EXPLORING COMPREHENSION MONITORING IN SENTENCE READING WITH ENGLISH LANGUAGE LEARNERS

by

CAMERON GRACE ODDONE

(Under the Direction of Scott Ardoin)

ABSTRACT

Past research results indicate a strong relationship between oral reading fluency (ORF) and reading comprehension for native English speaking (EL1) students; however, little is known about the relationship between ORF and comprehension for English Language Learners (ELLs). The current study examined differences in comprehension monitoring behaviors for ELL (n = 31) and EL1 (n = 29) elementary students who were matched on ORF scores. Eye tracking technology was utilized to examine underlying reading behaviors as students read sentences that were either consistent or inconsistent in meaning. Results indicated that comprehension monitoring techniques did not differ between ELL and EL1 participants; however, students did engage in disrupted reading when they encountered inconsistent information while reading. Finally, ELL students performed at a lower level than EL1 students on a comprehension question that followed each sentence.

INDEX WORDS: English Language Learners, Reading, Fluency, Comprehension, Eye tracking, Children

EXPLORING COMPREHENSION MONITORING IN SENTENCE READING WITH ENGLISH LANGUAGE LEARNERS

by

CAMERON GRACE ODDONE

B.A., Emory University, 2010

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2015

EXPLORING COMPREHENSION MONITORING IN SENTENCE READING WITH

ENGLISH LANGUAGE LEARNERS

by

CAMERON GRACE ODDONE

Major Professor: Scott Ardoin
Committee: Amy Reschly
Rebecca Lieberman-Betz

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
May 2015

ACKNOWLEDGEMENTS

The research reported here was supported by the University of Georgia's Graduate Dean's Award in Social Sciences. The opinions expressed are those of the authors and do not represent views of the University of Georgia.

TABLE OF CONTENTS

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION

The United States Department of Education reported that the number of public school students in the U.S. who were English Language Learners (ELLs) increased from the 2002-2003 school year (an estimated 4.1 million students, or 8.7%) to the 2011-2012 school year (an estimated 4.4 million students, or 9.1%). The largest number of ELL students resided in cities, making up an average of 14.2% of total public school enrollment in cities in 2011-2012 (National Center for Education Statistics, 2014). Due to the growing number of students who are ELLs in U.S. schools, it is important that researchers and educators understand how to best serve this population. One of the areas in which ELL students often struggle is development of reading skills in English. A difficult distinction for school personnel to make is whether the ELL student is struggling to learn to read in English due to language barriers or if a specific learning disability is present (Figueroa & Newsome, 2006; Liu, Ortiz, Wilkinson, Robertson, & Kushner, 2008).

With the 2004 amendments to the Individuals with Disabilities Education Act (IDEA), Response to Intervention (RTI) became an important component in identifying students with learning disabilities. Frequently, schools employing RTI models use oral reading fluency (ORF) measures both to identify students with deficits in reading and to monitor their response to reading interventions. The wide use of ORF measures is primarily due to the extensive research that exists demonstrating their strong relationship with measures of reading comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Reschly, Busch, Betts, Deno, & Long, 2009). But recent research suggests that this relationship may not be as strong for ELL students and thus, ORF measures may not be as appropriate for identifying students from this population in need of supplemental reading intervention services (Crosson & Lesaux, 2010; Quirk & Beem, 2012).

School professionals often have ample experience identifying and remediating reading difficulties in native English speaking (EL1) students. Unfortunately, due to their more limited experience with ELL students, these same professionals frequently over generalize their experiences with EL1 students' reading development to ELL populations (Figueroa & Newsome, 2006; Liu et al., 2008). Research indicates that key differences exist in how ELL and EL1 students develop reading skills in English. For example, comprehensive findings from the National Literacy Panel on Language Minority Children and Youth (August & Shanahan, 2006) suggest that with quality instruction, ELLs often match their EL1 peers in word-level abilities (e.g., phonological awareness, spelling, decoding); however, their skills are not equivalent in higher-level reading tasks such as reading comprehension (August & Shanahan, 2006, 2010). Recent research suggests that measures of ORF largely assess word-level abilities as opposed to higher-level reading skills (i.e., reading comprehension) in ELL students (Crosson & Lesaux, 2010; Quirk & Beem, 2012). Therefore, ELLs' ORF scores may be an inflated representation of their global reading skills. Consequently, it is essential to develop a better understanding of differences in reading abilities for ELL and EL1 populations and determine appropriate procedures for accurately identifying which ELL students need supplemental reading intervention services.

Researchers employed eye tracking procedures in the present study to explore the behaviors that underlie ELL and EL1 students' reading and comprehension monitoring behaviors, while ensuring that possible differences between groups were not due to differences in

ORF scores between groups. Comprehension monitoring is a metacognitive process that reflects a student's ability to be aware, while reading, of whether or not the text makes sense (Eme, Puustinen, & Coutelet, 2006). Resultant data should help us to better understand differences in the relationship between fluency and comprehension for ELL and EL1 students.

Relationship Between ORF and Reading Comprehension

Extant research, typically conducted with EL1 children, supports a strong relationship between ORF measures and reading comprehension measures. For example, Reschly et al. (2009) conducted a meta-analysis on the appropriateness of using Curriculum Based Measurement in Reading (CBM-R), an ORF measure, to assess the reading achievement of students in first through sixth grades. Results indicated that there was a moderately high correlation between CBM-R and standardized measures of global reading achievement (weighted average, $r = .67$). Additionally, findings suggested that CBM-R was a significant predictor of students' performance on measures of reading comprehension, vocabulary, decoding, and word identification (Reschly et al., 2009).

