Comprehension monitoring behaviour during reading of connected text in elementary school-children: Comparing eye-tracking and think-aloud methods

Stacey Chomey, Janelle Collett, Bradie Molzan

Supervisors: Melissa Skoczylas, PhD, Esther Kim, PhD, Monique Charest, PhD

Comprehension monitoring: eye-tracking and think-alouds

ABSTRACT

Comprehension monitoring while reading is the ability to recognize that a coherence problem has occurred and that a repair strategy is needed to resolve a discrepancy in understanding. This pilot study examined two methods for detecting incoherence: eye-tracking and think-aloud protocols. The objective of the present study was to investigate whether eye-tracking and think-aloud methodologies reveal similar evidence of a child's comprehension monitoring through inconsistency detection. The participants were three children between the ages of eight and ten, with typically developing reading and spoken-language skills. Both eye-tracking and think-aloud protocols showed whether or not the participant had comprehension monitoring abilities. This study revealed that for speech pathologists, teachers, literacy consultants, and other professionals, think-alouds reveal more immediate and meaningful information that could be used to assess a child's comprehension monitoring abilities in a more transparent, time-efficient, and cost-effective way than eye-tracking.

INTRODUCTION

There are multiple abilities involved in the complex task of reading, one of which is comprehension monitoring, a skill that school-aged children must develop in order to become successful readers (Baker, 1979; Owings, Petersen, Bransford, Morris, & Stein, 1980; Paris & Myers, 1981; Rapp, van den Broek, McMaster, Kendeou & Espin, 2007). Comprehension monitoring is the ability to recognize that a discrepancy in understanding during reading has occurred and that it is necessary to apply a repair strategy in order to resolve the discrepancy (Kamhi & Catts, 2012). When reading, a skilled comprehender continually evaluates whether their understanding of the text is coherent (van der Schoot, Reijntjes, & van Lieshout, 2012).

When readers read text, they form a situation model in their minds. A situation model refers to when the reader crafts a mental representation of the characters, setting, plot, and other story details, from text they have read (Zwaan, Radvansky, Hilliard, & Curiel, 1998). If a reader comes across information that violates their situation model and recognizes this violation, it may be referred to as successful inconsistency detection (van der Schoot et al., 2012). An inconsistency while reading can either be local, meaning that the potential source of the inconsistency is still within the short term memory, or global, meaning the inconsistency occurs when information that was previously presented is no longer in short term memory (Albrecht & O'Brien, 1993; McKoon & Ratcliff, 1992).

When evaluating a child's reading comprehension, it is possible to examine both the products and processes of comprehension. The former is traditionally measured by performance on reading comprehension questions, while the latter evaluates *how* the child arrives at this level of understanding by assessing their ongoing cognitive processes, including

comprehension monitoring.

One way to evaluate comprehension monitoring processes is to use an inconsistency detection task to determine whether child realizes there is an error in the text, that is, that something in the text does not fit in with their situation model. This task evaluates a child's ability to find inconsistencies placed in text. For example, Vorstius, Radach, Mayer, and Lonigan (2013) used eye-tracking to determine if fifth grade students could detect such semantic inconsistencies as "Daniel was shivering because he was hot" by analyzing time spent re-reading. Inconsistency detection tasks allow analysis of a child's response to an error in the text, either verbally, in a think-aloud paradigm (Scott, 2008), or nonverbally, in an eye-tracking paradigm (van der Schoot et al., 2012).

After successfully detecting an inconsistency, a skilled reader will employ comprehension repair strategies in an attempt to integrate the new information into their situation model (van der Schoot et al., 2012). This may include looking backwards or forwards in the text to search for information that could be a possible source of inconsistency (Kamhi & Catts, 2013; van der Schoot et al., 2012). In contrast, a poor reader who did detect the inconsistency may also fail at this stage by neglecting to attempt repair (van der Schoot et al., 2012). Failure to detect the inconsistency could reflect an insufficiently rich situation model, as the reader does not recognize that the new information is at odds with the previous information.

The purpose of the present study was to examine two process measures of comprehension monitoring, eye-tracking and think-aloud protocols, in order to determine if they both reveal similar evidence of a child's comprehension monitoring through inconsistency

detection. According to Rapp et. al. (2007), there are subgroups of struggling readers; therefore, there is a need to be able to identify these subtypes as individualized remediation calls for knowledge of individual struggles in reading comprehension. Rapp et al. (2007) proposed that when evaluated through product measures only, struggling readers can be indistinguishable from one another, but it is possible that there is more than one way to arrive at a given level of comprehension performance. That is to say, speech-language pathologists and teachers working with children need reliable measures of comprehension monitoring to apply in reading comprehension assessment and intervention. The two examples of process measures of comprehension monitoring are eye-tracking and think-aloud tasks.

Eye-tracking

Eye-tracking for comprehension monitoring involves recording or tracking participants' eyes while they are reading text. In order to track eye-movements, extensive equipment and technology are needed, including but not limited to: a camera, computer hardware, and eye-tracking software. It has been suggested that these eye movements are a window into the hidden mental processes of readers (Just & Carpenter, 1980).

