MONITORING EYE MOVEMENTS TO MEASURE READING BEHAVIOR AND INSTRUCTIONAL EFFECTS

by

TORI ELLA FOSTER

(Under the Direction of Scott P. Ardoin)

ABSTRACT

Recent developments in eye movement recording have led to its increased use in research with previously "hard-to-track" individuals, including young children. In addition, in light of its unique potential for capturing concurrent behaviors and mechanisms underlying the reading process, reading researchers have demonstrated increased interest in using the technology to investigate reading skill development and instructional effects. However, due to its newness in the field of applied reading research, the use of eye-tracking methodology is certainly not without its limitations, and substantial gaps remain within the existing scientific literature. The first study in this dissertation examined the technical adequacy (i.e., test-retest reliability, alternate-form reliability, and concurrent criterion-related validity) of recently utilized eye movement measures among 175 second-grade students. The second study investigated the influences of age and word frequency on eye movement patterns exhibited by 72 unskilled readers (36 adult literacy learners and 36 students in Grades 2–5 matched on broad achievement level) during four consecutive rereadings of the same text. Taken together, results from both empirical studies affirm the link between students' eye movements during reading and their level of reading skills (e.g., broad reading achievement, as measured by more traditional assessment tasks) and extend upon and clarify previous eye movement research with adults and children.

INDEX WORDS: Reading, Eye tracking, Eye movements, Technical adequacy, Reliability, Validity, Test-retest reliability, Alternate-form reliability, Criterion-related validity, Reading achievement, Reading fluency, Repeated reading, Repetition, Rereading, Word frequency, Elementary students, Adult literacy learners

MONITORING EYE MOVEMENTS TO MEASURE READING BEHAVIOR AND INSTRUCTIONAL EFFECTS

by

TORI ELLA FOSTER

B.A., University of Richmond, 2008

M.A., University of Georgia, 2012

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2016

MONITORING EYE MOVEMENTS TO MEASURE READING BEHAVIOR AND

INSTRUCTIONAL EFFECTS

by

TORI ELLA FOSTER

Major Professor: Scott P. Ardoin

Committee: A. Michele Lease
Jennifer H. Lindstrom
Amy L. Reschly

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2016

ACKNOWLEDGEMENTS

This research was supported by the Institute of Education Sciences, U.S. Department of Education (through Grant R305A100496 to the University of Georgia) and the Society for the Study of School Psychology. The opinions expressed are those of the author and do not represent views of these organizations.

This dissertation would be incomplete without acknowledgement of the many contributions made by faculty members, colleagues, family, and friends. First, I would like to thank my mentor, Dr. Scott Ardoin, for his incredibly generous support, his patience and insight during several difficult challenges, and the high standards and respect with which he has trained me throughout my graduate career. I would also like to express sincere gratitude to Drs. Kathy Binder, Michele Lease, Jennifer Lindstrom, and Amy Reschly for challenging me to improve my work and to aspire to higher levels of scientific skepticism and scholarly responsibility. I am particularly grateful to my colleagues who assisted with data collection (especially Andrea Zawoyski, Patrick Morin, and Cameron Oddone) and my friends from UGA and NICPP who motivated me to keep writing (and taking deep breaths) throughout the past year. Finally, I owe the deepest appreciation to my family members and loved ones for their unfailing care and encouragement and their gracious understanding throughout the past several years. They have trusted in and promoted my best even when encountering me at my worst, and I could not have completed this process without them.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Eye tracking is a unique technology that allows researchers to observe and examine multiple behaviors and processes underlying reading (Rayner, 1998). Specifically, eye movement research is based on the assumption that eye movement patterns consisting of rapid movements (*saccades*) and pauses (*fixations*) reflect otherwise unobservable information processing involved in reading (LaBerge & Samuels, 1974; Rayner, 2009). Past research indicates that individuals extract textual information during fixations, then direct their eyes to new locations to gather subsequent input from areas of visual acuity. During saccades (the movements between fixations), vision is suppressed, and readers do not extract information (Rayner, Chace, Slattery, & Ashby, 2006). Occasionally, readers make backward saccades (*regressions*), which seem to reflect further processing of previously encountered information or correction for misdirected movements (Just & Carpenter, 1980; Rayner et al., 2006).

Characteristics of readers' eye movements (e.g., length, frequency, and duration) have been noted to change as a function of the level of difficulty associated with reading a particular text; increased difficulty, due to a reader's skill level and/or the difficulty level of a passage, is associated with lengthier fixations, shorter saccades, and more regressions (Rayner et al., 2006). Thus, eye movement patterns provide researchers with a dynamic record of information processing that allows them to pinpoint reading ease or difficulty within a text. As a result, by recording and analyzing the magnitude, count, and temporal extent (i.e., duration) of the aforementioned movements, researchers can gain a sense of how and when different textual variables (e.g., word length, frequency, predictability, etc.) aid or disrupt processing. For example, studies indicate that readers spend more time fixating on and, thus, require more time in processing uncommon, difficult, important, and long words as compared to familiar, short, and function words (Just & Carpenter, 1980; Rayner, 1983; Rayner, 2009). Due to established relationships between eye movement characteristics and textual processing, reading researchers can regard eye movement patterns as behavioral indicators of problems or skills with encoding and comprehending text (Rayner et al., 2006).

Not only does eye tracking allow researchers to localize effects of textual variables on reading, but it also presents researchers with numerous technological advantages. Given its high temporal resolution (i.e., recording capabilities at the millisecond level), eye tracking yields information regarding the time course of reading behaviors and is recognized as an ideal method for inferring moment-to-moment processes (Rayner, 1998, 2009; Rayner et al., 2006). Eye tracking also allows researchers to observe natural reading directly without relying on secondary tasks (Rayner, 1998) or conflating silent reading behavior with other phenomena like recall or articulation (Rayner et al., 2006). Finally, because of the highly detailed records it produces, eye tracking enables examination of multiple processes (e.g., word recognition, higher-level processing) and behaviors (e.g., fixations, saccades, regressions) within a single sample of data (e.g., a single reading of a given passage; Just & Carpenter, 1980; Rayner, 1998).

In addition to enabling empirical investigation of behaviors underlying reading and yielding theoretical insight regarding textual processing, eye-tracking research has valuable implications for the practice of school psychology. Unlike commonly utilized reading measures that simply capture overall changes in outcomes (e.g., performance on broad measures of academic skill; changes in fluency measured in words read correctly in a minute, WRCM), eye monitoring allows researchers to investigate instructional effects with precision and to identify improvements specific to certain words or passage characteristics. Eye movement studies by Foster, Ardoin, and Binder (2013) and Ardoin, Binder, Zawoyski, Foster, and Blevins (2013) suggest that commonly used fluency-based intervention procedures primarily improve students' reading of low-frequency words. Eye-tracking research also has the potential to clarify the developmental trajectories of and relationships among component skills of reading (i.e., phonemic awareness, phonics, fluency, vocabulary, and comprehension; National Institute of Child Health and Human Development [NICHD], 2000). For example, recent eye movement studies have indicated that careful rereading may facilitate comprehension but adversely impact fluency (Vorstius, Radach, Mayer, & Lonigan, 2013) and that phonemic awareness skills contribute to later silent reading fluency and word recognition (Ashby, Dix, Bontrager, Dey, & Archer, 2013). Such studies have the potential to yield greater insight into students' reading behavior during assessment and intervention tasks (Miller & O'Donnell, 2013), to help identify appropriate targets for intervention, and to inform effective instructional practices.

Coupled with the advantages outlined above, recent developments in eye-tracking technologies have led to their increased use in reading research, broadening the array of extant studies in the scientific literature on eye movements. In contrast with eye trackers utilized during the 1980s and 1990s, modern eye-monitoring systems better account for head movement and, thus, do not require absolute head stabilization dependent on apparatuses like headrests and bite bars. As a result, researchers can conduct eye tracking with a broader range of participants than previously possible, including individuals wearing eyeglasses and young children who have difficulty remaining still for extended periods of time (Blythe & Joseph, 2011; Miller & O'Donnell, 2013; Rayner, Ardoin, & Binder, 2013). Modern eye-tracking systems also are more “user-friendly” and easier to master, require less effort on the part of experimenters (e.g., do not require hand scoring of data; Rayner et al., 2013), and offer users advanced capabilities (e.g., improved software and sampling rates) and increased portability at lower costs (Miller & O’Donnell, 2013).

Alongside technological advances that have improved the feasibility of conducting eye movement research with broader ranges of participants, researchers’ growing interest in skill development has yielded an increasing number of eye movement studies on similarities and differences between skilled and unskilled reading behavior. Whereas earlier cognitive studies focused on the “end state” of successful reading acquisition (Miller & O’Donnell, 2013) and were conducted with proficient adult readers (i.e., college students or graduates), recent research has examined how the eye movements of adults and children resemble and/or differ from each other with regard to both the eye movements themselves (i.e., oculomotor functioning) and the patterns they constitute, which are thought to reflect linguistic processing (e.g., Blythe, Häikiö, Bertram, Liversedge, & Hyönä, 2011; Blythe, Liversedge, Joseph, White, & Rayner, 2009; Joseph & Liversedge, 2013).

Studies of eye movement control among adults and children reveal several similarities between groups. Research on the perceptual span (i.e., the area from which useful information can be obtained during a fixation) indicates that, as early as the sixth grade, child and adult readers extract comparable amounts of word length information (Rayner, 1986), letter feature information (Rayner, 1986), and letter identity information (Häikiö, Bertram, Hyönä, & Niemi, 2009) during fixations. In addition, both beginning readers and adult readers exhibit asymmetrical perceptual spans, such that they are able to acquire more information to the right of a fixation than to the left (Rayner, 1986). Eye movement research also suggests that adults and children process information during reading in similar ways. Children and adults appear equally capable of processing words even after seeing them for extremely brief intervals (e.g., as short as 80 ms); by the age of 7 years, children capture visual information almost as quickly and efficiently as do adults (Blythe et al., 2009). Furthermore, both adults and children seem to adjust their allocation of visual attention based on different text characteristics. For example, elementary students and adult readers attend to and process information further to the right of a fixation when reading unified words (e.g., compound words that are spatially close together) versus separate words (Häikiö, Bertram, & Hyönä, 2010).

Despite similarities in the oculomotor control and eye movement behaviors of adults and children, research has also identified many differences between the eye movement patterns exhibited by each group. Compared to adults, children skip words (i.e., do not fixate on them) less frequently, *re*fixate on words more frequently (Blythe et al., 2011), and make a large number of short, “express” saccades (occupying less than a single letter position) that are not emitted during adult reading behavior (McConkie et al., 1991). Children also primarily focus their visual attention on words in the center of their fields of vision and thus extract less “preview” information from upcoming words (Rayner, 1986). Studies of binocular coordination indicate that, although both children and adults are able to coordinate distinct visual input from each eye into a unified percept, children exhibit greater disparity between eyes and make a higher proportion of crossed fixations (such that the left eye is further to the right than the right eye, and vice versa) than adults (Blythe et al., 2006; Kirkby, Webster, Blythe, & Liversedge, 2008). Given that the disparity between eyes does not vary as a function of processing difficulty, age-related differences in binocular coordination (i.e., how the two eyes move *in relation to each other*) seem to stem from developmental differences in eye motor control (Blythe et al., 2006;

Kirkby et al., 2008). However, research indicates that age-related differences in eye movement patterns during reading (i.e., how the eyes move *in relation to text*) primarily relate to developmental differences in linguistic processing. Eye movement studies have suggested that, compared to adults, children utilize similar yet slower mechanisms for syntactic processing (i.e., computing the structures of language; Joseph & Liversedge, 2013), are slowed down more when reading and processing long words (Joseph, Liversedge, Blythe, White, & Rayner, 2009), and are less skilled at integrating real-world, practical knowledge into sentence representations (Joseph et al., 2008).

Although the current chapter has mentioned only a small portion of the eye movement studies conducted with adults and children, this research literature suggests that, in general, group differences in eye movement patterns during reading relate to differences in reading skill and linguistic processing. That is, age-related differences during reading do not seem to result from differences in the actual mechanics of adults' and children's eye movements, but rather from differently developed reading-related skills (e.g., the ability to identify words efficiently, the ability to integrate practical information with textual information; Reichle et al., 2013). Adults and children appear to direct their eyes similarly with regard to *where* they move or land (e.g., where within a word to initially fixate), but their decisions about *when* to move their eyes are under cognitive control and are influenced strongly by linguistic processing (Reichle et al., 2013; Zang, Liang, Bai, Yan, & Liversedge, 2013). Furthermore, children and adults exhibit similar processing mechanisms, but those of children are typically less efficient and require longer courses of time (Blythe & Joseph, 2011). Of note, these general conclusions also align with findings from eye movement research conducted with children at different ages and grade levels. Although older and younger children demonstrate similar eye movement control (e.g., target saccades to certain landing positions similarly), more skilled readers exhibit higher average fixation counts and durations, longer saccades, and fewer regressions (Blythe & Joseph, 2011; Buswell, 1922; McConkie et al., 1991).

Despite aforementioned advances in the technology and breadth of research associated with the study of eye movements in reading, the use of eye-tracking methodology is certainly not without its flaws. Due to its nascence in the field of applied reading research, much remains to be uncovered regarding how eye movement characteristics may be influenced by external factors impacting reading performance (e.g., inattention, sleep deprivation, anxiety, learning and psychological problems). As with all assessment/data collection methods involving the sampling of behavior at a specific point in time, eye movement monitoring is susceptible to the influence of individual differences and environmental factors that may impact performance and compromise internal and external validity of specific results. For example, recent studies of "mindless reading" (i.e., "zoning out," such that readers are not aware of what they just read and/or are thinking about something other than the text; Reichle, Reineberg, & Schooler, 2010; Schad, Nuthmann, & Engbert, 2012) have begun to describe how inattention exerts clearly distinguishable effects on readers' eye movements compared to their behavior during typical mindful reading. Specifically, readers exhibit longer fixation durations, decreased sensitivity to lexical and linguistic variables, and more erratic eye movements even before they are aware of their own mind-wandering behavior (Reichle et al., 2010). Although such research has begun to elucidate how reading behavior might be affected by level of attention, recent studies investigating the effects of other individual differences on eye movement patterns associated with reading are scarce. Eye movement research also is adversely affected by other disadvantages, including the considerable cost, limited portability, vulnerability to technological malfunction, and difficulty of data collection/analysis (e.g., time-consuming data cleaning procedures) associated with use of eye-tracking technology. Although modern advances have mitigated these obstacles (Miller & O'Donnell, 2013), they continue to limit the feasibility of eye monitoring as well as its practical use. Furthermore, due to the fine-grained nature of eye movement data, research based on these reading behaviors may seem to lack applied significance and implications compared to other assessment methods that more closely approximate everyday educational procedures (e.g., paper-and-pencil measures).

Despite these limitations and flaws, it is expected that eye movement research has the potential to reveal underlying mechanisms of reading and thus may expand our current understanding of reading behavior. However, despite aforementioned progress, substantial gaps remain within the existing scientific literature. Two such gaps, which are particularly relevant within the context of this dissertation, involve the technical adequacy of eye movement data and the influence of variables like reading skill and age on developmental differences in eye movement patterns.

Researchers have asserted that eye-tracking methods have great utility in the investigation of reading behavior (Rayner et al., 2006) and how it changes along with improvements in accuracy, fluency, and comprehension (Rayner et al., 2013). Furthermore, the increasing use of eye-tracking methods in applied reading research has resulted in claims that eye movement studies have the potential to clarify developmental trajectories of reading acquisition, to guide development and validation of reading assessment methods, and to inform reading instruction and intervention (Miller & O'Donnell, 2013). Such statements are based on experimental cognitive research that assumes a representative relationship between reading processes and eye movement patterns; however, given that well-designed applied research (e.g., studies providing evidence for the use of particular instructional procedures) requires evidence of the reliability and validity of assessment measures (Seftor et al., 2011), research evaluating the technical adequacy of eye movement measures is imperative. Although studies involving both eye monitoring and administration of more traditional reading measures (e.g., standardized tests of reading skills) have suggested predictive relationships between test performance and eye movement patterns (e.g., Huestegge, Radach, Corbic, & Huestegge, 2009; Kuperman & Van Dyke, 2011), up-to-date research primarily focusing on and investigating the reliability and validity of eye movement measures of reading is lacking.