Although correlations between ORF and reading achievement measures are well established for EL1 students, much less is known regarding the relationship between ELL students' ORF and reading achievement. The use of ORF measures with ELL students is a highly debated topic, and findings regarding the validity of using ORF measures for assessing ELLs' reading achievement levels are largely inconsistent. For example, results from some studies suggest that CBM-R scores are reliable and valid for ELL students and that relationships between CBM-R and measures of reading achievement are comparable to these relationships for EL1 students (Baker & Good, 1995; McMaster, Wayman, & Cao, 2006; Muyskens, Betts, Lau, & Marston, 2009; Wiley & Deno, 2005). Conversely, findings from other large-scale studies reveal significant bias in using CBM-R scores to predict reading comprehension scores for ELL students due to the fact that the CBM-R scores are more representative of the ELL students' word reading ability than overall reading proficiency (Crosson & Lesaux, 2010; Klein & Jimerson, 2005; Quirk & Beem, 2012).

Support for using ORF measures with ELL students comes from Wiley and Deno (2005) who conducted a study to investigate the predictive power of CBM-R and a maze reading measure on the reading achievement of 36 third-grade and 33 fifth-grade students, approximately half of whom were ELLs. The ELL students in their sample spoke Hmong (80%), Somali (13%), or Spanish (7%) at home. Analyses conducted separately for the ELL and EL1 samples indicated that both CBM-R and maze measures were correlated with ELL and EL1 students' performance on the Minnesota Comprehensive Assessment in reading (Minnesota Department of Children, Families, and Learning, 2002). Specifically, ORF, as measured by CBM-R scores, was significantly correlated with reading achievement for both language groups at both grade levels (.61 and .69 for ELL students in third and fifth grades and .71 and .57 for EL1 students in third and fifth grades). Unfortunately, Wiley and Deno failed to directly compare the differences between the predictive relationships for ELL and EL1 students.

Conversely, Crosson and Lesaux (2010) suggested that differences exist in the relationship between ORF and higher-level reading abilities for ELL and EL1 students when moderator variables such as listening comprehension in English are considered. Crosson and Lesaux conducted their study with a sample of 76 ELL fifth graders from Spanish-speaking backgrounds and made comparisons to existing literature with EL1 students. Similar to results reported by Wiley and Deno (2005), Crosson and Lesaux found that ORF, as measured by the Gray Oral Reading Test – fourth edition (GORT; Weiderholt & Bryant, 2001), was strongly related to ELL students' reading comprehension performance. However, the impact of ORF on reading comprehension for the ELL students was largely moderated by their English listening comprehension performance. That is, a stronger relationship was evident between ORF and reading comprehension amongst students with better listening comprehension skills. For students with poorly developed listening comprehension, reading comprehension was low regardless of demonstrated ORF ability. According to Crosson and Lesaux, these findings indicate that the connection between ORF and reading comprehension is not as direct for ELL students as it is for native English speaking students; there are other contributing factors that must be considered (Crosson & Lesaux, 2010). But due to the fact that there were no EL1 students included in the study, the researchers cannot be certain that a similar interaction between ORF and listening comprehension in predicting reading comprehension scores does not exist for EL1 students.

Another study with findings in line with Crosson and Lesaux (2010) comes from Quirk and Beem (2012). Quirk and Beem utilized a sample of 171 Spanish-speaking ELL students in second, third, and fifth grades to examine the relationship between ORF and reading comprehension. As an overall measure of ORF, ELL students were administered AIMSweb R-CBM passages (Shinn & Shinn, 2002) in addition to the Sight Word Efficiency and Phonemic Decoding Efficiency subtests of the Test of Word Reading Efficiency (TOWRE; Pro-Ed Inc., 2008). Students were also administered the Comprehension subtest of the Gates-MacGinitie Reading Test – fourth edition (GMRT-4, Fourth Edition, Form S; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2009) as a measure of reading comprehension and the California English Language Development Test (CELDT; CTB/McGraw-Hill, 2008) as a measure of English language proficiency. Results indicated a similar, although slightly weaker, overall relationship between ORF and comprehension (*r* = .56) for ELL students in their sample as compared to prior studies with EL1 students. Interestingly, results also suggested that a majority (55%) of the ELL participants earned ORF scores that were significantly higher than would be predicted based on their low reading comprehension scores. Quirk and Beem therefore cautioned against using ORF scores as a proxy for reading comprehension measures for ELL students. Had ORF scores been used as the indicator that remedial reading services were needed (as is typical practice in the schools), 55% of their sample would have failed to be identified as needing intervention, despite their actual need. These students had adequate ORF scores, but poor reading comprehension skills that required intervention. Additionally, Quirk and Beem noted that the prevalence of ELL students with gaps between their ORF and reading comprehension scores varied by English language proficiency level, with those at the intermediate level evidencing the highest rates of fluency/comprehension gaps. The authors suggested that these findings are consistent with prior research (Nakamoto, Lindsey, & Manis, 2007) indicating that ELL students may develop decoding skills more rapidly than reading comprehension abilities, causing the largest gaps in these abilities to become evident once the student has passed the beginning stages of English language proficiency, but before they reach an advanced level.

Further support for the fluency/comprehension gap suggested by Quirk and Beem (2012) comes from studies indicating that with quality instruction, ELL students are able to master word-level reading abilities at the same rate as their native English speaking peers (Geva & Yaghoub Zadeh, 2006; Lesaux & Siegel, 2003). However, when a higher level of text comprehension is required, differences arise between these populations (Lesaux & Kieffer, 2010; Nakamoto et al., 2007). Nakamoto et al. (2007) conducted a longitudinal study following 261

Spanish-speaking ELL students from first through sixth grades. Results indicated that although early rapid growth of both ORF and reading comprehension allowed ELL students to stay on par with native English speaking peers until second grade, the rate of improvement in reading comprehension began to slow after second grade for ELL students. After second grade, ELL students continued with age-appropriate decoding and word reading abilities, but exhibited subpar reading comprehension abilities.

