Eye movements and reading comprehension performance: examining the relationships among test format, working memory capacity and reading comprehension

Corrin Moss[1] · Sharon Kwabi[1] · Scott P. Ardoin[2] · Katherine S. Binder[1]

Abstract

The ability to form a mental model of a text is an essential component of successful reading comprehension (RC), and purpose for reading can influence mental model construction. Participants were assigned to one of two conditions during an RC test to alter their purpose for reading: concurrent (texts and questions were presented simultaneously) and sequential (texts were presented first, then questions were shown without text access). Their eye movements were recorded during testing. Working memory capacity (WMC) and centrality of textual information were measured. Participants in the sequential condition had longer first-pass reading times compared to participants in the concurrent condition, while participants in the concurrent condition had longer total processing times per word. In addition, participants with higher WMC had longer total reading times per word. Finally, participants in the sequential condition with higher WMC had longer processing times in central regions. Even among skilled college readers, participants with lower WMC had difficulty adjusting their reading behaviors to meet the task demands such as distinguishing central and peripheral ideas. However, participants with higher WMC increased attention to important text areas. One potential explanation is that participants with higher WMC are better able to construct a coherent mental model of the text, and attending to central text areas is an essential component of mental model formation. Therefore, these results help clarify the relationship between the purpose for reading and mental model development.

Keywords Reading comprehension · Eye tracking · Working memory · Individual differences · Test characteristics

✉ Katherine S. Binder
  kbinder@mtholyoke.edu

1   Department of Psychology, Mount Holyoke College, South Hadley, USA

2   Department of Educational Psychology, University of Georgia, Athens, USA

Introduction

Reading comprehension (RC) is an essential skill throughout school and adult life, and many components interact to affect RC. These components include lower-level (e.g., word recognition, decoding) and higher-level (e.g., inference-making) processes to create a moment-to-moment construction of the text (Kendeou et al., 2014; van den Broek & Espin, 2012), which can be thought of as the process of comprehension (van den Broek & Espin, 2012). One widely endorsed theoretical perspective is the Construction Integration model of RC (Kintsch, 1988, 1994). According to this model, the construction of a model is largely bottom-up, relying on lower-level word decoding skills. To integrate new concepts into the representations, readers must hold recently read information in their working memory (WM) and integrate it with new information, as well as integrate it with background knowledge, on a moment-to-moment basis (Daneman & Carpenter, 1980; Kendeou et al., 2014). While reading, readers work on constructing a mental model of the text, which is "coherent" when it is high quality (van den Broek & Espin, 2012). This mental model can be thought of as the "product of comprehension." The quality of the mental model can be assessed in many ways, including through multiple-choice (MC) questions after reading (van den Broek & Espin, 2012). Successful comprehension largely depends on readers' ability to create a high-quality mental model of the text (Kendeou, 2014).

One question of particular interest is whether manipulating a reader's goals can encourage them to construct a better mental model of a text. In the context of RC tests, readers often have access to the text while they answer associated questions (concurrent test format). However, if readers cannot access the text while answering questions (sequential test format), they may alter their reading behavior to meet the increased demand of the task (Ferrer et al., 2017; Ozuru et al.). Prior research indicates that readers allocate more attention to areas of text that are central to understanding the passage when they are not given specific instructions about what is important, such as via comprehension questions (Schraw et al., 1993). Sequential presentation may also require readers to rely more on their working memory capacity (WMC) to reproduce details of the text, whereas concurrent presentation may demand different skills, such as test-taking strategies (Andreassen & Braten, 2010; Ozuru et al., 2007; Schroeder, 2011). Therefore, sequential test formats could encourage readers to form a more coherent mental model of the text. To date, few studies have examined how WMC and test presentation format influence the processes and products of comprehension (Clemens et al., 2020; Keenan et al., 2008; Schaffner & Schiefele, 2013). Eye-tracking methodology allows researchers to observe online reading processes as they unfold over time and to make inferences about underlying cognitive processing (e.g., lower- or higher-level) in addition to measuring RC response accuracy. The present study aims to address this gap in the literature by using eye-tracking methodology to examine participants' online reading behavior during a RC test and investigate how test format and WMC are related to reading processes and outcomes.

Comprehension test format

Different RC test formats require students to use different RC skills and reading strategies (Ferrer et al., 2017; Ozuru et al., 2007; Schaffner & Schiefele, 2013; Schroeder, 2011). Prior research suggests that the ability to develop a coherent mental model may be particularly important in sequential test formats (Kendeou et al., 2014), and that sequential RC tests require students to utilize more higher-level processes than concurrent test formats (Ferrer et al., 2017; Schaffner & Schiefele, 2013; Schroeder, 2011). Ferrer et al. (2017) used Read & Answer software to investigate the impacts of text access on reading behavior. They found that when students knew they would not have text access while answering questions, they re-read more before turning to the questions. The authors proposed that the sequential format encouraged students to read more carefully, perhaps prompting them to utilize more higher-level processes to create a coherent mental model of the text. Additionally, Schroeder (2011) investigated the relationship between WMC, reading strategies, and RC in adolescent readers using concurrent vs. sequential test formats, and found that the ability to create an accurate mental model of the text had a larger impact on RC performance in the sequential condition than in the concurrent condition. Schaffner and Schiefele (2013) found similar results when examining the relationship between cognitive and motivational factors and RC in approximately 450 eighth and ninth-grade students. In the sequential condition, reasoning ability and inferencing skills strongly predicted RC performance. However, in the concurrent condition, these same higher-level processing skills did not predict RC performance. The researchers posited that the sequential format require readers to rely more heavily on mental representations of texts.

Despite a general consensus that mental models differ in importance based on test format (Ferrer et al., 2017; Schroeder, 2011), the effect of test format on comprehension response accuracy is less clear. Schroeder (2011) reported no difference in MC question response accuracy between concurrent and sequential conditions. Conversely, Cerdán et al. (2021) and Ozuru et al. (2007) found that response accuracy on some types of MC items were lower in the sequential condition than the concurrent condition. Therefore, more research is needed to determine if test format affects MC question response accuracy.

WMC and RC

A substantial body of evidence suggests that WMC is an important component of RC, and it may become even more important when students answer comprehension questions without text access (Andreassen & Braten, 2010; Schroeder, 2011). When having to answer questions without text access, readers must be able to hold recently read information in their WM and update that information as they continue to read and process new information (Daneman & Merikle, 1996). Readers must be able to effectively manage both the storage and updating components of WM to achieve successful comprehension (Daneman & Merikle, 1996), while also making strategic decisions during reading such as spending more time on central text areas or

rereading to create a high-quality mental representation of the text. Daneman and Merikle (1996) conducted a meta-analysis of 77 studies investigating the relationship between WMC and various types of language comprehension tasks. WMC was moderately correlated with RC performance, and they posited that readers with lower WMC were at a disadvantage when integrating new ideas into their mental model. They found that elementary-age poor readers exhibited difficulty in modifying the contents in WM and struggled to control irrelevant information (Carretti et al., 2005).

WMC is also related to other higher-level RC processes. For example, Yeari (2017) found that university students with a higher WMC generated predictive inferences faster than lower-span participants, and that higher-span participants could better inhibit less-relevant inferences. Additionally, higher-span participants generated more bridging inferences than lower-span participants. Similarly, Dutke and Von Hecker (2011) reported that participants with a higher WMC were better able to adjust their mental model of the text based on new information, and that higher-span readers could more easily disregard "distractor" information that did not fit into their mental model. Clearly, WMC is an important higher-level cognitive resource that influences students' ability to construct a coherent mental model of the text.