Similarly, although studies (e.g., Joseph et al., 2008; Joseph et al., 2009; Joseph & Liversedge, 2013) have yielded preliminary findings regarding developmental differences between the eye movement patterns of adults and children, there is a need for further research that explicitly investigates the contributions that reading skill and age make to these differences. Given that eye movement studies with adults have largely examined the reading behavior of proficient readers (i.e., college students or graduates), it is unclear if observed differences between adults and children stem from age-related developmental differences in eye movements (i.e., differences between adults and children), differences in skill level (i.e., differences between skilled and unskilled readers), and/or between-group differences in linguistic processing (e.g., strengths and weaknesses in orthographic versus phonological processing; Greenberg, Ehri, & Perrin, 2002; Thompkins & Binder, 2003). As an example, research on rereading behavior has yielded disparate findings between studies conducted separately with adult participants and children (Ardoin et al., 2013; Foster et al., 2013). However, given the absence of controlled experiments directly comparing the rereading behavior of these groups, differences could relate to any of the aforementioned factors (i.e., age, skill level, linguistic processing profiles) and/or additional differences between studied adult and child readers (e.g., strategies/approaches to reading, level of attention, task persistence, motivation, etc.). Although extensive research quite beyond the scope of this dissertation is necessary to evaluate all of these factors, the relative contributions of age and skill level seem most relevant to investigate at the current time due to their established importance based on past research and their ability to be identified and controlled/manipulated. Examining these factors offers at least partial clarification of past findings and seems to be a crucial stepping stone in enabling further, needed evaluations of the relationships between eye movements and additional group/individual differences.

Purpose

The present studies sought to lessen identified gaps in the literature and to extend upon and clarify recent eye movement research by examining (a) the technical adequacy of recently utilized eye movement measures and (b) how observed developmental differences might be impacted by the variables of reading skill and age. A secondary purpose of both empirical studies was to integrate and build upon research findings from multiple fields (i.e., cognitive literature including eye movement studies, applied reading research, adult literacy research). The first study involved the analysis of collected data detailing children's eye movements during reading; Chapter 2 documents the investigation of relationships within and between these data and other collected assessment data indicating children's reading skills (e.g., standardized test performance, curriculum-based measurement, etc.). The second study involved observation and empirical investigation of eye movement patterns exhibited by unskilled adult readers and children during rereading; Chapter 3 details this examination of the influences of age and reading skill on reading behavior (as evidenced by eye movements).

References

Ardoin, S. P., Binder, K. S., Zawoyski, A. M., Foster, T. E., & Blevins, L.A. (2013). Using eye-tracking procedures to evaluate generalization effects: Practicing target words during repeated readings within versus across texts. *School Psychology Review*, *42*(4), 477-495.

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

Blythe, H. I., Häikiö, T., Bertram, R., Liversedge, S. P., & Hyönä, J. (2011). Reading disappearing text: Why do children refixate words? *Vision Research*, *51*, 84-92. doi: 10.1016/j.visres.2010.10.003

Blythe, H. I., & Joseph, H. S. S. L. (2011). Children's eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.). *The Oxford handbook of eye movements* (pp. 801-818). Oxford, Great Britain: Oxford University Press.

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, *49*, 1583-1591. doi: 10.1016/j.visres.2009.03.015

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., Findlay, J. M., & Rayner, K. (2006). The binocular coordination of eye movements during reading in children and adults. *Vision Research*, *46*, 3898-3908. doi: 10.1016/j.visres.2006.06.006

Buswell, G. T. (1922). *Fundamental reading habits: A study of their development*. Chicago: The University of Chicago.

Foster, T. E., Ardoin, S. P., & Binder, K. S. (2013). Underlying changes in repeated reading: An eye movement study. *School Psychology Review*, *42*(2), 140-156.

Greenberg, D., Ehri, L. C., & Perin, D. (2002). Do adult literacy students make the same word-reading and spelling errors as children matched for word-reading age? *Scientific Studies of Reading*, *6*(3), 221-243. doi: 10.1207/s1532799xssr0603_2

Häikiö, T., Bertram, R., & Hyönä, J. (2010). Development of parafoveal processing within and across words in reading: Evidence from the boundary paradigm. *The Quarterly Journal of Experimental Psychology*, *63*(10), 1982-1998. doi: 10.1080/17470211003592613

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology*, *102*, 167-181.

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, *49*, 2948-2959. doi: 10.1016/j.visres.2009.09.012

Joseph, H. S. S. L. & Liversedge, S. P. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLoS ONE*, *8*(1): e54141. doi: 10.1371/journal.pone.0054141

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*(5), 708-723. doi: 10.1080/17470210701400657

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research*, *49*, 2078-2086. doi: 10.1016/j.visres.2009.05.015

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354. doi: 10.1037/0033-295x.87.4.329

Kirkby, J. A., Webster, L. A. D., Blythe, H. I., & Liversedge, S. P. (2008). Binocular coordination during reading and non-reading tasks. *Psychological Bulletin*, *134*(5), 742-763. doi: 10.1037/a0012979

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*, 42-73. doi: 10.1016/j.jml.2011.03.002

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293-323. doi: 10.1016/0010-0285(74)90015-2

McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, R. B., & Wolff, P. M. (1991). Children's eye movements during reading. In J. F. Stein (Ed.), *Vision and visual dyslexia* (pp. 251-262). London, England: The Macmillan Press, Ltd.

Miller, B., & O'Donnell, C. (2013). Opening a window into reading development: Eye movements' role within a broader literacy research framework. *School Psychology Review*, *42*(2), 123-140.

National Institute of Child Health and Human Development (2000). Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Rayner, K. (1983). The perceptual span and eye movement control during reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 97-139). New York: Academic Press.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41*(2), 211-236. doi: 10.1016/0022-0965(86)90037-8

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. doi: 10.1037/0033-2909.124.3.372

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

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

Rayner, K., 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. doi: 10.1207/s1532799xssr1003_3

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*, 110-149. doi: 10.1016/j.dr.2013.03.001

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science*, *21*, 1300-1310. doi: 10.1177/0956797610378686

Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: Objective measures reveal mindless reading at different levels. *Cognition*, *125*, 179-194. doi: 10.1016/j.cognition.2012.07.004

Seftor, N., Constantine, J., Cody, S., Ponza, M., Knab, J., Deke, J., & Monahan, S. (2011). *What Works Clearinghouse: Procedures and standards handbook 2011* (NCEE 2011-XXXX). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Thompkins, A. C., & Binder, K. S. (2003). A comparison of the factors affecting reading performance of functionally illiterate adults and children matched by reading level. *Reading Research Quarterly*, *38*(2), 236-258. doi: 10.1598/RRQ.38.2.4

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.

Zang, C., Liang, F., Bai, X., Yan, G., & Liversedge, S. P. (2013). Interword spacing and landing position effects during Chinese reading in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 720-734. doi: 10.1037/a0030097

CHAPTER 2

RELIABILITY AND VALIDITY OF EYE MOVEMENT MEASURES OF CHILDREN'S READING[1]

[1] Foster, T.E. and S.P. Ardoin. To be submitted to *Reading Research Quarterly*.

Abstract

In light of the practical implications of applied eye movement research, there is a clear need to scrutinize the methodological and design characteristics of these studies with scientific objectivity. Although strong claims have been made regarding the educational utility of eye tracking (i.e., potential to inform theory, assessment, and instructional practices), such statements seem somewhat unfounded in the absence of clear evidence regarding the technical adequacy of eye movement data. Past studies have yielded direct and indirect evidence concerning the utility of eye movements as measures of reading, but even remotely recent research explicitly investigating their reliability and validity is lacking. Thus, the current study was designed to update and extend upon past research by investigating the test-retest reliability, alternate-form reliability, and concurrent criterion-related validity (relative to several current assessments of reading-related skills) of recently utilized eye movement measures of children's reading. Participants were 175 second-grade students whose eye movements were monitored during silent reading of experimenter-developed narrative text(s) at 2 points in time. Participants were also administered reading subtests from the Woodcock-Johnson III Tests of Academic Achievement, 3 curriculum-based measurement probes, and a measure of morphological awareness at the first point of data collection. Correlation analyses were used to evaluate the technical adequacy of eye movements across passage reading and for embedded high-frequency and low-frequency words. Results indicate adequate reliability and validity for passage-level measures of fixation duration but suggest that elementary students' reading behaviors relative to specific words are more variable, susceptible to change, and weakly associated with their normative levels of reading achievement. Implications for conducting and interpreting eye movement studies of reading are discussed.

Introduction

Although the use of eye tracking originated primarily within experimental cognitive studies, eye-monitoring methods have become increasingly common in applied reading research due to several methodological advantages. First, eye tracking allows for *direct* observation of natural reading processes. Specifically, by monitoring eye movements, researchers can observe the otherwise invisible process of silent reading without requiring participants to pronounce words aloud or to perform secondary actions (e.g., button presses) that might impact their normal reading behavior (Rayner, 1998; Rayner, Chace, Slattery, & Ashby, 2006). Second, eye tracking enables *precise* measurement of reading behaviors on both temporal and spatial levels. That is, eye-tracking systems record fleeting movements in real time, allowing researchers to pinpoint exactly when (to the millisecond) and where within a visual stimulus (e.g., a reading passage) the movements occur (Rayner, 1998, 2009; Rayner et al., 2006). The precise nature of eye movement recording has allowed researchers to discern the time course of behaviors underlying reading and to identify relationships between eye movement patterns and characteristics of given words within a text. For example, research indicates that readers spend more time fixating on uncommon, difficult, important, and long words as compared to familiar, short, and function words (Just & Carpenter, 1980; Rayner, 1983, 2009). Due to the highly detailed records it produces, eye tracking is also *versatile.* Specifically, eye-monitoring procedures allow researchers to investigate multiple processes (e.g., word recognition, higher-level processing) within a single sample of reading behavior (Just & Carpenter, 1980; Rayner, 1998). Eye movement records contain a wealth of data concerning multiple qualities (e.g., count/frequency, magnitude, duration) of multiple eye movements (e.g., fixations, saccades, regressions); as a result, data collected during a single reading trial can be aggregated and analyzed in several different ways to reveal information about concurrent behaviors and processes underlying reading.

Eye Movements Underlying Reading

During eye-tracking studies, reading researchers observe and record patterns consisting of two broad categories of eye movements: *fixations* (pauses) and *saccades* (rapid movements). Although readers generally feel like they are moving their eyes smoothly across lines of text, the reading process is actually characterized by a somewhat fragmented stream of eye movement behaviors. While making saccades, readers move their eyes from one point to another, and vision is suppressed. In contrast, during fixations, readers extract textual information from fixated points and surrounding areas of visual acuity (Rayner et al., 2006). After information extraction and processing is complete, readers direct their eyes to new locations to gather input for subsequent processing (Just & Carpenter, 1980). Although movements typically take the form of forward (i.e., rightward) saccades or *return sweeps*, which involve movement from the end of a line to the beginning of the next line (Just & Carpenter, 1980), readers occasionally make backward saccades (*regressions*), which are generally interpreted as additional processing of previously extracted information or correction for misdirected movements (e.g., overshooting the target location; Just & Carpenter, 1980; Rayner et al., 2006).

Eye movement studies are based on the assumption that eye movement patterns reflect cognitive processes involved in reading (LaBerge & Samuels, 1974; Rayner, 2009). Given that the durations and frequencies of fixations and saccades change as a function of the reader's skill level and/or the difficulty of the material being read (Rayner et al., 2006), researchers generally interpret these behavioral characteristics as indicators of ease or difficulty with reading a text. Fixations are thought to allow readers to encode words, activate their representations of words and corresponding concepts, assign case roles (e.g., agent, recipient, action) to encountered words/concepts, process word meaning information, and integrate information across clauses/phrases and sentences. Thus, fixation durations are assumed to reflect time spent executing these processes, such that longer fixations indicate longer processing times due to greater demands (Rayner et al., 2006). Studies indicating that readers make longer fixations on difficult (e.g., uncommon or long) words as opposed to familiar, short, and unimportant words (e.g., function words; Just & Carpenter, 1980; Rayner, 1983, 2009) support the idea that eye movements reflect skills or problems with reading and processing. Furthermore, researchers have associated increased reading difficulty—a function of individual differences *and* manipulable textual variables like word length, frequency, predictability, etc.—with shorter saccades and increased regressions as well as lengthier fixations (Rayner et al., 2006).

Commonly Utilized Eye Movement Measures

Given that characteristics of eye movements (e.g., magnitude, count/frequency, duration) reflect processing demands, they vary considerably between words or sections within a given reading passage. Rather than fixating once per word, readers typically fixate on only two thirds of the words they encounter, often skipping short function words and predictable words (Rayner et al., 2006). In contrast, they sometimes fixate multiple times on other words, namely, those associated with increased processing difficulty (e.g., long words, unpredictable or uncommon words, etc.). Readers' tendency to exhibit uneven distributions of fixations throughout texts requires researchers to measure eye movements in ways that capture variability; merely averaging fixation time across all words in a passage would result in over- or underestimation of fixation time for certain words (Rayner, 1998).

Researchers commonly utilize multiple eye movement measures that summarize fixation behavior in different ways and are thought to reflect distinct aspects of textual processing (e.g., Blythe, Liversedge, Joseph, White, & Rayner, 2009; Joseph & Liversedge, 2013). To indicate time spent processing individual words, researchers use measures of *single fixation duration* (fixation duration when only one fixation is made on a word); *first fixation duration* (the duration of the first fixation on a word, regardless of the number of fixations made on the word); *gaze duration* (the summed duration of all of the fixations on a word prior to moving to another word); and *total fixation time* (the summed duration of all fixations, including regressions, on a word). Given that the first three measures account for fixations made only during the initial pass through a word (i.e., before the eyes move on to another word for the first time), they are thought to reflect early stages of word processing (e.g., word identification or retrieval). In contrast, total fixation time is thought to reflect later stages of processing (e.g., higher-level, integrative processes) due to its inclusion of rereading time and/or fixations made during multiple passes through a word. Measures conveying the probabilities of skipping or fixating words and the frequencies of regressions in and out of words are also useful indicators of lexical processing. To reflect processing beyond the word level, researchers simply aggregate eye movement data collected across a wider target area; they commonly use measures like *first-pass reading time* (the summed duration of all fixations within a given region prior to moving to another region) and *total reading time* (the summed duration of all fixations, including regressions, within a given region) to represent sentence- or discourse-level processing (Rayner et al., 2006).

Eye Tracking in Applied Reading Research

Due to recent technological advances in eye-monitoring apparatuses, the use of eye-tracking technology has become increasingly common in applied reading research (Miller &

O'Donnell, 2013). Whereas eye-tracking systems used in the 1980s and 1990s required the use of headrests and bite bars to ensure absolute head stabilization, modern eye trackers are more capable of handling and correcting for participant head movement. As a result, improvements in eye-tracking technology have made it easier to conduct research with "hard-to-track" individuals, including people wearing eyeglasses and young children who have difficulty remaining still (Blythe & Joseph, 2011; Miller & O'Donnell, 2013; Rayner, Ardoin, & Binder, 2013). Modern eye-monitoring systems also offer improved software and sampling rates and are easier to operate, more portable, and more affordable (Miller & O'Donnell, 2013; Rayner et al., 2013). Consequently, eye tracking as a research method has become more accessible and more likely to be utilized by researchers outside of the cognitive field in which its use began.

The growing use of eye tracking in applied research may also relate to increased interest in using the technology to investigate reading development and behavior among unskilled readers. Influenced by the cognitive revolution, earlier eye movement studies focused on the "end state" of successful reading and, as a result, predominantly involved skilled adult readers (i.e., undergraduate students or college graduates; Miller & O'Donnell, 2013). In contrast, recent research reflects a growing interest in examining the eye movements of developing readers. Recent studies with children and unskilled adult readers have sought to uncover how the eye movements of these individuals resemble and/or differ from those of skilled readers (e.g., Joseph et al., 2008; Joseph, Liversedge, Blythe, White, & Rayner, 2009) and how component skills of reading (i.e., phonemic awareness, phonics, fluency, vocabulary, and comprehension; National Institute of Child Health and Human Development [NICHD], 2000) develop over time and relate to each other and eye movement control (e.g., Ashby, Dix, Bontrager, Dey, & Archer, 2013; Vorstius, Radach, Mayer, & Lonigan, 2013). As researchers have demonstrated increased interest in conducting eye tracking with unskilled readers, participant samples in eye movement research have begun to more closely resemble participants in applied research on reading instruction.