Considering evidence from the aforementioned studies, the relationship between ORF and reading comprehension may not be as direct for ELL students as it is for EL1 students. Consequently, the data provided by ORF measures may not accurately represent an ELL student's overall reading skills (Crosson & Lesaux, 2010; Klein & Jimerson, 2005; Quirk & Beem, 2012). However, we can only speculate about the reasons that the relationship between ORF and reading comprehension may vary for ELL and EL1 students because these studies rely solely on outcome scores (e.g., word reading rate and accuracy). An alternative method of investigating reading fluency and comprehension is to utilize eye tracking technology in order to examine the underlying behaviors that produce the outcomes measured within previous research.

Examining the Underlying Behaviors Related to Reading Comprehension

Eye movement data can inform our understanding of students' textual processing by providing details regarding reading behaviors not available by simply measuring reading outcomes. For example, eye tracking technology allows for analysis of differing amounts of time spent reading certain words within texts, number of times a student returns to reread certain words, and time spent on initial readings versus repeated readings of certain words. Whereas outcome data can reveal that two students performed equally well on a test of comprehension, eye movement data can expose the distinct processes used by the students to earn the same score.

Eye movement data have revealed that as individuals read, they make rapid eye movements from one point to another called *saccades*; however, new information is only encoded during the pauses between saccades referred to as *fixations*. Backwards saccades, generally denoted as *regressions,* are saccades to previously encountered material. Although eye movement research has greatly improved our understanding of the processes that underlie effective reading, to date the majority of eye tracking research has examined the reading behavior of skilled adult readers. Fortunately, recent advances in technology allow for easier and more accurate collection of eye tracking data with children (Rayner, Ardoin, & Binder, 2013).

The current study employed eye tracking procedures to evaluate possible differences between ELL and EL1 students' comprehension monitoring. Comprehension monitoring reflects a individual's ability to evaluate their understanding of text while reading (Eme et al., 2006). Although there is an extensive research base for skills needed for good reading comprehension outcomes, few have explored the basic processes at work in comprehension monitoring behaviors (Vorstius, Radach, Mayer, & Lonigan, 2013). Rayner, Chace, Slattery, and Ashby (2006) investigated comprehension monitoring behaviors of adults while they were reading inconsistent texts on an eye tracker. Findings indicated that adults fixated longer on regions where inconsistencies were present and also made more regressive eye movements in this condition.

Based upon an extensive review of the literature, to date only three studies have examined children's eye movements during comprehension monitoring (Joseph et al., 2008; van der Schoot, Reijntjes, & van Lieshout, 2012; Vorstius et al., 2013), and none of these studies included ELL students. In two eye movement studies with fifth- and sixth-grade students, van der Schoot et al. (2012) examined comprehension monitoring behaviors in a story reading task for students with good versus poor reading comprehension scores. Results suggested that both good and poor comprehenders spent more time reading the target sentence in text that had inconsistencies regarding a protagonist's character description and their actions. But poor comprehenders had greater difficulty than good comprehenders detecting inconsistencies in text when there was a long distance between the original information and an inconsistent target sentence (global inconsistency). Similarly, Joseph et al. (2008) utilized eye tracking technology to investigate the differences between child readers (ages 7-12) and adult readers as they read sentences containing plausible, implausible, and anomalous thematic relations. Both the children and adults demonstrated immediately disrupted reading (i.e., longer gaze durations on a target word) for the anomalous sentences where the inconsistencies were very obvious. But the children and adults in this sample differed in the efficiency with which they were able to integrate real-world knowledge, as evidenced by children's delayed effects on the implausibility sentences in comparison to adults (Joseph et al., 2008).

In an extension of previous work, Vorstius et al. (2013) examined the eye-movement behavior of 76 fifth-grade students as they read sentences that varied in regards to consistency and polarity. Sentence consistency changed when the meaning in the first clause of the sentence either matched or did not match the meaning in the second clause of the sentence (e.g., Daniel was shivering because he was cold. vs. Daniel was shivering because he was hot.). Polarity changed through the use of "because" vs. "although" between sentence clauses (e.g., Daniel was shivering because he was cold. vs. Daniel was shivering although he was hot.). Findings from this study indicated that inconsistency in sentence meaning played a minor role and that polarity had a more dominant effect on comprehension monitoring behaviors. The researchers also reported that sentences with negative polarity were particularly difficult for fifth-grade students to comprehend (i.e., sentences with "although" between clauses), with less than 40% of responses correct in the negative polarity conditions to a question asking whether or not the sentence made sense.

Purpose and Hypotheses

The current eye movement study was designed to examine the comprehension monitoring behaviors of fourth- and fifth-grade native English speakers as well as ELL students, a population with whom similar research has yet to be conducted. This study extended Vorstius et al. (2013) in order to evaluate comprehension monitoring behaviors as elementary students read consistent and inconsistent sentences. Based on the literature review, it was hypothesized that because EL1 students typically demonstrate more advanced reading comprehension abilities, EL1 students would respond more accurately to reading comprehension questions and engage in more comprehension monitoring behaviors during sentence reading than ELL students. Possible group differences in comprehension monitoring behaviors would be detected through global analysis of fixations and regressions within the sentences as well as fixations and regressions for target words that were central to sentence meaning. Based on results from Vorstius et al. as well as an adult study conducted by Rayner et al. (2006), we hypothesized that comprehension monitoring behaviors would vary across sentence conditions, with inconsistent sentences resulting in greater fixation counts, longer fixation durations, and more inter-word regressions.