There are three common types of eye movements that may be analyzed in eye-tracking: saccades, fixations, and regressions (Rayner, 1998). Saccades pertain to the rapid eye movements that occur while reading; these saccades are movements that happen during the scanning of a passage. A fixation is a pause in eye movement. Fixations occur when the reader takes a longer period of time to process a particular stimulus. Regressions refer to the reader's eyes going back into the passage and re-reading.

There are several informative measures that make use of information from saccades,

fixations, and and regressions, including first pass fixation, total fixation duration, first fixation duration, and regression duration (Rayner, Chace, Slattery, & Ashby, 2006). First pass fixation refers to the average time a reader spends the first time their eye fixates on a target word or area of text. Total fixation duration is the overall time that a reader looks at a particular word or area and first fixation duration is the duration of the first fixation event, before any regressions occur. Regression duration, which refers to the difference between the total fixation duration and the first fixation duration, may give insight into the process of re-reading for comprehension (Rayner, 1998).

Eye movements of adults may also be responsive to inconsistencies within the text, as readers may fixate longer on the inconsistent portion of the passage (Rayner et al., 2006). Additionally, when an inconsistency occurs a short distance from the previously presented information, a reader may make more regressions (Rayner et al., 2006). Eye movements of the reader appear to be an indication of comprehension monitoring, particularly with respect to regressions and fixations.

Eye movements have also been used to study the comprehension monitoring of children during reading tasks (Mason, Tornatoro, & Pluchino, 2013; Vorstius et al., 2013; van der Schoot et al., 2012). Researchers have found that first pass duration may be indicative of comprehension repair processes such as double checking, or thinking of possible resolutions for the contradiction (van der Schoot et al., 2012; Mason et. al., 2013). As well, a higher frequency of regressions may indicate greater likelihood that the reader experienced an inconsistency in the text (van der Schoot et al., 2012).

As there is evidence that eye movements may give insight into ongoing cognitive

processes of readers, eye-tracking technologies may be used by clinicians as a method of determining the region of text that cause problems for the child by being able to track, on a finer scale, the time the reader spends hesitating on a portion of text, or regresses to re-read (Vorstius et al., 2013). However, access to such technology in institutions, such as schools, where reading comprehension is assessed, may be limited due to funding. In addition, collection and analysis of such data requires specialized expertise, which may not be readily available in all settings. Thus, eye-tracking may not be feasible, and an alternative, more accessible, measure for assessing reading comprehension may be warranted.

Think-Alouds

Think-aloud protocols, or think-alouds, involve a reader verbalizing what would have been covert thoughts and ideas about a text while reading. Think-aloud tasks have been widely used to gain insight into reading comprehension processes, particularly in terms of the reader's comprehension (McClintock, Pesco, & Martin-Chang, 2014; Scott, 2008; Liang & Khami, 2002). Unlike eye-tracking technology, materials required for think-aloud protocols can be minimal, and may include materials that are already found in schools, such as texts for reading, and recording equipment.

Readers' verbalizations of their thoughts regarding a text passage can, as with eye-tracking, be examined through inconsistency detection tasks, where errors are embedded in the text. During a think-aloud inconsistency detection task, the reader is presented with text and asked to verbalize if there is anything abnormal in the text during their silent reading of the text (Scott, 2008). Thus, the reader is primed for the error detection task (Ehrlich, Remond, & Tardieu, 1999). The reader's comments about the potential inconsistencies in the text may be

referred to as monitoring comments. Once the reader's comments are recorded, they may be coded for analysis. For example, codes may include 1) errors were identified, 2) non-error words were identified as errors, and 3) errors were not identified (Scott, 2008). It has been suggested

that think-aloud tasks, and these types of monitoring comments, may make readers' thought processes more transparent (McClintock et al., 2014).

Think-alouds may also help to determine whether a comprehension breakdown has occurred, as well as provide insight into the cause of these breakdowns (Liang and Khami, 2002). Monitoring comments may give researchers an indication of whether the reader is able to monitor their comprehension by identifying the inconsistency aloud.

Although using think-alouds appears to be a cost-effective solution to measuring the process of reading comprehension as compared to eye-tracking technology, there may be limitations to their use. It is possible that think-aloud tasks may reduce performance due to the additional cognitive load of verbalizing one's thoughts, which may not be natural during silent reading. Further, think-aloud processes can only tap into those processes of which the reader is consciously aware (van den Broek, Beker, & Oudega, 2015). This is compared to eye-tracking, which may tap into comprehension monitoring processes of which the reader is unaware. However, It has been found that using think-alouds as comprehension strategies helped facilitate reading comprehension in both children with language impairments and those with normal language (McClintock et al., 2014). This suggests that think-aloud processes may actually alter the underlying processes by improving reading comprehension. Thus, further investigation seems warranted into the relative merits of the eye-tracking and think-aloud

approaches and the correspondence between the information that the measures return.