Centrality of textual information

One skill that may interact with RC test format and WMC is the ability to identify and retain important textual information. Some areas of text, such as main ideas, are more important to understanding the whole passage than components such as supporting details (McCrudden & Schraw, 2007). McCrudden and Schraw (2007) define important textual ideas as "essential… to understand[ing] the text" (p. 114) and as "author defined and… cued by various characteristics internal to the text" (p. 114). In this study, we refer to important ideas as *central* text areas and less important text areas as *peripheral* areas. Prior research (Schraw et al., 1993) suggests that readers automatically judge the centrality of text components and devote more attention to central than peripheral areas unless given specific relevance instructions. Given this tendency to attend to central text areas, it follows that identifying and integrating central text segments is an important piece of creating a coherent mental model of the text (Kendeou et al., 2014).

However, the influence of centrality on RC varies based on the purpose for reading, including whether participants receive specific instructions that guide their attention before reading (McCrudden & Schraw, 2007; Schraw et al., 1993). Prior work has described varying effects of centrality based on purposes for reading. For example, Yeari and colleagues (Yeari & Lev, 2021; Yeari et al., 2015) reported varying results about the relationship between centrality and text processing times, which may be the result of different reading purposes. In Yeari and Lev (2021) adult participants were instructed to recall text ideas and identify central text areas after reading. Both good and poor comprehenders had longer reading times on central information than peripheral information. Interestingly, these longer reading times only appeared during rereading—there was no difference in first-pass reading time

between central and peripheral information. Conversely, Yeari et al. (2015) manipulated participants' purpose for reading (such as reading for pleasure versus reading to answer test questions), and results indicated that participants spent more time reading central information than peripheral information, but only during initial reading. When participants knew they would have to answer open-ended or MC questions after reading (both test formats were sequential), they spent more time re-reading the peripheral information, which effectively canceled out the initial centrality effect. The authors suggest that re-reading peripheral information may be a corrective strategy to prepare for unknown questions. However, participants remembered more central information than peripheral (measured by a surprise MC test after reading) regardless of the purpose for reading. Evidently, the purpose for reading influences how and when readers devote cognitive resources to processing central information.

A significant gap that remains in the literature is whether more subtle differences in reading task demands also influence readers' behavior on central information. College-age readers routinely complete both concurrent (i.e., standardized tests) and sequential (i.e., reading a paper to discuss in class) reading tasks, yet no research to date has investigated whether altering test format influences reading behavior on central information. Allocating increased attention to central areas of the text is one indicator that readers are attempting to create a coherent mental model of the text, which is a behavior that becomes increasingly important when reading tasks are difficult. It is important to know whether skilled readers adjust their reading behaviors according to task demands to better understand the types of instruction that may be beneficial for pre-college students. The present study investigated this gap by measuring processing times on central text areas during sequential and concurrent RC test formats.

Current study

The current study investigated the roles of test format, WMC, and text centrality in students' online reading processes and RC-test performance. Online RC is difficult to study because it is a silent task, and many studies (e.g., Anmarkrud et al., 2013; Farr et al., 1990) have utilized student self-reports to investigate reading behaviors. These self-reports were potentially inaccurate and may have disrupted students' natural reading processes (Cordón & Day, 1996). Other studies, such as Ferrer et al. (2017), used a moving-window paradigm to measure processing time, which allows readers to see one word or sentence at a time. Although readers can sometimes return to previous portions of text using self-paced reading, only one section of text appears at a time, which is unlike natural reading. Eye-tracking more closely mimics real-world reading, as participants can freely move between sections of the text, and it is a reliable and valid procedure that allows researchers to observe online RC processes with minimal interruption (Rayner et al., 2013). Additionally, eye-movement data allows us to make inferences about underlying cognitive processes during reading. Measures such as first fixation duration and gaze duration are thought to represent lower-level processing (such as decoding and word recognition), while

regressions and total reading time are thought to indicate higher-level processing (such as inference generation; Rayner et al., 2013). Examining several eye-movement measures allows us to observe how processing unfolds over time. The present study adds nuance to the literature by using eye tracking to examine how WMC and text centrality may impact processing behavior in concurrent vs. sequential test formats throughout the RC process.

College-aged participants' eye movements were monitored as they completed the Nelson Denny Comprehension Test (Brown, 1960). Half of the participants saw the passage and questions concurrently, and half saw the passage, then saw the questions without text access (sequential). Participants' WMC was assessed. Additionally, with the help of independent raters, we identified central and peripheral regions of the passage and compared the reading times for these section types.

We investigated several research questions. First, we examined whether WMC and test format affected RC response accuracy. Because WMC is an important component of RC (Daneman & Merikle, 1996), we predicted that WMC would be positively correlated with RC response accuracy. Due to the conflicting findings regarding the effect of test format on response accuracy (Ozuru et al., 2007; Schroeder, 2011), we also examined the relationship between test format and response accuracy. We predicted that individuals with higher WMC would have higher response accuracy than lower WMC participants, and that this effect would be larger in the sequential condition than in the concurrent condition. WMC may become more important when readers answer comprehension questions without text access because they must rely solely upon their mental model of the text to answer questions, whereas readers in the concurrent condition can utilize non-memory-based strategies to answer MC questions (Ozuru et al., 2007; Schroeder, 2011).

We also examined how test format was related to eye-movement measures that reflect both lower-level processing skills (i.e., word recognition) and higher-level processing skills (i.e., integrating information, forming inferences, updating the mental model). We expected that participants in the sequential condition would engage in more higher-level processing, which would be reflected in longer total reading times and more regressions, because the sequential condition likely requires readers to develop a more coherent mental model of the text (Ferrer et al., 2017; Schaffner & Schiefele, 2013; Schroeder, 2011).

Additionally, we investigated how text centrality influenced reading behavior across test format and WMC. We expected that participants would have longer initial reading times and more regressions on central areas than peripheral areas because prior research has shown that readers attend more to central than peripheral information during first-pass reading (Yeari et al., 2015). Although our construct of initial reading is slightly different from that in Yeari et al. (2015), both constructs include reading that occurred before participants answered RC-test questions. Additionally, we predicted that centrality and test format would interact; we expected that participants in the sequential condition would show more regressions and have longer total reading times in central text areas than in peripheral areas, but that there would be no difference in reading time between central and peripheral areas in the concurrent condition. Readers show a preference for central information when they have no other instructions regarding what is important in the text (Schraw

et al., 1993), so participants in the sequential condition likely devoted substantial attention to processing central information. However, participants in the concurrent condition may have used other information to guide their reading, such as information in the questions or answer options. We also expected that participants would have longer total reading times and more regressions in central areas than peripheral areas, and that the differences would be greater for participants with higher WMC than lower WMC. Participants with higher WMC are likely better able to identify, attend to, and retrieve information from central areas of text, whereas lower WMC participants may not have the cognitive resources to utilize such higher-level processes (Yeari & Lev, 2021).

Method

Participants

We recruited 90 students from a small liberal arts college in the Northeast region of the United States, aged 18–36 years ($M = 20.4$, SE = 2.4). Participants formed a diverse racial pool with 29 White, 30 Asian, 24 Black, 1 Native American, 4 Hispanic, 1 Middle Eastern, and 1 who declined to identify race. Participants were screened for reading disabilities before participation. Participants were randomly assigned to either the concurrent ($n = 44$) or sequential condition ($n = 46$). If necessary for reading, they were allowed to wear glasses or contact lenses. They were compensated with either course credit or a raffle entry for a \$25 gift card.

Apparatus

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. The camera was placed 65 cm from the SR Research chinrest to minimize head movements. 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 24-in. monitor placed 93 cm from the SR Research chinrest. The passage was 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.

