers (Rupp et al., 2006) and school districts (e.g., Rock Hill Schools; Leone, J., n.d.) frequently teach and encourage the QF strategy, extensive research examining the effects of reading the QF versus reading the PF during multiple-choice reading comprehension tests has produced inconsistent results. For example, some research shows that using the QF strategy results in lower accuracy scores or poorer understanding of the text than the PF strategy (Cerdán et al., 2009; Guo et al., 2016). However, other work suggests that there is no difference in accuracy scores between these two test-taking strategies (Bayrak Karsli et al., 2020; Daneman & Hannon, 2001; Reed et al., 2019), yet others propose that under certain conditions, the QF strategy leads to higher response accuracy (Yeari et al., 2021).

Despite mixed findings on the impact of PF and QF strategies on response accuracy, research related to their impact on other aspects of the reading process is more consistent with evidence suggesting the PF strategy is the more beneficial strategy. For example, Cerdán et al. (2009) showed that college students had a deeper understanding of the text in the PF condition. Cerdán et al. argued that by first reading the passage thoroughly, participants formed a better mental model of the text, strengthening their understanding of the passage. In addition, contrary to frequently recommended testing advice (Rupp et al., 2006), QF strategies may not improve test-taking efficiency (Bayrak Karsli et al., 2020; Cerdán et al., 2009; Yeari et al., 2021). Efficiency can be characterized using many dimensions of reading time. For example, using technology that required students to click on sections of the text that they wished to read, Cerdán et al. (2009) found that participants spent more time reading the passage in the PF condition than in the QF condition; however, there was no difference in efficiency as total-task time (time reading the passage plus time answering associated questions) did not differ between strategy conditions. Conversely, Yeari et al. (2021) found that the PF condition was less efficient as total-task time was longer in the PF condition.

Alternatively, efficiency can be measured using the total time spent rereading and answering questions, independent of the time it took participants to complete an initial read of the passage. Prior research demonstrates that the PF strategy results in participants (a) spending less time reading questions and answer options and (b) making fewer regressions back to the text when answering questions (Bayrak Karsli et al., 2020; Yeari et al., 2021). Yeari et al. (2021)

noted that the QF condition was characterized by “scattered reading and degraded comprehension” (p. 15) and resulted in readers more frequently alternating between looking in the passage and question areas. Bayrak Karsli et al. (2020) posit that creating a good mental model of the text during reading may facilitate question-answering, leading to fewer regressions—rereading—to the text when answering questions. Finally, efficiency can be assessed by measuring the time spent reading relevant sections of text relative to time spent on peripheral ideas in the text. Cerdán et al. (2009) found that participants spent more time reading areas of text relevant to answering the questions in the PF condition than in the QF condition.

There is also evidence to suggest that the QF strategies may present unique challenges for struggling readers. Struggling readers who find text challenging or uninteresting are more likely to search the text for answers and avoid reading (Farr et al., 1990). Less-skilled readers are especially likely to attempt to avoid reading (Nguyen et al., 2014), and using a QF strategy may prompt them to use a search-and-find strategy instead of acting as a goal-focusing tool. Additionally, Bayrak Karsli et al. (2020) suggest that the QF condition may tax readers’ short-term memory and that some readers may be unable to remember the questions and answer options after reading the whole passage. Thus, it is also essential to examine how reader characteristics, frequently measured as part of psychoeducational evaluations (e.g., reading achievement and working memory capacity), interact with students' reading strategies during reading comprehension tests.

Text Relevance and Goal-Oriented Reading

Theories of task-oriented reading (e.g., the TRACE model) posit that readers create a task model, allowing them to determine which parts of the text are relevant to their reading goals and allocate their attentional resources accordingly (Rouet & Britt, 2011). TRACE model builds on McCrudden and Schraw’s (2007) work, demonstrating that readers allocate more attention to text areas explicitly deemed relevant by instructions before reading. For instance, in a study examining differences in readers' eye movements, Kaakinen et al. (2015) randomly assigned second-, fourth-, and sixth-graders to one of two goal-directed conditions. One condition directed students to read the passage to comprehend the text, and the other directed students to read the passage to answer a "why" question that required the integration of information presented throughout the text. As expected and consistent with prior research (Vorstius et al., 2014), regardless of condition, students' eye movements differed across age groups with younger, less-skilled readers making longer first fixations on words. Interestingly, differences were also observed as a function of goal directions, and the differences varied as a function of developmental level. Second graders increased the amount of time that they spent on words during their first reading. In contrast, older students increased their regressions to previously read words. Findings by Kaakinen et al. (2015) might suggest that having students use a QF strategy might increase their efforts to read for comprehension. However, the goal a student establishes from reading a single open-ended question designed to promote the integration of information is likely to differ from pre-reading a set of multiple-choice questions (i.e., QF strategy).

Lewis and Mensink (2012) argue that QF strategies provide explicit relevance instruction and should lead readers to devote more attention to information that will help them answer reading comprehension test questions. Prior research suggests that relevance instructions help readers to better learn and remember relevant information (Lewis & Mensink, 2012; McCrudden & Schraw, 2007). However, the relationship between relevance instructions and reading time is less clear. For example, although McCrudden et al. (2005) reported no difference in reading times between text relevant to instruction and less-relevant text areas, Lewis and Mensink (2012) found longer reading times on relevant areas of text than non-relevant areas. The differences in results between the two studies may be related to differing methods: Lewis and Mensink (2012) utilized eye-tracking data, whereas McCrudden et al. (2005) used a moving-window paradigm (software that displays one word or region of the text at a time and is paced by the reader) that prohibited participants from rereading the text. The present study used eye-tracking procedures to investigate differences in reading time on relevant versus less relevant areas during both initial reading and rereading.

Additionally, past research suggests that individual differences, such as working memory capacity (WMC) and reading skills, interact with relevance instructions and are related to processes that occur while reading (Kaakinen & Hyönä, 2002; Kaakinen et al., 2003). For example, poor comprehenders may be more easily distracted by irrelevant information. In a moving-window paradigm study with middle-school students, Cerdán et al. (2011) found that good and poor comprehenders were equally likely to read distractor information in the text following a misleading question. Good comprehenders were, however, able to discard the distractor information and choose the correct answer, but poor comprehenders were not. Poor comprehenders also invested more time reading the distractor information than skilled comprehenders. A QF strategy may present a similar challenge as incorrect response options are meant to be distractors, which may hinder poor comprehenders' reading time and response accuracy. Furthermore, in eye-movement studies, WMC and relevance instructions interacted in their relationship with reading time (Kaakinen & Hyönä, 2002; Kaakinen et al., 2003). For example, Kaakinen and Hyönä (2002) demonstrated that readers with higher WMC had longer reading times on relevant areas (relevance effects) during both initial reading and rereading. In contrast, readers with lower WMC showed relevance effects only during rereading. This result suggests that WMC is an important student characteristic that interacts with text-relevant instructions. However, the participants in Kaakinen and Hyönä (2002) were college students with advanced reading skills. Research is needed to understand the effects of relevance instructions and the interactions between text relevance and reader characteristics in younger students. Information gleaned from such research has the potential to inform not only the test-taking strategies that teachers encourage students to use when taking high-stakes tests but also how teachers establish students' reading goals when teaching reading comprehension. Furthermore, the findings have potential implications for which tests school psychologists choose when assessing students' reading comprehension skills. Evidence of differences in student outcomes between test-taking strategies might suggest the need for school psychologists to use tests that encourage or potentially prevent students from using one of the strategies.

Age is another important variable related to reading comprehension. Vocabulary, background knowledge, and semantic knowledge increase with age (Cain & Oakhill, 2011; Smith et al., 2021; Wassenberg et al., 2008); consequently, older children have higher reading comprehension accuracy than younger children. Minimal research has examined the effect of age on reading time at the word level during reading comprehension tests. Ardoin et al. (2019) found that younger children (2nd graders) had longer first-fixation and gaze durations than older children (fifth graders), and Kaakinen et al. (2015) demonstrated that children made more frequent and longer rereads of the text than adults. In addition, it is quite possible that the types of strategies children use change over developmental periods based on the advice given to them by teachers.