CHAPTER 2

METHOD

Participants

Participants were 61 fourth- and fifth-grade students recruited from two elementary schools in a Southeastern school district. An a priori power analysis was conducted to determine the necessary sample size based on results from Geva and Yaghoub Zadeh (2006) and Klein and Jimerson (2005). The sample included 31 ELL students (38.7% female, mean age = 10.55 years) and 30 EL1 students (60.0% female, mean age = 10.52 years) who attended the same schools and were drawn from the same classes. A chi-square test confirmed that there were no significant differences in gender between the ELL and EL1 groups (p = .096) and an independent samples *t*-test confirmed that there were no significant differences in age between the ELL and EL1 groups (p = .881). The native language for all ELL students participating in this study was Spanish, and they all identified as Hispanic. Of the students in the EL1 group, 76.7% were Caucasian, 16.7% were African-American, and 6.7% were Multiracial. Consent forms were sent home in English and Spanish to the parents of ELL students and in English only to the parents of EL1 students. Only students with written consent from a parent or legal guardian and a signed student assent form participated in the study.

A brief survey regarding the frequency of Spanish and English spoken at home was included on the Spanish consent form for the parents of the ELL students to complete. Of the ELL students participating in the study, 16.1% spoke only Spanish at home, 32.3% spoke Spanish and a little English, 35.5% spoke Spanish and English equally, and 16.1% spoke English and a little Spanish. Although all students in the ELL group came from Spanish-speaking homes, the participants in this group were immersed in general education classrooms with English-speaking teachers. Their teachers indicated that they had a high level of English proficiency and were generally capable of understanding oral and written English.

The schools participating in this study administer the Formative Assessment System for Teachers (FAST; Christ et al., 2014) three times each academic year. This study utilized the scores from the winter administration of FAST CBM-R, a standardized measure of ORF in English. Scores ranged from 76 to 208 words read correctly per minute (WRCM; $M = 126.54$; $SD = 31.07$). Matched pairs within schools and grades were created between the ELL and EL1 groups based on ORF scores to ensure that differences in reading behaviors between groups were not due to differences in ORF scores. WRCM scores for ELL (M =126.16 WRCM; range = 67-208; $SD = 31.92$) and EL1 students (M = 126.93 WRCM; range = 67-189; $SD = 30.70$) did not differ ($p = .924$). Likewise, FAST Adaptive Reading (aReading; Christ et al., 2014) scores from the two samples (ELL $M = 497.55$, range = 469-522, $SD = 14.96$; EL1 $M = 500.93$, range = 479-521, $SD = 12.63$) did not differ statistically ($p = .344$). aReading is a computer adapted measure of students' print awareness, phonemic awareness, phonics, vocabulary, and comprehension.

Apparatus

Eye movement data were collected while students read sentences displayed individually on a 19-in. ViewSonic VG930m LCD monitor. Data collection was conducted using an SR Research Eyelink 1000 system, with a sampling rate of 500 Hz, resolution of 0.05° of visual angle, and a range of 32° horizontally and 25° vertically. Real-time eye tracking data were available for the experimenter to view on a separate computer screen, allowing the experimenter to give feedback to the student to facilitate accurate tracking. The Eyelink 1000 system defaults to recording eye movements from the right eye, but recordings were occasionally switched to the left eye due to tracking difficulties. Although the eye tracking system only follows one eye, participants' vision was binocular. Data collection was conducted in a softly lit room in each participant's school, and the brightness of the room was dimmed to minimize tracking losses. Participants used a Microsoft Sidewinder Plug and Play game pad to notify experimenters when they had finished reading the sentences and to answer a question about each sentence without head movement.

Materials

School screening measures. The schools participating in this study provided FAST CBM-R and aReading scores for all students with parental consent. The scores from the winter 2013 school administration of these measures were utilized in the present study to assess the reading abilities of students participating in the study. FAST CBM-R is an evidenced-based assessment of ORF in children (Christ et al., 2014), and scores from this assessment were used to match students in the EL1 and ELL groups. FAST aReading is a computer adaptive measure of broad reading ability, which employs a question and response format that is similar to many state standardized assessments (Christ et al., 2014).

Eye tracking materials. A set of two practice sentences and 28 experimental sentences were presented on the computer screen one at a time for the participants to read silently. Text was seen as black letters against a white background in 20-point Times New Roman font, with standard upper and lowercase letters. Sentences were developed by the experimenters to prevent the possibility that students had previous exposure to the reading material. Sentences were each seven or eight words long, with a mean length of 7.26 words per sentence. Fifth grade teachers were consulted to improve the likelihood that the fourth- and fifth-grade participants would understand the instructions and the wording of the sentences. *The American Heritage Word Frequency Book* (Carroll, Davies, & Richman, 1971) was used to assess the frequency per million running words of each word used in the experiment sentences. Low-frequency words, defined as occurring at a rate of less than 10 instances per million running words (Carroll, Davies, & Richman, 1971), were not used in this experiment.

Methodological issues regarding poor response to comprehension questions associated with the Vorstius et al. (2013) study were addressed by first piloting sentences to increase the probability that participants could in fact understand the sentences to be employed in the study. The 28 sentences presented within this experiment were chosen from a larger group of 100 experimenter-created sentences based on student responses in a pilot test. For a sentence to be included in the final set of 28 for the experiment, at least 85% of students in the pilot test had to respond correctly to the comprehension question, suggesting that it was not too difficult for the students to understand. Two fifth-grade classrooms that were not included in the main experiment participated in the pilot test. The students who returned signed parent consent forms and completed student assent (n = 52), read the 100 randomly ordered sentences and were asked to circle "yes" or "no" to indicate if each sentence made sense or not.