Eye-tracking and Think-Alouds

The importance of reading for later success in life has been highlighted among researchers (Rapp et al., 2007). Eye-tracking technologies are one way to access reading comprehension process information online as it occurs. However, this is only an indirect measure, as eye movements must be interpreted. Think-aloud protocols are one way to provide insight into a child's cognitive process of reading comprehension, keeping in mind the limitations mentioned above. In fact, the process of thinking aloud in itself can improve a child's comprehension of a text. It has also been suggested that these two methodologies could be used together to provide insight into children's reading comprehension processes.

A study by Kaakinen and Hyönä (2005) contributed to this area of questioning: they aimed to determine whether think-aloud protocols are a valid measure of text comprehension by measuring eye movements of 36 university students during a think-aloud task. Kaakinen and Hyönä (2005) concluded that think-aloud responses reflecting more "deep" processes (e.g., self-explanations), were preceded by longer first-pass re-reading times (indicative of a reader's need to immediately re-read a sentence) than for "shallower" processes (e.g., associations, paraphrases).

While Kaakinen and Hyönä (2005) overlaid eye-tracking upon think-alouds to explore aspects reading comprehension, the present study seeks to compare eye-tracking and think-alouds with consistent and inconsistent reading passages to determine whether both can capture the online process of comprehension monitoring. It was hypothesized that monitoring comments in think-alouds would show evidence of comprehension monitoring when the child

verbalizes that they have detected an inconsistency in a text. It was also hypothesized that eye-tracking would show similar, nonverbal evidence through increased gaze times within the zone of the inconsistency within the text both when the eye first fixates on this region and any time spent on the same region after possible regressions to previously read sections of the text.

METHODS

Participants

Six children participated in this study, five female and one male. For two participants, eye-tracking was unsuccessful; thus, they did not complete the think-aloud protocol. For another participant, eye-tracking data was missing due to poor calibration, so analysis of the participant's data was not possible. The sample for analysis included a total of three participants, all of whom were female. The children, who were in grades three to four, were aged between eight and ten. The mean age of the participants was 9.2 years (*SD* = 0.376). All participants had normal or corrected to normal vision and hearing, as reported by their parents. We recruited a convenience sample of participants through personal invitation. There was no compensation offered for participation in this study.

Table 1. Participants

Participant	Age (Years; Months)	School Grade
P1	9;2	4
P2	9;8	4
P6	8;9	3

The participants were determined to have typically-developing language based on four screening tests, listed in Table 2 below.

Table 2. Screening tests

Target Skill	Task
Oral language skills	*CELF-5*[1] Screening test
Memory	*CELF-4*[2] Number repetition
Reading comprehension	*CELF-5*[1] Reading comprehension subtest
Word-level reading	*KTEA-II*[3] Letter and word recognition subtest

[1] *Clinical Evaluation of Language Fundamentals--5th Edition (Wiig, Semel & Secord, 2013)*
[2] *Clinical Evaluation of Language Fundamentals--4th Edition (Wiig, Semel & Secord, 2003)*
[3] *Kaufman Test of Educational Achievement --2nd Edition (Kaufman & Kaufman, 2004)*

Materials

The stimuli used for both the eye-tracking and the think-aloud were originally developed in a previous CSD 900 project (Cartmell, Coutts, & Mitchell, 2013), and were modelled after Rayner et al. (2006). The stimuli consisted of 24 paragraphs in six different conditions, for a total of 144 different stimulus items. Each paragraph contained an antecedent and an anaphor. An antecedent is a mention of a person, place, thing, or event, and an anaphor is a reference to, or a second mention of, the initial person, place, thing, or event. In this particular study, only nouns were used for the antecedents and anaphors. Additionally, in the present study, some of the vocabulary used in Cartmell et al. (2013), which was originally developed for adult readers, was altered in order for the passages to be readable for children aged eight to twelve (e.g., portage → trip).

In order to create the six conditions per paragraph, Cartmell et al. (2013) made half of the antecedents and anaphors consistent, and half inconsistent. They also varied the distance in words between the antecedent and the anaphor. There were three distances between the antecedent and the anaphor used to create the conditions: near (6-7 words between the antecedent and the anaphor), intermediate (approximately 15 words between the antecedent and the anaphor), and far (approximately 30 words between the antecedent and the anaphor) (Cartmell et al., 2013). The six conditions were equally distributed across six different sets of 24 stimuli, in order for each participant to receive a different set. See Appendix for sample stimuli.

Apparatus

During the eye-tracking procedure participants read passages and proceeding questions from a 24 inch flat-panel monitor with a resolution of 1920×1080 pixels. The text was presented in size 18 Courier New font in white against a black background. Then chin rest, which minimized head movements, was positioned 90 cm from the monitor. The EyeLink 1000+ system recorded movements of the left eye using a sample rate of 500 Hz (one sample every 2 ms). Before beginning, and after every fourth trial (a passage and related question pair), a 13-point calibration was performed, followed by a validation to confirm the accuracy of the gaze position achieved in calibration. Before each trial a drift-correct was performed to ensure the calibration parameters were maintained. An additional calibration was performed if there was error greater than 5 degrees. Following passage reading, participants answered a yes/no reading comprehension question. The purpose of this question was to encourage the participants to read carefully. In the eye-tracking condition, comprehension questions were answered using gaze position data which was collected by the Eyelink 1000 in real time to

create a gaze-control paradigm. The participant fixated their gaze for two seconds on either "yes" or "no" options on the screen to progress to the next screen. This ensured that participants kept their heads still and directed towards the computer monitor.