Eye tracking has become increasingly common in reading research investigating skill development and educational practices (e.g., Ardoin, Binder, Zawoyski, Foster, & Blevins, 2013; Joseph, Nation, & Liversedge, 2013). Whereas experimental studies primarily explored basic characteristics of eye movements and effects of linguistic variables on eye movement patterns, applied research focuses on implications of these phenomena for assessment and instruction. In recent applied studies, researchers have used eye tracking to: (a) investigate differences in the reading behaviors of individuals at different stages of reading acquisition, (b) scrutinize relationships among component reading skills, (c) examine reading behavior during assessment tasks, and (d) identify instructional/intervention effects with precision (Ashby et al., 2013; Foster, Ardoin, & Binder, 2013; Joseph et al., 2013; Valle, Binder, Walsh, Nemier, & Bangs, 2013; Vorstius et al., 2013). According to eye movement researchers, such studies have the potential not only to clarify the developmental trajectories of and relationships among reading skills, but also to inform instructional targets, assessment strategies, and intervention practices (Miller & O'Donnell, 2013; Rayner et al., 2013).

In light of the increased practical implications of eye movement research, there is an obvious need for researchers to scrutinize the methodological and design characteristics of these studies with scientific objectivity. Before conclusions from eye movement research can be translated into practical guidelines affecting the everyday lives of students, careful attention must be given to the sampling and selection of data collection instruments and procedures inherent in each study. Due to individual differences in development, behavioral functioning (e.g., attention, motivation, psychosocial strengths and vulnerabilities), educational history, and social/environmental factors (e.g., socioeconomic status and distribution of resources), results achieved with a particular sample may not generalize appropriately to students in different contexts. Likewise, results are tied to the specific procedures that beget them. Of particular importance is the technical adequacy of measures/instruments utilized in research; in general, "the higher the stakes associated with a given test use, the more important it is that test-based inferences are supported with strong evidence of technical quality" (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). For eye monitoring to inform educational theory, development of reading assessment methods, and effective instructional practices (Miller & O'Donnell, 2013), researchers must first inspire confidence in the use and application of eye-tracking technologies and eye movement measures. In line with definitions of "well-designed" educational research (as outlined by organizations like the U.S. Department of Education's Institute of Education Sciences), researchers need to provide clear support for the reliability and validity of assessment measures, including eye movement data (Seftor et al., 2011).

Properties Indicating Technical Adequacy

Reliability. Under the classical true score model, reliability coefficients, or correlations between parallel measures (i.e., alternate forms or repeated administrations of the same measure), are thought to indicate the proportion of variability in observed performance that is due to true performance (Crocker & Algina, 1986). Thus, higher reliability coefficients reflect a lesser degree of error. According to Crocker and Algina (1986), there are no absolute guidelines defining a minimally acceptable level of reliability; however, many standardized test publishers report coefficients of equivalence or stability around .70 to .90. Similarly, Nunnally (1978) recommended that measures yield reliability coefficients of .80 in basic research or .90 to .95 in applied settings, in which higher-stakes decisions are made based on assessment results. Although many researchers use Nunnally's (1978) definition of "modest reliability"—indicated by a coefficient of .70—as a cutoff criterion, "satisfactory" cutoffs for reliability depend largely on the purpose/use of measures (Lance, Butts, & Michels, 2006). Furthermore, researchers evaluating test-retest reliability must carefully take into account how the amount of elapsed time between assessment administrations might affect reliability coefficients (Crocker & Algina, 1986). In the context of educational research, it is particularly important to consider and balance possible influences of events/changes unfolding during this time frame (e.g., maturation, instructional/intervention effects) and potential retest/practice effects.

Validity. Criterion-related validity is indicated by validity coefficients, or correlation coefficients between scores/performance on the assessment being evaluated and the criterion measure (Crocker & Algina, 1986). When considering minimally acceptable levels of criterion-related validity, clear standards do not exist. According to Cronbach (1970), the highest possible validity coefficients are desirable (i.e., "the bigger, the better"); however, coefficients rarely rise above .60, and even low positive correlations (e.g., .20) still indicate that a given assessment can predict criterion performance above chance levels. Another indicator of criterion-related validity is a measure's ability to *improve* prediction of criterion performance above that afforded by a known predictor variable. When evaluating criterion-related validity, assessment consumers must also consider several potential problems including selection of a suitable criterion, small sample sizes, "contamination"/influence of criterion performance on predictor performance, restriction of range, and low reliability of measures (Crocker & Algina, 1986).

Research Evaluating Eye Movement Measures

To date, only a handful of publications have addressed the reliability and validity of eye movement measures directly. More commonly, studies involving both eye tracking and administration of more traditional reading measures (e.g., standardized tests of reading skills) have hinted at the utility of eye movement data; results indicating predictive relationships between test performance and eye movement patterns (e.g., Huestegge, Radach, Corbic, & Huestegge, 2009; Kuperman & Van Dyke, 2011a) provide indirect support for the validity of eye movement measures. The following section will review the extant literature —including both direct and indirect investigations of reliability and validity—and will outline limitations of past research, which the current study seeks to overcome.

Direct examinations of reliability and validity. A review of the eye movement literature reveals that, although researchers have investigated the reliability and validity of measures like fixation duration and frequencies of fixations and regressions, even remotely up-to-date research examining the technical adequacy of eye movement measures is lacking. A thorough review of the eye tracking literature yielded only seven publications directly evaluating the utility of eye movements as measures of reading, with all seven publications (Table 2.1) being in the 1930s (Eurich, 1933a, 1933b; Futch, 1934; Litterer, 1932; Tinker, 1933, 1936; Tinker & Frandsen, 1934). In general, these studies investigated the consistency of eye movements during short samples of reading and the validity of eye movement records using participants' reading speed and comprehension performance as criteria. Whereas some of the researchers correlated eye movement measures with performance on standardized reading assessments (e.g., Chapman-Cook Speed of Reading Test, Minnesota Speed of Reading Test, Iowa Silent Reading Test, Minnesota Reading Examination for College Students; Eurich, 1933a, 1933b; Litterer, 1932; Tinker, 1936), others related observed eye movements to experimental measures of comprehension (Eurich, 1933b; Futch, 1934) or other eye movement measures representing reading speed (Futch, 1934; Tinker & Frandsen, 1934). Similarly, methods for calculating reliability varied, with some researchers utilizing split-half and odd-even methods for determining internal consistency (e.g., Tinker & Frandsen, 1934) and others correlating data collected across different passages (e.g., Litterer, 1932). Despite methodological differences, reliability and validity evidence was comparable across studies.

Reported reliability coefficients (corrected for the full length of reading material, when applicable) were generally high, ranging from .78 to .92 (Litterer, 1932), .73 to .91 (Eurich, 1933b), .86 to .91 (Eurich, 1933a), .55 to .93 (Tinker & Frandsen, 1934), and .80 to .94 (Futch, 1934). Taken together, these results suggested that eye movement measures captured reading behavior in a consistent manner throughout and across different types of reading material. Thus, researchers deemed eye movement measures adequately reliable for making relative comparisons based on collected data (Tinker, 1936) and at least as reliable as well-regarded standardized tests (Eurich, 1933a).

In contrast, studies revealed limited criterion-related validity, as reflected by correlations between eye movement data and participants' performance on assessments of reading speed, vocabulary, paragraph reading, comprehension, and broad reading achievement. Reported validity coefficients indicated only moderate to fair relationships between eye movements and tested reading skills, ranging from −.316 to −.617 (Litterer, 1932), −.02 to −.25 (Eurich, 1933b), −.18 to −.55 (Eurich, 1933a), .01 to −.34 (Futch, 1934), and .01 to −.66 (Tinker, 1936). Results suggested that eye movement measures lacked sufficient validity to function satisfactorily as diagnostic measures of reading rate or comprehension (Eurich, 1933a). That is, eye movement measures appeared to capture some aspect of reading performance distinct from rate or comprehension as measured by standardized tests (Eurich, 1933b). As a result, researchers posited that students earning similar criterion scores might exhibit markedly different eye movement patterns during reading (Futch, 1934).

In several of the studies cited above, researchers also analyzed the intercorrelations between eye movement measures (i.e., perception time, fixation frequency, regression frequency, and pause/fixation duration) to investigate their relative utility and the relationships between them (Eurich, 1933a, 1933b; Futch, 1934; Tinker & Frandsen, 1934). Across studies, results revealed that different eye movement measures did not appear to measure the same reading functions (Tinker & Frandsen, 1934). Whereas high intercorrelations between fixation frequency and regression frequency (e.g., .81 to .94) suggested that they captured comparable reading skills (Eurich, 1933a), lower intercorrelations between both measures of frequency and fixation duration (e.g., .19 to .38 for fixation frequency, and .10 to .41 for regression frequency) indicated that fixation duration measured a distinct aspect of reading (Eurich, 1933a). Employing perception time—a proxy for reading speed, calculated by summing fixation time—as a criterion against which to judge other eye movement measures, researchers labeled fixation frequency a “satisfactory” measure of reading, regression frequency a “fair” measure of reading, and fixation duration a “poor” measure of reading (Futch, 1934; Tinker & Frandsen, 1934). Of note, these conclusions were consistent with earlier results suggesting greater criterion-related validity for fixation frequency than for fixation duration (Eurich, 1933b). Thus, although aforementioned validity coefficients indicated non-optimal criterion-related validity of eye movement measures in general, researchers noted differences in the technical adequacy and utility associated with each measure (Tinker & Frandsen, 1934). Furthermore, Futch (1934) noted that, despite the lack of strong criterion-related validity evidence for eye movement measures, eye tracking could provide researchers with distinct information about reading (e.g., ongoing processes, as opposed to static performance).

In addition to examining the consistency and criterion-related validity of eye movements as indicators of reading, Eurich (1933a, 1933b) and Futch (1934) investigated relationships between observed eye movements and performance on other criterion assessments. Specifically, they scrutinized correlations between eye movement patterns and scores representing IQ, general academic achievement, and comprehension of other languages. Detailed findings are not presented here; however, consistent with aforementioned conclusions specific to reading, results indicated limited criterion-related validity.

In 1936, Tinker broadened the scope of reliability and validity studies by evaluating the test-retest reliability of eye movement measures and increasing the similarity between eye-tracking (predictor) tasks and criterion measures. Specifically, Tinker examined correlations between participants' eye movements during two separate sittings and compared performance on a reading speed test administered both traditionally (i.e., with a paper version at a table) and in front of an eye-monitoring camera. Results indicated that the test-retest reliability of eye movement measures was lower than the consistency of measures within a single sitting. Reliability coefficients calculated for perception time on paragraphs read during the same day ranged from .74 to .76, whereas corresponding coefficients for paragraphs read on different days ranged from .60 to .72. Corrected correlations between recorded eye movements and earned test scores revealed high criterion-related validity for measures of fixation frequency (−.80 to −.99) and perception time (−.87 to −.90), but indicated inadequate validity for fixation duration (−.10 to −.31). Results also suggested that criterion-related validity of eye movement measures increased with greater similarity between experimental and criterion tasks. Specifically, validity coefficients relating eye movement measures to reading achievement scores ranged from −.41 to −.69 for eye movements during test completion, but from .01 to −.66 for eye movements during paragraph reading.

Summary. Direct evaluations of eye movement measures from the 1930s suggested that such measures were adequately reliable for making relative comparisons based on collected data (Tinker, 1936) and at least as reliable as well-regarded standardized tests (Eurich, 1933a). According to one study (Tinker, 1936), the test-retest reliability of eye movement measures was lower than the consistency of the same measures during a single sitting. Validity coefficients indicated that eye movement measures lacked sufficient criterion-related validity to function as diagnostic measures of reading rate or comprehension (Eurich, 1933a). However, all measures were not of equal quality; fixation frequency and regression frequency appeared to possess greater validity than fixation duration (Eurich, 1933b; Futch, 1934; Tinker & Frandsen, 1934). Furthermore, criterion-related validity of eye movement measures was shown to increase with greater similarity between experimental and criterion tasks (Tinker, 1936). Despite reports of limited criterion-related validity, studies suggested that eye movement measures had the potential to reflect distinct aspects of reading processes that were not captured by standardized reading assessments (Eurich, 1933b; Futch, 1934).

Indirect evidence of reliability and validity. Although researchers have not explicitly investigated the technical adequacy of eye movement measures since the 1930s, more recent studies involving eye tracking and other assessments of reading-related skills provide indirect support for the reliability and validity of eye movements as measures of reading (e.g, Ashby et al., 2013; Kuperman & Van Dyke, 2011a, 2011b). Such studies fail to provide definitive evidence of technical adequacy but, importantly, build upon previously described research by suggesting relationships between observed eye movement patterns and test performance reflecting *specific* reading skills. Whereas the aforementioned studies from the 1930s (e.g., Eurich, 1933a, 1933b; Futch, 1934) primarily examined associations between eye movements and reading speed, comprehension, and broad reading abilities, more recent research has involved assessment of skills like phonological processing and phonemic awareness, word identification, silent and oral reading fluency, spelling, phonological memory, rapid naming, and working memory (Table 2.2).

Recent eye movement studies involving correlation and regression analyses (Ashby et al., 2013; Kuperman & Van Dyke, 2011a, 2011b) suggest that eye movement measures capture and reflect a wide spectrum of verbal and reading-related skills rather than a single general ability (Kuperman & Van Dyke, 2011a). In two studies conducted with 71 non-college-bound young adults (aged 16 to 24 years), Kuperman and Van Dyke (2011a, 2011b) related performance on 17 or 18 measures of sub-lexical, word-level, and sentence-level reading skills (i.e., tests of phonological awareness, phonological memory, rapid naming, word and non-word reading, working memory, and reading and listening comprehension) to eye movement characteristics during sentence reading (e.g., first fixation duration, single fixation duration, gaze duration, total fixation time). In the first study, performance on rapid naming and word identification tests significantly predicted behavior across all eye movement measures, such that higher test performance was associated with shorter fixation durations, lower cumulative fixation time, and fewer regressions/refixations. Furthermore, regression coefficients revealed that the predictive effects of rapid naming and word identification on eye movements were comparable in strength to well-established effects of word length and frequency. Results also indicated that performance on measures assessing working memory, phoneme reversal, and reading and listening comprehension significantly predicted observed eye movement patterns; however, only rapid naming and word identification performance consistently explained unique variance in eye movements (Kuperman & Van Dyke, 2011a). In the second study, which investigated readers' use of morphological information, Kuperman and Van Dyke (2011b) noted that performance on tests assessing word and non-word segmentation and reading comprehension skills significantly predicted fixation durations indicating differences in morphological processing.

More recently, Ashby et al. (2013) noted multiple associations between elementary students' eye movements and other indicators of reading performance. Specifically, they related eye movement behavior during assessments of phonemic awareness and receptive spelling to characteristics of silent and oral reading during the same year and 1 year later (i.e., reading rate/fluency, fixation count). They also found that oral reading rate accounted for a significant proportion of variance in total fixation time and fixation count during silent reading 1 year later. Taken together, this study and those by Kuperman and Van Dyke (2011a, 2011b) suggest that eye movement measures and more traditional assessments of specific reading skills capture shared aspects of reading behavior.

Research documenting developmental differences in eye movement patterns also suggests that eye movement parameters adequately reflect changes in reading skills as measured by tests of broad reading, word identification, reading rate/fluency, and comprehension. Eye movement studies conducted with children at different ages and grade levels (Buswell, 1922; McConkie et al., 1991; Valle et al., 2013) indicate that along with improved performance on standardized reading assessments (e.g., subtests of the Iowa Tests of Basic Skills and Woodcock-Johnson III Tests of Academic Achievement), readers exhibit wider recognition spans, longer saccades, fewer and shorter fixations per word (on average), and fewer regressions and refixations. Furthermore, studies indicate that developmental differences in children's eye movements reflect changes in linguistic skills (i.e., those assessed by traditional reading measures) rather than changes in oculomotor control (e.g., Huestegge et al., 2009; Hutzler & Wimmer, 2004). For example, Huestegge et al. (2009) found that, whereas eye movements during non-reading tasks did not significantly relate to eye movements during oral sentence reading, measured reading skills (assessed by tests of word/sentence comprehension, naming speed, speeded word/sentence reading, and pseudoword reading) significantly predicted total reading time on target words.

Similar to Huestegge et al. (2009), researchers studying the eye movements of dyslexic and average readers have suggested that eye movement measures of reading reflect differences in skill level, not eye movement control. Although some research indicates similar eye movement patterns across groups (e.g., Hyönä & Olson, 1995), most studies reveal differences in the reading behaviors of children with dyslexia—as determined by performance on standardized achievement tests measuring component reading skills such as phonemic awareness. Studies indicate that, compared with average readers, readers with dyslexia make more and significantly longer fixations (Bayram, Camnalbur, & Esgin, 2012; De Luca, Di Pace, Judica, Spinelli, & Zoccolotti, 1999; De Luca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002; Hutzler & Wimmer, 2004), skip fewer words (Hutzler & Wimmer, 2004), are more sensitive to word length effects (Bayram et al., 2012; De Luca et al., 1999), and exhibit different sizes and frequencies of saccades and regressions (De Luca et al., 1999; De Luca et al., 2002; Hutzler & Wimmer, 2004). However, both groups demonstrate similar eye movement control during non-reading tasks (e.g., experimental tasks requiring them to make saccades toward and away from visual targets), suggesting that aforementioned differences reflect dyslexic readers' underdeveloped reading skills rather than impaired oculomotor functioning (e.g., erratic eye movements). Together with evidence indicating high correlations between silent reading rate during eye monitoring and traditionally assessed oral reading rate (.86 to .87), these findings suggest that eye movements during reading are valid indicators of reading skills and difficulties (Hutzler & Wimmer, 2004).