Of the 100 sentences administered as part of the pilot test, 50 were consistent with the parameters set by Vorstius et al. (2013) and 50 were modeled after the sentences on the Reading Fluency subtest of the Woodcock-Johnson III Tests of Achievement (WJ III; Woodcock, McGrew, & Mather, 2001). Specifically, the sentences modeled after Vorstius et al. (2013) had four versions each to reflect changes in sentence consistency and polarity while sentences modeled after the WJ III Reading Fluency subtest had two versions each to reflect changes in consistency. Overall, students in the pilot responded correctly more often to the sentences modeled after the WJ III Reading Fluency subtest; therefore, this type of sentence was used for the majority of the final experiment sentences (24 of the final 28), and four additional sentences modeled after the Vorstius et al. (2013) study were included at the end of each participant's session for purposes related to a different study. All analyses for the current study focused on student behaviors while reading sentences modeled after the WJ III Reading Fluency subtest. Each sentence had two versions, one that had consistent meaning (e.g., The airplane was flying in the sky.) and one that had inconsistent meaning (e.g., The airplane was flying in the ground.).

Procedure

Eye tracking data were collected with participants sitting in a chair with their chins placed on a chin rest and foreheads against a bar to reduce head movements during reading. The chin rest and forehead bar were height adjustable and secured to the table at a distance of 50-55cm from the computer screen. Participants were informed that they would be silently reading a series of sentences from the computer screen while the camera recorded their eye movements. The students were given instructions on how to use the game pad and were told that it should be used to indicate when they were done reading each sentence and to answer the post-sentence question.

Before eye movement data were collected, students were exposed to two practice sentences on the computer screen, one with consistent meaning and one with inconsistent meaning, to ensure that they understood the directions and how to use the game pad to clear the sentence screens and answer the comprehension questions. The experimenters provided corrective feedback to students on their answers to the practice questions and clarified any points of confusion for the students. Following the practice trials, a 9-point calibration and validation was completed to ensure accurate tracking. The calibration and validation process required the students to follow a dot on the screen with their eyes. Finally, 28 experimental trials were presented, 24 that were modeled after the WJ III Reading Fluency subtest, followed by 4 modeled after the Vorstius et al. (2013) study. Half of the sentences that each student read had consistent meaning, and the remaining half had inconsistent meaning. The sentences were randomized by sentence type (consistent vs. inconsistent) as well as order of presentation. Before each sentence trial, the experimenter asked participants to fixate on a dot on the left side of the screen, which was the location where they would need to begin reading on the next screen. Each session lasted approximately 15-20 min. After the session, students received a small prize and were escorted back to their classroom.

Data Analysis

Three types of analyses were conducted: analysis of question response accuracy, global analyses of reading behavior at the sentence level (i.e., eye movements across all words within the sentence), and target word analyses based on reading of the final word in the sentence. The final word of each sentence was central to the sentence meaning because it determined if the sentence had consistent or inconsistent meaning. Definitions of all dependent measures are provided in Table 1. For all analyses, fixations shorter than 120 ms or longer than 800 ms were omitted, which is generally consistent with prior eye movement research on reading behaviors (Binder, 2003; Raney & Rayner, 1995). Repeated measures analyses of variance (ANOVAs) were used to examine the effects of consistent and inconsistent sentence conditions (within subjects factor) as well as ELL status (between subjects factor) on reading behaviors and question response accuracy.

In a replication of the Vorstius et al. (2013) study, all students read consistent and inconsistent sentences and answered a comprehension question following each sentence. One sentence was removed from analyses based on lower question response accuracy (70.5% correct responses) in comparison to the other experiment sentences. The remaining 23 sentences ranged in response accuracy from 78.7% correct to 98.4% correct. One EL1 student was removed from all analyses based on atypical eye movement data indicating that the student did not fixate on the final word of any sentence.

Table 1

Dependent Measures

Parameter	Definition
Question response accuracy	Whether the student was correct or incorrect when answering the comprehension question
First fixation duration	The duration of the first fixation on a word before a saccade is made to another part of the same word or a different word
Gaze duration	The duration of all fixations made on a word before movement to another word
Total fixation time	The duration of all fixations on a word, including those following regressions
Fixation count	Total number of fixations on a word
Inter-word regressions	Total number of regressions made between words (global analyses only)
Regression out count	Total number of regressions made out of a word to a previously read word (target word analyses only)

CHAPTER 3

RESULTS

Question Response Accuracy

A 2 (sentence consistency) x 2 (ELL status) repeated measures ANOVA was conducted to analyze question response accuracy. ELL students differed significantly from EL1 students on question response performance. Despite being matched on ORF scores, ELL students responded correctly to significantly fewer (p = .047) of the comprehension questions than their EL1 counterparts. However, there was surprisingly not a significant effect for sentence consistency on question response. That is, across all students, question response accuracy was not significantly different for consistent and inconsistent sentences.

Global Analyses

Global analyses averaging eye movement parameters across all words in the sentences included first fixation duration, gaze duration, total fixation time, fixation count, and inter-word regressions. A 2 (sentence consistency) x 2 (ELL status) repeated measures ANOVA was conducted for each variable. Means and test statistics for global measures are presented in Table 2.

ELL students did not differ significantly from EL1 students on any of the global eye movement measures. Analyses to determine the effects of sentence consistency on eye movement behaviors indicated that overall, gaze duration was significantly lower (p = .022) for consistent sentences than inconsistent sentences. But when total fixation time (all fixations on words, including those following regressions) was analyzed, there were no significant effects for sentence consistency. Similarly, significant differences were not indicated for first fixation duration, total fixation count, or inter-word regressions with regards to sentence consistency.

Target Word Analyses

Target word analyses were conducted on the final word of each sentence, and measures included first fixation duration, gaze duration, total fixation time, fixation count, and regression out count. A 2 (sentence consistency) x 2 (ELL status) repeated measures ANOVA was conducted for each variable. Means and test statistics for target word measures are presented in Table 3.