Current Study

We investigated the impact of manipulating reading strategy—QF or PF—during reading comprehension tests and explored how reading strategy was related to student characteristics. Although researchers have investigated reading comprehension processes during testing (Kaakinen & Hyönä, 2002; Kaakinen et al., 2003), relatively few studies have examined student strategies throughout the entire testing process, including while answering associated multiple-choice questions. Additionally, this study focused on the effect of PF/QF strategies used by school-aged participants; this line of research has overwhelmingly focused on adult populations (Bayrak Karsli et al., 2020; Cerdán et al., 2009; Yeari et al., 2021). It is crucial to empirically examine test-taking strategies to understand how test format might impact test outcomes, as well as to ensure that practitioners are not providing incorrect advice to students.

The present study had several research aims. First, we examined whether response accuracy on multiple-choice items differed between the PF and QF strategy conditions. Much prior research (Bayrak Karsli et al., 2020; Cerdán et al., 2009; Reed et al., 2019; Yeari et al., 2021) failed to find a difference in response accuracy between PF and QF conditions; however, findings from several studies suggest that the PF condition fosters deeper comprehension (Bayrak Karsli et al., 2020; Cerdán et al., 2009). Therefore, we predicted that response accuracy would be higher in the PF condition. In addition, we expected this effect would be greater in students with lower WMC, as they have greater difficulty updating the contents of their WMC (Caretti et al., 2005) and therefore reading the questions before reading the text (QF condition) may provide too much distractor information, taxing their WMC by creating too much cognitive load. We also expected that this effect would be greater for students with lower reading achievement as the QF strategy would signal that their goal is to answer questions, not read the text for comprehension, and thus they would be more likely to search passages for answers and not read passages for comprehension (Farr et al., 1990).

We also examined whether strategy was related to the length of time that students spent reading the passage and completing the total reading and responding task. Prior research reported that participants spent more time on specific tasks in the QF condition, including completing the whole reading comprehension test (Bishop, 2001), reading the questions (Bayrak Karsli et al., 2020), and rereading while searching for answers (Yeari et al., 2021). Additionally, findings suggest that the QF condition can lead to distracted reading (Yeari et al., 2021). However, other studies have reported participants spend more time reading the text (but not while completing other parts of the task) in the PF condition (Cerdán et al., 2009; Yeari et al., 2021). Therefore, we predicted that participants would have longer average total reading times on words in the passage in the PF condition but that total-task time would be longer in the QF condition. We also predicted that this effect would be greater for participants with lower WMC and participants with lower reading achievement, as the QF condition may tax their cognitive resources and lead to longer total-task times (Bayrak Karsli et al., 2020). We also examined the relative proportion of distracted reading during the QF condition by examining the number of instances in which the readers read the text in its entirety in the QF condition compared to alternating between text and questions. We predicted that the QF condition would lead to more distracted reading. That is, participants would be more likely to read only part of the passage before searching for the answer instead of reading the entire passage to develop a better understanding (Yeari et al., 2021).

Finally, we investigated whether strategy use might influence students' time reading the question-relevant text sections. Prior research suggests that readers strategically devote attention to portions of text that align with their purpose for reading (e.g., McCrudden & Schraw, 2007). The QF condition may prompt students to spend more time searching for answers to specific questions. Therefore, we predicted that participants would have longer reading times on question-relevant text when primed with the questions (i.e., QF) than in the PF condition. However, because prior research indicates that PF may lead to readers developing a better mental model of the passage (Cerdán et al., 2009), we predicted that the PF condition would be associated with greater precision when students searched the passage for answers. That is, they will be able to target their search to the area containing question-relevant information instead of rereading the entire text or searching it in an unorganized way.

Method

Participants and Setting

Participants included 150 students in grades three (n = 49), five (n = 54), and eight (n = 47) attending one of four districts in the Southern United States. These grades were chosen because they spanned the range of grades in which reading comprehension tests were administered to students in the state where the study was conducted. These grade levels also allowed us to capture students across at least two stages of reading development (Kuhn & Stahl, 2003). The percentage of students receiving free and reduced lunch for the three districts serving students in grades three and five was 14.85%, 45.63%, and 57.32%. The percentage of eighth-grade students from two school districts receiving free and reduced lunch was 14.85% and 77.33%, respectively. Students were majority female (56%) and Caucasian (74%). See Table 1 for further demographic information.

Research assistants conducted testing with students in quiet spaces within the schools (e.g., conference rooms). Although testing locations varied in size, the testing apparatus was consistent across schools and the locations were all free from outside distractions. Eye-tracking data collection was conducted across two sessions lasting approximately 30 min. Reading achievement and WMC measures were administered immediately before or following the collection of eye-tracking data. Occasionally, a third testing session was necessary to complete all testing elements.

Power Analysis

We used a repeated measures design because we wanted to optimize power in our experiment. Optimal Design Plus (Raudenbush et al., 2011) was used to estimate power for the linear mixed effects analyses. For a 2-level (passage- and student-level) linear mixed effects model, a minimum detectable effect size of .20, .80 power level, and an estimated 40% of the variance accounted for by covariates in the model, indicated that we would need 43 students per grade level for the study. More participants were included in the study than required because it was expected that a portion of the eye movement data would not be suitable for analysis due to technical difficulties.

Materials

Participants were administered aReading (FastBridge Learning, 2017) to assess their reading achievement and Lucid Recall (Lucid Innovations Limited, 2013) to assess their WMC. The aReading assessment is an item-response theory-based computer-adaptive measure designed to measure broad reading skills and predict students' overall reading achievement. Items administered to elementary students target print awareness, phonics, vocabulary, and comprehension, whereas middle school items are meant to assess orthography, morphology, vocabulary, and comprehension. Students were required to respond to 30 items with completion times generally ranging from 10–15 min for elementary students and 20–30 min for middle school students. Criterion-related validity is provided through correlations between aReading and Gates MacGinitie Reading Tests–4th Edition (GMRT–4; MacGinitie et al., 2000). Correlation between composite scores ranges from .64 to .83 across grades.

To assess WMC, we administered the Pattern Recall and Word Recall subtests of the Lucid Recall computerized assessment, designed for children 7 years, 0 months to 16 years, 11 months old. For the Word Recall subtest, the computer program audibly presents a series of words and then presents a 3 x 3 set of words. The student must click on the words just presented audibly in the order presented. Level 1 requires recalling two items, and the final level requires recalling six items. The Pattern Recall subtest presents a model matrix of white and black boxes. After the model matrix disappears, a matrix of all white boxes appears, and students must turn the boxes that were black in the model matrix black by clicking on the corresponding boxes. Level 1 is a set of four squares; the final level consists of 24 squares. Trials for both subtests are discontinued after a student misses three items within a level. Test-retest reliabilities range from 0.68–0.77 for Pattern and Word Recall subtests. For evidence of criterion validity, the technical manual provides correlations for the Word and Pattern Recall measures with the reading (range: .35–.50), writing (range: .29–.58), and mathematics (range: .31–.46) measures of the National

Curriculum Levels. Students completed the subtests independently on a laptop with an experimenter nearby to answer questions. Administration time was approximately 25 min.

Passages

To select grade-appropriate passages, we searched for passages within existing third, fifth, and eighth grade published reading comprehension tests. The length and type of all passages were categorized, and passage difficulty was assessed (see Table 2 for passage lengths). Using those data, we developed two passage sets, each consisting of a passage at each length (short, medium, and long) for each passage type (narrative and expository). All short passages had four associated multiple-choice questions, and each medium and long passage had six associated questions. For each passage, half of the questions were literal and half were inferential. Each question had four response options. See supplemental materials for additional information on passage selection.

Apparatus

Eye-tracking programs, constructed using SR Research Experiment Builder, were developed to first provide students with an opportunity to practice reading a passage and answer multiple-choice questions in an identical fashion to the administration of experimental passages and questions. Using the program, eye movement data were collected using an SR Research Eyelink 1000 Plus system, with a sampling rate of 2000 Hz, a resolution of 0.01 degrees of visual angle, and a range of 32 degrees horizontally and 25 vertically. By default, eye movements were recorded via the right eye, but students' viewing of the computer monitor was binocular. Stimuli (text and questions) were presented on a flat 24-in. ViewSonic LCD monitor, placed 93 cm from the SR Research chinrest. The camera was placed 65 cm from the chinrest. To stabilize their heads during eye tracking, participants positioned their heads on a chin rest. They used a standard USB mouse to transition between texts and respond to questions. The texts were presented in standard upper- and lowercase paragraph format, using Times New Roman font, 1.5 spaced black text on a white background on a computer screen.