Materials

The Nelson–Denny comprehension subtest (Brown, 1960) Form G was used to assess participants' comprehension skills. Six of the seven expository passages were used, as they were of similar length ranging from 15 to 21 lines of text. Each passage had five multiple-choice comprehension questions, which were a mix of literal ($n = 14$) and inferential questions ($n = 16$), and participants were required to answer all questions. Passages ranged in length from 2 to 4 paragraphs ($M = 2.67$), were

composed of 8–13 sentences ($M = 11.5$), and had 165–236 words ($M = 203.17$). Flesh-Kincaid reading-grade levels for five of the six passages ranged from 9.3 to 10.9, while the remaining passage was a bit of an outlier at 14.7. The passages were also similar in terms of difficult/complex vocabulary. The passages ranged from 15.6% to 18.14% of words that were considered complex, again with an outlier of 25.75%, which was the same passage as above. See Fig. 1 for an example of the concurrent and sequential formats.

Automated operation span task

After the RC test, an E-prime 2.0.8 automated Operation task was administered to measure WMC (Unsworth et al., 2005). Participants solved a math problem, picked the correct answer, then memorized a letter that appeared after the equation. For example, the first screen had "5+9," the second screen had "10? True or False," and the last screen had an "A." After a random number of equations and letters, participants were asked to recall all the letters in the order they saw them. The automated Operation task program generates a score which is the sum of the total number of

Concurrent – Passage and Questions

Fig. 1 Passage and question presentation in the concurrent and sequential conditions

N. Scott Momaday was born in Lawton, Oklahoma. His father was a Kiowa Indian, and his mother was the great-granddaughter of a Cherokee. Both parents taught at Indian reservations. He spent his boyhood on a number of reservations in the Southwest, acquiring a close knowledge of American Indian culture and history. He graduated from the University of New Mexico in 1958 and went on to obtain a Ph.D. in English in 1963 from the Stanford University. He became a professor of English at Stanford.

Momaday's first book, *House Made of Dawn* (1968), dealing with a young Indian's attempt to reconcile the values of an old way of life with those of the modern world, was awarded the Pulitzer Prize for fiction. Momaday had earlier written a historical study of the Kiowa tribe, and later he enlarged this work to include his impressions of the contemporary history and culture of his people. This expanded book, published under the title *The Way to Rainy Mountain*, is of the most eloquent and perceptive accounts of American Indian life and culture yet to appear. Momaday, who has said that he considers himself primarily a poet, is the author of two books of poems. These volumes reflect his interest in American Indian subjects, but they also contain poems in philosophical and meditative vein. In 1976 he published *The Names*, a memoir about his parents' lives and his boyhood on various reservations.
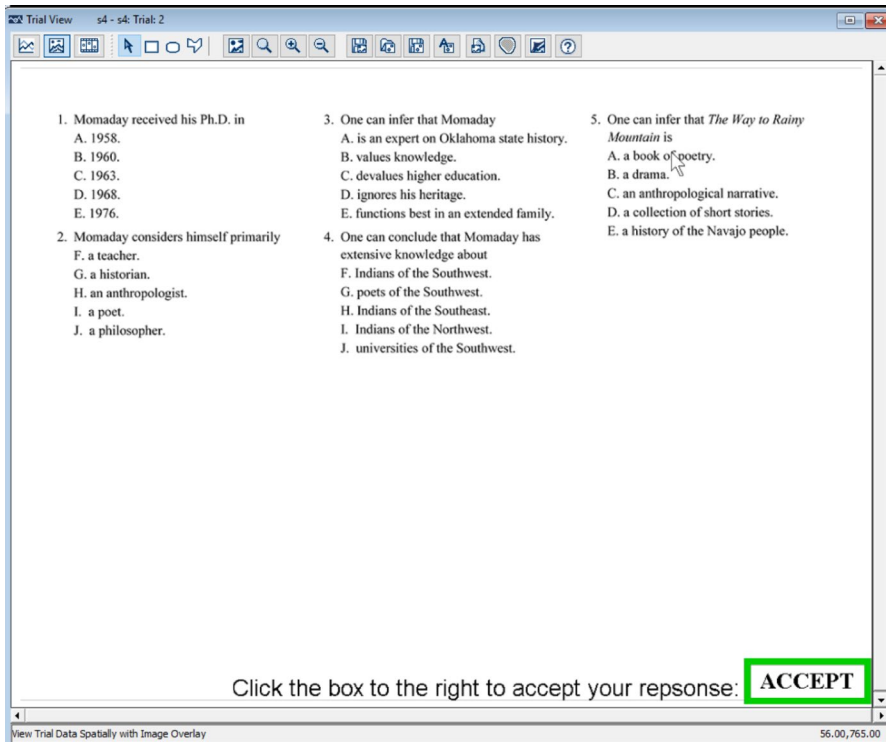
Sequential – Passage

Fig. 1 (continued)

correct letters recalled in the correct position. The automated operation span task has a test–retest reliability of 0.83 and good internal consistency ($\alpha = 0.78$), thus is an accurate measure of WMC for college students (Unsworth et al., 2005). In addition, Redick et al. (2012) conducted a meta-analysis ($N = 6000$) on automated complex span tasks (including automated operation span) and reported strong reliability and validity across all tasks.

Procedure

We obtained IRB approval before recruiting participants, and we sought informed consent from participants before beginning the experiment. After signing the consent form, participants were fitted to the eye-tracking apparatus. The camera was focused on one of the participant's eyes, and the system was calibrated. Participant viewing was binocular. The participant then completed a practice passage and two questions in the format consistent with their assigned condition (sequential or concurrent), so they were aware of their test format before beginning the experimental trials. Six passages in either the sequential or concurrent format were presented to the participant in a randomized order. Participants in the concurrent condition saw both the passage and questions on one screen, allowing them to read the passage and

Sequential – Questions

Fig. 1 (continued)

questions in their preferred order and refer to the passage when responding to questions. In contrast, participants in the sequential condition were informed that they would not see the questions until clicking a button signifying that they had read the passage, and could not refer to the text when answering questions. In both conditions, participants responded to questions by selecting their answers with a mouse click. Responses could be changed before progressing. Once students answered all questions, they clicked a box to proceed to the next screen. The eye-tracking program automatically recorded and saved all answers. Following the RC test, participants completed the WMC test. Combined, the RC test and WMC task took approximately 30 min. Participants were tested individually.

Centrality judgements

A norming study was conducted to identify the sections of the passages that were most and less central to the meaning of the passage. Twelve participants who did not participate in the eye-movement study were recruited for the norming study. Researchers have used anywhere from 3 (Yeari et al., 2015; Yeari et al., 2017) up to

20 raters (Miller & Keenan, 2009; Yeari & Lantin, 2021; Yeari & Lev, 2021). Participants were given the same six passages and asked to highlight sections that they found most central to the overall meaning of the passage. Participants did not see the RC questions. When eight or more of the 12 participants made the same judgment, the section was deemed central. All other text was classified as peripheral. We calculated Cronbach reliability coefficients for these ratings for each passage, and coefficients ranged from 0.80 to 0.98 across the six passages, with an average of $\alpha = 0.92$. These reliability coefficients are comparable to other studies (e.g., Miller & Keenan, 2009 $\alpha = 0.92$; Yeari et al., 2015 and 2017, $\alpha = 0.80$; Yeari & Lantin, 2021 and Yeari & Lev, 2021, $\alpha = 0.93$.). On average, about 11% of the information contained within the text was classified as central. The average length of this information was 4.79 contiguous words (SD = 2.67). Additionally, these regions were spread across the texts. As mentioned above, the participants who rated centrality never saw the RC questions, but we did examine how well the central regions lined up with the questions. Surprisingly, very few of the RC questions addressed the ideas that were deemed central: for two of the passages, none of the central items were asked about in the questions, for three passages, one question related to the central material, and for the remaining passage, two questions were related to the central material.