Similar to findings from global measures across all words, no significant differences were present between ELL and EL1 students in target word eye movement analyses. But differences did exist between consistent and inconsistent sentences. Target word analyses indicated that students spent less time ($p < .001$) on their first pass reading of consistent sentences as opposed to inconsistent sentences for the final word in the sentence (gaze duration). Additionally, total fixation time (including fixations following regressions) ($p < .001$) and total fixation count ($p < .001$) were significantly lower for the last word of consistent sentences than inconsistent sentences. Thus, students, regardless of ELL status, spent longer on initial and total reading times and fixated more times on the last words of inconsistent sentences than consistent sentences. In contrast, sentence consistency did not have a significant effect on first fixation duration or regressions out of the last word in the sentences.

Table 2

Summary of Global Eye Movement Parameters Across Sentence Type by ELL Status

Measure	Consistent *M (SD)*	Inconsistent *M (SD)*
First fixation duration (ms) (*N* = 60)	249 (29)	252 (33)
ELL (*n* = 31)	255 (30)	255 (33)
EL1 (*n* = 29)	243 (26)	250 (32)
Between groups: $F(1, 58) = 1.29$, $p = .260$		
Main effect (Consistency): $F(1, 58) = 2.17$, $p = .146$		
Interaction (Consistency x ELL status): $F(1, 58) = 1.53$, $p = .219$		
Gaze duration (ms) (*N* = 60)	322 (61)	332 (57)
ELL (*n* = 31)	333 (52)	342 (53)
EL1 (*n* = 29)	311 (56)	322 (61)
Between groups: $F(1, 58) = 2.30$, $p = .135$		
Main effect (Consistency): $F(1, 58) = 5.51$, $p = .022$*		
Interaction (Consistency x ELL status): $F(1, 58) = 0.01$, $p = .911$		
Total fixation time (ms) (*N* = 60)	576 (180)	587 (165)
ELL (*n* = 31)	604 (182)	600 (169)
EL1 (*n* = 29)	547 (175)	573 (163)
Between groups: $F(1, 58) = 0.98$, $p = .327$		
Main effect (Consistency): $F(1, 58) = 0.63$, $p = .432$		
Interaction (Consistency x ELL status): $F(1, 58) = 1.16$, $p = .287$		
Fixation count (#) (*N* = 60)	2.02 (0.59)	2.05 (0.52)
ELL (*n* = 31)	2.08 (0.58)	2.05 (0.53)
EL1 (*n* = 29)	1.96 (0.61)	2.06 (0.52)
Between groups: $F(1, 58) = 0.15$, $p = .700$		
Main effect (Consistency): $F(1, 58) = 0.56$, $p = .457$		
Interaction (Consistency x ELL status): $F(1, 58) = 1.87$, $p = .177$		

Table 2 Continued

Inter-word regressions (#) (*N* = 60)	0.52 (0.18)	0.49 (0.16)
ELL (*n* = 31)	0.50 (0.18)	0.47 (0.17)
EL1 (*n* = 29)	0.53 (0.18)	0.51 (0.15)
Between groups: $F(1, 58) = 0.63$, $p = .430$		
Main effect (Consistency): $F(1, 58) = 2.14$, $p = .149$		
Interaction (Consistency x ELL status): $F(1, 58) = 0.21$, $p = .647$		

*Significant effect

Table 3

Summary of Target Word Eye Movement Parameters Across Sentence Type by ELL Status

Measure	Consistent *M (SD)*	Inconsistent *M (SD)*
First fixation duration (ms) (*N* = 60)	273 (60)	287 (63)
ELL (*n* = 31)	285 (65)	290 (59)
EL1 (*n* = 29)	261 (53)	284 (67)
Between groups: $F(1, 58) = 1.18$, $p = .282$		
Main effect (Consistency): $F(1, 58) = 3.62$, $p = .062$		
Interaction (Consistency x ELL status): $F(1, 58) = 1.42$, $p = .239$		
Gaze duration (ms) (*N* = 60)	340 (101)	404 (128)
ELL (*n* = 31)	347 (95)	431 (136)
EL1 (*n* = 29)	333 (107)	376 (115)
Between groups: $F(1, 58) = 1.73$, $p = .193$		
Main effect (Consistency): $F(1, 58) = 20.65$, $p < .001$*		
Interaction (Consistency x ELL status): $F(1, 58) = 2.01$, $p = .162$		
Total fixation time (ms) (*N* = 60)	542 (239)	713 (291)
ELL (*n* = 31)	578 (243)	723 (297)
EL1 (*n* = 29)	502 (233)	704 (290)
Between groups: $F(1, 58) = 0.57$, $p = .454$		
Main effect (Consistency): $F(1, 58) = 38.42$, $p < .001$*		
Interaction (Consistency x ELL status): $F(1, 58) = 1.05$, $p = .311$		
Fixation count (#) (*N* = 60)	1.66 (0.90)	2.29 (1.01)
ELL (*n* = 31)	1.69 (0.86)	2.20 (0.91)
EL1 (*n* = 29)	1.63 (0.95)	2.39 (1.10)
Between groups: $F(1, 58) = 0.08$, $p = .778$		
Main effect (Consistency): $F(1, 58) = 64.92$, $p < .001$*		
Interaction (Consistency x ELL status): $F(1, 58) = 2.69$, $p = .107$		

Table 3 Continued

Regression out count (#) (*N* = 60)	1.05 (0.44)	1.14 (0.39)
ELL (*n* = 31)	1.05 (0.47)	1.09 (0.39)
EL1 (*n* = 29)	1.05 (0.41)	1.20 (0.39)
Between groups: $F(1, 58) = 0.34$, $p = .562$		
Main effect (Consistency): $F(1, 58) = 2.89$, $p = .094$		
Interaction (Consistency x ELL status): $F(1, 58) = 1.09$, $p = .301$		

*Significant effect

CHAPTER 4

DISCUSSION

Extant research suggests that for EL1 students, a strong relationship exists between scores on measures of ORF and scores on measures of reading comprehension (Fuchs et al., 2001; Reschly et al., 2009). Therefore, ORF measures are commonly used in schools as a quick and easy gauge of students' reading ability. But recent research findings suggest that the relationship between ORF scores and reading comprehension scores may not be as strong for ELL students as it is for EL1 students (Crosson & Lesaux, 2010; Quirk & Beem, 2012). Although these studies call into question the use of ORF measures with ELL students, without knowledge of the underlying reading behaviors at play, we can only speculate about the reasons that the relationship between ORF and reading comprehension may vary for ELL and EL1 students. In order to better understand potential differences between ELL and EL1 students' reading behaviors, the current study employed eye tracking technology and individually presented sentences that were either consistent or inconsistent in meaning.