Procedures

Prior to students taking part in the study, researchers obtained parental consent and student assent using forms approved by the institutional review board of the second author. Participants' eye movements were monitored as they read 12 passages and answered multiple-choice questions. During one testing session, six trials presented the passage screen first (without access to the questions). During another session, six trials presented the question screen first (without access to the passage). After first presenting students with the passage (PF) or questions (QF) screen, the passage and questions were presented to students on the same screen. Research assistants also measured students' reading achievement and WMC. Because the PF and QF conditions were delivered on different days for each participant, we did not think participants would be confused about the directions for each day. Furthermore, holding participants' attention to read six passages per day as opposed to 12 passages seemed more reasonable. Participants were randomly assigned to the strategy condition order (PF-Day 1, QF-Day 2; QF-Day 1, PF-Day 2) and the passage set (Set 1 or Set 2) they would complete under each strategy condition. Research assistants informed students that they would need to keep their heads steady on the chin rest and use the computer mouse to answer questions and move between screens. Then, researchers read students a set of directions consistent with the condition they were completing that day (see supplemental material for a copy of the script).

Passage first (PF). Participants were first informed that they would need to read the passage in its entirety before clicking on a right arrow button which would present a new screen with the same text and the multiple-choice questions. They were also informed that they could move between screens containing the text and questions using left and right-pointing arrows positioned at the bottom of the screens. Below each arrow was either the word back (left) or next (right). After the researcher read the directions, students were allowed to read a practice passage and answer questions in a fashion identical to the presentation of the PF strategy condition. The practice session consisted of three screens. Screen 1 was the passage only. Screen 2 had the passage and questions one and two. Screen 3 had the passage and questions three and four. After answering questions associated with the PF practice passage, eye-movement calibration and validation procedures were conducted. The eye tracker was calibrated and then validated using a nine-point grid that extended across the monitor, which is consistent with eye-tracking studies that display passages on the monitor (e.g., Ardoin et al., 2019; Foster et al., 2018; Tremblay et al., 2021). Then, the PF program presented six passages (2 types and 3 lengths) in randomized order with their associated questions. Question order was not randomized. For each passage, the first screen of the PF condition included the passage with no questions and a right-pointing arrow (next) at the bottom right corner of the screen. Screen 2 consisted of the passage in the exact location presented in Screen 1 and the first two multiple-choice questions. Screen 3 consisted of the passage and the third and fourth questions. A fourth screen had the passage and questions five and six for medium and long passages. Screens one and two had next arrows only. Screens three and four had back and next arrows, allowing students to move back and forth between questions or move to the following passage after the student answered all questions.

Questions first (QF). Participants were informed that they would need to read all questions associated with a passage before clicking on the right (next) arrow button that would present them with the passage and questions on the same screen. Researchers also informed students how to use the left and right-pointing arrows to move between questions of the same passage. After researchers read the directions, the program presented students with the practice questions associated with the practice passage; the passage was not on Screen 1. After reading the questions on Screen 1, students clicked on the next arrow to proceed to Screen 2, which included the practice passage and questions one and two. They could then click the next arrow to access Screen 3, containing the passage and questions three and four. After answering all questions associated with the QF practice passage, calibration and validation procedures were completed. Then the QF procedures were implemented with six passages (2 types and 3 lengths) and their associated questions. Passage order but not question order was randomized. For each passage, the first screen of the QF condition presented all questions of the associated passage but not the passage itself. Screens 2 and 3 (short) and 2, 3, and 4 (long) of each passage for the QF condition were identical to the corresponding screens of the PF condition, with both the text and questions presented concurrently and arrows at the bottom of each screen to allow for movement between screens.

Dependent measures

In addition to measuring the products of reading comprehension through evaluating students' response accuracy on multiple-choice reading comprehension questions, we measured (a) average total time on words in the passage, (b) total-task reading time (passage and question reading time), (c) time on question relevant text, (d) incidence of distracted reading for the QF condition, and (e) rereading precision. Descriptive statistics for response accuracy and eye movement measures can be found in Table 3. Foster et al. (2018) established reliability and validity estimates for children for the eye movement measures that are used in this study. Test-retest reliability was quite high, ranging from .73 - .82 for time-based eye movement measures. Alternate-form reliability coefficients for the eye movement measures ranged from .77 to .89. Concurrent validity estimates were assessed by correlating eye movement measures with participants' standard scores on WJ-III ACH reading subtests, composite Broad Reading standard scores, median scores across three CBM-R probes. Significant moderate to high correlations were obtained for the relationships between the time-based eye movement measures and the other language-based assessments.

Response Accuracy.

Response accuracy was based on the percentage of correct responses for each passage. The short passages had four questions, and the medium and long passages each had six questions.

Average Total time on Words in the Passages.

Researchers examining eye movements during reading define a fixation as a pause the eyes make on a word that allows the reader to extract visual information from the text. Readers can make multiple fixations on a word before leaving the word, and they also return to and refixate on previously read words. One of the primary dependent eye movement measures used in the current study was the average total fixation time on words in the passage, which is the duration, in ms, of all fixations made on a word and is thought to represent higher-level text processing (Foster et al., 2018). For the average total time on words in the passage, total fixation time was calculated for each word in the passage, and then averaged across the number of words in the passage.

Total-Task Reading Time.

Total-task reading time included the cumulative time on the passage and the questions, and any time students spent rereading the passage while answering the questions.

Distracted Reading.

Incidences of distracted reading were evaluated for the QF condition. We coded whether the reader made a complete read of the passage before answering the question or if they read part of the passage and then went to the questions before completing a full read. Thus, the test-taking strategy used by students was coded as either a complete or incomplete read.

Question Relevant Reading Times.

We examined reading time on the question-relevant sections of the text relative to the non-relevant sections. Question-relevant texts were those phrases/sentences identified by two experts[1] in reading comprehension as providing information necessary for answering multiple-choice reading comprehension questions. Because question-relevant and non-relevant regions differed in length, we summed up the total reading times in these areas and divided them by the number of characters to yield a ms by character measure.

Precision.

We also used eye movement recordings to code whether students returned to the passage after reading a question and the precision of their regressions back to the passage. High precision was defined as students immediately regressing back to the sentence containing information that would allow them to answer the question or one sentence prior to or following the sentence containing information before reading the actual sentence containing the relevant information. Low precision was defined as students reading more than one sentence just prior to or just following the sentence containing information that related to answering the question before reading the actual sentence containing the information. No precision was defined as students rereading the passage but never looking at the sentence containing information that would allow them to answer the question. No rereads was defined as students never returning to the passage after reading the question. We examined this behavior for every question the participant answered. That is, each participant read 12 passages. Eight of the passages each had six questions, and the remaining four passages each had four questions. Thus, each participant had 64 precision observations, unless we were missing data for that participant.

Data Preparation

Files were prepared prior to analysis using techniques applied in previous eye movement studies with children (e.g., Ardoin et al., 2016; Tremblay et al., 2021; Zawoyski et al., 2015).

[1] One expert was a full professor in special education who specialized in dyslexia and reading comprehension and had over 50 peer refereed articles related to reading comprehension. The second was an assistant professor with numerous peer-refereed publications examining components of reading comprehensions, quality of screening measures, and multiple years of experience working with the Florida Center for Reading Research.

Specifically, fixations were aligned with regions of interest (i.e., passage words or questions and answer choices) to ensure that they would be included in data analyses. Individuals preparing the data completed training on these procedures and demonstrated reliability by achieving accuracy on at least 90% of data points compared with established reliable coders. Fixations outside the range of 120 ms and 800 ms were removed because children are not likely to have fixation lengths outside of this range during reading (Morrison, 1984).