Summary. Recent studies involving eye tracking and other assessments of reading-related skills provide indirect support for the reliability and validity of eye movements as measures of reading. In particular, three lines of research suggest that eye movement measures consistently reflect reading skills captured by more traditional assessments of phonological processing, word identification, fluency, comprehension, and broad reading achievement. These include: (a) studies examining relationships between eye movement measures and other assessments of reading-related skills, (b) studies investigating developmental differences in eye movement patterns, and (c) studies documenting differences in the reading behaviors of children with dyslexia and average readers (e.g., Ashby et al., 2013; Bayram et al., 2012; Valle et al., 2013).

Limitations of extant research. Although past studies have yielded direct and indirect evidence concerning the technical adequacy of eye movement measures of reading, recent research explicitly investigating their reliability and validity is lacking. Eye-tracking technologies have changed significantly since the reliability and validity studies conducted in the 1930s (i.e., Eurich, 1933a, 1933b; Futch, 1934; Litterer, 1932; Tinker, 1933, 1936; Tinker and Frandsen, 1934). At the time of these early investigations, eye-tracking methods differed significantly from modern video-based systems and required attachment of recording apparatuses to participants' eyes, experimenter observation of eye movements, or photography in unnatural lab settings involving bright lights and absolute head stabilization (Tinker, 1933, 1936). Due to technological advances (Miller & O'Donnell, 2013) and modern-day children's increased experience with reading from a display as opposed to paper material, contemporary eye-tracking methods are much more naturalistic than those utilized in past research. In addition, reading assessments employed as criteria in past studies (e.g., standardized tests of reading achievement) have undergone significant changes and updates since the 1930s. Although recent studies have related observed eye movements to performance on current reading measures (e.g., Kuperman & Van Dyke, 2011a; Valle et al., 2013), researchers have not explicitly examined the criterion-related validity of eye movement measures using current reading achievement tests. Given that approximately 80 years have passed since researchers evaluated the technical adequacy of eye movement measures during reading, past researchers' statements expressing the need for reliability and validity evidence in light of the growing use of eye tracking (e.g., an increase "during the last 15 years"; Tinker, 1933, p. 381) ring even more true in the present day.

Extant studies evaluating the quality of eye movement measures of reading are also limited with regard to the eye movement parameters they investigate. Research from the 1930s examined fixation duration in general, as opposed to contemporary measures that summarize fixation duration in multiple ways (e.g., first fixation duration, single fixation duration, gaze duration, total fixation time). Furthermore, eye movement records in these early studies summarized reading across entire passages rather than indicating behaviors specific to particular target words (e.g., high-frequency versus low-frequency words).

Finally, given that only a small number of studies have directly examined the technical adequacy of eye movement measures of reading, definitive replicated evidence regarding the reliability and validity of such measures is lacking. To date, only one researcher (Eurich, 1933a, 1933b) has evaluated eye movement measures of reading among children and unskilled adult readers, as opposed to skilled adult readers. Similarly, only one study (Tinker, 1936) has addressed the test-retest reliability of eye movement measures of reading.

Purpose and Hypotheses

The current study sought to investigate the alternate-form reliability, test-retest reliability, and criterion-related validity of commonly utilized eye movement measures of reading (i.e., first fixation duration, gaze duration, total fixation time, number of regressions, proportion of words initially skipped, and fixation count). By examining multiple indicators of reliability and validity, evaluating a wider variety of eye movement parameters, and employing several current assessments of reading-related skills (i.e., morphological processing, reading rate/fluency, decoding/word identification, comprehension, and broad reading) as criteria against which to judge eye movement measures, the current study overcomes aforementioned limitations and improves upon past research. By utilizing a large participant sample of second-grade students, it is designed to extend upon findings from past research conducted with older (Eurich, 1933a) and fewer (Ashby et al., 2013) children. Furthermore, through analysis of eye movements specific to high- and low-frequency target words, this study evaluates the consistency of eye movement measures of reading in a novel manner.

It was hypothesized that, consistent with findings in past research, eye movement measures would demonstrate adequate alternate-form and test-retest reliability, with coefficients of equivalence being greater than coefficients of stability. In light of advancements in eye-tracking technology, it was hypothesized that current results would indicate greater reliability and criterion-related validity compared to findings from past research.

Method

Participants and Setting

Participants were 175 second-grade students with a mean age of 7 years, 8 months (range = 6 years, 11 months to 8 years, 8 months) at the initial point of data collection. This sample was comprised of 82 boys and 93 girls and included children of White (87%), Black or African American (3%), Hispanic or Latino (3%), Asian (2%), and multiracial (5%) ethnicities. Students were drawn from two elementary schools serving K–5 students and one primary school serving K–2 students located in the Southeastern United States. Across the three schools, the percentage of students eligible to receive free or reduced-price meals ranged from 23% to 32%. Students with a first language other than English and students receiving reading instruction outside of the general education classroom (e.g., special education or gifted education) were excluded from the current study due to reasons associated with the larger study for which these data were collected (i.e., a randomized controlled study involving 9 to 10 weeks of one-to-one reading intervention and pre- and posttest eye movement data collection). At the initial time of data collection, participants demonstrated Low Average to Superior broad reading skills (range = 85 to 130; M = 105.22), as indicated by scores summarizing their performance on three reading subtests from the Woodcock-Johnson Tests of Academic Achievement – Third Edition, Form A (WJ-III ACH; Woodcock, McGrew, & Mather, 2001). Of note, as part of the larger study for which these data were collected, 112 participants received 9 to 10 weeks of one-to-one reading intervention (Repeated Reading n = 56; Wide Reading n = 56; Ardoin, Binder, Foster, & Zawoyski, in press), and 63 participants received only typical classroom instruction between points of data collection. However, one-way analyses of variance revealed no significant group differences in achievement or eye movement data as a function of intervention condition; thus, it was expected that intervention procedures did not yield any effects foreseeably impacting correlation results.

Apparatus

Eye movements were measured using an SR Research EyeLink 1000 system. By default, eye movements were recorded from the right eye, but tracking issues occasionally necessitated recording from the left eye. Although eye movements were recorded from one eye, viewing was binocular. Reading text was displayed on either a 19-inch or 22-inch LCD monitor, which was adjusted to a comfortable level of brightness that remained constant throughout testing. Eye movement monitoring was conducted in a dimly illuminated room in each participant's school. The brightness of these classrooms was adjusted occasionally to minimize track losses. Participants were provided with a Microsoft Sidewinder Plug and Play game pad, which allowed them to answer comprehension questions and to indicate when they were finished reading displayed text without head movement.

Materials

Silent reading passages. During eye tracking sessions, participants read two experimenter-developed narrative passages. One text was adapted from *Anansi and the Talking Melon* by Eric A. Kimmel, and the other passage was an original story concerning a colorful dragon and her best friend. The experimenter-created passage was developed using examples from second- and third-grade-level reading textbooks and was designed to include a distinct beginning, middle, and end and a conflict or challenge. The first passage ("Sammy") consisted of 157 words in 12 sentences, and the second ("Emma") contained 162 words in 16 sentences. The reading level of both stories fell at Grade 3 (range = 3.18 to 3.53), according to the Spache (1953) readability formula. Embedded within both texts were five to six low-frequency target words and five to six high-frequency target words. Low-frequency words had a frequency of U = 10 or less, and high-frequency words had a frequency of U = 40 or above, with U indicating the number of instances of that word per million running words, according to *The American Heritage Word Frequency Book* (Carroll, Davies, & Richman, 1971). Both passages were displayed as black text against a white background and were formatted in standard upper- and lowercase letters and 20-point Times New Roman font. Passages were presented individually as one page of 1.5-spaced text occupying 12 to 13 lines. The maximum line width of each passage was 84 and 87 characters, respectively, with 3.7 characters equaling 1 degree of visual angle.

Criterion measures of reading skills. Assessments measuring participants' reading achievement, oral reading fluency, and morphological awareness were administered individually.

Woodcock-Johnson Tests of Academic Achievement – Third Edition, Form A (WJ-III ACH). Participants were administered four reading-related subtests from the WJ-III ACH (Woodcock et al., 2001), a comprehensive norm-referenced test of academic achievement in the broad areas of reading, mathematics, and writing. The first subtest, *Letter-Word Identification*, required participants to read lists comprised of individual letters and words aloud. During the second subtest, *Reading Fluency*, participants were given 3 min to silently read and indicate the veracity of as many printed sentences as possible. The third task, *Passage Comprehension*, required participants to read sentences/short texts each containing a missing word and to provide the word orally (e.g., "The ___ barked, frightening the children.") During the last subtest, *Word Attack*, participants were asked to read lists of nonsense words (e.g., glerz) aloud. Raw scores earned for each subtest and a Broad Reading composite score (an age-based standard score summarizing performance on the first three subtests) served as data for the current analyses.

Oral reading fluency. Participants' oral reading fluency was assessed using three curriculum-based measurement probes drawn from Formative Assessment Instrumentation and Procedures for Reading (FAIP-R) materials. Past research has demonstrated that the technical adequacy of FAIP-R passages is comparable to that of other curriculum-based measures (Christ, Ardoin, & Eckert, 2010; Ardoin, Eckert, et al., 2013). Following universal screening procedures for curriculum-based measurement in reading (CBM-R; Shinn, 1998), participants individually read FAIP-R probes aloud to an examiner who monitored performance, marked errors, and supplied words after hesitations of 3 s. Performance across all three probes was used to calculate median scores in words read correctly in a minute (WRCM) for each participant.

Morphological awareness. The *Test of Morphological Structure* (Carlisle, 2000) was administered to assess participants' ability to use morphemic knowledge (i.e., knowledge of base words and suffixes) to decompose and produce derived forms of words. During this task, an examiner orally read a target word followed by a sentence missing its final word, then prompted the participant to provide the correct form of the target word to complete the sentence (e.g., "Driver. Children are too young to ___"; "Farm. My uncle is a ___"). Correct responses (e.g., "drive," "farmer") earned 1 point each, and the test was discontinued after six cumulative incorrect responses. Raw scores based on 28 test items were used in the current analyses.

Procedure

Eye movement recording was conducted individually, with each participant sitting approximately 50 to 55 cm from the display monitor and placing his/her chin on a chin rest used to minimize head movement during eye tracking. While experimenters made slight adjustments to the chin rest and camera setup (e.g., changing the height of the chin rest, tilting the camera to better capture the eye, adjusting image thresholds), participants were informed that they would silently read passages from the display monitor while a camera recorded their eye movements. They were also instructed on how to use the game pad. Before practice and reading trials, the eye-tracking system was calibrated using a nine-point grid extending across the display screen; experimenters explained the calibration process as a "game" requiring participants to follow a displayed dot with their eyes. Upon successful calibration, another nine-point grid was used to validate the accuracy of tracking. Following a practice trial, which acquainted participants with silently reading information from the monitor and using the game pad appropriately, experimenters reminded participants that they would be reading multiple passages silently and answering a comprehension question after each passage. After delivering these instructions, experimenters recalibrated the eye-tracking system and presented each reading passage once participants' eyes were positioned appropriately (i.e., fixating on a target in the upper left corner of the display). Upon reading each passage, participants used the game pad to clear the text and to answer the resulting comprehension question.

In addition to the procedures outlined above, head movement and other tracking issues occasionally necessitated repetition of the calibration process. Across reading trials, participants were instructed to do their "best reading" and to try to read each word without assistance. They were also instructed to read the entire passage each time. Participants were not informed whether their answers to comprehension questions were correct, and their responses were not scored or analyzed; questions were provided primarily to ensure that participants were reading for comprehension rather than solely focusing on reading rate.

The current study involved analysis of eye movement data collected during pre- and post-intervention assessment for a larger project involving 9 to 10 weeks of reading intervention. Thus, eye-tracking sessions yielded more data than are reported here. Each pretest session (lasting approximately 5 to 10 min) included a practice trial and one reading trial, and each posttest session (approximately 15 to 20 min in length) included a practice trial, two reading trials, and four rereading trials. Current analyses involved eye movement data recorded during participants' pretest sessions and only the first reading and rereading trials of their posttest sessions.

Criterion measures of reading skills were administered on a separate day (i.e., not on the same day that eye tracking was conducted) during both pre- and post-intervention assessment periods. However, current analyses involved only pretest assessment data. All measures were administered individually, with each participant working with a trained examiner (i.e., graduate student trained in the administration of the WJ-III ACH, CBM-R, and morphological measures) for approximately 45 min in an unoccupied classroom.

Results

Eye movement data for multiple measures of fixation duration (i.e., first fixation duration, gaze duration, and total fixation time), the proportion of words initially skipped, number of regressions per word, and average number of fixations per word were gathered from pre- and posttest eye movement records. Correlation coefficients and patterns of statistical significance were used to investigate the reliability and validity of eye movement measures across passage reading and for particular types of words (i.e., embedded high-frequency words and low-frequency words). For both types of analyses, Pearson correlations were calculated and evaluated using two-tailed tests to determine if associations between measures were statistically significant. For variables where skewness and/or kurtosis exceeded statistical conventions for normality (i.e., ±2; Cameron, 2004), non-parametric correlations were evaluated; specifically, Spearman's rho was selected for its robustness to non-normality.

In total, there were missing data for 28 participants; pretest WJ-III and CBM-R data were missing for one participant due to school absence, and eye movement data were missing for 10 participants at pretest and 26 participants at posttest (16 for "Sammy" and 22 for "Emma") due to school absence (n = 1), track losses (n = 9 at pretest and 19 at posttest), and/or skipping or non-reading behavior (n = 6 at posttest). Track losses occurred when participant movement or technological malfunction resulted in lost calibration/tracking. In addition, data were lost when participants completely skipped reading material (e.g., used the game pad to end a reading trial before having finished reading) either inadvertently or purposely. Non-reading behavior (reliably identified by trained coders) was characterized by shorter gaze duration and total fixation time, increased skipping of words, and a generally more erratic pattern of eye movements compared to on-task reading behavior; see Nguyen, Binder, Nemier, and Ardoin (2014) for additional information on this behavior.

Descriptive Statistics

Means and standard deviations for all pre- and posttest eye movement measures (i.e., first fixation duration, gaze duration, total fixation time, average number of regressions per word, average proportion of words initially skipped, and average number of fixations per word) and pretest criterion measures (i.e., WJ-III ACH reading subtests, Broad Reading composite scores, CBM-R median scores, and raw scores on the *Test of Morphological Structure*) are reported in Tables 2.3–2.5. There was considerable variability in target-word data, with the standard deviations of eye movement measures on high- and low-frequency words occasionally approaching or exceeding the means. In contrast, standard deviation values indicated a greater degree of consistency among participants' eye movements when averaged across all words in a passage. Consistent with past eye movement research documenting the impact of word frequency on how long readers look at individual words (Just & Carpenter, 1980; Rayner, 2009), mean gaze durations were longer on low-frequency target words (493.05 ms to 786.54 ms; Table 2.5) than on high-frequency target words (384.12 ms to 489.73 ms; Table 2.4). Finally, participants' scores on criterion measures indicated consistently age-appropriate achievement on standardized reading tasks, with mean standard scores on the WJ-III subtests/composite ranging from approximately 99 to 108 (*SD* = 7.06 to 9.00; Table 2.3).

Correlation Analyses

Intercorrelations between criterion measures are provided in Table 2.6. Consistent with past research demonstrating technical adequacy comparable to that for other curriculum-based reading measures (Ardoin, Eckert, et al., 2013; Christ, Ardoin, & Eckert, 2010), participants' median scores on FAIP-R CBM-R probes were significantly and highly correlated with their standard scores on WJ-III ACH subtests (*r* = .52 to .83; Table 2.6) and the Broad Reading composite (*r* = .86; Table 2.6). As expected, participants' individual WJ-III ACH scores were also significantly and highly intercorrelated (*r* = .51 to .92; Table 2.6). Participants' raw scores on the *Test of Morphological Structure* (Carlisle, 2000) were moderately to highly correlated at the .01 level with CBM-R and WJ-III ACH scores (*r* = .31 to .51; Table 2.6). Of note, two measures of orthographic awareness (Cassar & Treiman, 1997) were administered at the time of testing for reasons associated with the larger study for which data were collected; however, due to weak (*r* = .02 to .16) and nonsignificant associations between these measures and other criterion measures of reading achievement, these data were omitted from the current study.