During the think-aloud procedure, participants read passages and questions from a 24 inch flat-panel monitor with a resolution of 1920×1080 pixels. In this condition, the text was also presented in size 18 Courier New font in white against a black background in Microsoft Powerpoint, in order to remain as consistent as possible with the stimulus presentation in the eye-tracking condition. The participant's chair was positioned wherever the child was comfortable. Participants' think-alouds were recorded in audio and video by a Sony Handycam HDR-XR550 video camera. To answer the comprehension questions, participants used the mouse to click on either "yes" or "no" and researchers recorded each response.

Data Collection

Participants completed the pre-testing, eye-tracking, and think-aloud procedures with breaks in between each. When necessary, due to time constraints of participants, procedures were administered across two days. Recorded eye-tracking data included total fixation duration, first fixation duration, and regression duration. The eye-tracker also recorded the participant's response to the reading comprehension questions. Participants were given a response accuracy score of 1 for each correct response, and a score of 0 for each incorrect response. Participants received a total response accuracy score out of 24.

Think-aloud data were recorded and coded for monitoring comments. Comments were divided into two categories: real world consistency (e.g., "It says they are having brownies for supper") and lexical consistency (e.g., "That should be roses instead of daisies"). Similar to the

protocol used by Scott (2008), participants were given a consistency score of 1 for a think-aloud if they responded appropriately (i.e., with a comment about lexical consistency). If the participant made a comment in their think-aloud about the perceived real-world consistency without finding the lexical inconsistency, they were given a consistency score of 0. Participants received a total consistency score out of 24. Data were collected in response to the comprehension question following each passage. Participants were given a score of 1 for each correct response, and a score of 0 for each incorrect response. Participants received a total response accuracy score out of 24.

Procedure

Screening Measures

A test battery was used to determine if the participants had adequate language-related skills, in order to be considered typically developing and to be included in the present study. All tests used to determine inclusion in the present study were norm referenced and a cut-off score of one standard deviation from the mean was used to determine inclusion criteria.

The screening measures (and the relevant domain being assessed) included: the KTEA-II Letter and Word Recognition subtest (single-word reading ability), the CELF-4 Number Repetition subtest (working memory), the CELF-5 Screening Test (oral language), and the CELF-5 Reading Comprehension subtest (reading comprehension). Two examiners alternated administering the subtests in order for the participant's scores to be calculated in a time-efficient manner to confirm they met the inclusion criteria.

Eye-tracking

Prior to beginning the eye-tracking procedure, instructions were presented on the

screen and by researchers. Participants were told that some of the passages might seem "strange". Researchers did not refer to the inconsistencies as "errors" in order to discourage error-searching behaviour. The researchers assured participants that the task was not a test in order to keep their anxiety low and reading more natural. After receiving directions and having had the opportunity to ask questions, a calibration was performed, followed by a validation. After breaks, another calibration was performed and validated. To confirm accuracy, calibration was repeated every fourth trial and a drift correction was performed after all other trials.

First, participants were presented with three practice trials to familiarize them with the task and procedure. Next, they proceeded to work through the passages and corresponding questions at their own pace. After each passage, including the three practice trials, participants answered a comprehension question. After all the passages had been read, participants were guided to room where the think-aloud portion of the study was to be completed.

Think-Alouds

The think-aloud condition involved participants reading a similar series of consistent and inconsistent passages on a computer screen. They read the passages in a similar manner as above; however, no eye movements were recorded. Rather, video recording captured the participant verbalizing thoughts and ideas about the text they silently read. Prior to reading the passages, participants were told: "If you notice something strange, tell me. If the passage is okay, just tell me it's okay." Again, participants were again asked to answer a comprehension question about the passage they had just read. Participants were given the same opportunities for breaks after every four passages or upon request.

Three practice trials were presented to familiarize participants with the think-aloud task. One consistent and two inconsistent passages were presented to give participants an opportunity to practice thinking aloud. Children are not usually asked to talk about what they are reading as they read, so the practice trials ensured participants were comfortable with the task. If participants thought a practice passage was consistent, they were prompted to tell researchers the passage was okay. If participants failed to notice an inconsistency in the practice trials, the researchers alerted them to the inconsistency. Participants were not given any feedback on their responses once the practice trials were completed.

Measures

Eye-tracking

There were three eye-movement measures and one accuracy measure in the eye-tracking portion of the study. The first dependent variable, *total fixation duration*, represented the sum of all time, in milliseconds, that a participant gazed at the words within the interest area, including regressive fixations. For our purposes, the interest area was defined as the anaphor, one word before and two words after. The second dependent variable, *first fixation duration*, represented the duration, in milliseconds, of the first fixation event within the interest area of the anaphor *not* including regressive fixations. The third dependent variable *regression duration*, represented the difference between total fixation duration and first fixation duration. This value was derived by subtracting the mean first fixation duration from the mean total fixation duration for each participant. This measure assessed the mean time participants spent gazing at the interest area in regressive eye movements, that is, the time they spent looking at the interest area of the anaphor *after* the first fixation event.