Results

Nelson Denny response accuracy and multiple eye-movement measures (first fixation duration, gaze duration, total time, and regressions) were the dependent variables. Fixations times for each word in the passage were used in the analyses. We averaged across all words in the passage to obtain average eye-movement measures (e.g., for first fixation duration, we measured the amount of time on each initial fixation on the word, and then took an average of all words in the passage to generate the measure for a participant for a particular passage). Linear mixed-effects (LME) regression models were run for each dependent measure, using the MIXED command in SPSS and REML estimation procedures, but we used a generalized linear mixed-effect model using a Poisson distribution for the regression eye-movement data since this dependent variable was a count variable. The appeal of these type of regression models was partially due to the ability to incorporate by-subject and by-item influences as random effects into the same analysis (Baayen et al., 2008; Bates et al., 2018). The intercepts for subjects and items were included as random effects, and the fixed effects in all models for global analyses included test format (concurrent vs. sequential) and WMC scores, the interaction between test format and WMC. Test format was a categorical variable, and WMC was a continuous variable. We conducted analyses on four eye-movement measures, so while the family-wise alpha level was $p = 0.05$ we corrected that to $p = 0.01$ to accommodate for all of our dependent measures. Furthermore, Bonferroni corrections were used in post-hoc analyses. WMC scores were centered by subtracting the mean score from each participant's individual score. The uncentered mean for the WMC task was 59.26, SD = 12.57. See Tables 1 and 2 for means and standard deviations for all conditions.

Table 1 Descriptive statistics for nelson Denny response accuracy scores and WMC by WMC groups

		Higher WMC	Lower WMC
Nelson Denny response accuracy	Mean	83.47	79.84
	SD	10.81	9.86
	Skewness	− .62	− .49
	Kurtosis	− .92	− .61
	25% percentile	73.30	73.30
	50% percentile	86.70	83.30
	75% percentile	90.00	86.70
WMC	Mean	68.65	48.63
	SD	4.77	9.93
	Skewness	− .35	− .93
	Kurtosis	− .49	1.03
	25th percentile	66.00	45.00
	50th percentile	69.00	51.00
	75% percentile	90.00	56.00
Correlation between ND and WMC		.40 (p < .001)	.27 (p < .001)

Table 2 Descriptive statistics for global eye-movement measures

Dependent measure	Concurrent high WMC mean (SD)	Concurrent low WMC mean 7 (SD)	Sequential high WMC mean (SD)	Sequential low WMC mean (SD)
First fixation	229.11 (21.89)	235.71 (23.63)	239.67 (27.74)	242.76 (27.74)
Gaze duration	258.21 (32.68)	260.19 (30.63)	274.20 (42.14)	288.11 (43.60)
Total time	376.89 (80.78)	333.93 (56.42)	338.27 (156.53)	243.83 (65.02)
Regressions	0.36 (.20)	0.30 (0.15)	0.40 (0.20)	0.26 (0.14)

First fixation, gaze duration, and total time are reported in msec

In another set of analyses, we examined eye-movement measures on central and peripheral text areas. See Table 3 for means and standard deviations for both Test Format conditions across WMC groups. For those analyses, test format, WMC, centrality, test format*WMC, test format*centrality, WMC*centrality, and test format*WMC*centrality were fixed factors. Data were evaluated for normality prior to analyses, and all dependent measures had skew and kurtosis values within accepted limits. There was no difference between WMC across the two format conditions (F (1, 88) = 1.40, $p = 0.24$; $M_{concurrent} = 57.74$ and $M_{sequential} = 60.85$).

Table 3 Descriptive statistics for eye-movement measures and centrality

Dependent measure	Concurrent high WMC mean (SD)	Concurrent low WMC mean (SD)	Sequential high WMC mean (SD)	Sequential low WMC mean (SD)
Central				
First Fixation	27.78 (9.08)	31.68 (8.32)	30.91 (8.14)	31.10 (8.61)
Gaze Duration	31.12 (10.44)	35.92 (10.29)	36.32 (10.23)	37.08 (12.35)
Total Time	80.70 (49.29)	71.54 (31.91)	81.97 (60.21)	59.07 (29.60)
Regressions	0.44 (0.28)	0.36 (0.18)	0.13 (0.10)	0.11 (0.07)
Peripheral				
First Fixation	31.18 (9.48)	35.57 (7.75)	33.92 (7.47)	34.31 (8.86)
Gaze Duration	34.30 (10.94)	39.82 (9.72)	39.02 (9.76)	39.57 (11.50)
Total Time	76.16 (32.59)	69.20 (24.33)	60.99 (25.34)	55.88 (23.38)
Regressions	0.04 (0.05)	0.03 (0.03)	0.01 (0.03)	0.02 (0.14)

Measure for first fixation, gaze duration and total time are reported in ms/char

Research question 1: how do test format and WMC relate to RC response accuracy?

Participants in the concurrent condition had higher response accuracy ($M = 86\%$, $SE = 1.55$) than participants in the sequential condition ($M = 77\%$, $SE = 1.74$). WMC was also a significant predictor in the model, ($B = 0.016$), with higher WMC associated with higher response accuracy scores. The interaction between WMC and test format was not significant.

Research question 2: how do test format and WMC influence eye-movement measures?

Several eye-movement measures can be extracted from a single reading record, with some measures thought to reflect lower-level cognitive processing such as word recognition, and others thought to reflect higher-order processing such as text comprehension. First fixation duration and gaze duration served as the lower-level eye-movement measures; total time and regressive eye movements served as higher-order text processing measures. For all reading measures, an average was obtained across all words per passage. See Table 4 for *F*s and *p*s for all of the fixed effects.

Table 4 Fixed effects for question 2

	Test Format		WMC		Test Format*WMC	
	F	*p*	*F*	*p*	*F*	*p*
Accuracy	15.59	<.001	15.61	<.001	1.64	.20
First fixation	5.41	.02	0.86	.36	0.10	.75
Gaze duration	8.47	.005	0.70	.41	0.07	.80
Total time	14.79	<.001	6.41	.01	1.69	.20
Regressions	3.21	.07	14.35	<.001	3.95	.05

Average first fixation and gaze duration

First fixation duration is the length of the first fixation on a word, regardless of the number of fixations on a word. Gaze duration is the sum of consecutive fixations made on a word before leaving the word. Participants in the sequential condition had longer average first fixation and gaze durations (first fixation: $M = 246.64$ ms, $SE = 3.83$; gaze: $M = 285.43$ ms, $SE = 5.32$) than participants in the concurrent condition (first fixation: $M = 234.23$ ms, $SE = 3.95$; gaze: $M = 264.41$ ms, $SE = 5.51$). WMC was not a significant predictor, (first fixation: $B = -0.196$; gaze: $B = -0.241$). There was no interaction between test format and WMC on first fixation or gaze duration.

Total time spent on a word

Total reading time on words is the sum of all fixations on a word and includes regressions back to the word, and any rereading on that word. Participants in the concurrent condition spent more total time on words ($M = 362.65$, $SE = 14.60$) than participants in the sequential condition ($M = 293.22$, $SE = 14.13$). WMC was also a significant predictor in the model, ($B = 1.85$). Participants with higher WMC spent more total time on words than participants with lower WMC. There was no significant interaction between WMC and test format.

Regressive eye movements

Regressions are the number of times participants return to a word after initially leaving that word. There were no significant differences in regressions between conditions. WMC was a significant predictor in the model, ($B = 0.003$); participants with higher WMC had more regressions to a word than participants with lower WMC. There was no interaction between WMC and test condition.