Test-retest reliability. Stability of eye movement measures (akin to test-retest reliability) was assessed by examining correlation coefficients between participants' eye movement characteristics during two readings of the same passage ("Sammy") separated by an interval of 9 to 10 weeks. Of note, although some participants (*n* = 112) received 9 to 10 weeks of one-to-one reading intervention between assessment administrations whereas others received only typical classroom instruction, one-way analyses of variance revealed no significant group differences in global or target-word data as a function of intervention condition; thus, it was expected that intervention procedures did not yield any effects foreseeably impacting correlation results. Reliability coefficients are reported in Table 2.7 for eye movement across all words within the passages (global analyses) and in Table 2.8 for eye movements on embedded high- and low-frequency words.

When considering participants' global eye movements, correlation coefficients for all but one eye movement measure (i.e., average proportion of words initially skipped) were significant at the .01 level. Furthermore, reliability estimates for all three measures of fixation duration (i.e., first fixation duration, gaze duration, and total fixation time) met the cutoff criterion of .70 for modest reliability (Nunnally, 1978), and the measure of fixation count (r_s = .66; Table 2.7) approached the criterion for modest reliability. In contrast, none of the correlation coefficients for eye movements specific to target words met criterion, despite the statistical significance of nearly all test-retest correlations (five of six measures on high-frequency words; all measures on low-frequency words); in general, estimates of stability were slightly higher for high-frequency words (range = .19 to .57; Table 2.8) compared to low-frequency words (range = .18 to .40; Table 2.8).

Alternate-form reliability. Equivalence (akin to alternate-form reliability) was evaluated by examining correlation coefficients between corresponding eye movement measures on different passages read at the same sitting (i.e., readings of "Sammy" and "Emma" at posttest assessment). Reliability coefficients are reported in Table 2.7 for eye movements across all words within the passages (global analyses) and in Table 2.8 for eye movements on embedded high- and low-frequency words.

All six global correlation coefficients were significant at the .01 level. In addition, correlations obtained for all but one eye movement measure (i.e., average proportion of words initially skipped) met the cutoff criterion of .70 for modest reliability (Nunnally, 1978). In contrast, none of the correlation coefficients for eye movements specific to target words met criterion, despite the majority of them (i.e., five of six measures on high-frequency words; four of six measures on low-frequency words) being statistically significant. With the exception of that calculated for average number of regressions per word, estimates of equivalence (i.e., alternate-form reliability coefficients) were generally higher for high-frequency words (range = .09 to .55; Table 2.8) compared to low-frequency words (range = .13 to .46; Table 2.8).

Concurrent criterion-related validity. Concurrent validity estimates were obtained by correlating eye movement measures with participants' standard scores on WJ-III ACH reading subtests, composite Broad Reading standard scores, median scores across three CBM-R probes, and raw scores on the *Test of Morphological Structure* (Carlisle, 2000) at approximately the same point in time (i.e., pretest assessment). The results for validity estimates can be found in Table 2.7 for global eye movements and in Table 2.8 for eye movements on embedded high- and low-frequency words.

Significant moderate to high correlation coefficients were obtained with concurrent CBM-R ($r = -.49$ to $-.77$; Table 2.7) and WJ-III ACH measures of Letter-Word Identification ($r = -.35$ to $-.56$; Table 2.7), Reading Fluency ($r = -.43$ to $-.63$; Table 2.7), Passage Comprehension ($r = -.30$ to $-.43$; Table 2.7), and Broad Reading ($r = -.40$ to $-.63$; Table 2.7) for global measures of fixation duration (i.e., first fixation duration, gaze duration, and/or total fixation time) and average number of fixations per word. Significant correlations were also obtained with participants' scores on the WJ-III ACH Word Attack subtest as well as the morphological measure for these same eye movement parameters, though the strength of these relationships was weak (r = –.21 to –.28; Table 2.7). In contrast, none of the correlations involving global eye movement measures of skipping and regressions were significant.

Similar patterns of associations were found for participants' eye movements specific to embedded target words; in particular, significant moderate to high correlation coefficients were obtained with CBM-R and WJ-III ACH measures of Letter-Word Identification, Reading Fluency, and Broad Reading for measures of gaze duration on both high- and low-frequency words (r_s = –.43 to –.63 for high-frequency words; r_s = –.30 to –.54 for low-frequency words; Table 2.8), total fixation time on high-frequency words (r_s = –.39 to –.63; Table 2.8), and average number of fixations per word on high-frequency words (r_s = –.31 to –.54; Table 2.8). Gaze durations on high-frequency words were also significantly and moderately correlated with participants' scores on WJ-III ACH Passage Comprehension and Word Attack subtests (r = .38; Table 2.8). Additional significant moderate to high correlation coefficients (r_s = –.34 to –.53; Table 2.8) were obtained between participants' total fixation times and average number of fixations per word on embedded low-frequency words and select criterion measures (i.e., CBM-R and WJ-III Reading Fluency for both, and WJ-III ACH Broad Reading for total fixation time). However, validity estimates specific to low-frequency words were generally weaker in degree of correlation compared to those specific to high-frequency words.

Discussion

Due to technological advances and increased interest in the development of reading behavior, the use of eye-tracking technology (which was previously limited to cognitive studies with proficient adult readers) has become increasingly common in applied reading research with unskilled readers (e.g., children; Miller & O'Donnell, 2013). The heightened practical implications of such research necessitate added scrutiny and caution regarding the technical adequacy of data yielded by eye monitoring procedures. Specifically, there is a clear need for scientific evidence regarding the reliability and validity of the eye movement measures being utilized in recent educational research. Although recent studies (e.g., Ashby et al., 2013; Kuperman & Van Dyke, 2011a, 2011b) have noted associations between eye movements during reading and more established, technically sound measures of reading-related skills (e.g., phonemic awareness, spelling, rapid naming, etc.), explicit investigations of the reliability and validity of eye movement measures have not been conducted since the 1930s with either adult or child readers (e.g., Eurich, 1933a, 1933b; Futch, 1934). Furthermore, only one of those studies (Eurich, 1933a, 1933b) evaluated eye movement measures of reading among children and unskilled adult readers, as opposed to skilled adult readers. As a result, research regarding the quality of eye movement measures of reading is painfully in need of update with regard to the use of contemporary eye-tracking technology, recently employed eye movement parameters (i.e., more contemporary measures, target-word data), and the population for whom current applied reading research is applicable (i.e., unskilled readers). Thus, the purpose of the current study was to directly evaluate the test-retest reliability, alternate-form reliability, and criterion-related validity of recently utilized eye movement measures among a sample of participants comparable to those in recent educational eye movement research (i.e., elementary students).

As hypothesized and consistent with past research (Tinker, 1936), the results of this investigation indicated adequate reliability (equivalence) estimates for multiple measures of global eye movements, with fixation durations appearing more reliable than fixation count and regression count, in that order. Estimates of equivalence (alternate-form reliability coefficients) for these measures ranged from .77 to .89 (Table 2.7), representing satisfactory reliability around the same level as that reported by many standardized test publishers (Crocker & Algina, 1986). Also aligning with hypotheses based on the only previous study to address the test-retest reliability of eye movement measures of reading (Tinker, 1936), estimates of stability (i.e., test-retest reliability coefficients) were numerically lower than the consistency of measures given during the same sitting (i.e., alternate-form reliability coefficients), ranging from .56 to .82 (Table 2.7). Nonetheless, all three measures of fixation duration met cutoff criteria for adequate test-retest reliability (Nunnally, 1978). Thus, despite differences in participant sample (i.e., elementary students as opposed to college students), eye movement parameters of interest, and changes in eye-monitoring systems over the past 80 years, current findings were comparable to those in previous studies that demonstrated sufficient reliability for making relative comparisons based on collected data (Tinker, 1936).

Also as hypothesized, estimates of concurrent criterion-related validity reflected improvement in the quality of global eye movement measures compared to those obtained in past research (e.g., Tinker, 1936). For example, whereas previous researchers found "exceedingly low or negligible" relationships between eye movements and performance on criterion measures of reading (e.g., Eurich, 1933b), current data indicated significant moderate to high correlations for measures of fixation duration and count with standardized measures of reading fluency, basic reading, comprehension, and overall reading skills. Thus, although contemporary eye movement measures appear to be weakly associated with specific reading skills such as nonsense word identification (WJ-III ACH Word Attack) and morphological awareness, they do appear to be capturing a considerable degree of variability in elementary students' reading fluency and broad reading achievement.

In stark contrast with previous research (Eurich, 1933b; Futch, 1934; Tinker & Frandsen, 1934), measures of fixation duration were shown to exhibit greater criterion-related validity than the count/frequency of fixations or regressions. However, use of different criterion measures (i.e., focus on time-based measures of reading speed in the above-cited previous studies) may help explain this inconsistency; it may be that the count/frequency of eye movements is more strongly tied to absolute reading speed (without regard for accuracy), whereas measures of fixation duration (which are thought to reflect time spent encoding and processing text) may better capture accuracy-based achievement. Average proportion of words initially skipped, which had not been investigated previously, was not shown to be reliable or significantly related to any criterion measure of reading achievement. Thus, it may be that young readers' skipping behavior may reflect individual/environmental differences or some other distinct skill or aspect of behavior (e.g., level of engagement) impacting reading performance. Furthermore, the skipping measure in the current study simply indicated the proportion of words that children *initially* skipped (e.g., passed over momentarily before returning to fixate on them) rather than *completely* skipped. Even participants displaying high levels of initial skipping exhibited overall fixation behavior typical for children (e.g., multiple fixations per word, on average), suggesting that this measure may lack utility in representing children's overall reading behavior and should be interpreted with caution.

The current study also expanded upon previous research (Eurich, 1933a, 1933b; Futch, 1934; Litterer, 1932; Tinker, 1933, 1936; Tinker and Frandsen, 1934) by evaluating the technical adequacy of eye movements relative to embedded high- and low-frequency target words. None of the obtained correlation coefficients for target-word data met criterion for modest reliability. However, despite inconsistencies across passages and over time, measures of participants' fixation durations (especially gaze durations) and/or count on target words were moderately to highly correlated with scores related to basic reading, reading fluency, and broad reading achievement (for both types of target words), as well as passage comprehension and word attack (for high-frequency words only).

Limitations and Future Directions

Results from the current study should be interpreted with careful consideration of several limitations. First, current participants were drawn from a single high-performing school district and did not include students receiving reading instruction outside of the general education setting or English to Speakers of Other Languages (ESOL) instruction. As a result, generalizability of current findings to the wider school-age population may be limited by the fact that participants in the current study were not representative of students who are likely to receive targeted reading instruction. However, given that participants were young students capable of developing greater reading proficiency, current findings bear more relevance to this population than previous technical adequacy research with skilled adult readers. Although the specificity of the current participant sample raises concerns regarding possible restriction of range, the significance and strength of obtained correlation results would be expected to persist or improve with use of a less restricted sample (e.g., one including lower-performing and higher-performing students, such as those receiving special education or gifted instruction). Nonetheless, it bears noting that current results are most applicable to elementary students exhibiting typically developed reading skills and receiving general education instruction.

In addition, obtained correlation coefficients may be influenced by the specific selection and nature of assessment tasks utilized in the current study. For example, Tinker (1936) documented how the difficulty and degree of similarity between material read during eye movement recording and criterion measures impacted reliability and validity estimates, such that easier and more comparable tasks yielded higher correlations. However, given that the silent reading passages utilized during eye monitoring fell at a higher reading level than participants' grade level and were considerably dissimilar from paper-and-pencil criterion measures, it is unlikely that these factors inflated reliability and validity estimates. An additional limitation is that current "test-retest reliability" coefficients differ from traditional estimates of test-retest reliability in that assessment administrations were separated by a 9- to 10-week period during which a portion of the participant sample received one-to-one reading intervention while remaining participants received typical instruction. However, statistical analyses revealed no significant between-group differences as a function of intervention condition at either point of data collection, so participants' exposure to intervention seems unlikely to have affected current results. Thus, true test-retest reliability of eye movement measures (unaffected by intervention procedures) would be expected to be comparable or higher in significance and strength compared to current estimates.

Results specific to high- and low-frequency words should be interpreted with caution due to the selection of a small number of embedded target words, particularly in light of the high degree of variability in participants' reading behaviors on these words. To limit the degree to which frequency-related analyses are skewed by these differences (e.g., participants' skipping of target words), it likely would be beneficial for future researchers to examine the *degree* of frequency for all words in a passage rather than manipulating this variable based upon dichotomous grouping of a handful of high- and low-frequency words. Given that it is standard for textual variables to be manipulated and examined in this manner in eye movement research (i.e., through selection and classification of a small proportion of embedded target words), current results indicating low reliability for target-word data also indicate the need for caution in interpreting previous and future findings that are based on a small sample of target words.

Although the current study served its purpose in providing much-needed evidence regarding the reliability and validity of commonly utilized eye movement measures of reading, future studies are needed to expand upon these findings. Specifically, future research involving multiple and hierarchical regression would be useful to determine how much variance in criterion scores (e.g., WJ-III ACH subtest and Broad Reading scores) is predicted by eye movement measures, both individually and in combination, and to investigate whether eye movement measures add predictive value to models based solely on known predictors (e.g., oral reading fluency median scores). Finally, evaluating the technical adequacy of these measures with a larger and more diverse participant sample will be essential in determining the generalizability of current findings and how the reliability and validity of eye movement measures may vary as function of participants' achievement and skill levels.

Conclusions and Implications

Despite the aforementioned limitations, findings from the current study provide much-needed evidence indicating that recently utilized eye movement measures of reading, particularly passage-level measures of fixation duration: (a) demonstrate stability over time, (b) display consistency across passages read at the same point in time, and (c) appear to be capturing a considerable degree of variability in elementary students' levels of reading fluency and broad reading achievement. In contrast, elementary students' reading behaviors relative to specific words (especially low-frequency words) appear to vary considerably, to be susceptible to change over time, and to be more weakly associated with their normative levels of reading achievement. Thus, current findings suggest that conclusions based on very specific data (e.g., eye movements occurring on small subsets of target words) should be interpreted with caution.

In order to avoid overgeneralization of these findings, it is imperative to emphasize that the purpose of this study was *not* to evaluate eye movement measures as standalone assessment tools (e.g., diagnostic measures) but rather to determine the degree to which these fairly novel behavioral indicators reliably explain variance in traditionally measured reading skills. When considering readers' eye movements across larger sections of text (e.g., entire passages), variability on these measures appears to correspond with variability on other, more commonly utilized measures of reading (e.g., WJ-III ACH, CBM-R). This finding inspires increased confidence in recent eye movement research (e.g., Foster et al., 2013; Zawoyski, Ardoin, & Binder, 2014), suggesting that changes/facilitation seen in these studies actually are indicative of improvements in reading skill rather than representing arbitrary changes in motor eye movements. In contrast, analyses of children's eye movements relative to specific words may reveal very fine-grained changes in behavior that will not necessarily be reflected by traditional reading assessments, especially standardized measures not scaled to detect small amounts of growth. Furthermore, the current study suggests that observed eye movements on specific words are likely to be highly variable and may be more susceptible to the influence of lexical characteristics (e.g., frequency, length) and individual/environmental differences.

Given that the use of eye-tracking technology has become increasingly common in applied research with practical implications, the current study sought to investigate the relationship between elementary students' behavior during experimental eye monitoring methods (i.e., eye movement recording during silent passage reading) and their performance on trusted measures of reading achievement. Again, this purpose should be carefully distinguished from evaluation of eye tracking as a screening or diagnostic tool. In addition to inspiring a higher degree of confidence in past eye movement research, current results revealed key differences in the quality/utility of different types of eye movement data. Findings suggest that measuring students' eye movements at the passage level may offer valuable information regarding their relative/normative levels of general reading achievement, whereas monitoring these behaviors relative to specific words is likely to be more useful in revealing absolute progress/growth (e.g., fine-grained changes in fluency) and differential effects based on manipulated variables (e.g., word frequency). It is important for conductors and consumers of eye movement studies to bear in mind how specific research findings may be influenced by these different levels of analysis (e.g., how global analyses may obscure changes specific to certain types of words) and to critically consider how observed changes in eye movements may be influenced by passage- and word-level variables. Although the current study provides an important and promising update regarding the reliability and validity of eye movement measures, the nascent status of eye tracking in the field of applied reading research demands continued scrutiny and caution in interpreting such data.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Ardoin, S. P., Binder, K. S., Foster, T. E., & Zawoyski, A. M. (in press). Repeated versus wide reading: A randomized control design study examining the impact of fluency interventions on underlying reading behavior. *Journal of School Psychology*.