The final measure for eye-tracking was reading comprehension accuracy. Scores were determined based on participant's responses to the comprehension questions after each passage. Participants were assigned a score out of twenty-four which represented the number of comprehension questions they answered correctly.

Think-Alouds

Data were collected on two measures during the think-aloud task: inconsistency detection accuracy and reading comprehension accuracy. Inconsistency detection accuracy was measured by the number of inconsistencies the participants correctly identified aloud while reading the paragraphs. This yielded information about the participant's ability to identify the errors and articulate the inconsistent information in the paragraphs. Reading comprehension accuracy scores were collected based on the answers to comprehension questions after each passage.

RESULTS

Reading Comprehension Accuracy Scores

Although there was some individual variability in accuracy scores for the eye-tracking and think-aloud protocols, average reading comprehension accuracy was the same for both conditions (22/24, 91.7%), as seen in Table 3.

Table 3. Eye-tracking and Think-Aloud Reading Comprehension Accuracy Scores

Participant	Eye-tracking Accuracy Score (/24)	Think-Aloud Accuracy Score (/24)
P1	20	24
P2	24	21
P6	23	22
Average	22	22

Eye-tracking

Total Fixation Duration

Due to a small sample size, all results were analyzed by visual inspection rather than by statistical analysis. The mean total fixation duration appears to be greater for inconsistent than consistent trials. The total fixation duration accounts for eye movements that are made during the first fixation event, as well as regressive eye movements. The results, shown in Figure 1, were ordered as expected, as it appeared that participants, on average, gazed longer at the interest area for inconsistent trials than for consistent trials.

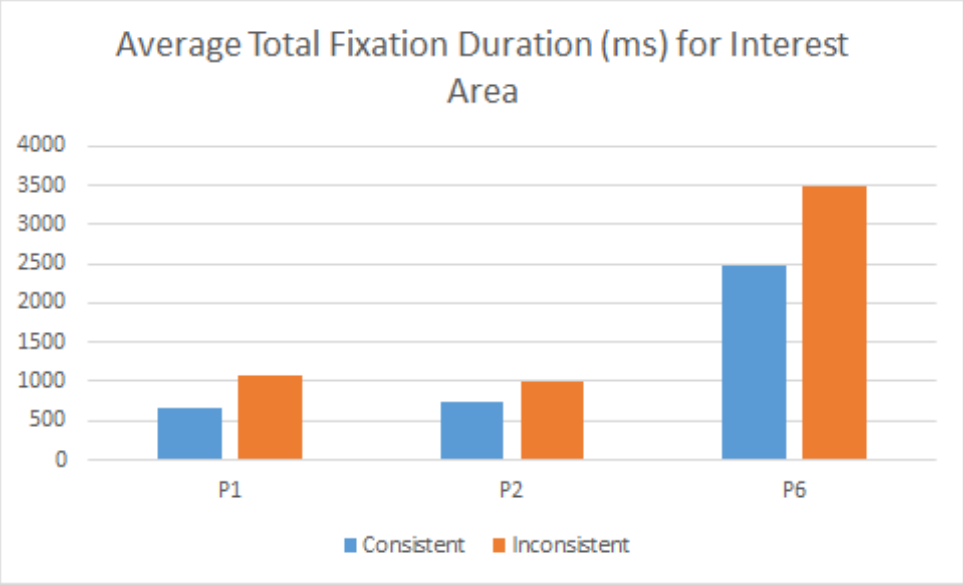

Figure 1. Average Total Fixation Duration for Interest Area

First Fixation Duration

The mean first fixation duration appears to be slightly greater for inconsistent than consistent trials, which is shown in Figure 2. The first fixation duration accounts for eye movements that are made during the first fixation event, not including regressive eye movements. The results were ordered as expected, as it appears that participants gazed, on average, longer at the interest area for inconsistent trials than for consistent trials.