Analytic Approach

We conducted linear mixed effects models for response accuracy, average total time per word, total-task reading time, and question relevancy reading time. In these models, the following effects were entered as fixed effects: strategy condition (PF vs. QF), reading achievement, word recall, pattern recall, and the interactions between strategy condition and the three student characteristics (reading achievement, word recall, and pattern recall). Strategy condition was a categorical variable, while the others were continuous predictors. The continuous variables were all centered. When one of the continuous variables interacted with one of the categorical variables, two groups were created to explain the interaction. Scores of zero and above were placed in a “higher” category, while all scores falling below zero were in the “lower” category. For the question relevancy reading time analyses, the fixed effects of relevancy (categorical variable), the interaction of that variable with strategy condition, and each of the student characteristics were also fixed effects. In each linear mixed effects model, the intercepts for participants and passage were included as random effects. For distracted reading and reading precision, χ^2 analyses were run since the dependent variables were categorical. For the linear mixed effects model, effect sizes are reported as the Snijders and Bosker (1994) proportional reduction in variance as effect sizes (R^2 [S&B]). Effect sizes for the χ^2 are reported as □c. Assumptions for the linear mixed effects models were examined. We examined the residuals for each dependent measure for normality by examining the QQPlots. We also examined whether there were constant variance/covariance matrices between the residuals and each predictor variable. We also examined the distributions associated with the random effects by plotting empirical Bayes estimators (Eblups). Our data met the assumptions associated with linear mixed effects models.

Results

While there was no data loss for response accuracy, there was missing data for the eye-tracking measures. Specifically, the data loss was due to track losses, which occurred when participants had a large head movement or a technological malfunction that resulted in a loss of calibration or tracking. Overall, track losses affected 2.50% of the data (third grade: 2.66%; fifth grade: 4.17%; eighth grade: 0.37%). These track losses occurred equally often in the PF (1.08%) and the QF condition (1.42%). The track losses often only happened for one of the 12 trials for a participant. Thus, we were able to use the remaining data for that participant. Linear mixed effect models can easily accommodate missing data.

Response Accuracy

The first dependent measure examined was response accuracy scores. The statistics associated with the fixed effects for each grade level are presented in Table 4.

Third Grade. Higher reading achievement scores were associated with higher response accuracy (B = 8.42; R^2 [S&B] = 0.11). No other effects (strategy condition, word recall, pattern recall, and the interactions between strategy condition and the three student characteristics) were significant.

Fifth Grade. Higher reading achievement (B = 8.11; R^2 [S&B] = 0.08) and higher word recall scores (B = 2.16; R^2 [S&B] = 0.06) were associated with higher response accuracy for the fifth graders. No other effects were significant.

Eighth Grade. Similar to the fifth graders, higher reading achievement (B = 3.61; R^2 [S&B] = 0.09) and higher word recall scores (B = 2.89; R^2 [S&B] = 0.09) were associated with higher response accuracy scores for the eighth graders. In addition, response accuracy scores in the QF condition (*M* = 90.39) were higher than in the PF condition (*M* = 86.47; R^2 [S&B] = 0.12). No other effects were significant.

Average Total time on Words in the Passages

Third Grade. Participants had a longer average total time per word in the QF condition (*M* = 679 ms) compared to the PF condition (*M* = 595 ms; R^2 [S&B] = 0.05). No other effects were significant.

Fifth Grade. In contrast to the third graders, fifth graders had a longer total time per word in the PF condition (*M* = 568 ms) compared to the QF condition (*M* = 534 ms; R^2 [S&B] = 0.03). No other effects were significant.

Eighth Grade. Like the fifth graders, eighth graders spent more total time per word when reading the passage in the PF condition (*M* = 475 ms) compared to the QF condition (*M* = 429 ms; R^2 [S&B] = 0.07). No other effects were significant.

Total-Task Reading Time

Third Grade. Students with higher reading achievement scores had shorter total-task time than students with weaker reading achievement (*B* = -18,328.95; R^2 [S&B] = 0.03). Students had longer total-task time in the QF (*M* = 156,773 ms) condition relative to the PF condition (*M* = 139,004 ms; R^2 [S&B] = 0.04). No other effects (word recall, pattern recall, and the interactions between strategy condition and the three student characteristics) were significant.

Fifth Grade. Students' total-task time was longer in the QF condition (*M* = 189,131 ms) relative to the PF condition (*M* = 172,726 ms; R^2 [S&B] = 0.06). There was also a significant 2-way interaction between strategy condition and word recall (R^2 [S&B] = 0.02). To investigate this interaction, we divided the participants into higher and lower word recall groups and examined differences across strategy conditions. Higher memory individuals spent more total-task time in the QF condition relative to the PF condition F =30.20, $p < .001$, while no differences were found for lower word recall participants across strategy conditions, $F = 0.75$, $p = 0.39$. See Figure 1 top panel for means and standard errors. No other effects were significant.

Eighth Grade. There was a significant interaction between strategy condition pattern recall (R^2 [S&B] = 0.01). We divided the participants into higher and lower recall groups to investigate this interaction and examined differences across strategy conditions. As can be seen in the bottom panel of Figure 1, eighth graders with higher memory scores spent longer in the PF condition relative to the QF condition, $F = 10.18$, $p < .001$, but there were no differences across strategy conditions for lower memory participants, $F = 1.35$, $p = 0.25$.

Distracted Reading

As mentioned earlier, students' strategies during the QF condition were coded as a complete or incomplete read, and a χ^2 analysis was conducted for each grade level. Third graders were more likely to do a complete ($n = 184$) than incomplete ($n = 90$) reading of the passage $\chi^2 = 32.25$, $p < .001$, □$c = 0.34$. There was no difference in the number of fifth graders who chose between complete ($n = 141$) and incomplete ($n = 163$) strategies, $\chi^2 = 1.59$, $p = .21$, □$c = 0.07$. Eighth graders were more likely to choose an incomplete ($n = 209$) than complete reading strategy ($n = 66$), $\chi^2 = 74.36$, $p < .001$, □$c = 0.52$.

Question Relevant Reading Times

For analyses of question-relevant reading times, the fixed effects were relevancy, text type, text length, strategy condition, reading achievement, and the two memory measures. Since relevancy is the new variable in these analyses, only effects that implicate relevancy are reported. However, all effects are provided in Table 5.

Third Grade. As hypothesized, third graders spent longer time on the text sections that were relevant ($M = 96.5$ ms/char) to answering the questions than on non-relevant text ($M = 87.5$ ms/char; R^2 [S&B] = 0.02). No other effects were significant.

Fifth Grade. Consistent with the third graders, fifth graders spent more time reading question-relevant information (M = 93.5 ms/char) than non-relevant information (M = 73.8 ms/char; R^2 [S&B] = 0.09). In addition, there were significant interactions between strategy condition and relevance R^2 [S&B] = 0.12. As presented in Figure 2, fifth graders spent more time reading the question-relevant regions compared to the non-relevant regions, but the magnitude of the differences was larger for the QF relative to the PF condition, QF: $F = 74.71$, $p < 0.001$; PF: $F = 57.30$, $p < 0.001$.

Eighth Grade. Like third and fifth graders, eighth graders also spent more time reading question-relevant information (M = 59.0 ms/char) than non-relevant information (M = 52.8 ms/char; R^2 [S&B] = 0.02). No other effects were significant.

Precision Measures

We measured the precision with which students searched the text for question-relevant information. The precision categories were as follows: High Precision, Low Precision, No Precision, and No Rereading during question answering. Frequency counts are found in Table 6.

Third Grade. There was no significant association between strategy condition and precision level for the third-grade readers ($\chi 2 = 0.57$, $p = .90$, □$c = 0.02$).

Fifth Grade. There was a significant association between strategy condition and precision for fifth graders ($\chi 2 = 22.59$, $p < .001$, □$c = 0.09$). Post-hoc analyses show that when readers were in the PF condition, they had more Low Precision searches than in the QF condition ($z = 2.8$), and there were more No Precision searches in the QF condition relative to the PF condition ($z = 3.9$). There were no other significant differences across strategy conditions.

Eighth Grade. There was a significant association between strategy condition and precision for these students ($\chi^2 = 15.55$, $p < .001$, □$c = 0.08$). Similar to the fifth-grade data, post-hoc analyses show that when readers were in the PF condition, they had more Low Precision searches than in the QF condition ($z = 2.7$), and there were more No Precision searches in the QF condition relative to the PF condition ($z = 3.4$). There were no other significant differences across strategy conditions.