Ardoin, S. P., Binder, K. S., Zawoyski, A. M., Foster, T. E., & Blevins, L.A. (2013). Using eye-tracking procedures to evaluate generalization effects: Practicing target words during repeated readings within versus across texts. *School Psychology Review*, *42*(4), 477-495.

Ardoin, S. P., Eckert, T. L., Christ, T. J., White, M. J., Morena, L. S., January, S. A., & Hine, J. F. (2013). Examining variance in reading comprehension among developing readers: Words in context (Curriculum-Based Measurement in Reading) versus words out of context (word lists). *School Psychology Review*, *42*(3), 243-261.

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

Bayram, S., Camnalbur, M., & Esgin, E. (2012). Analysis of dyslexic students' reading disorder with eye movement tracking. *Cypriot Journal of Educational Sciences*, *7*(2), 129-148.

Blythe, H. I. & Joseph, H. S. S. L. (2011). Children's eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.). *The Oxford handbook of eye movements* (pp. 801-818). Oxford, Great Britain: Oxford University Press.

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, *49*, 1583-1591. doi: 10.1016/j.visres.2009.03.015

Buswell, G. T. (1922). *Fundamental reading habits: A study of their development*. Chicago: The University of Chicago.

Cameron, A. C. (2004). Kurtosis. In *The SAGE Encyclopedia of Social Science Research Methods*. (Vol. 1, p. 543). M. S. Lewis-Beck, A. E. Bryman, & T. F. Liao (Eds.). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781412950589.n467

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, *12*(3), 169-190. doi: 10.1023/A:1008131926604

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

Cassar, M. & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, *89*(4), 631-644. doi: 10.1037/0022-0663.89.4.631

Christ, T. J., Ardoin, S. P., & Eckert, T. L. (2010). *Formative Assessment Instrumentation and Procedures for Reading (FAIP-R): Oral reading passages*. University of Minnesota. Retrieved from http://www.fast.cehd.umn.edu

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart and Winston.

Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. *Brain and Language*, *80*, 617-626.

De Luca, M., Di Pace, E., Judica, A., Spinelli, D. & Zoccolotti, P. (1999). Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologica*, *37*, 1407-1420.

Eurich, A. C. (1933a). Additional data on the reliability and validity of photographic eye-movement records. *Journal of Educational Psychology*, *24*(5), 380-384. doi: 10.1037/h0070240

Eurich, A. C. (1933b). The reliability and validity of photographic eye-movement records. *Journal of Educational Psychology*, *24*(2), 118-122. doi: 10.1037/h0071126

Foster, T. E., Ardoin, S. P., & Binder, K. S. (2013). Underlying changes in repeated reading: An eye movement study. *School Psychology Review*, *42*(2), 140-156.

Futch, O. (1934). The reliability and validity of photographic eye-movement records in the reading of Latin. *Journal of Educational Psychology*, *25*(8), 620-629. doi: 10.1037/h0073657

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, *49*, 2948-2959. doi: 10.1016/j.visres.2009.09.012

Hutzler, F. & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain and Language*, *89*, 235-242.

Hyönä, J. & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1430-1440. doi: 10.1037/0278-7393.21.6.1430

Joseph, H. S. S. L. & Liversedge, S. P. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLoS ONE*, *8*(1): e54141. doi: 10.1371/journal.pone.0054141

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*(5), 708-723. doi: 10.1080/17470210701400657

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research*, *49*, 2078-2086. doi: 10.1016/j.visres.2009.05.015

Joseph, H. S. S. L., Nation, K., & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, *42*(2), 207-222.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354. doi: 10.1037/0033-295x.87.4.329

Kuperman, V. & Van Dyke, J. A. (2011a). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*, 42-73. doi: 10.1016/j.jml.2011.03.002

Kuperman, V. & Van Dyke, J. A. (2011b). Individual differences in visual comprehension of morphological complexity. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings* *of the 33rd annual meeting of the Cognitive Science Society* (pp. 1643-1648). Austin, TX: Cognitive Science Society.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293-323. doi: 10.1016/0010-0285(74)90015-2

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*(2), 202-220. doi: 10.1177/1094428105284919

Litterer, O. F. (1932). An experimental analysis of reading performance. *The Journal of Experimental Education*, *1*(1), 28-33. doi: 10.2307/20150251

McConkie, G. W., Zola, D., Grimes, J., Kerr, P. W., Bryant, R. B., & Wolff, P. M. (1991). Children's eye movements during reading. In J. F. Stein (Ed.), *Vision and visual dyslexia* (pp. 251-262). London, England: The Macmillan Press, Ltd.

Miller, B. & O'Donnell, C. (2013). Opening a window into reading development: Eye movements' role within a broader literacy research framework. *School Psychology Review*, *42*(2), 123-140.

National Institute of Child Health and Human Development (2000). Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

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

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Rayner, K. (1983). The perceptual span and eye movement control during reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 97-139). New York: Academic Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. doi: 10.1037/0033-2909.124.3.372

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

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

Rayner, K., 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. doi: 10.1207/s1532799xssr1003_3

Seftor, N., Constantine, J., Cody, S., Ponza, M., Knab, J., Deke, J., & Monahan, S. (2011). *What Works Clearinghouse: Procedures and standards handbook 2011* (NCEE 2011-XXXX). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Shinn, M. R. (Ed.). (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford Press.

Spache, G. (1953). A new readability formula for primary-grade reading materials. *Elementary School Journal*, *55*, 410-413.

Tinker, M. A. (1933). Uses and limitations of eye-movement measures of reading. *Psychological Review*, *40*(4), 381-387. doi: 10.1037/h0074534

Tinker, M. A. (1936). Reliability and validity of eye-movement measures of reading. *Journal of Experimental Psychology*, *19*(6), 732-746. doi: 10.1037/h0060561

Tinker, M. A. & Frandsen, A. (1934). Evaluation of photographic measures of reading. *Journal of Educational Psychology*, *25*(2), 96-100. doi: 10.1037/h0074523

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

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

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

Zawoyski, A. M., Ardoin, S. P., & Binder, K. S. (2014). Using eye tracking to observe differential effects of repeated readings for second-grade students as a function of achievement level. *Reading Research Quarterly*, *50*(2), 171-184. doi: 10.1002/rrq.91

Table 2.1
Direct Examinations of the Reliability and Validity of Eye Movement Measures

			Eye Movement Measures							Reliability			Validity		
Author	*N*	Age	First Fixation Duration	Gaze Duration	Total Fixation Time	Average Fixation Duration	Perception Time (fixation count × duration)	Regression Count	Fixation Count	Internal Consistency	Alternate-Form	Test-Retest	Speed	Comprehension	Other
Litterer (1932)	~ 76	adults					✓	✓	✓	✓	✓		✓	✓	
Eurich (1933b)	173	adults					✓	✓	✓		✓		✓	✓	✓
Eurich (1933a)	> 100	children (Grades 4-5)				✓		✓	✓		✓		✓	✓	✓
Tinker & Frandsen (1934)	50-216	adults				✓	✓	✓	✓	✓			✓		
Futch (1934)	27	children (Grade 9)				✓	✓	✓	✓		✓			✓	✓
Tinker (1936)	77	adults				✓	✓	✓	✓		✓	✓	✓	✓	✓

Table 2.2

Indirect Evidence of the Validity of Eye Movement Measures

			Eye Movement Measures								Criterion Measures								
Author	*N*	Age	First Fixation Duration	Gaze Duration	Total Fixation Time	Average Fixation Duration	Perception Time (fixation count × duration)	Regression Count	Fixation Count	Other	Phonological Processing	Word Reading / Accuracy	Rate / Fluency	Comprehension	Broad Reading	Spelling	Rapid Naming	Memory	Other
Buswell (1922)	186	children (all grades) and adults				✓		✓		✓		✓		✓					
McConkie et al. (1991)	> 200	children (Grades 1-5)				✓				✓		✓	✓	✓					
Hyönä & Olson (1995)	42	children (8-16)	✓	✓	✓			✓	✓			✓		✓		✓			✓
De Luca et al. (1999)	51	children (10-17)				✓		✓	✓	✓	✓				✓				✓
De Luca et al. (2002)	22	children (11-16)				✓		✓		✓					✓				✓
Hutzler & Wimmer (2004)	22	children (tested in Grades 3 and 7)	✓			✓		✓	✓				✓			✓			
Huestegge et al. (2009)	21	chidren (tested in Grades 2 and 4)	✓	✓	✓			✓		✓	✓		✓	✓			✓	✓	✓
Kuperman & Van Dyke (2011a)	71	young adults (16-24)	✓	✓	✓			✓		✓	✓	✓		✓			✓	✓	✓
Kuperman & Van Dyke (2011b)	71	young adults (16-24)	✓	✓	✓					✓	✓	✓		✓			✓	✓	
Bayram et al. (2012)	30	children (*M* age = 10)																	
Ashby et al. (2013)	10	children (*M* age = 7)			✓				✓		✓		✓			✓			
Valle et al. (2013)	90	children (6-8)	✓	✓	✓			✓	✓						✓				

Table 2.3
Descriptive Statistics for Global Eye Movement and Criterion Measures

Measure	*n*	Min	Max	*M*	*SD*
		Eye Movement Measures			
Pretest ("Sammy")					
First fixation duration (ms)	165	201.35	366.51	287.55	35.21
Gaze duration (ms)	165	222.90	761.22	438.12	101.23
Total fixation time (ms)	165	289.18	1396.57	640.66	182.91
Average number of regressions per word	165	.10	.82	.34	.13
Average proportion of words initially skipped	165	.08	1.00	.43	.21
Average number of fixations per word	165	.29	4.68	1.86	.66
Posttest ("Sammy")					
First fixation duration (ms)	159	183.70	380.28	276.81	34.50
Gaze duration (ms)	159	212.50	689.20	391.61	86.73
Total fixation time (ms)	159	233.01	1320.50	550.92	168.53
Average number of regressions per word	159	.06	.78	.31	.13
Average proportion of words initially skipped	159	.08	1.00	.42	.19
Average number of fixations per word	159	.68	4.07	1.70	.52
Posttest ("Emma")					
First fixation duration (ms)	153	186.79	356.89	276.52	34.39
Gaze duration (ms)	153	217.85	646.24	407.62	90.13
Total fixation time (ms)	153	243.68	1132.84	575.77	158.13
Average number of regressions per word	153	.03	.77	.30	.12
Average proportion of words initially skipped	153	.07	.96	.34	.19
Average number of fixations per word	153	.63	3.17	1.80	.51
		Criterion Measures			
CBM-R median (WRCM)	174	20.00	221.00	86.28	34.10
WJ-III ACH Broad Reading (SS)	174	85.00	130.00	105.22	8.22
WJ-III ACH Letter-Word Identification (SS)	174	89.00	129.00	108.15	7.06
WJ-III ACH Reading Fluency (SS)	174	76.00	135.00	103.99	8.67
WJ-III ACH Passage Comprehension (SS)	174	76.00	121.00	98.54	9.00
WJ-III ACH Word Attack (SS)	174	87.00	126.00	105.11	7.30
Morphological Structure (raw)	174	1.00	24.00	12.12	5.81

Table 2.4

Descriptive Statistics for Eye Movement Measures on High-Frequency Target Words

Measure	*n*	Min	Max	*M*	*SD*
		Eye Movement Measures			
Pretest ("Sammy")					
First fixation duration (ms)	144	159.80	562.75	301.16	68.66
Gaze duration (ms)	144	171.25	1466.00	489.73	172.75
Total fixation time (ms)	144	177.75	2158.67	725.53	360.46
Average number of regressions per word	144	.00	1.67	.33	.28
Average proportion of words initially skipped	144	.00	1.00	.33	.28
Average number of fixations per word	144	.67	7.67	2.25	1.08
Posttest ("Sammy")					
First fixation duration (ms)	154	167.50	450.75	280.18	59.93
Gaze duration (ms)	154	167.50	1940.00	446.82	188.13
Total fixation time (ms)	154	115.83	2120.33	597.81	260.27
Average number of regressions per word	154	.00	1.00	.20	.21
Average proportion of words initially skipped	154	.00	1.00	.31	.25
Average number of fixations per word	154	.50	6.17	1.99	.81
Posttest ("Emma")					
First fixation duration (ms)	146	157.75	467.40	261.78	59.68
Gaze duration (ms)	146	157.75	1027.00	384.12	161.27
Total fixation time (ms)	146	164.00	1744.00	600.09	260.42
Average number of regressions per word	146	.00	2.00	.27	.28
Average proportion of words initially skipped	146	.00	1.00	.16	.26
Average number of fixations per word	146	.16	9.00	2.17	1.05

Table 2.5
Descriptive Statistics for Eye Movement Measures on Low-Frequency Target Words

Measure	*n*	Min	Max	*M*	*SD*
		Eye Movement Measures			
Pretest ("Sammy")					
First fixation duration (ms)	143	177.67	504.75	301.01	67.44
Gaze duration (ms)	143	191.4	1583.00	579.13	244.20
Total fixation time (ms)	143	315.67	4009.17	879.85	438.02
Average number of regressions per word	143	.00	.80	.21	.20
Average proportion of words initially skipped	143	.00	1.00	.24	.25
Average number of fixations per word	143	.83	12.83	2.67	1.38
Posttest ("Sammy")					
First fixation duration (ms)	150	159.60	493.17	291.09	65.77
Gaze duration (ms)	150	159.60	1489.00	493.05	205.78
Total fixation time (ms)	150	161.40	1999.33	719.04	321.21
Average number of regressions per word	150	.00	.67	.10	.16
Average proportion of words initially skipped	150	.00	1.00	.20	.22
Average number of fixations per word	150	.00	6.00	2.25	.98
Posttest ("Emma")					
First fixation duration (ms)	146	175.80	635.80	297.49	76.09
Gaze duration (ms)	146	187.75	2298.33	786.54	435.81
Total fixation time (ms)	146	187.75	4531.20	1220.32	652.05
Average number of regressions per word	146	.00	1.00	.29	.30
Average proportion of words initially skipped	146	.00	.80	.08	.17
Average number of fixations per word	146	.20	13.20	3.85	1.88

Table 2.6
Correlation Matrix of Criterion Measures

		1	2	3	4	5	6
1	CBM-R median score						
2	WJ-III ACH Broad Reading standard score	.86**					
3	WJ-III ACH LWI standard score	.78**	.92**				
4	WJ-III ACH RF standard score	.83**	.81**	.73**			
5	WJ-III ACH PC standard score	.68**	.85**	.73**	.67**		
6	WJ-III ACH WA standard score	.52**	.63**	.70**	.51**	.54**	
7	Morphological raw score	.41**	.51**	.47**	.38**	.50**	.31**

* $p < .05$, ** $p < .01$

Table 2.7
Reliability and Validity Estimates for Global Eye Movement Measures

			Concurrent Validity						
			CBM-R	WJ-III ACH (SS)					Morphological
Eye Movement Measures	Test-Retest Reliability	Alternate-Form Reliability	(WRCM)	Broad Reading	LWI	RF	PC	WA	Structure
First fixation duration	.82 **	.86 **	–.54 **	–.47 **	–.42 **	–.43 **	–.24 **	–.14	–.24 **
Gaze duration	.82 **	.89 **	–.77 **	–.63 **	–.56 **	–.63 **	–.43 **	–.28 **	–.24 **
Total fixation time	.73 **	.87 **	–.68 **	–.56 **	–.49 **	–.57 **	–.39 **	–.25 **	–.21 **
Average proportion of words initially skipped	.15	.37 **	–.09	–.07	–.03	–.04	–.07	–.06	–.04
Average number of regressions per word	.56 **	.77 **	–.11	–.09	–.05	–.09	–.09	–.03	–.05
Average number of fixations per word	.66 **	.83 **	–.49 **	–.40 **	–.35 **	–.44 **	–.30 **	–.19 *	–.24 **

* $p < .05$, ** $p < .01$

Table 2.8

Reliability and Validity Estimates for Eye Movement Measures on Target Words

				Concurrent Validity						
					WJ-III ACH (SS)					
	Eye Movement Measures	Test-Retest Reliability	Alternate-Form Reliability	CBM-R (WRCM)	Broad Reading	LWI	RF	PC	WA	Morphological Structure
High-Frequency Words	First fixation duration	.19 *	.27 **	–.15	–.16	–.12	–.03	–.03	–.19 *	–.06
	Gaze duration	.44[a] **	.43[a] **	–.63[a] **	–.53[a] **	–.47[a] **	–.43[a] **	–.38[a] **	–.38[a] **	–.22[a] **
	Total fixation time	.57[a] **	.55[a] **	–.63[a] **	–.45[a] **	–.39[a] **	–.47[a] **	–.25[a] **	–.21[a] *	–.18[a] *
	Average proportion of words initially skipped	.24 **	.18[a] *	–.26 **	–.26 **	–.26 **	–.22 **	–.18 *	–.17 *	–.28 **
	Average number of regressions per word	.10[a]	.09[a]	–.01[a]	–.02[a]	–.02[a]	.02[a]	.04[a]	.05[a]	.05[a]
	Average number of fixations per word	.51[a] **	.43[a] **	–.54[a] **	–.37[a] **	–.31[a] **	–.42[a] **	–.23[a] **	–.14[a]	–.12[a]
Low-Frequency Words	First fixation duration	.18 *	.17	–.24 **	–.23 **	–.17 *	–.14	–.03	–.07	–.12
	Gaze duration	.33[a] **	.37[a] **	–.54[a] **	–.35[a] **	–.30[a] **	–.34[a] **	–.17[a] *	–.25[a] **	–.10[a]
	Total fixation time	.40[a] **	.46[a] **	–.53[a] **	–.34[a] **	–.29[a] **	–.45[a] **	–.21[a] *	–.21[a] *	–.13[a]
	Average proportion of words initially skipped	.21 *	.13[a]	–.23 **	–.23 **	–.29 **	–.18 *	–.17 *	–.13	–.19 *
	Average number of regressions per word	.21[a] *	.24[a] **	–.04	–.04	–.05	–.06	–.03	–.04	–.08
	Average number of fixations per word	.33[a] **	.41[a] **	–.42[a] **	–.24[a] **	–.20[a] *	–.36[a] **	–.17[a] *	–.17[a] *	–.07[a]

[a] Spearman correlation coefficients

* $p < .05$, ** $p < .01$

CHAPTER 3

EXAMINING UNSKILLED READERS' EYE MOVEMENTS DURING REPEATED READINGS: A COMPARISON OF ADULT LITERACY LEARNERS AND CHILDREN[2]

[2] Foster, T.E. and S.P. Ardoin. To be submitted to *Reading Research Quarterly*.