We want to note that to more directly compare the concurrent and sequential conditions, we examined reading times before participants started answering questions for the concurrent condition. However, participants in the concurrent condition could have used different passage reading strategies, thus, we examined the strategies used in the concurrent condition. Participants typically read

each passage completely before reading the questions (64.4%—the percentages in this section refer to the number of passages since some participants varied the strategy they used across the six passages). Next, participants read some of the passage, but then read some of the questions (30.3%). We defined "reading the questions" as at least three fixations from left to right in the new region. Very few participants read the questions prior to the passage (3% read all questions before the passage, while 2.3% read some of the questions before the passage). When participants read some of the passage, then went to the questions (30.3%, referenced above), they often then read the passage completely (78.4% of the 30.3%). Very few participants never read the whole passage (only 21.6% of the 30.3% referenced above). Even when participants did not completely read the passage, they, on average, read 88% of the passage. Thus, nearly all participants, most of the time, read the majority of the passage. In addition, when they read the questions before finishing the passage, most of the time (65% of the 30.3%), participants only read one question before returning to the passage. If the participant left the passage, read one question, then returned to the passage, their reading time once they returned to the passage was included in the initial reading data. Because the data in the concurrent condition are complicated by these strategies, we ran another analysis on the total word reading time, but this time, only the 64.4% of the data that fell into the "passage first" strategy for the concurrent condition were included in these analyses. Interestingly, the above reported patterns were maintained. Participants had longer total word reading times in the concurrent condition ($M = 355.39$, SE $= 16.02$) compared to the sequential condition ($M = 293.22$, SE $= 14.13$). WMC was also a significant predictor in the model, ($B = 2.79$, $F = 4.85$, $p = 0.03$), with higher WMC individuals spending more time on individual words relative to participants who had lower WMC. The interaction was not significant, $F = 2.18$, $p = 0.14$.

Research question 3: analysis of eye-movement measures across test condition, WMC, and centrality of information

Independent raters identified sections of each passage that were central to the overall meaning. Within each of these sections, the times and counts were summed and then divided by the number of characters in the section, since these regions differed in size. See Table 5 for all Fs and ps for the fixed effects.

Average first fixation and gaze duration

Participants had longer first fixations and gaze durations on peripheral text regions (first fixation: $M = 33.71$ ms/char, SE $= 0.95$; gaze: $M = 38.19$ ms/char, SE $= 1.20$) than central regions (first fixation: $M = 30.36$ ms/char, SE $= 0.95$; gaze: $M = 35.15$, SE $= 1.20$). WMC was not a significant predictor. Test format was not significant, nor were any of the interactions.

Table 5 Fixed effects for question 3

	F	P
First fixation		
Centrality	113.80	<.001
Test format	0.24	.63
WMC	0.02	.88
Test format*WMC	0.49	.49
Test format*Centrality	0.96	.33
WMC*centrality	0.35	.55
Test Format*WMC*centrality	0.30	.59
Gaze duration		
Centrality	65.43	<.001
Test format	1.13	.29
WMC	0.001	.97
Test format*WMC	0.001	.97
Test format*centrality	1.89	.17
WMC*centrality	0.98	.32
Test format*WMC*centrality	1.06	.30
Total time		
Centrality	21.73	<.001
Test format	23.04	<.001
WMC	14.44	<.001
Test format*WMC	5.19	.03
Test format*centrality	13.78	<.001
WMC*centrality	10.02	.002
Test format*WMC*centrality	8.98	.003
Regressions		
Centrality	563.41	<.001
Test format	172.95	<.001
WMC	14.35	<.001
Test format*WMC	3.95	.06
Test format*centrality	179.04	<.000
WMC*centrality	20.17	<.001
Test format*WMC*centrality	2.55	.12

Total time per word

Participants in the concurrent condition ($M = 43.24$ ms/char, SE $= 4.16$) spent significantly less total time on words than those in the sequential condition ($M = 65.87$ ms/char, SE $= 4.12$). Participants with higher WMC had longer total times ($B = 0.59$). In addition, participants spent significantly more time reading central sections ($M = 58.55$ ms/char, SE $= 3.61$) than peripheral sections ($M = 50.56$ ms/char, SE $= 3.61$). While there were significant two-way interactions between WMC and centrality, centrality and test format, and centrality and WMC, these effects were

qualified by a significant three-way interaction among test format, WMC, and centrality. To investigate this three-way interaction, we divided our sample into two WMC categories, with higher WMC participants having WMC centered scores 0 or above and lower WMC participants having centered scores below 0, and separated the data by test format and WMC category. We then ran additional post-hoc LME models where centrality was the fixed effect. See Fig. 2 for means and standard errors.

For the concurrent condition, there were no differences between reading times for central and peripheral regions across WMC categories (higher: $F=0.92$, $p=0.34$, and lower: $F=0.57$, $p=0.45$. However, for the sequential condition, higher WMC individuals spent more total time in central areas compared to peripheral areas, $F=27.37$, $p<0.001$. No differences were observed for lower WMC individuals, $F=1.61$, $p=0.21$.

Regressions

Participants in the concurrent condition ($M=0.22$, SE$=0.02$) had significantly more regressions than participants in the sequential condition ($M=0.07$, SE$=0.02$). In addition, participants had significantly more regressions into central sections ($M=0.26$, SE$=0.02$) than into peripheral sections ($M=0.02$, SE$=0.02$). There was a significant interaction between test format and centrality. See Fig. 3 for means and standard errors.

Although participants always made more regressions to central regions compared to peripheral regions, the difference was larger for the concurrent condition than the sequential condition (concurrent: $F=755.32$, $p<0.001$; sequential: $F=220.02$, $p<0.001$). There was also an interaction between WMC and
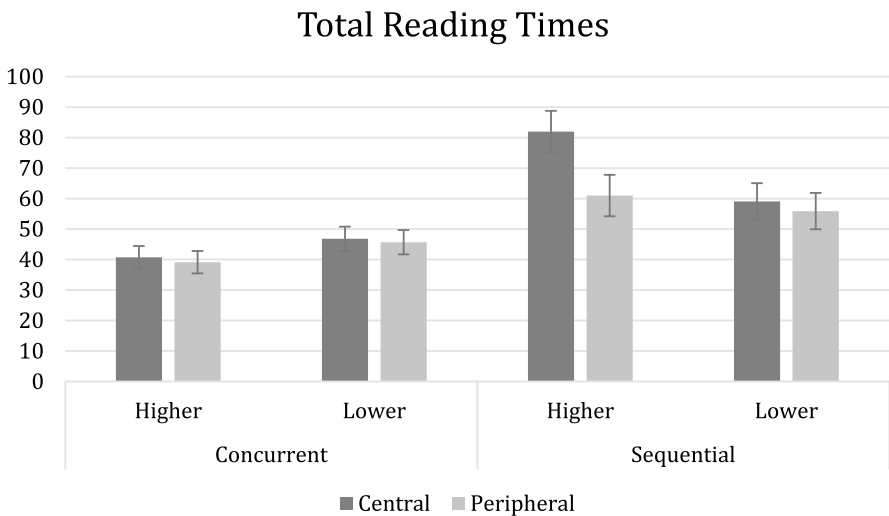