Discussion

This study aimed to examine the impact of QF and PF reading strategies on third, fifth, and eighth-grade students' multiple-choice question response accuracy, text processing, and question processing. We also examined how strategy condition was related to various student characteristics (e.g., reading achievement and WMC). Although strategy condition only affected response accuracy for the eighth-grade students with QF associated with higher response accuracy, strategy condition did influence a number of the processing measures, which speaks to test-taking efficiency. Consistent with past research (Cerdán et al., 2009), fifth and eighth-grade students showed a more careful reading of the passage in the PF condition compared to the QF condition by having longer average total times per word. However, when we examined the total-task time when strategy condition had an effect, it was QF that resulted in greater total-task time. This was true for third graders, but WMC interacted with strategy condition for fifth and eighth graders. There were no differences between strategy conditions for either grade for students with lower WMC, but for fifth graders with higher WMC, the QF condition was longer than the PF condition, and for the eighth graders, higher WMC students took less time in the QF than the PF condition. In addition, we examined the amount of distracted reading in the QF condition, and this seemed to increase by grade level, such that (a) third graders were more likely to do a complete read of the passage before beginning to answer questions, (b) fifth graders were evenly split between a complete read and distracted reading, and (c) eighth graders were more likely to begin answering questions before completely reading the passage.

The final question addressed by the current study examined differences in students rereading precision across strategy condition. Results indicate that strategy condition was related to the precision with which fifth- and eighth-grade students searched the text. Although no differences were found for many of the precision categories, students in the QF condition seemed to search with less precision compared to the PF condition by having fewer low-precision searches and more no-precision searches. Thus, the majority of the processing measures demonstrate that the QF condition leads to less efficient test-taking processes and less reading for comprehension. Thus, encouraging students to use a QF condition might not only result in greater test-taking time, which requires more effort on the part of students—but also result in test scores that reflect less of the construct that we intend to measure, reading comprehension. These findings will be discussed in detail below.

Response Accuracy

Consistent with past research (Bayrak Karsli et al., 2020; Kaakinen et al., 2015; Reed et al., 2019), strategy condition did not affect students' response accuracy, with the exception that eighth graders achieved higher response accuracy scores under the QF condition relative to the PF condition. This finding is consistent with Yeari et al. (2021), but in that study, a QF advantage was only found when there was a time limitation. We suspect that by the eighth grade, students have had much more practice with a QF strategy, and perhaps this increased exposure has allowed them to master this strategy. Although strategy condition had a limited effect in this study on response accuracy, not surprisingly, student characteristics were consistently related to response accuracy. Reading achievement was positively correlated with response accuracy across the grades, and WMC was also positively related for the fifth and eighth graders. Thus, student characteristics appear to be more robust predictors than strategy condition.

Building a Mental Model

Consistent with Cerdán et al. (2009), fifth and eighth graders had longer average total reading time per word in the passage when reading in the PF condition relative to the QF condition. Cerdán et al. argued that longer reading times on the passage reflected students developing a more coherent mental model of the text during their initial reading. Our findings support this suggestion, and this finding may begin to help build a case for more efficient test-taking processing. That is, if readers spend more time initially reading and understanding the text, that extra time may pay off when students turn to answer the questions. It is puzzling that the third graders had the opposite effect: they had longer average total reading times per word in the QF condition relative to the PF condition. As was mentioned in the introduction, however, past studies (e.g., Bayrak Karsli et al., 2020) have speculated that the QF condition may create a greater cognitive load for less experienced readers. Although we hypothesized this effect would be demonstrated with an interaction of strategy condition with our measures of reading achievement and/or WMC, it could simply be the inexperience of the third graders across a variety of factors related to reading comprehension tests such as reading skills and strategy use that leads to this effect. In fact, we also found that in the QF condition, there seemed to be a developmental trend in that third graders were much more likely to complete an entire read of the passage before turning back to the question, fifth graders seemed to be evenly split between complete and incomplete reads of the passage in this condition, but eighth graders displayed a preference for incomplete reads over complete reads. Thus, we argue that third graders may not have the necessary skills to engage in a QF strategy, which resulted in a longer average total time per word on the passage in the QF condition as this strategy was unfamiliar to them.

Searching the Mental Model

Our conclusion above that the PF condition facilitated the development of a more coherent mental model is further supported when examining differences across strategy conditions for total-task time. For all grades, there was a main effect of strategy condition such that students took longer in the QF condition relative to the PF condition, providing further evidence that PF leads to greater test-taking efficiency. Although student characteristics did not interact with strategy condition for the third graders, they did for the fifth and eighth graders. Specifically, WMC interacted with strategy condition, but the effects were not entirely consistent across grades. What was consistent was that for individuals with lower WMC, they had longer total-task times, and there was no evidence of a difference between strategy conditions. Thus, readers with lower WMC, a characteristic often associated with students with learning disabilities (Smith-Spark & Fisk, 2007), did not seem to have the necessary resources to take advantage of different strategies or respond to differences in task demands. The data pattern for the fifth graders with higher WMC confirmed our hypothesis. They had shorter total-task time in the PF condition relative to the QF condition. In fact, the higher WMC students' QF total-task time was indistinguishable from the times associated with the lower WMC students. Surprisingly, the eighth grade higher WMC students had longer total-task time in the PF condition relative to the QF condition. Recall that eighth graders also had greater accuracy in the QF condition compared to the PF condition. It may be that for high-performing eighth graders, QF truly directs their reading, and their practice with QF allows them to read with the purpose of answering questions not necessarily reading for meaning.

Time Reading Relevant Text

Consistent with past research (e.g., McCrudden & Schraw, 2007), we found that all students at all grade levels were sensitive to the relevancy of the information they were reading. That is, students spent more total time reading the question relevant information than the non-relevant information. Although we hypothesized that strategy condition would focus readers' attention on text relevant to the questions, such that we expected larger effects of relevancy in the QF condition, we only obtained that pattern for the fifth graders. That is, the magnitude of the difference in reading time between question-relevant and non-relevant areas was larger in the QF condition than in the PF for fifth graders. These findings are consistent with research reporting that readers allocate more attentional resources to text areas that help them achieve a defined goal (e.g., McCrudden & Schraw, 2007). To extend prior research and explore this question, we looked closer at the precision with which students searched the passage for answers. For our precision analyses, we examined where a participant landed when they left a particular question and went back to the passage to find the answer. Although we found no differences across strategy conditions for third graders, we did have some interesting effects for fifth and eighth graders, which support the idea that QF leads to issues with searching the text for answers. For both grade levels, PF was associated with more low precision searches—keep in mind that a low precision search still means that the reader eventually read the question-relevant information—and QF led to more no precision searches—searched but failed to find the question-relevant text-relative to PF. Thus, it seems that when students use a QF strategy, they have a less coherent mental model of the text and thus, have a more difficult time finding the answers. This may be because they engaged in more distracted reading patterns (i.e., eighth graders in the QF condition were more likely to do an incomplete read of the passage before returning the questions).

Limitations and Future Directions

Results of the current study are interesting and have both theoretical and practical applications, but as with most studies, they must be interpreted in the context of the associated limitations. One limitation to consider is the manner in which reading comprehension was measured. Although we examined both response accuracy on a multiple-choice test and the ongoing processing associated with developing an understanding of the passages, the assessment of the “end product” is far from perfect. We used the multiple-choice format because it is a widely used form of reading comprehension assessment used by schools to make educational decisions and by researchers to evaluate intervention effects. Thus, we thought it would be useful to better understand what children do when they are engaged in this sort of task. Like other researchers, we believe that we are tapping into the quality of the mental model that students develop during the course of the task. However, there are several limitations associated with this form of assessment, and currently, the field lacks clear indicators of high-quality mental models formed during and post-reading.

Additionally, reading comprehension tests are prone to design challenges, both in passage stimuli and assessment formats. Future research should entertain the idea of using multiple forms of mental model assessments to remedy some of these limitations. In terms of the actual assessment, open-ended or constructed response questions could be used in addition to multiple-choice questions. For example, students could be asked to provide a retell of the passage or could be asked to provide a summary of the passage. Critically, multiple forms of assessment will help us to better understand which variables are related to which aspects of mental model formation, and importantly, how strategy use influences different forms of assessment. Finally, while we were successful in recruiting from schools that had economic diversity, our sample was not as racially and ethnically diverse as we would hope. Future studies must strive to ensure more diversity in their samples so that we can be more confident in the generalizations we make from the collected data. Finally, procedural integrity data were unfortunately not collected, but given that the administration of the practice text and questions and actual experimental text and questions were administered by the computer program, there could be no variation from the designated experimental procedure.