Abstract

Previous eye movement research has demonstrated how consecutive rereading of the same text (such as that involved in instructional procedures like Repeated Reading) facilitates both adults' and children's reading fluency. Specifically, separate studies conducted with skilled adult readers and children indicate that reading a passage multiple times facilitates textual processing and results in increased reading speed and accompanying changes in eye movements (i.e., decreases in fixation counts, fixation durations, and number of regressions). These studies also shed light on how textual characteristics (e.g., word frequency) can influence improvements related to RR. Although past research clearly indicates differences in rereading effects among adults versus children, it is unclear how observed differences between participant samples relate to confounded variables of age, reading skill level, and/or additional characteristics such as linguistic processing abilities and how RR might differentially impact the reading behavior of different aged and/or differently skilled groups. Thus, the current study was designed to investigate and compare rereading behavior among two distinct groups of similarly skilled but different aged developing readers (i.e., 36 elementary students and 36 adult literacy learners). Results of the current study suggest that RR improves the efficiency of word- and passage-level reading behavior for unskilled, elementary-level readers at a variety of ages. Findings support the idea that eye movements during reading—and the manner in which these change in response to instructional procedures—ultimately reflect and are predominantly influenced by individuals' levels of reading skill and achievement. Implications for reading research and implementation of RR are discussed.

Introduction

Reading fluency, or the ability to read with accuracy, automaticity, and proper expression, was recognized as an essential instructional target over a decade ago (National Institute of Child Health and Human Development [NICHD], 2000). However, educational statistics since then indicate that many students continue to lack fluency and fail to attain basic-level reading skills. In 2002, the U.S. Department of Education reported that approximately 40% of fourth-grade students were not fluent readers (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). Similarly, scores on the National Assessment of Educational Progress (NAEP) administered between 2003 and 2013 revealed that 32% to 37% of fourth graders and 22% to 27% of eighth graders failed to demonstrate fundamental reading skills (U.S. Department of Education, 2014). Despite a strong emphasis on reading in our nation's schools and curricula, it is clear that many students continue to lack crucial skills needed to function and participate in a literate society. Thus, current instructional strategies targeting reading fluency warrant further scrutiny and evaluation.

Not only is reading fluency an important component of reading in its own right, but it also is an essential skill due to its strong connection with comprehension—the ultimate goal of reading instruction (Kuhn, 2004; NICHD, 2000). Fluency enables automatic word recognition, allowing readers to devote attention and cognitive resources to the meaning of text rather than the decoding of individual words (Bos, 1982; LaBerge & Samuels, 1974). Given that fluency accounts for a significant portion of comprehension performance (Rasinski et al., 2005), there is a need for continued research evaluating intervention strategies designed specifically to increase students' reading fluency.

Repeated reading (RR), a method involving the rereading of short passages of text (LaBerge & Samuels, 1974), is the oldest and most widely known approach for improving students' reading fluency (Kuhn, 2004; Meyer & Felton, 1999). Meta-analyses of RR-based interventions have linked RR with improved reading speed, accuracy, and comprehension for a broad range of students, including students with and without disabilities at different ages and skill levels (Meyer & Felton, 1999; NICHD, 2000; Therrien, 2004). In the studies reviewed, effect sizes varied across component skills, with gains highest for accuracy, smaller for fluency, and lowest (but still significant) for comprehension. However, results indicated that RR significantly improved the reading achievement of nondisabled students through Grade 4 and students with reading problems through high school (overall weighted effect size average = 0.41; NICHD, 2000). Furthermore, despite the time-limited nature of RR procedures (e.g., 45 or fewer 15-min sessions in most reviewed studies), results revealed possible improvements in students' reading and comprehension of untrained passages, suggesting potential for generalization effects (Therrien, 2004). Based on observed instructional effects, Therrien (2004) identified three essential components of effective RR-based intervention: (a) providing students with cues regarding their purpose in rereading, (b) having students read aloud to adults who can provide them with error correction and feedback, and (c) requiring students to reread passages three or four times.

Although extensive research supports the use of RR, the majority of empirical studies on RR (e.g., those reviewed in the three meta-analyses mentioned above) demonstrate that it is effective without explaining why or how it yields improved reading skills. Researchers have repeatedly shown that RR allows students to read passages more quickly, assessing fluency using outcome measures like reading rate in words read correctly in a min (WRCM; e.g., Ardoin, McCall, & Klubnik, 2007; Martens et al., 2007). However, such indicators are inexact; they only indicate absolute gains and fail to reflect how a given change in reading speed (as indicated by increased WRCM) might be the result of different effects on underlying reading behaviors. With repeated practice, some students may feel less of a need to go back and reread certain words or sections, whereas other students may simply read particular types of words (e.g., high-frequency or low-frequency words) faster. Thus, different changes in reading behavior could result in the same observed improvement in reading fluency (i.e., the same quantitative increase in rate; Hyönä & Niemi, 1990; Raney & Rayner, 1995). Given that outcome measures do not fully encapsulate the nature of reading improvements, reading researchers are increasingly relying on more precise and dynamic measures of reading behavior to identify instructional effects (e.g., Foster, Ardoin, & Binder, 2013).

Eye Movement Research on RR

Eye tracking is a unique technology that allows for thorough examination of intervention effects (e.g., effects of RR). In contrast with global measures of reading fluency, eye movement measures capture multiple behaviors during the same episode of reading (Rayner, 1998). For example, a single eye movement record indicates the amount of time a reader spent on particular words, the number of times the reader returned to previously read content, and how many times the reader paused to extract information from a given word or section of text. Eye tracking also allows researchers to pinpoint when and where within a text instructional effects occur and to discern how manipulated factors (e.g., word length or frequency) impact behavior within the context of natural silent reading (Just & Carpenter, 1980; Rayner, 1998, 2009; Rayner, Chace, Slattery, & Ashby, 2006). Furthermore, different eye movement measures are thought to reflect distinct stages of lexical processing (e.g., surface-level word recognition, higher-level semantic processing), potentially providing researchers with a means for assessing multiple component skills of reading within a single sample of reading behavior (Just & Carpenter, 1980; Rayner, 1998).

Extant studies detailing behavioral changes across multiple readings of the same text have provided insight regarding the mechanisms underlying RR, an intervention fundamentally based on rereading (e.g., Hyönä & Niemi, 1990; Raney & Rayner, 1995). Research with adults generally suggests that reading a passage multiple times increases readers' familiarity with that passage, facilitates textual processing, and results in increased reading speed and accompanying changes in eye movements. Specifically, rereading typically yields a decreased number of regressions and decreases in the number and average duration of fixations (Hyönä & Niemi, 1990; Kaakinen & Hyönä, 2007; Raney & Rayner, 1995; Rayner, Raney, & Pollatsek, 1995; Schnitzer & Kowler, 2006). Based on these observed changes, researchers have posited that rereading decreases readers' need for fixation time, increases the amount of textual information they can process during each fixation, and reduces their need to revisit previously fixated text (Hyönä & Niemi, 1990; Kaakinen & Hyönä, 2007; Raney & Rayner, 1995; Schnitzer & Kowler, 2006). Rereading appears to improve higher-level textual processing as well as surface-level reading, as suggested by research indicating that rereading facilitates integrative "wrap-up" processing at the ends of sentences and increases readers' familiarity with passage-level topical structures (Hyöna, 1995; Kaakinen & Hyöna, 2007).

Rereading research with adults also sheds light on how textual characteristics (e.g., word frequency) might influence the effects of RR. One study examining the simultaneous effects of rereading and word frequency (Raney & Rayner, 1995) indicated that, although rereading facilitated textual processing, its effects were not sufficient to overcome processing difficulties imposed by word frequency. Rather, rereading effects were consistent across high- and low-frequency words, so frequency effects remained intact even after rereading (as evidenced by persisting differences in eye movements on high- versus low-frequency words). In contrast, Rayner et al. (1995) suggested that rereading resulted in the eventual "disappearance" of word frequency effects. Specifically, rereading continued to facilitate reading of low-frequency words after no longer exerting an effect on the reading of high-frequency words.

Although eye movement research with adults has yielded insight regarding the effects of rereading (i.e., the foundation of RR), more recent studies conducted with children have provided greater clarification concerning how RR impacts reading behavior. In particular, two recent publications detailed changes in children's eye movements across four consecutive readings of the same passage, utilizing methods analogous to RR procedures (Ardoin, Binder, Zawoyski, Foster, & Blevins, 2013; Foster, Ardoin, & Binder, 2013). These studies transcended rereading research with proficient adult readers by more closely approximating real-life implementation of RR. First, given that fluency instruction (including RR) is most likely to benefit students reading at a first- to third-grade level (Therrien, Gormley, & Kubina, 2006), children and unskilled readers represent the population for which RR is intended. Second, the rereading procedures utilized by Ardoin, Binder, et al. (2013) and Foster et al. (2013) were better aligned with typical RR procedures, as compared to those utilized in adult studies (i.e., two readings, a delayed third reading, or four nonconsecutive readings; Hyönä & Niemi, 1990; Kaakinen & Hyönä, 2007; Raney & Rayner, 1995; Schnitzer & Kowler, 2006; Shebilske & Fisher, 1980).

Results of Foster et al. (2013) and Ardoin, Binder, et al. (2013) revealed both similarities and differences in the effects of rereading observed among children versus adults. Consistent with adult participants in aforementioned rereading studies (e.g., Hyönä & Niemi, 1990; Kaakinen & Hyönä, 2007; Raney & Rayner, 1995; Schnitzer & Kowler, 2006), second-grade students exhibited significantly fewer fixations and regressions and spent significantly less time fixating on words (on average) after four consecutive readings of the same text (Foster et al., 2013). Furthermore, like the adult participants in Raney and Rayner's (1995) study, children demonstrated a persistent sensitivity to word frequency even after rereading. They continued to make more fixations and to fixate longer on low-frequency words as compared to high-frequency words (Ardoin, Binder, et al., 2013; Foster et al., 2013). Despite these similarities, both studies with children revealed significant interactions between rereading and word frequency (Ardoin, Binder, et al., 2013; Foster et al., 2013). In addition, child participants in Foster et al.'s (2013) study continued to devote roughly the same amount of time to initial fixations on low-frequency words even after rereading. Thus, whereas rereading yielded consistent improvement in skilled readers' processing of high- and low-frequency words, it appeared primarily to reduce unskilled readers' need for additional processing time on low-frequency words as opposed to high-frequency words (Ardoin, Binder, et al., 2013; Foster et al., 2013; Raney & Rayner, 1995).

Although past research clearly indicates differences in rereading effects among adults versus children, what remains unclear is the source of these differences. Given that adults and children differ from each other in myriad ways, observed differences between adults' and children's rereading behavior could relate to multiple variables, including but not limited to chronological age, reading skill level, and linguistic processing abilities. Participants in adult rereading studies (e.g., Hyönä & Niemi, 1990; Raney & Rayner, 1995) were university students. Thus, they differed both in age and in reading level from the second-grade students participating in studies by Foster et al. (2013) and Ardoin, Binder, et al. (2013). Furthermore, research indicates that, even when matched for broad reading level, adults and children exhibit different linguistic processing profiles (e.g., Greenberg, Ehri, & Perin, 1997, 2002; Thompkins & Binder, 2003). For example, children demonstrate stronger phonological decoding skills than do functionally illiterate adults, and adults are more likely than children to rely on visual or orthographic knowledge and context to compensate for reading difficulties (Thompkins & Binder, 2003).

In the absence of research directly comparing rereading effects among readers of varying ages and skill levels, it is difficult to determine how chronological age and reading skill might moderate intervention effects associated with RR. Observed differences between skilled adult readers (Raney & Rayner, 1995) and unskilled elementary readers (Ardoin, Binder, et al., 2013; Foster et al., 2013) cannot be attributed cleanly to either variable. Thus, based on these results alone, it is impossible to predict how RR might differentially impact the reading behavior of different aged and/or differently skilled groups, including but not limited to: skilled (e.g., gifted) and struggling elementary students, proficient adult readers and adult literacy learners, and unskilled adult readers and schoolchildren exhibiting comparable reading skills. However, past eye movement research conducted with these groups might offer clues as to their potential responses to RR.

Eye Movement Research Evaluating Age- and Skill-Based Developmental Differences

Eye movement studies examining developmental differences in reading behavior provide preliminary insight into potential age- and skill-based differences in *re*reading behavior. Consequently, three lines of research with particular relevance to the current study are briefly described here. These include: (a) studies investigating differences in the eye movements of adults and children, (b) studies comparing the reading behavior of age-matched individuals with differing skill levels, and (c) studies conducted with unskilled adult readers (e.g., Binder & Borecki, 2008; Blythe et al., 2006; Hutzler & Wimmer, 2004; Joseph & Liversedge, 2013).

Although eye movement research investigating rereading has been limited to separate studies with adults (e.g., Hyönä & Niemi, 1990; Raney & Rayner, 1995) and children (Ardoin, Binder, et al., 2013; Foster et al., 2013), several researchers have directly compared the eye movement behavior of adult and child participants within single studies (e.g., Blythe, Häikiö, Bertram, Liversedge, & Hyönä, 2011; Blythe, Liversedge, Joseph, White, & Rayner, 2009; Joseph & Liversedge, 2013). However, as mentioned previously, eye movement research involving skilled adult readers like college students and graduates does not allow for the parsing of age-based versus skill-based differences. In fact, studies have suggested that observed differences in the eye movements of adults and children actually might relate to both sources (e.g., Joseph et al., 2008; Kirkby, Webster, Blythe, & Liversedge, 2008). Past studies of binocular coordination revealed that, although both children and adults were able to coordinate distinct visual input from each eye into a unified representation, children exhibited greater disparity and made more crossed fixations due to age-related differences in oculomotor development (Blythe et al., 2006; Kirkby et al., 2008). Similarly, research on the perceptual span (i.e., the area from which useful information can be obtained during a fixation) indicated that its size increased with age (Rayner, 1986). In contrast, studies specifically examining the reading behavior of adults and children linked differences in eye movement patterns to differences in reading-related skills (e.g., syntactic processing; integrating practical knowledge into sentence representations; Joseph & Liversedge, 2013; Joseph, Liversedge, Blythe, White, & Rayner, 2009). In general, extant research conducted with adults and children suggests that oculomotor control and visual perception develop with age, but that reading-related eye movement patterns vary as a function of linguistic processing skills and efficiency (Blythe & Joseph, 2011; Reichle et al., 2013; Zang, Liang, Bai, Yan, & Liversedge, 2013).