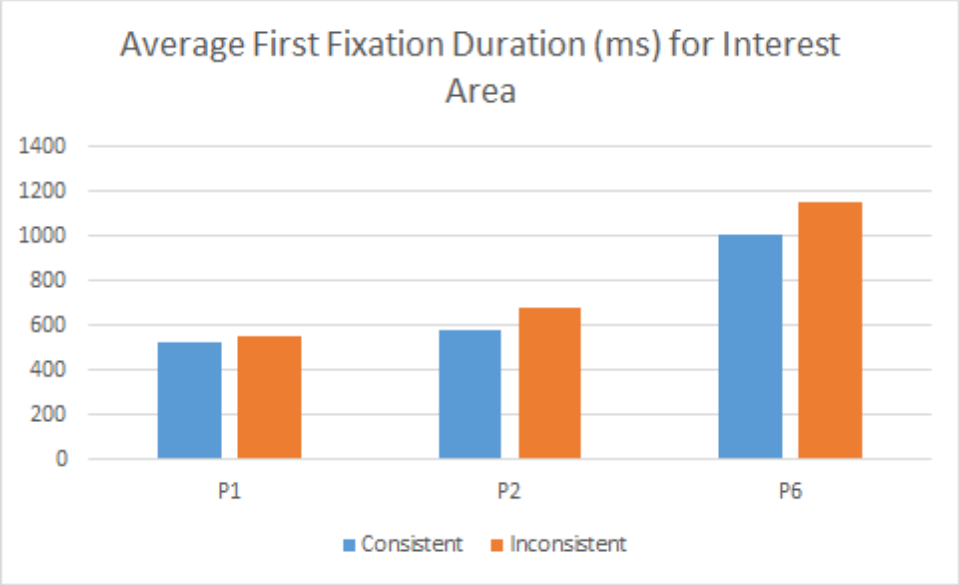

Figure 2. Average First Fixation Duration

Regressive Eye Movements

The amount of time that each participant spent in regressive fixations to the anaphor appears to be greater for inconsistent than consistent trials. These are eye fixations that are made after the eye has left the interest area after the first fixation event and then returns to the interest area. The results were ordered as expected, as shown in Figure 3; it appeared that on average, participants gazed longer in a regressive pattern at the interest area for inconsistent trials than for consistent trials.

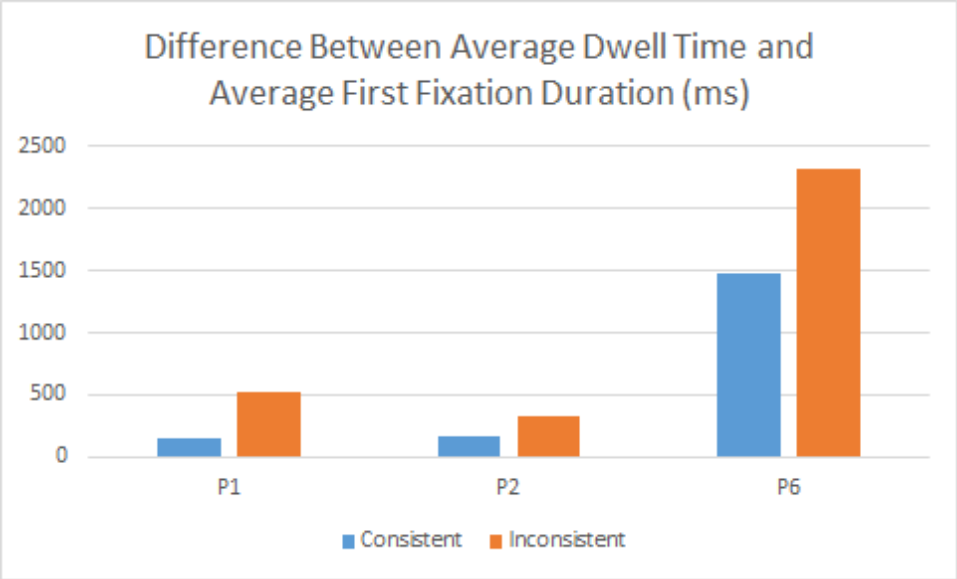

Figure 3. Difference Between Average Total Fixation Duration and Average First Fixation Duration

Think-Alouds

Inconsistency Detection Accuracy Scores

For the think-aloud portion of the study, the average inconsistency detection accuracy score for each participant was calculated by determining how many inconsistent trials were identified with appropriate monitoring comments. A percentage accuracy was calculated by dividing the number of correctly identified inconsistencies divided by the total number of inconsistent trials present in each of the participants' stimulus sets (11, 13, and 14 for P1, P2, and P6, respectively).