Summary

Despite strategy condition having a limited influence on students' response accuracy, it had a pronounced influence on the efficiency with which students processed the passages. Reading the passage first (PF) mostly resulted in students spending more time on the individual words in the passages. They were more successful in their searches when returning to the passages while answering questions (i.e., fewer no-precision searches). In contrast, the QF condition was associated with more inefficient processing. Students generally had longer total-task time under the QF condition, and there appeared to be developmental differences associated with the amount of distracted reading in the QF condition. Given these results and similar findings in the literature (Bayrak Karsli et al., 2020; Daneman & Hannon, 2001; Reed et al., 2019), we are hard-pressed to imagine situations in which we should encourage the teaching of the QF strategy to students.

When interpreting student outcomes on reading comprehension tests, administrators, teachers, and school psychologists assume that a student's test results represent their ability to engage in the many processes associated with reading comprehension. It is not clear that QF strategies promote reading comprehension. If this strategy is characterized by scattered reading (Yeari et al., 2021), it would be difficult for a practitioner to know how to help a student, and it would not be clear which component(s) of reading comprehension were failing the student. Thus, if one of the essential uses of reading comprehension tests is to identify which students need additional intervention, it behooves us to measure the actual skill of reading and not just searching through a text. We realize that for some tasks, students are asked to search for information, but in such cases, knowing what to search for is essential, and under these conditions, promoting QF would be useful. For most other learning situations, we believe it is important to stress to students to read for understanding, which would be consistent with the PF approach.

School psychologists spend a considerable portion of their job administering achievement tests, interpreting the outcomes of those tests, and providing recommendations. They must understand the processes that tests are intended to measure as well as how students might use test-taking strategies that result in test outcomes not reflecting the constructs intended to be measured. Furthermore, if they provide students with test-taking strategies, those strategies must be evidenced-based and help not hinder student performance. Findings from the current study provide evidence to suggest that encouraging students to use a QF strategy is likely to have more negative than positive consequences, especially for younger students and students with poor working memory. Not only is it likely to result in their test outcomes not representing their ability to comprehend, but the strategy will likely require them to engage in more effort in a task that is already challenging for them.

References

Ardoin, S. P., Binder, K. S., Zawoyski, A. M., Nimocks, E., & Foster, T. E. (2019). Measuring the behavior of reading comprehension test takers: What do they do, and should they do it? *Reading Research Quarterly, 54*(4), 507-529. http://doi.org/048/10.1002/rrq.246

Ardoin, S. P., Binder, K. S., Foster, T., & Zawoyski, A. M. (2016). Repeated versus wide reading: A randomized control trial study examining the impact of fluency interventions on underlying reading behavior. *Journal of School Psychology, 59,* 13-38. http://doi.org/10.1016/j.jsp.2016.09.002

Ashby, J., Dix, H., Bontrager, M., Dey, R., & Archer, A. (2013). Phonemic awareness contributes to text reading fluency: Evidence from eye movements. *School Psychology Review*, *42*(2), 157-170.

Bayrak Karsli, M., Demirel, T., & Kurşun, E. (2020). Examination of different reading strategies with eye tracking measures in paragraph questions. *Hacettepe University Journal of Education, 35(1),* 92-106. https://doi.org/10.16986/HUJE.2019051160

Bishop, N. S. (2001, April). *The validity of reading comprehension test scores: Evidence of generalizability across different test administration conditions.* Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, Washington.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, *44*(5), 431-443. https://doi.org/10.1177/0022219411410042

Carretti, B., Cornoldi, C., De Beni, R., & Romanò, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal Experimental Child Psychology, 91(1),* 45–66. https://doi.org/10.1016/j.jecp.2005.01.005

Cerdán. R. Gilabert, R., & Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences, 21(2),* 201–205. https://doi.org/10.1016/j.lindif.2010.11.007

Cerdán, R., Vidal-Abarca, E., Martinez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*(1), 13-27. https://doi.org/10.1016/j.learninstruc.2007.12.003

Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General*, *130*(2), 208. https://doi.org/10.1037/0096-3445.130.2.208

Davis, C.J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods, 37(1),* 65–70. https://doi.org/10.3758/BF03206399

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple- choice reading comprehension test. *Journal of Educational Measurement*, *27*(3), 209-226. https://doi.org/10.1111/j.1745-3984.1990.tb00744.x

FastBridge Learning. (2017, June 26). Norms and benchmarks. www.fastbridge.org/

Foster, T. E., Ardoin, S. P., & Binder, K. S. (2018). Reliability and validity of eye movement measures of children's reading. *Reading Research Quarterly*, *53*(1), 71-89. https://doi.org/10.1002/rrq.182

Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn to process coherence relations in narrative and expository text. In A. P. Sweet and C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82 – 98). Guilford Publications.

Guo, Q., Kim, Y. S. G., Yang, L., & Liu, L. (2016). Does previewing answer choice options improve performance on reading tests? *Reading and Writing*, *29*(4), 745–760. https://doi.org/10.1007/s11145-016-9626-z

Kaakinen, J. K., & Hyönä, J. (2002). Perspective effects on online text processing. *Discourse Processes, 33(2),* 239–257. https://doi.org/10.1207/S15326950DP3302_03

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2003). How prior knowledge, working memory capacity, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 447. https://doi.org/10.1037/0278-7393.29.3.447

Kaakinen, J. K., Lehtola, A. & Paattilammi, S. (2015). The influence of a reading task on children's eye movements during reading. *Journal of Cognitive Psychology*, *27*(5), 640–656. http://doi.org/10.1080/20445911.2015.1005623

Kuhn, M. R., & Stahl, S. (2003). Fluency: A review of developmental and remedial strategies. *The Journal of Educational Psychology*, *95(1),* 1–19. https://doi.org/10.1037/0022-0663.95.1.3

Leone, J. (n.d.). *Test-taking strategies for reading - Rock Hill Schools*. Rock-Hill.k12.us. https://www.rock-hill.k12.sc.us/cms/lib/SC01000464/ Centricity/Domain/2283/NHSPE-reading.pdf

Lewis, M. R., & Mensink, M. C. (2012). Prereading questions and online text comprehension. *Discourse Processes*, *49*(5), 367–390. http://doi.org/10.1080/0163853X.2012.662801

Lucid Innovations Limited (2013). *Lucid Recall Administrator's Manual.* Lucid Research Ltd.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests, Fourth Edition, Level 3 Form S.* The Riverside Publishing Company.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, *19*(2), 113–139. https://doi.org/10.1007/s10648-006-9010-7

McCrudden, M. T., Schraw, G., & Kambe, G. (2005). The Effect of Relevance Instructions on Reading Time and Learning. *Journal of Educational Psychology*, *97*(1), 88–102. https://doi.org/10.1037/0022-0663.97.1.88

Morrison, R.E. (1984). Manipulations of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception & Performance*, *10(5)*, 667-682. https://doi.org/10.1037/0096-1523.10.5.667

Nguyen, V.T., Binder, K.S., Nemier, C., & Ardoin, S.P. (2014). Gotcha! Catching kids during mindless reading. *Scientific Studies of Reading, 18(4)*, 274-290. http://doi.org/10.1080/10888438.2013.876424

Raudenbush SW, Spybrook J, Congdon R, Liu X, Martinez A, Bloom H, Hill C. Optimal design plus empirical evidence (Version 3.0) 2011. http://www.wtgrantfoundation.org/resources/optimal-design.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62(8),* 1457-1506. http://doi.org/10.1080/17470210902816461

Rayner, K., Ardoin, S. P., & Binder, K. S. (2013). Children's eye movements in reading: A commentary. *School Psychology Review, 42(2)*, 223-233.

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

Reed, D. K., Stevenson, N., & LeBeau, B. C. (2019). Reading comprehension assessment: The effects of reading the items aloud before or after reading the passage. *The Elementary School Journal*, *120*(2), 300-318. https://doi.org/10.1086/705784

Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Information Age Publishing.

Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing, 23(4)*, 441-474. http:// doi.org/10.1191/0265532206lt337oa

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, *42*(3), 214-240. https://doi.org/10.1080/02702711.2021.1888348

Smith-Spark, J. H., & Fisk, J. E. (2007). Working memory functioning in developmental dyslexia. *Memory, 15*(1), 34-56. https://doi.org/10.1080/09658210601043384

Snijders, T. & Bosker, R. (1994). Modeled variance in two-level models. *Sociological Methods & Research. 22 (3)*, 342-363. https://doi.org/10.1177/0049124194022003004

Tremblay, K. A., Binder, K. S., Ardoin, S. P., Talwar, A., & Tighe, E. L. (2021). Third graders' strategy use and accuracy on an expository text: an exploratory study using eye movements. *Journal of Research in Reading*, *44(4)*, 737– 756. https://doi.org/10.1111/1467-9817.12369

Valle, A., Binder, K. S., Walsh, C. B., Nemier, C., & Bangs, K. E. (2013). Eye movements, prosody, and word frequency among average and high skilled second grade readers. *School Psychology Review*, *42*(2), 171-190.

Vorstius, C., Radach, R., & Lonigan, C. J. (2014). Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition, 22*(3-4), 458-485. https://doi.org/10.1080/13506285.2014.881445

Wassenberg, R., Hurks, P., Hendriksen, J. G. M., Feron, F. J. M., Meijs, C. J. C., Vles, J. S. H., & Jolles, J. (2008). Age-related improvement in complex language comprehension: Results of a cross-sectional study with 361 children aged 5 to 15. *Journal of Clinical and Experimental Psychology, 30(4)*, 435–448. https://doi.org/10.1080/13803390701523091.

Yeari, M., Schlesinger, L. M., & Moshka, E. (2021). The role of time constraints and domain knowledge in reading comprehension tests: The case of text- first versus questions- first strategies. *Reading Research Quarterly, 57*(3), 913-936. https://doi.org/10.1002/rrq.452

Zawoyski, A. M., Ardoin, S. P., & Binder, K. S. (2023). The impact of test-taking strategies on eye movements of elementary students during reading comprehension assessment. *School Psychology, 38(1)*, 59-66. https://doi.org/10.1037/spq0000526

Zawoyski, A. M., Ardoin, S. P., & Binder, K. S. (2015). Using eye tracking to observe differential effects of repeated readings for second-grade students as a function of achievement level. *Reading Research Quarterly*, *50(2)*, 171-184. https://doi.org/10.1002/rrq.91

Table 1

Participant Demographic Information

		Third grade *n* = 49	Fifth grade *n* = 54	Eighth grade *n* = 47
Age		8.67 – 9.85 years	10.67 – 12.08 years	13.00 – 14.67 years
Gender				
	Males	44.90%	48.15%	36.96%
	Females	55.10%	51.85%	63.04%
Race or Ethnicity				
	White	75.51%	81.48%	65.22%
	Black	12.24%	1.85%	15.22%
	Hispanic	0%	0%	10.87%
	Asian	4.08%	5.56%	4.35%
	Mixed/Biracial	8.16%	11.11%	2.17%
	Other	0%	0%	2.17%

Note. Missing demographic data for two eighth grade participants.

Table 2

Passage Word Lengths

	Third grade		Fifth grade		Eighth grade	
	M	Range	*M*	Range	*M*	Range
Short	84.25	71-90	121.5	118-126	173	170-177
Medium	160	156-165	284	283-285	363	358-366
Long	229.75	222-237	401.5	399-405	425.25	424-426

Table 3

Descriptive Statistics for Dependent Measures across Strategy Condition

	Third grade *n* = 49	Fifth grade *n* = 54	Eighth grade *n* = 47
Accuracy			
PF	79.62 (2.19)	85.32 (1.72)	86.47 (1.63)
QF	80.94 (2.20)	84.02 (1.72)	90.39 (1.62)
Total Time per Word			
PF	595.13 (47.27)	567.65 (24.33)	474.99 (17.11)
QF	678.69 (47.31)	534.34 (24.33)	428.68 (17.03)
Total-Task Time			
PF	139003.57 (14905.56)	172725.68 (17948.78)	152343.85 (15213.26)
QF	156772.66 (14912.77)	189130.86 (17948.41)	149721.73 (15194.55)
Relevancy			
Relevant Text			
PF	96.70 (7.32)	93.92 (7.19)	65.95 (3.31)
QF	96.39 (7.33)	93.08 (7.20)	52.06 (3.29)
Less Relevant Text			
PF	89.20 (7.32)	78.55 (7.19)	61.04 (3.31)
QF	85.85 (7.33	69.01 (7.20)	44.54 (3.29)

Note. Standard errors are in parentheses.

Table 4

Fixed Effect for Response Accuracy and Eye Movement Measures

	Third Grade		Fifth Grade		Eighth Grade	
	F	*p*	*F*	*p*	*F*	*p*
Accuracy						
Strategy Condition	0.922	0.337	0.802	0.371	12.770	<0.000
Reading Achievement	30.486	<0.000	31.101	<0.000	21.522	<0.000
Word Recall	0.366	0.549	7.317	0.009	4.685	0.037
Pattern Recall	0.248	0.621	2.647	0.110	1.586	0.215
Strategy Condition * Reading Achievement	1.334	0.249	2.159	0.142	1.544	0.215
Strategy Condition * Word Recall	0.677	0.411	1.840	0.175	1.600	0.207
Strategy Condition * Pattern Recall	0.059	0.808	0.452	0.501	2.684	0.102
Average Total Time per Word						
Strategy Condition	27.268	<0.000	13.106	<0.000	37.706	<0.000
Reading Achievement	4.033	0.051	3.242	0.078	3.541	0.067
Word Recall	0.104	0.749	1.160	0.287	0.385	0.539
Pattern Recall	1.316	0.258	0.001	0.973	0.147	0.703
Strategy Condition * Reading Achievement	0.100	0.752	0.669	0.414	0.232	0.630
Strategy Condition * Word Recall	0.009	0.923	2.951	0.086	0.767	0.381
Strategy Condition * Pattern Recall	0.113	0.737	0.137	0.712	0.460	0.498
Total-Task Time						
Strategy Condition	18.964	<0.000	18.247	<0.000	0.487	0.486
Reading Achievement	4.722	0.035	3.587	0.064	2.353	0.133
Word Recall	0.377	0.542	0.308	0.582	0.648	0.425
Pattern Recall	1.042	0.313	0.004	0.952	0.173	0.679
Strategy Condition * Reading Achievement	0.107	0.743	0.872	0.351	0.402	0.527
Strategy Condition * Word Recall	0.006	0.941	11.037	0.001	0.114	0.736
Strategy Condition * Pattern Recall	0.686	0.408	0.818	0.366	4.876	0.028

Table 5

Fixed Effects for Eye Movement Measures for Relevancy Analyses

	Third Grade		Fifth Grade		Eighth Grade	
	F	***p***	***F***	***p***	***F***	***p***
Question Relevant Reading Time						
Strategy Condition	0.722	0.396	7.394	0.007	136.505	<0.000
Relevance	18.379	<0.000	115.741	<0.000	23.998	<0.000
Reading Achievement	3.074	0.087	2.352	0.132	3.617	0.064
Word Recall	0.354	0.555	0.470	0.496	0.700	0.408
Pattern Recall	0.211	0.648	0.043	0.837	0.229	0.635
Strategy Condition * Relevance	0.497	0.481	5.533	0.019	1.099	0.295
Strategy Condition * Relevance * Reading Achievement	1.717	0.162	0.336	0.799	0.166	0.920
Strategy Condition * Relevance * Word Recall	0.401	0.753	1.418	0.236	0.437	0.726
Strategy Condition * Relevance * Pattern Recall	0.458	0.711	0.209	0.890	0.848	0.468