Fig. 2 mean total reading times on central and peripheral text sections between working memory categories
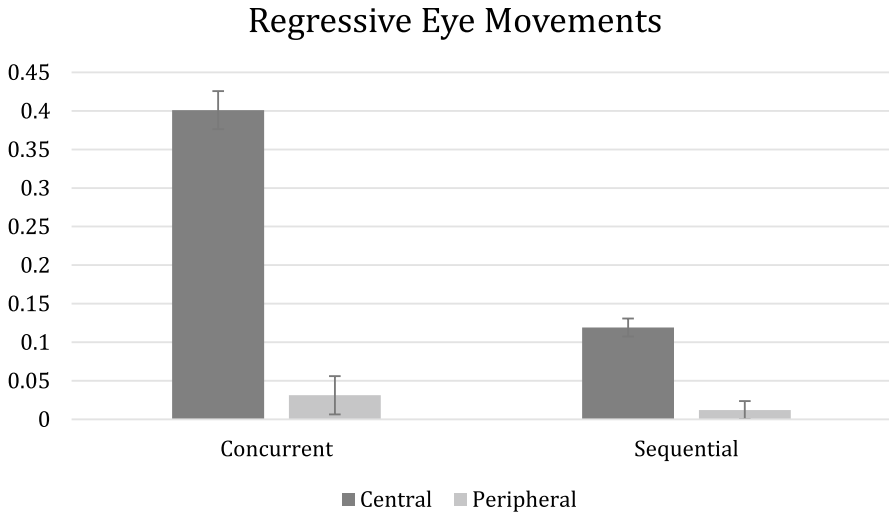
Regressive Eye Movements

Fig. 3 mean regressive eye movements in central and peripheral text sections between concurrent and sequential test formats

centrality, $F = 19.89$, $p < 0.001$. See Fig. 4 for means and standard errors. Although participants always made more regressions to central regions compared to peripheral regions, the difference was larger for higher WMC participants than lower WMC participants (high WMC: $F = 401.98$, $p < 0.001$; low WMC $F = 342.94$, $p < 0.001$).
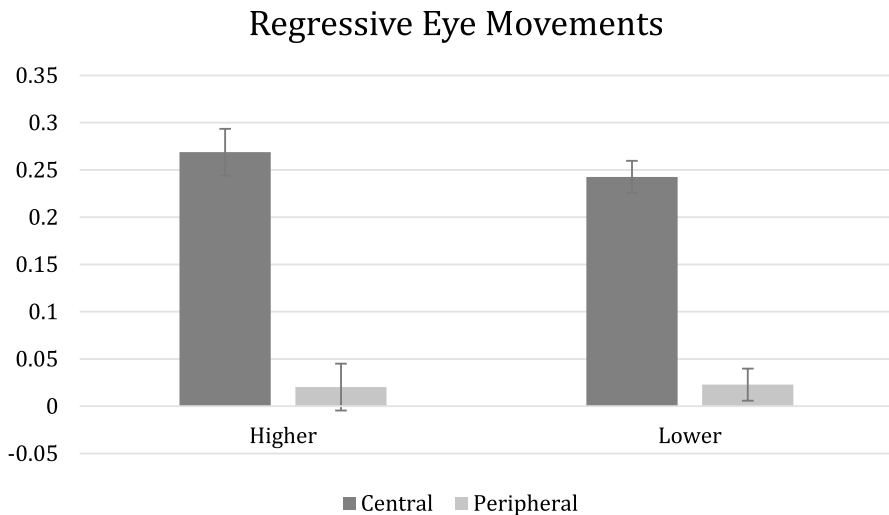
Regressive Eye Movements

Fig. 4 Regressive eye movements in central and peripheral sections of text between working memory categories

Discussion

This study examined how test format (concurrent or sequential), WMC, and centrality of textual information affect reading processes and RC test outcomes, which may influence the formation of a mental model of the text. We examined eye-movement records of college-aged participants' reading behavior to better understand the underlying cognitive processing that may be associated with creating a mental model of the text. To directly compare reading times between the test format conditions, we only analyzed initial passage reading time, or reading time before participants began to answer MC questions. That is, although participants in the concurrent condition could have re-read the passage while answering the questions, this analysis only includes their initial reading time on the passage (including regressions back to sections of text before reading and answering questions). It is important to better understand the factors that influence the formation of a mental model of the text so we can assist readers in creating high-quality mental models across a variety of reading settings.

The present study yielded several important findings. First, test format impacted comprehension response accuracy and processing times. Participants in the concurrent condition had higher response accuracy than participants in the sequential condition. Participants in the concurrent condition also had longer processing times on measures thought to reflect higher-level processes than participants in the sequential condition, which contrasts with our hypotheses. However, participants in the sequential condition had longer first fixation and longer first gaze durations, which are measures thought to represent lower-level processing. This suggests that when participants know they will not have text access while answering questions, they slow down when they are initially reading each word. Second, WMC was related to both response accuracy and processing time. Higher WMC participants had higher response accuracy than lower WMC participants. Interestingly, participants with higher WMC also had longer total reading times on words, which is a measure thought to reflect higher-level processing. However, WMC was not related to eye-movement measures thought to reflect lower-level processing. Contrary to the hypotheses, there was no interaction between WMC and test format on response accuracy. Finally, participants engaged in different reading behaviors depending on text centrality. Surprisingly, participants had longer first fixation and first gaze durations on peripheral information than central information. However, as hypothesized, participants had longer total reading times on words and more regressions in central sections than peripheral sections, suggesting that they used more higher-level processes while reading central sections. Further analyses of interaction effects showed that participants with higher WMC in the sequential condition had longer reading times on words and more regressions in central areas of text, but there was no difference in processing times in central and peripheral areas for participants with lower WMC or in the concurrent condition.

Test format and reading comprehension

Consistent with prior research (Ozuru et al., 2007), participants in the concurrent condition had higher response accuracy than those in the sequential condition. Conversely, our test format results for the reading time data are in contrast with most prior research (Ferrer et al., 2017; Schaffner & Schiefele, 2013; Schroeder, 2011), which suggested that readers utilize more higher-level processes in the sequential condition than the concurrent condition. Our findings also contrast with O'Reilly et al. (2018) and Wang et al. (2017), which found that readers had shorter reading times during a concurrent-format RC test than when they answered a summary question after reading (in other words, when they read to form a coherent mental model of the text). In our study, participants in the concurrent condition exhibited longer total word reading times, which is indicative higher-level processing, compared to participants in the sequential condition. However, participants in the sequential condition had longer first fixation and first gaze durations than in the concurrent condition. Therefore, participants had slower initial reading times (possibly indicative of more carefully encoding text information) during the sequential format than the concurrent, but we had expected this pattern for the total reading time measure. Additionally, as predicted, participants with higher WMC in the sequential condition had longer total reading times on words in central sections than peripheral, but there was no difference for participants in the concurrent condition or participants with lower WMC.

The longer total reading times for words in the concurrent condition compared to the sequential condition was surprising. One explanation is that the sequential condition may tax skills we did not measure. For example, Ozuru et al. (2007) found that background knowledge was the most important variable for success in the sequential condition, whereas the concurrent condition required readers to strategically utilize the text. In the present study, the concurrent condition may have prompted readers to use higher-level processes to answer MC questions; however, the sequential condition may have relied on skills such as background knowledge or metacognition.

Another possible explanation for the longer total word reading times in the concurrent condition is that in the concurrent condition, readers may have been distracted by seeing both the questions and text, leading to slower reading times. To investigate this possibility, we conducted a supplementary analysis comparing participants in the concurrent condition who read the text entirely before reading any questions and those who read at least one question before completing the passage. Participants who moved between the text and questions, even once, had longer total times on words and more regressions than participants who read the entire passage first ($F = 7.72$, $p = 0.006$). This is consistent with prior research suggesting that reading the entire passage first (vs. reading the questions first) may facilitate test-taking efficiency (Bayrak Karsli et al., 2020; Yeari et al., 2021), and that reading the questions first may tax memory (Bayrak Karsli et al., 2020). Therefore, the longer reading times in the concurrent condition may indicate distracted reading rather than deep text processing. However, to be clear, even when we included only the participants who read the entire passage first before reading any questions, total word reading times were still longer for the concurrent condition than the sequential condition.

It could be the mere presence of the questions is enough to distract, even though readers have not yet turned to the questions.