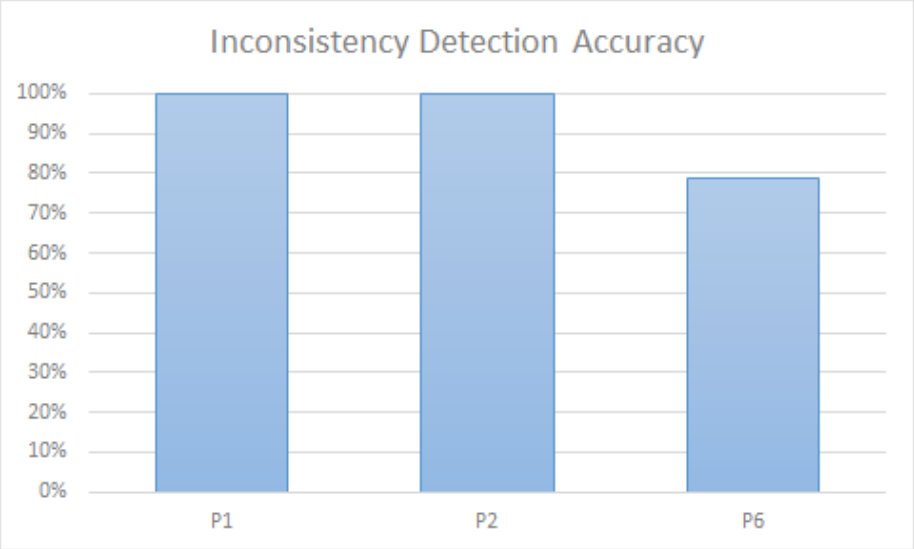

Figure 4. Inconsistency Detection Accuracy

DISCUSSION

This study examined two methods of looking at comprehension monitoring during the reading of connected text in school-aged children. In particular, the study aimed to determine whether the two methods, eye-tracking and think-aloud protocols, could each detect whether a child is monitoring their understanding of what they are reading. The purpose of the study was to give insight into whether eye-tracking and think-aloud protocols would both yield reliable information about the comprehension monitoring of school-aged children. It was hypothesized that the two methods would both yield equally reliable information about the comprehension monitoring skills of the participants. In the current study, there was evidence of comprehension monitoring for both eye-tracking and think-alouds.

Based on the results obtained with the eye-tracking methodology, there was evidence that all three participants were able to monitor their comprehension. It appeared that during

the first fixation event participants gazed longer (on average), at an anaphor when it was inconsistent with its antecedent as compared to when it was consistent. It also appeared that participants spent longer looking at the anaphor regressively, indicating that they may have looked back in the passage in order to re-read and resolve the inconsistency before returning to the anaphor. It was inferred by the researchers that each participant monitored their comprehension as there was evidence of increased time spent gazing at the anaphor both progressively and regressively for inconsistent trials as compared to consistent trials. This was reflected in an increased total fixation duration, and the difference between the total fixation duration for consistent versus inconsistent trials.

By looking at average fixation durations across all trials for each participant the researchers could infer generally whether a participant had the ability to monitor their comprehension. However, it was not transparent on a trial-by-trial basis whether or not the participants detected the individual inconsistencies; while there were overall trends showing greater fixation times for inconsistent trials, there is no specific fixation duration that can unequivocally separate detected from undetected inconsistencies. Because inconsistency detection is not transparent on a trial-by-trial basis, the accuracy of a participant's comprehension monitoring cannot be determined. This has clinical implications, as eye-tracking may not be a good indicator of the skills level of a child's comprehension monitoring.

The think-aloud condition also appeared to yield evidence of comprehension monitoring for all three participants. Evidence of comprehension monitoring was observed in the comments made by the participants. For inconsistent trials, only two types of comments were made. If the participant failed to detect the inconsistency, they reported that the passage was

okay. If the participant detected the inconsistency, they reported what word was incorrect (e.g., “It says juice here, but up here it says coffee”). These two types of comments made it transparent whether or not the participant had detected the inconsistency. This allowed an accuracy score to be assigned for each passage, rather than a general sense of comprehension monitoring as in the eye-tracking condition.

While the data collected during the think-aloud condition is not normed, each participant’s inconsistency detection accuracy score can be compared to the other participants’ scores. For example, P6’s accuracy score was lower than the other two participants, which may be interpreted as P6’s comprehension monitoring abilities being not as developed as the other two participants. As P6 was younger than the other two participants, it is plausible that her comprehension abilities may not be as developed. However, based on P6’s fixation times and regressions during eye-tracking, there is evidence of some comprehension monitoring.

Eye-tracking and think-alouds were both found to detect comprehension monitoring, although each rendered a different kind of data. Although eye-tracking is useful in that it reveals information about the unconscious processes in reading comprehension, think-alouds yield more immediate and meaningful data that could be used to determine a child’s level of performance in comprehension monitoring. Think-alouds give more of a sense of skill level for each passage compared to eye-tracking. Think-alouds are also more plausible in schools because they are considerably more affordable since special equipment and training is not required. Eye-tracking requires extensive time and training in the operation of the equipment, programming of the passages, and analysis of data.

Limitations

One of the primary limitations of the present study was the eye-tracking portion and the technical challenges involved in collecting eye-gaze data. There are limitations regarding the eye-tracking technology in itself, such as the need for repeated calibration throughout the test procedure to ensure the camera is capturing the necessary movements. Recalibration is also required if the initial calibration fails, which can be time consuming. For two potential participants, the eye-tracker was unable to detect their pupil, so they were unable to complete the study. It was believed that eye shape was a factor in the eye-tracker's ability to detect their pupil, making this a limitation of eye-tracking, especially with children. For another participant, much of the data needed for data analysis was missing due to calibration errors, rendering this data unusable. Overall, the technical challenges may limit feasibility for large-scale individualized use.

Second, the time required by participants was too long. Children appeared to be restless after the screening tests and were then required to read 24 passages in each condition. Because the skilled readers had difficulty sitting through the entire protocol, it is unlikely that unskilled readers or those with a language impairment would be able to proceed through all parts of the study and perform to the best of their abilities. If this is further explored in a population of unskilled readers, demands would have to go down (e.g., limiting the screening measures for the pre-assessment, limiting the number of passages). Additionally, in terms of the outcomes of the present study, the think-aloud condition appeared to be more feasible with school-aged children than the eye-tracking condition.

Finally, this study had a small sample size. While this size was warranted because it was

a pilot study, statistical analyses were unable to be used to compare the two methodologies.