Through the use of eye tracking technology, prior studies have successfully examined elementary students' reading through investigation of comprehension monitoring behaviors. However, this is the first study to directly compare comprehension monitoring behaviors of ELL and EL1 students utilizing eye tracking technology. Comprehension monitoring reflects a student's ability to be aware of text meaning while reading and is typically identified through global analysis of fixations and regressions within the text as well as fixations and regressions for target words that were central to meaning (Rayner et al. 2006; Vorstius et al., 2013).

Therefore, these eye movement variables were analyzed for the fourth- and fifth-grade ELL and EL1 students in our sample in addition to outcome data on response to comprehension questions following the text.

Consistent with previous findings that ELL students often fall behind their EL1 peers on higher-level reading tasks such as reading comprehension (August & Shanahan, 2006, 2010; Lesaux & Kieffer, 2010; Nakamoto et al., 2007), analysis of question response accuracy indicated that ELL participants performed worse than EL1 participants on comprehension questions. Despite the fact that ELL and EL1 students in this study had the same ORF ability, ELL students were not as adept as EL1 students at recognizing whether the experiment sentences made sense or not. This finding corroborates results from past studies that suggest that ORF scores may not be as directly related to reading comprehension scores for ELL students as they are for EL1 students (Crosson & Lesaux, 2010; Quirk & Beem, 2012).

Although significant differences were indicated between ELL and EL1 students on the outcome measure of comprehension, these differences were not observed in analyses of eye movement variables. Had EL1 students engaged in more comprehension monitoring behaviors than ELL students, we would have expected to see greater fixation counts, longer fixation durations, and more inter-word regressions for EL1 students when inconsistencies in text were detected (Rayner et al. 2006; Vorstius et al., 2013). However, ELL students did not perform significantly differently from EL1 students on eye movements across all words within the sentences or on eye movements for the final words in the sentences that were essential to the sentence meaning. Failure to observe differences may be due to similarities between ELL and EL1 students, such as attendance at the same schools and instruction from the same teachers.

Across all students, differences were observed for eye movement variables for consistent versus inconsistent sentences, indicating that overall, the students were engaging in comprehension monitoring behaviors. In global analyses investigating reading behaviors across all words in the sentences, gaze duration was significantly lower for consistent sentences than inconsistent sentences. Therefore, students spent a longer amount of time reading words in inconsistent sentences before moving to another word than they did in consistent sentences. Other global eye movement parameters were not significantly different for consistent and inconsistent sentences, which could be due to the diluting effect of averaging across all words in the sentence as opposed to focusing on words that were central to the sentence meaning.

Target word analyses were conducted to gain a more concentrated look at the final word in the sentences, which determined whether the sentence had consistent or inconsistent meaning. Differences were found on measures examining fixations on the last word, but not for those evaluating regressions. Students fixated more times and had longer fixation times (gaze duration and total fixation time) on the last word of inconsistent sentences than consistent sentences. These findings indicate that when participants encountered a word that was inconsistent with the meaning of the rest of the sentence, they fixated more times on that word and spent more time on initial and total readings of that word than they did for consistent sentences. Although students engaged in the expected comprehension monitoring behaviors with regards to fixations, differences in regressions out of the last word to reexamine prior information were not observed based on sentence consistency. Failure to observe differences in regressions may be due to the short length of sentences employed within the study. It may have been possible for students to hold all of the sentence information in working memory without needing to return to previously read words when inconsistencies were discovered.

Limitations and Implications

The results from the present study should be interpreted with consideration of several limitations. First, English language proficiency scores and data on the amount of time students had lived in the U.S. were not available for the participants. Therefore, it was difficult to determine the degree to which the ELL students in our sample had mastered the English language. Data from our home language survey indicated that the ELL participants were regularly exposed to Spanish at home, with 48% of parents noting that they either never or infrequently spoke English and another 36% noting that they spoke English and Spanish equally. Despite differences in the ELL and EL1 participant samples due to the ELL students receiving more limited exposure to English in the home, similarities existed in their school environment. All students attended school in the same district and thus were exposed to the same reading curriculum and English instruction. A second limitation is that all ELL students spoke Spanish, and thus study results should not be generalized beyond the Spanish-speaking population.

In addition to limitations pertaining to the sample, the current study demonstrates certain limitations relating to the materials and procedures. A limitation common in the eye tracking research is that findings reflect students' behavior across silent readings of sentences, thus word reading accuracy cannot be measured. Although experimenter created sentences were beneficial to prevent possible prior exposure to the materials, these materials were limited in that no reliability or validity information was available. Two final limitations relate to the students' responses to the comprehension questions following the sentences. Although the pilot test was conducted to ensure that students did not have difficulty understanding the sentences employed in the experiment, the scarcity of errors meant that we were unable to conduct separate analyses for reading behaviors based on accuracy of question response. Additionally, the ease with which students understood the sentences may have limited potential differences in comprehension monitoring behaviors.