Working memory and reading comprehension

The relationship between WMC and comprehension response accuracy was somewhat consistent with past research (e.g., Daneman & Merikle, 1996). Daneman and Hannon (2001) suggest that creating a mental model of the text is harder for lower WMC individuals, and this affects their performance. Correspondingly, our results suggest that even in the skilled college population, participants with higher WMC may be able to better encode important information in WM and integrate it in their mental model. First, lower WMC participants achieved lower response accuracy than higher WMC participants. Additionally, participants with lower WMC spent less time on measures thought to reflect higher-level processes than participants with higher WMC; however, there was no difference in lower-level processing between WMC groups. This suggests that readers with lower WMC utilized different strategies than readers with higher WMC. These results are in line with prior work suggesting that readers with lower WMC struggle with many aspects of the RC process (Carretti et al., 2005; Dutke & Von Hecker, 2011; Yeari, 2017), which may be negatively related to their overall RC performance. Together, these findings suggest that readers with lower WMC may have struggled to construct a mental model of the text, and that WMC is particularly important for the higher-level processing needed to form a coherent mental model of the text, which is consistent with past work by Dutke and Von Hecker (2011) and Yeari (2017).

One reason participants with higher WMC may have utilized more higher-level processing and achieved higher response accuracy is that they may also have better metacognitive skills. Metacognitive skills in reading are one's ability to plan, monitor and regulate their reading behaviors (Burin et al., 2020). Readers with good metacognitive skills will set their goals before reading, evaluate their comprehension throughout the process, and if they capture inconsistencies or breakdowns in comprehension, they will repair them by re-planning and using a different strategy (Zargar et al., 2020). Touron et al. (2010) suggested that WMC and metacognitive skills are closely related. Their work found that individuals with higher WMC were better at using particular strategies for learning, regulating cognition, and monitoring problem-solving processes (Touron et al., 2010). Research also suggest that metacognition is an important variable for RC (see Baker, 1989 for a review; Burin et al., 2020; Soto et al., 2019). Therefore, it is possible that participants with higher WMC also had better metacognitive skills, which helped them build and update their mental models of the texts. Importantly with respect to the current findings, higher WMC participants made more regressive eye movements and subsequently had longer total word reading times on central regions compared to more peripheral ideas, but only in the sequential condition. This suggests that higher WMC participants adjusted their strategies (a metacognitive component) while they were initially reading the passages,

knowing they would not have access to the passage while answering questions. There was no such evidence of this type of strategy use for participants with lower WMC.

Centrality of textual information

We also observed impacts of text centrality on cognitive processing that were consistent with prior work. Participants had longer total reading times per word and more regressions in central areas than peripheral, under certain conditions. This result is consistent with Yeari et al. (2015), which found that participants had longer initial passage reading times on central areas than peripheral areas. However, participants had longer first fixations and first gaze durations on peripheral than central areas, which we did not expect. This finding contrasts with past literature (Yeari et al., 2015; Yeari & Lev, 2021), which found that there was no difference in first-pass reading times between central and peripheral areas of text. The present study differs from Yeari and colleagues' work in that it utilized differing RC test formats to manipulate task difficulty and purpose for reading. Because both test formats utilized a standard RC test with associated MC questions, readers may have attended more to peripheral information to prepare themselves for a wide variety of MC questions.

Although all participants had more regressions to central areas of text, the effect was greater for participants with higher WMC. Higher-achieving readers often have greater WMC, and therefore better executive functioning to direct important processes such as WM updating skills, (Daneman & Merikle, 1996) and are also typically better able to identify and integrate central text areas into their mental model (Yeari & Lev, 2021). Our results seem to support that individuals with higher WMC may be better at integrating central ideas into their mental model. However, while participants had more regressions in the central sections than peripheral sections in both conditions, the effect was larger in the concurrent condition, which contrasts with prior findings (Ferrer et al., 2017; Schaffner & Schiefele, 2013; Schroeder, 2011). This may again be due to possible distracted reading in the concurrent condition.

Importantly, we found that participants with higher WMC had longer total reading times in central areas of text, but only in the sequential condition. The sequential condition likely required readers to rely more on their mental model of the text, and attending to central areas of text may be indicative of attempting to form a coherent mental model. This finding suggests that good readers strategically devote attention to important information to help them build a coherent mental model of the text when it is most needed. Perhaps college-age readers with lower WMC do not have the necessary metacognitive skills to identify that they need to change their reading behavior in the sequential condition, or they do not have enough cognitive resources available to change their strategy. Future research should investigate additional reader characteristics that may facilitate successful RC during differing task demands.

Limitations and directions for future research

Results of this study should be evaluated in the context of its limitations. First, the literal MC questions that accompany the Nelson-Denny passages used here rely almost exclusively on peripheral information (although most inferential questions rely on integration of multiple ideas in the passage). Therefore, performance on the associated questions does not necessarily indicate how much central information participants remembered. Additionally, attending to peripheral information may have been a successful strategy to answer MC questions correctly. However, prior research suggests that participants remember central information better than peripheral information regardless of purpose for reading, and that attending to central information is an essential part of forming a coherent mental model (Yeari & Lev, 2021; Yeari et al., 2015). Creating a coherent mental model of the text is an important skill that influences RC test performance, regardless of whether questions address central or peripheral details. Future research should use MC questions specifically designed to measure memory of central details to further investigate the questions posed in this study. In addition, past research has indicated that the location of the central and peripheral information (e.g., beginning, middle, end of the text) may influence how that information is recognized and processes (Swett, et al., 2013). Since we used materials from an established RC battery, we were unable to manipulate the location of the central information. Thus, future work should investigate how placement of central ideas interacts with the other variables we did examine.

Additionally, this study investigated the relationship between centrality of text information and cognitive processing, but it did not analyze the influence of text relevance. In the context of RC tests that utilize MC questions, text relevance refers to areas of passages that contain the answers to associated questions. Prior research (e.g., McCrudden & Schraw, 2007; Yeari et al., 2015) indicates that centrality and relevance of textual information influence cognitive processing in different ways, and relevance also influences the construction of a mental model of a text. Completing a RC test demands both the ability to recognize and encode central information and the ability to recognize and utilize the information necessary to answer the associated questions. If readers with greater WMC do build a better mental model of the text, we would expect them to be better able to locate relevant information, integrate it into their mental model, and utilize it to answer associated MC questions. Future research should therefore evaluate whether WMC is related to processing and utilizing text-relevant information, and if other factors such as test format affect this relationship.

Additionally, the present study used a between-subjects design, and although there were no WMC differences across test format conditions, a within-subjects design would be optimal. Future research should utilize a within-subjects design to address this issue. In addition, as we noted in the Results, assessing initial passage reading times in the concurrent condition was challenging since readers could make their own decisions as to how to approach the task (i.e., passage first or question first). Future work should compare sequential and concurrent conditions by first instructing the participants in both conditions to read the entire passage, then in

the concurrent condition allowing them to refer back to the passage after they view the questions. This might allow for a more direct comparison between the conditions. However, we do want to note that our design allowed for a more naturalistic approach for the readers, and it showed what may complicate the RC assessment in the concurrent condition. That is, students become distracted by the mere presence of the questions, even if they do not read them until after the passage is read.

Finally, the present study did not measure any student characteristics besides WMC. Prior work (e.g., Ozuru et al., 2007; Touron et al., 2010; Yeari, 2017) has shown that many higher-order skills, including background knowledge, inferencing, and metacognitive skills, contribute to successful RC. Future research should examine how other reader characteristics and skills are related to centrality effects and creating a coherent mental model of the text. In addition, there is a great deal of work that has examined how well readers generate inferences during reading. We suspect that had we incorporated a way of assessing the inferences that readers make, that may have revealed differences between higher and lower WMC participants.