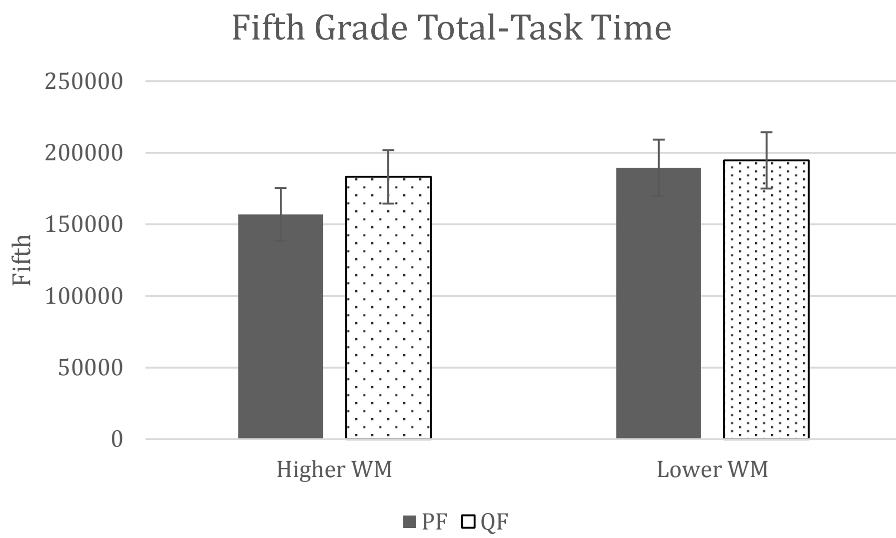
Table 6

Frequency Counts of Precision Category by Strategy Condition and Grade Level

Precision Category	Strategy Condition	
	PF	QF
Third Grade		
High Precision	214	225
Low Precision	118	113
No Precision	196	191
No Rereads	817	796
Fifth Grade		
High Precision	307	339
Low Precision	274	217
No Precision	200	278
No Rereads	778	728
Eighth Grade		
High Precision	352	360
Low Precision	312	274
No Precision	200	284
No Rereads	447	482

Figure 1

Total-task time in MS by Strategy Condition and Word Recall



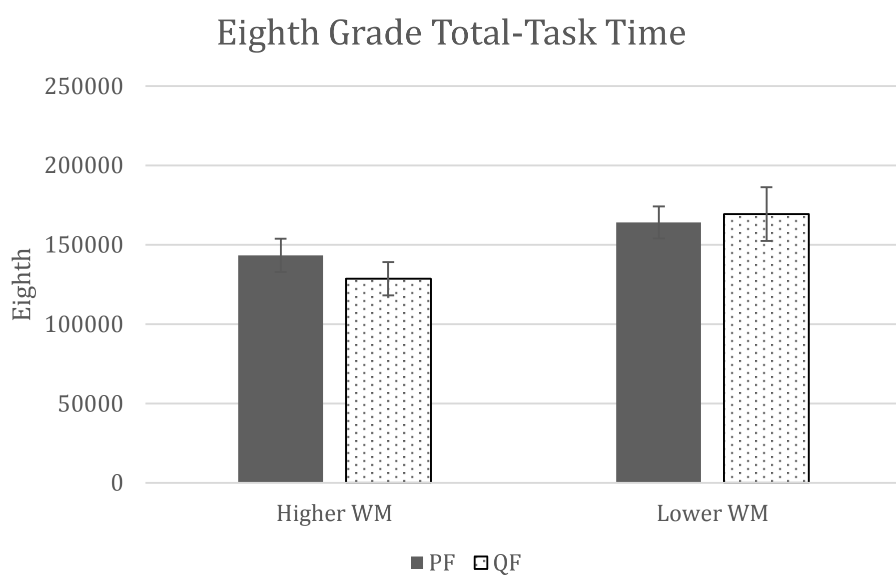
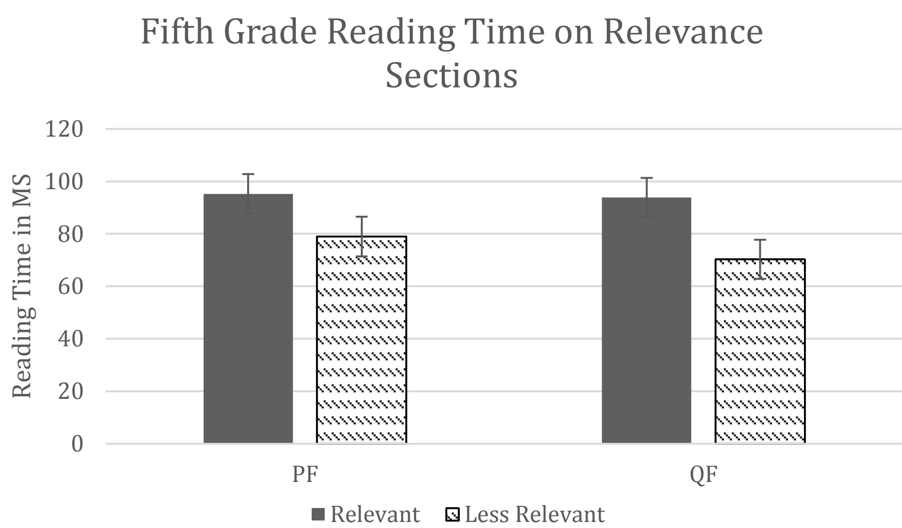


Figure 2

Time on Relevance Sections in MS by Strategy Condition



EXPO RC Study 2 Eye Tracking Directions

Passage First

General Directions	**Today we're going to have you read some stories and answer some questions. For each story you will first see a screen with the passage, and you need to first read the entire passage. After you have read the entire passage click next and go to the next screen where you can read the passage and answer the questions.** **You will need to have your head on the chinrest so this camera can record your eye movements. While you're doing this, try to sit as still as you can and keep resting your head against the bar. We might talk to you sometimes, but don't look at us – we want you to keep your head straight the whole time.** Click mouse on bulldog screen, hit esc on black keyboard
Practice Story	**On each story, you can use the next and back arrows to move between the questions. The computer will not let you move on to the next story until all the questions have been answered.** Click the next button. **Go ahead and read this practice story.** During practice story, make camera adjustments. Make sure eye is in focus. Use the up/down and +/- keys to adjust the image threshold so that the center dot takes up the entire pupil and that there is no blue inside the eye. Once student has read the practice story, hand them the mouse.
Calibrate and Validate	Click camera set up, then calibrate. **Now we're going to play the follow the dot game. The dot that you see at the center of the screen is going to move around. Your job is to stare at the dot until it moves to its next location. Do not try to guess where it will go next, stare at it until it moves. After the dot game you will start reading the rest of the stories.** Validate.
Eye Tracking	After a student has read the first display screen (watch host to ensure this) provide them with the mouse. If the student does not read the entire passage, say, **"You need to read the entire passage before I can provide you with the mouse."** Once all questions have been answered for a passage, remove the mouse before conducting the single calibration dot.
Open-Ended Questions	**Good reading! You can take your head of the chinrest now.** If second ET condition: **Now I want you to tell me everything you can remember about the last passage you just read. Give me as many details as possible, lie you were trying to tell a friend about what you just read.** **Is there anything else?** **Now I want you to tell me everything you can remember about the passage you read just before the last one. The one about** (story title)**. Give me as many details as possible, like you were trying to tell a friend about what you just read.** **Is there anything else?**

EXPO RC Study 2 Eye Tracking Directions

Question First

General Directions	**Today we're going to have you read some stories and answer some questions. For each story you will first see a screen with all of the question, you need to read all of the questions on that screen. After you have read all of the questions click next and go to the next screen where you can read the passage and answer the questions.** **You will need to have your head on the chinrest so this camera can record your eye movements. While you're doing this, try to sit as still as you can and keep resting your head against the bar. We might talk to you sometimes, but don't look at us – we want you to keep your head straight the whole time.** Click mouse on bulldog screen, hit esc on black keyboard
Practice Story	**On each story, you can use the next and back arrows to move between the questions. The computer will not let you move on to the next story until all the questions have been answered.** Click the next button. **Go ahead and read these practice questions.** During practice story, make camera adjustments. Make sure eye is in focus. Use the up/down and +/- keys to adjust the image threshold so that the center dot takes up the entire pupil and that there is no blue inside the eye. Once student has read the practice story, hand them the mouse.
Calibrate and Validate	Click camera set up, then calibrate. **Now we're going to play the follow the dot game. The dot that you see at the center of the screen is going to move around. Your job is to stare at the dot until it moves to its next location. Do not try to guess where it will go next, stare at it until it moves. After the dot game you will start reading the rest of the stories.** Validate.
Eye Tracking	If the student does not read all of the questions, (MC answers do not have to be read) say, **"You need to read all of the questions before I can provide you with the mouse."**
Open-Ended Questions	**Good reading! You can take your head of the chinrest now.** If second ET condition: **Now I want you to tell me everything you can remember about the last passage you just read. Give me as many details as possible, lie you were trying to tell a friend about what you just read.** **Is there anything else?** **Now I want you to tell me everything you can remember about the passage you read just before the last one. The one about** (story title)**. Give me as many details as possible, like you were trying to tell a friend about what you just read.** **Is there anything else?**