Future Directions

The data collected from this study, namely in terms of the think-aloud portion of the study, has the potential to impact the way reading comprehension is assessed in elementary-school children. Think-alouds may reveal information regarding better ways to support students based on their individual needs, in a more financially and time conservative manner, in comparison to eye-tracking methodologies. Think-alouds may also have positive implications for educators and clinicians supporting struggling readers who may not have yet developed the strategies to resolve such a coherence problem, but who may be coached to learn how to do so.

It would be worth comparing the eye-tracking and think-aloud methodologies in future studies with a larger sample size and a larger sample of skilled readers in order for the data to be statistically analyzed. Further, future studies could explore the eye-tracking and think-aloud methodologies with skilled and struggling readers.

APPENDIX

Sample stimuli

The following passages exemplify consistent and inconsistent passages, with varying distances between the antecedent and the anaphor (near, intermediate, far), as used in both the eye-tracking and think-aloud conditions.

Near - Consistent

a) Nancy started every day off with a cup of coffee and a donut. Nancy went without her coffee and a donut on Tuesday, because she was running late. She was disappointed for the rest of the day because of the change in her normal routine.

Near - Inconsistent

b) Nancy started every day off with a cup of coffee and a donut. Nancy went without her juice and a donut on Tuesday, because she was running late. She was disappointed for the rest of the day because of the change in her normal routine.

Intermediate - Consistent

c) Nancy started every day off with a cup of coffee and a donut. On Tuesday, Nancy was running late, so she went without her coffee and a donut. Nancy was disappointed for the rest of the day because of the change to her normal routine.

Intermediate - Inconsistent

d) Nancy started every day off with a cup of coffee and a donut. On Tuesday, Nancy was running late, so she went without her juice and a donut. Nancy was disappointed for the rest of the day because of the change to her normal routine.

Far - Consistent

e) Nancy started every day off with a cup of coffee and a donut. Nancy was a creature of habit, and did not like when her routines were thrown off. On Tuesday, Nancy was running late, so she went without her coffee and a donut. She was not happy.

Far - Inconsistent

f) Nancy started every day off with a cup of coffee and a donut. Nancy was a creature of habit, and did not like when her routines were thrown off. On Tuesday, Nancy was running late, so she went without her juice and a donut. She was not happy.

Yes/No Comprehension Question

Did Nancy start every day with a donut? Answer: Yes

REFERENCES

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1061–1070. http://doi.org/10.1037/0278-7393.19.5.1061

Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. *Journal of Literacy Research*, *11*(4), 365–374.

Cartmell, M., Coutts, B., & Mitchell, M. (2013). Examining cognitive mechanisms of text comprehension using eye-tracking and working memory measures. Unpublished manuscript, Department of Speech-Pathology and Audiology, University of Alberta, Edmonton, Canada.

Coté, N., & Goldman, S. R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. In H. van Oostendorp & S. R. Goldman (Eds.), The construction of mental representations during reading (pp. 169–193). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25*, 1–53.

Ehrlich, MF., Remond, M. & Tardieu, H. (1999) Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing 11*(29). doi:10.1023/A:100799650237

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Kaakinen, J. K., & Hyona, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. *Discourse Processes*, *40*(3), 239–257. http://doi.org/10.1207/s15326950dp4003_4

Kamhi, A. G., & Catts, H. W. (2012). *Language and reading disabilities* (3rd ed). Boston: Pearson.

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*(1), 95–109. http://doi.org/10.1016/j.compedu.2012.07.011

McClintock, B., Pesco, D., & Martin-Chang, S. (2014). Thinking aloud: effects on text comprehension by children with specific language impairment and their peers: Thinking aloud effects on text comprehension by children with SLI. *International Journal of Language & Communication Disorders*, *49*(6), 637–648. http://doi.org/10.1111/1460-6984.12081

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–466.

Owings, R. A., Petersen, G. A., Bransford, J. D., Morris, C. D., & Stein, B. S. (1980). Spontaneous monitoring and regulation of learning: a comparison of successful and less successful fifth graders. *Journal of Educational Psychology*, *72*(2), 250–256.

Paris, S. G., & Myers, M. (1981). Comprehension monitoring, memory, and study strategies of good and poor readers. *Journal of Reading Behavior*, *13*(1), 5–22.

Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, *11*(4), 289–312.

Rayner, K. (1998) Eye movements in reading and information processing. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255. http://doi.org/10.1207/s1532799xssr1003_3

van den Broek, P., Beker, K., & Oudega, M. (2015). Inference generation in text comprehension: Automatic and strategic processes in the construction of a mental representation. In E. J. O'Brien, A. E. Cook, & R. F. Lorch (Eds.), *Inferences during reading* (pp. 94–121). Cambridge, UK: Cambridge University Press.

van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, *25*(7), 1665–1690. http://doi.org/10.1007/s11145-011-9337-4

Vorstius, C., Radach, R., Mayer, M. B., & Lonigan, C. J. (2013). Monitoring local comprehension monitoring in sentence reading. *School Psychology Review*, *42*(2), 191–206.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading*, *2*(3), 199–220. https://doi.org/10.1207/s1532799xssr0203_2.