Conclusion

The current study found that test format, WMC, and centrality of text information were related to the processes and products of RC. By measuring participants' eye-movements while reading texts and answering questions, we found that participants with higher WMC and participants in the concurrent condition had longer processing times per word and higher response accuracy on MC questions. In addition, participants with higher WMC in the sequential condition spent more time reading the central areas of text, but readers in the concurrent condition and with lower WMC (regardless of test format) did not show this effect. Therefore, WMC is an important skill related to identifying important regions and integrating those ideas into the mental model of the text. However, in general, the sequential test format may not encourage the formation of a more coherent mental model. Future research should investigate factors that can influence how readers build mental models of text to encourage reading for comprehension. In addition, our findings indicate that lower WMC individuals did not spend more time on the central regions of the text. Future work might investigate the possibility of instructing struggling readers on how to identify important concepts and then assess if this instruction helps with RC processes and outcomes.

Declarations

Conflict of interest None.

References

Andreassen, R., & Bråten, I. (2010). Examining the prediction of reading comprehension on different multiple-choice tests. *Journal of Research in Reading, 33*(3), 263–283. https://doi.org/10.1111/j.1467-9817.2009.01413.x

Anmarkrud, Ø., McCrudden, M. T., Bråten, I., & Strømsø, H. I. (2013). Task-oriented reading of multiple documents: Online comprehension processes and offline products. *Instructional Science, 41*(5), 873–894. https://doi.org/10.1007/s11251-013-9263-8

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baker, L. (1989). Metacognition, comprehension monitoring, and the adult reader. *Educational Psychology Review, 1*(1), 3–38. https://doi.org/10.1007/bf01326548

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2018). Parsimonious mixed models. ArXiv:1506.04967v2

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

Brown, J. I. (1960). *The Nelson-Denny reading test*. Houghton Mifflin.

Burin, D. I., Gonzalez, F. M., Barreyro, J. P., & Injoque-Ricle, I. (2020). Metacognitive regulation contributes to digital text comprehension in E-learning. *Metacognition and Learning, 15*(3), 391–410. https://doi.org/10.1007/s11409-020-09226-8

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

Cerdán, R., Máñez, I., & Serrano-Mendizábal, M. (2021). Reading from multiple documents: The role of text availability and question type. *Reading Research Quarterly, 56*(1), 209–220. https://doi.org/10.1002/rrq.380

Clemens, N. H., Hsiao, Y.-Y., Lee, K., Martinez-Lincoln, A., Moore, C., Toste, J., & Simmons, L. (2020). The differential importance of component skills on reading comprehension test performance among struggling adolescent readers. *Journal of Learning Disabilities, 54*(3), 155–169. https://doi.org/10.1177/0022219420932139

Cordon, L. A., & Day, J. D. (1996). Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology, 88*(2), 288–295. https://doi.org/10.1037/0022-0663.88.2.288

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

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, 130*(2), 208–223. https://doi.org/10.1037/0096-3445.130.2.208

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422–433. https://doi.org/10.3758/BF03214546

Dutke, S., & von Hecker, U. (2011). Comprehending ambiguous texts: A high reading span helps to constrain the situation model. *Journal of Cognitive Psychology, 23*(2), 227–242. https://doi.org/10.1080/20445911.2011.485127

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

Ferrer, A., Vidal-Abarca, E., Serrano, M. -Á., & Gilabert, R. (2017). Impact of text availability and question format on reading comprehension processes. *Contemporary Educational Psychology, 51*, 404–415. https://doi.org/10.1016/j.cedpsych.2017.10.002

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281–300. https://doi.org/10.1080/10888430802132279

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. https://doi.org/10.1111/ldrp.12025

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 294–303. https://doi.org/10.1037/0003-066X.49.4.294

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

Miller, A. C., & Keenan, J. M. (2009). How word decoding skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia, 59*(2), 99. https://doi.org/10.1007/s11881-009-0025-x

O'Reilly, T., Feng, D. G., Sabatini, D. J., Wang, D. Z., & Gorin, D. J. (2018). How do people read the passages during a reading comprehension test? The effect of reading purpose on text processing behavior. *Educational Assessment, 23*(4), 277–295. https://doi.org/10.1080/10627197.2018.1513787

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction, 25*(4), 399–438. https://doi.org/10.1080/07370000701632371

Rayner, K., Ardoin, S. P., & Binder, K. S. (2013). Children's eye movements in reading: A commentary. *School Psychology Review, 42*(2), 223–233. https://doi.org/10.1080/02796015.2013.12087486

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*(3), 164–171. https://doi.org/10.1027/1015-5759/a000123

Schaffner, E., & Schiefele, U. (2013). The prediction of reading comprehension by cognitive and motivational factors: Does text accessibility during comprehension testing make a difference? *Learning and Individual Differences, 26*, 42–54. https://doi.org/10.1016/j.lindif.2013.04.003

Schraw, G., Wade, S. E., & Kardash, C. A. (1993). Interactive effects of text-based and task-based importance on learning from text. *Journal of Educational Psychology, 85*(4), 652–661. https://doi.org/10.1037/0022-0663.85.4.652

Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology, 103*(4), 877–896. https://doi.org/10.1037/a0023731

Soto, C., Gutiérrez de Blume, A. P., Jacovina, M., McNamara, D., Benson, N., & Riffo, B. (2019). Reading comprehension and metacognition: The importance of inferential skills. *Cogent Education, 6*(1), 1–20. https://doi.org/10.1080/2331186x.2019.1565067

Swett, K., Miller, A. C., Burns, S., Hoeft, F., Davis, N., Petrill, S. A., & Cutting, L. E. (2013). Comprehending expository texts: The dynamic neurobiological correlates of building a coherent text representation. *Frontiers in Human Neuroscience, 7*, 853. https://doi.org/10.3389/fnhum.2013.00853

Touron, D. R., Oransky, N., Meier, M. E., & Hines, J. C. (2010). Metacognitive monitoring and strategic behaviour in working memory performance. *The Quarterly Journal of Experimental Psychology, 63*(8), 1533–1551. https://doi.org/10.1080/17470210903418937

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505. https://doi.org/10.3758/bf03192720

van den Broek, P., & Espin, C. A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review, 41*(3), 315–325. https://doi.org/10.1080/02796015.2012.12087512

Wang, Z., Sabatini, J., O'Reilly, T., & Feng, G. (2017). How individual differences interact with task demands in text processing. *Scientific Studies of Reading, 21*(2), 165–178. https://doi.org/10.1080/10888438.2016.1276184

Yeari, M. (2017). The role of working memory in inference generation during reading comprehension: Retention, (re)activation, or suppression of verbal information? *Learning and Individual Differences, 56*, 1–12. https://doi.org/10.1016/j.lindif.2017.04.002

Yeari, M., & Lantin, S. (2021). The origin of centrality deficit in text memory and comprehension by poor comprehenders: A think-aloud study. *Reading and Writing, 34*(3), 595–625. https://doi.org/10.1007/s11145-020-10083-9

Yeari, M., & Lev, N. (2021). Processing and memory of central and peripheral ideas in reading comprehension by poor comprehenders. *Scientific Studies of Reading, 25*(3), 215–233. https://doi.org/10.1080/10888438.2020.1759073

Yeari, M., van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and Writing, 28*(8), 1071–1097. https://doi.org/10.1080/10888438.2020.1759073

Zargar, E., Adams, A. M., & Connor, C. M. (2020). The relations between children's comprehension monitoring and their reading comprehension and vocabulary knowledge: An eye-movement study. *Reading and Writing, 33*(3), 511–545. https://doi.org/10.1007/s11145-019-09966-3