Despite the limitations of this study, findings indicate that Spanish-speaking ELL students performed at a lower level on a sentence comprehension measure than EL1 students with the same ORF abilities. This suggests that ELL students may have a delay in acquiring reading comprehension skills in comparison to their ORF skills. Therefore, caution should be taken when using ORF scores to extrapolate to overall reading achievement scores for ELL students. Additionally, the use of eye tracking technology provided insight into comprehension monitoring behaviors as students were reading consistent and inconsistent sentences. Findings from these data revealed that students did in fact engage in disrupted reading when they encountered inconsistent information in sentence reading, but the comprehension monitoring techniques did not differ between ELL and EL1 participants. Additional research is therefore needed to examine why EL1 students may outperform ELL students on measures of reading comprehension despite similar comprehension monitoring abilities.

REFERENCES

August, D., & Shanahan, T. (2006). *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language Minority Children and Youth*. Mahwah, NJ: Erlbaum.

August, D., & Shanahan, T. (2010). Response to a review and update on Developing literacy in second-language learners: Report of the national literacy panel on language minority children and youth. *Journal of Literacy Research, 42*, 341-348.

Baker, S. K., & Good, R. H. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. *School Psychology Review, 24*, 561-578.

Binder, K. S. (2003). Sentential and discourse topic effects on lexical ambiguity processing: An eye movement examination. *Memory & Cognition, 31*, 690-702.

Carroll, J. B., Davies, P., & Richman, B. (1971). The American Heritage word frequency book. Boston: Houghton Mifflin Company.

Christ, T. J., Arañas, Y. A., Kember, J. M., Kiss, A. J., McCarthy-Trentman, A., Monaghen, B. D., Newell, K. W., Van Norman, E. R., White, M. J. (2014). Formative Assessment System for Teachers Technical Manual: earlyReading, CBMReading, aReading, aMath, and earlyMath. Minneapolis, MN: Formative Assessment System for Teachers.

Crosson, A. C., & Lesaux, N. K. (2010). Revisiting assumptions about the relationship of fluent reading to comprehension: Spanish-speakers' text-reading fluency in English. *Reading and Writing, 23*, 475-494.

CTB/McGraw-Hill (2008). California English Language Development Test. Monterey, CA: Author.

Eme, E., Puustinen, M., & Coutelet, B. (2006). Individual and developmental differences in reading monitoring: When and how do children evaluate their comprehension? *European Journal of Psychology of Education, 21*, 91-115.

Figueroa, R. A., & Newsome, P. (2006). The Diagnosis of LD in English Learners: Is It Nondiscriminatory? *Journal of Learning Disabilities, 39*, 206-214.

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

Geva, E., & Yaghoub Zadeh, Z. (2006). Reading Efficiency in Native English-Speaking and English-as-a-Second-Language Children: The Role of Oral Proficiency and Underlying Cognitive-Linguistic Processes. *Scientific Studies of Reading, 10*, 31-57.

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., Gathercole, S. E., & Rayner, K. (2008). Children's and adults' processing of anomaly and implausibility during reading: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology, 61*, 708-723.

Klein, J. R., & Jimerson, S. R. (2005). Examining Ethnic, Gender, Language, and Socioeconomic Bias in Oral Reading Fluency Scores among Caucasian and Hispanic Students. *School Psychology Quarterly, 20*, 23-50.

Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal, 47*, 596-632.

Lesaux, N. K., & Siegel, L. S. (2003). The Development of Reading in Children Who Speak English as a Second Language. *Developmental Psychology, 39*, 1005-1019.

Liu, Y.-J., Ortiz, A. A., Wilkinson, C. Y., Robertson, P., & Kushner, M. I. (2008). From early childhood special education to special education resource rooms: Identification, assessment, and eligibility determinations for English language learners with reading-related disabilities. *Assessment for Effective Intervention, 33*, 177-187.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2009). Gates MacGinitie Reading Tests, Fourth Edition. Rolling Meadows, IL: Riverside.

McMaster, K. L., Wayman, M. M., & Cao, M. (2006). Monitoring the reading progress of secondary-level English learners: Technical features of oral reading and maze tasks. *Assessment for Effective Intervention, 31*, 17-31.

Minnesota Department of Children, Families, and Learning (2002). Minnesota Comprehensive Assessments: Grade 3 reading test specifications.

Muyskens, P., Betts, J., Lau, M. Y., & Marston, D. (2009). Predictive validity of curriculum-based measures in the reading assessment of students who are English language learners. *California School Psychologist, 14*, 11-21.

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing, 20*, 691-719.

National Center for Education Statistics. (2014). *The condition of education 2014 – English language learners*, Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Pro-Ed Inc. (2008). Test of Word Reading Efficiency. Austin, TX: Author.

Quirk, M., & Beem, S. (2012). Examining the relations between reading fluency and reading comprehension for English language learners. *Psychology in the Schools, 49*, 539-553.

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 49*, 151-173.

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

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye Movements as Reflections of Comprehension Processes in Reading. *Scientific Studies of Reading, 10*, 241-255.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-Based Measurement Oral Reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*, 427-469.

Shinn, M. M. & Shinn, M. R. (2002). AIMS web training workbook: Administration and scoring of reading curriculum-based measurement (R-CBM) for use in general outcome measurement. Bloomington, MN: Pearson, Inc.

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*, 1665-1690.

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

Weiderholt, J. L., & Bryant, B. R. (2001). Gray Oral Reading Test, Fourth Edition (GORT–4). Austin, TX: PRO-Ed.

Wiley, H. I., & Deno, S. L. (2005). Oral Reading and Maze Measures as Predictors of Success for English Learners on a State Standards Assessment. *Remedial and Special Education, 26*, 207-214.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson III Tests of Academic Achievement. Itasca, IL: Riverside Publishing.