Differences in the eye movement patterns of dyslexic and nondyslexic readers matched for chronological age have provided even stronger support for the idea that eye movements during reading are predominantly influenced by skill level. Past studies indicated that the sizes/durations and counts of fixations, saccades, and regressions made during reading by children with dyslexia differed from those of same-age average readers (Bayram, Camnalbur, & Esgin, 2012; De Luca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002; De Luca, Di Pace, Judica, Spinelli, & Zoccolotti, 1999; Hutzler & Wimmer, 2004). Furthermore, compared to readers demonstrating typically developed skills, readers with dyslexia skipped fewer words and exhibited greater sensitivity to word length (Bayram et al., 2012; De Luca et al., 1999; Hutzler & Wimmer, 2004). According to Pavlidis (1981), observed differences suggested that dyslexia is characterized by oculomotor impairment and that erratic eye movements are the source of reading difficulties. However, the majority of research comparing dyslexic and nondyslexic readers indicated that both groups demonstrated similar eye movement control during non-reading tasks (e.g., De Luca et al., 1999). Thus, between-group differences in eye movement patterns during reading are primarily thought to reflect dyslexic children's underdeveloped reading skills (e.g., De Luca et al., 1999; Hutzler & Wimmer, 2004).

Similar to studies conducted with disabled and nondisabled child readers, research with adult readers of varying skill levels has revealed differences in reading behavior due to differences in reading skill development. Within the oft-studied undergraduate population, Chace, Rayner, and Well (2005) investigated readers' use of phonological preview information and found that differences in skill level (as determined by percentile rank on the Nelson-Denny

Reading Test) explained differences in eye movement patterns and fixation times underlying reading. Similarly, Kuperman and Van Dyke (2011a, 2011b) discovered predictive relationships between reading-related skills (particularly rapid naming, word identification, segmentation, and reading comprehension) and eye movement patterns among a lesser-studied adult sample, namely non-college-bound young adults. Research comparing the reading behavior of college students and Adult Basic Education (ABE) students also indicated skill-based differences among similarly aged readers. Specifically, compared to college students, ABE students exhibited slower target-word naming speeds and relied more heavily on orthographic information during reading (Binder & Borecki, 2008; Binder, Chace, & Manning, 2007). However, eye movement studies also indicated behavioral similarities amongst adults; both undergraduate and ABE students benefited from predictive contextual information and generated inferences during passage reading (Binder & Borecki, 2008; Binder et al., 2007). Based on these results, Binder et al. (2007) and Binder and Borecki (2008) posited that, in some regards (e.g., context usage), unskilled adult readers more closely resemble skilled adult readers than children with comparable reading skills. Thus, research with proficient and unskilled adult readers suggests that both age- and skill-based differences might explain between-group differences in eye movement patterns during reading.

Limitations of Extant Eye Movement Research

Although past studies have demonstrated how reading behaviors (and thus, rereading behaviors) are likely to vary as a function of chronological age and reading skill, such studies merely suggest that readers of differing skill levels are likely to respond to RR in different ways, and that unskilled adult readers may respond more similarly to other adults than to children. The studies reviewed here are not nearly as informative as would be direct comparisons of rereading effects among skilled and unskilled readers of varying ages. Specifically, they provide no insight into *how* such readers might respond behaviorally to RR. In addition, studies examining the reading behaviors of adults sometimes confound age and reading skill. In eye movement research comparing skilled adult readers and elementary students, groups differed by both age and skill (e.g., Joseph & Liversedge, 2013; Joseph et al., 2009). Likewise, some of the unskilled adult readers in past research exhibited higher levels of reading skills than those of young elementary students (e.g., reading levels of Grades 2 through 8 in Binder et al., 2007); thus, differences in reading skill could partially explain why the eye movement patterns of unskilled adult readers have resembled those of skilled adult readers more than those of young children. Finally, research with age-matched participants has provided valuable support for skill-based differences in reading behavior (e.g., De Luca et al., 1999; Hutzler & Wimmer, 2004), but researchers have not similarly controlled for reading skill level to investigate potential age-based differences among unskilled readers.

Purpose and Hypotheses

The current study was designed to overcome limitations of past research on rereading/RR and age- and skill-based group differences by directly comparing rereading behavior among children and adults with similar reading skills. More specifically, it sought to determine if response to RR (i.e., four consecutive readings of the same passage) differs between unskilled adult and child readers. Given that differences in reading-related skills and linguistic processing efficiency have been shown to effect significant differences in the eye movements of adults and children (e.g., Blythe & Joseph, 2011; Reichle et al., 2013), this study aimed to control for skill level differences and to avoid confounding of age and skill effects. By involving participants similar in age to those in past eye movement studies on rereading (i.e., elementary students and adults), this study also sought to clarify and extend upon past findings (e.g., Ardoin, Binder, et al., 2013; Foster et al., 2013; Raney & Rayner, 1995).

In addition to bridging an identified gap in the eye movement research literature, the current study was designed to yield potential findings with practical implications for the implementation of RR. Specifically, it sought to discover how RR might differentially impact the reading behavior of two distinct groups of struggling readers (i.e., elementary students and adult literacy learners). Given that relatively little eye movement research has been conducted with unskilled adult readers (compared to proficient adult readers), this study sought to provide insight into the reading behavior of an understudied population and to investigate whether such individuals might potentially benefit from RR-based intervention. In light of the fact that many children continue to lack reading fluency (Daane et al., 2005) and 14% of American adults demonstrate below basic prose literacy (U.S. Department of Education, 2003), findings regarding the effectiveness of RR methods have important implications for both elementary education and ABE practices. Knowledge regarding these individuals' responses to RR and the factors influencing them could prove valuable for researchers and practitioners seeking to design and implement effective intervention. By utilizing an adult participant sample that represents the adult population most likely to benefit from methods like RR (e.g., struggling readers, high school dropouts, etc.), this study improves upon past research conducted with undergraduate students and potentially increases external validity of findings.

The current study focused on the following questions:

1) Does RR appear to facilitate children's textual processing, as indicated by changes in eye movements across consecutive readings of a passage? That is, do results replicate those in prior research with similar participants (e.g., Foster et al., 2013)?

2) Does RR facilitate unskilled adult readers' processing of text, as indicated by changes in eye movements across consecutive readings?
3) Do effects of RR (i.e., behavioral changes referred to in questions 1 and 2) differ significantly between participant groups, as evidenced by interactions between group and rereading effects?
4) For each group, do RR effects interact with word frequency effects? Specifically, after RR, do unskilled adult readers and/or children continue to require additional processing time on low-frequency words as compared to high-frequency words? Do observed behavior changes differ across high- and low-frequency target words?
5) Are there group differences in the degree to which observed RR effects are specific to words of a particular frequency (high- or low-frequency words)? Are results consistent with past observations during separate studies with adults and children (e.g., Foster et al., 2013; Raney & Rayner, 1995)?

It was hypothesized that, among both groups of readers, RR would yield significant changes in eye movements indicating facilitation of word recognition and higher-level processing of text (i.e., decreases in fixation counts, fixation durations, and number of regressions). Global effects of RR (i.e., behavioral changes measured across the entire passage) were not expected to differ significantly between groups. In addition, it was hypothesized that both groups of readers would exhibit continued sensitivity to word frequency (i.e., additional fixation time on low- versus high-frequency words) even after RR. In contrast with observed differences between adults' (Raney & Rayner, 1995) and children's (Foster et al., 2013) behavioral responses to rereading, it was expected that RR would primarily impact reading of low-frequency words for both groups; similarities were expected due to consistency of reading skill level across participants. However, it was also considered that, in the case of potential interactions involving RR and age group, improvements due to RR might occur more rapidly among unskilled adult readers versus children due to age-related differences in reading experience and context usage (Binder & Borecki, 2008; Binder et al., 2007).

Method

Participants and Settings

Participants included 36 second- to fifth-grade students (Grade 2 $n = 18$; Grade 3 $n = 5$; Grade 4 $n = 6$; Grade 5 $n = 7$) with a mean age of 9 years, 2 months (range = 7 years, 7 months to 11 years, 3 months) who were enrolled in two public elementary schools. This elementary student subsample consisted of 16 males and 20 females identifying as White (75%), Black (11%), multiracial (11%), or Hispanic (3%). Participants also included 36 adult literacy learners with a mean age of 41 years, 8 months (range = 19 years, 2 months to 74 years, 11 months) enrolled in four adult education programs located in the Southeastern United States. This adult subsample consisted of 14 males and 22 females identifying as Black (72%), White (19%), Hispanic (3%), Asian (3%), or American Indian (3%). English language learners (i.e., those receiving English for Speakers of Other Languages instruction) in both types of settings and elementary students receiving special education services for learning disabilities or gifted education were excluded from participation in the current study.

All participants were administered aReading (Adaptive Reading), a brief computer-adaptive measure of broad reading achievement yielding Rasch unit or RIT scores that estimate achievement using item difficulty values; additional details on aReading are provided in the Materials and Measures section. Consistent with the purpose of the current study, participants were included based on their performance on this measure in order to control for level of broad reading skills. More specifically, all RIT scores fell within the range expected for at-risk to grade-level performance appropriate for Grades 2 through 5 for both child (*M* = 488.78, range = 435 to 546, *SD* = 23.59) and adult participants (*M* = 489.06, range = 435 to 544, *SD* = 24.07). Furthermore, adult and child participants were matched based on aReading scores (mean difference = –.28; range of differences = –15 to 5), such that participant subgroups based on age did not differ significantly in terms of broad reading achievement, $t(70) = .05$, $p = .961$.

Apparatus

Eye movements were measured using a desktop-mounted SR Research EyeLink 1000 system, which uses an Ethernet connection between the eye tracker and a display computer for real-time transfer of eye movement data. The system has a sampling rate of 1000 Hz, a resolution of 0.01 degrees of visual angle, and a range of 32 degrees horizontally and 25 degrees vertically. By default, eye movements were recorded from the right eye; tracking from the left eye occurred only when necessary (i.e., when tracking issues did not allow for recording from the right eye). Although the movements of only one eye were recorded, participants' viewing of reading material was binocular throughout assessment.

Reading material was displayed on one of two ViewSonic LCD display monitors: a 19-inch (48.26 cm) VG930m model or a 22-inch (55.88 cm) VX2268wm model. Each monitor was adjusted to a comfortable level of brightness that remained constant throughout testing. Eye movement monitoring was conducted in a dimly illuminated room in a familiar location (i.e., an unused room in participants' places of education), with the brightness of the rooms occasionally adjusted to minimize track losses. Participants used a USB mouse to answer questions and to indicate when they were finished reading displayed text, which allowed them to communicate without speaking and/or moving their heads.

Materials and Measures

aReading (Adaptive Reading). Participants were administered aReading, a browser-based computer-adaptive measure of broad reading that assesses several reading-related skills including knowledge of concepts of print, phonemic awareness, phonics, orthography and morphology, reading comprehension, and vocabulary. aReading is intended for use with students at the kindergarten through twelfth-grade levels and is comprised of 30 items presented in conventional standardized testing formats (i.e., multiple choice, fill in the blank), with auditory and visual stimuli accompanying each question (FAST Research and Development, 2014). aReading is considered to be a robust measure of reading achievement and has demonstrated high levels of validity, reliability, and diagnostic accuracy as a screening tool (Center on Response to Intervention at American Institutes for Research, n.d.). Specifically, the majority of reliability coefficients for aReading are greater than .80, with alternate-form reliability and internal consistency coefficients equal to .95. Ongoing investigations suggest that aReading also demonstrates adequate content, construct, and predictive validity, with coefficients ranging from .56 to .84 (Christ, Monaghen, Van Norman, Kember, & White, 2014). Administration and scoring of aReading are fully automated, yielding Rasch unit or RIT scores that estimate achievement using item difficulty values. aReading score reports were used to determine participants' reading levels in line with grade-specific benchmark/criterion standards and to match adult and child participants for statistical analyses of collected eye movement data.

Silent reading passage. During eye-tracking sessions, participants read an experimenter-developed narrative passage adapted from a Grade 3 FAIP-R curriculum-based measurement probe (Christ, Ardoin, & Eckert, 2010). The reading level of the passage was 3.05, according to the Spache (1953) readability formula. An example of the text is provided in Appendix A.

Embedded in the text were eight target words of low frequency and eight target words of high frequency. Low-frequency words had a frequency of U = 10 or less, and high-frequency words had a frequency of U = 50 or above, with U indicating the number of instances of that word per million running words. Target words were determined to meet specified frequency standards based on corpora of both adult-level and child-level written/spoken language; that is, to be selected as a target word, a word was required to meet the above standards according to both sources. Child-level frequency norms were referenced using *The Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995), a corpus based on American textbooks, works of literature, and popular works of fiction and nonfiction for primary and secondary students; published statistics specific to Grades 2 through 5 were consulted. Adult-level frequency norms were referenced using the $\text{SUBTLEX}_{\text{US}}$ corpus (Brysbaert & New, 2009), which is based on subtitles from U.S. films from 1900 to 2007 and U.S. television series. The $\text{SUBTLEX}_{\text{US}}$ corpus was selected because it is recognized as a better approximation of real-life word exposure and a better predictor of lexical decision times when compared to other frequently used corpora (e.g., Kŭcera and Francis, 1967). In light of well-documented effects of word length on eye movement patterns (e.g., Just & Carpenter, 1980), high- and low-frequency target words also were matched for length to prevent confounding of frequency and length. Specifically, high-frequency words had a mean length of 7.25 characters (range = 5 to 10 characters), and low-frequency words had a mean length of 7.88 characters (range = 5 to 9 characters). Target words and their frequencies and lengths are presented in Table 3.1. Target word locations also were consciously manipulated to exclude the first and last sentences of paragraphs and the ends of sentences, given documented differences in processing time based on sentence location (e.g., extra "wrap-up" processing time at the ends of sentences; Just & Carpenter, 1980). High- and low-frequency words were also controlled for part of speech (such that they were all verb forms) and embedded within parallel phrase structures (e.g., infinitives, participles).

The reading passage was displayed as black text against a white background and was formatted in standard upper- and lowercase letters in 14-point Century Gothic font. The passage was presented in its entirety as one page of 1.5-spaced text occupying 23 lines, with a maximum line width of 74 characters.

Procedure

aReading was individually administered to participants prior to eye movement data collection, to best ensure that potential participants exhibited appropriate-level reading skills (i.e., broad reading achievement consistent with benchmark scores for Grades 2 through 5). During administration of aReading, experimenters briefly explained the task to participants, readied the assessment (i.e., opened the program in a browser, selected the appropriate participant), and proctored the test. Headphones were utilized to ensure adequate volume of auditory stimuli. Testing lasted approximately 15 to 30 min, and participants were given short breaks between tasks to prevent fatigue.

Eye movement monitoring was conducted individually, with each participant sitting approximately 50 to 55 cm from the display monitor and placing his/her chin in a chin rest to minimize head movement. During setup, participants were informed about the eye-tracking task—namely, that they would read a passage from the display monitor silently four consecutive times while a camera recorded their eye movements. They also were instructed on how to use the computer mouse to indicate when they were finished reading and to answer questions without speaking. Before reading trials, the eye-tracking system was calibrated using a nine-point grid extending across the display screen; experimenters explained this calibration process as a “game” requiring participants to follow a displayed dot with their eyes without moving their heads. Upon successful calibration, another nine-point grid was used to validate the accuracy of tracking. After calibration and validation were complete, participants completed a practice trial, which was designed to familiarize them with silently reading information from the monitor and using the mouse as directed. After the practice trial, participants were given a short break before the experimental trials.

Once participants were ready to continue with the task, experimenters reminded them that they would next complete a series of trials requiring them to read the experimental passage silently four consecutive times and to answer a unique comprehension question after each reading. Following recalibration and validation of the eye-tracking system, experimenters presented the experimental passage once participants’ eyes were positioned appropriately (i.e., fixated on a target displayed in the upper left corner of the monitor). After reading the passage, participants used the computer mouse to clear the passage and to answer the resulting comprehension question. They also were provided with verbal feedback regarding how long it took them to read the passage. This process was repeated for a total of four readings and four comprehension questions. Before each reading trial, participants were instructed to do their “best reading” without assistance and to read the entire passage.

Individual eye-tracking sessions, including the practice trial and four reading trials and questions, lasted approximately 10 to 15 min. Head movement and other tracking issues occasionally necessitated recalibration, additional breaks, and/or adjustment of the camera or lighting in the room. Of note, participants did not receive feedback as to whether their answers to comprehension questions were correct, and their responses were not scored or analyzed. Rather, questions primarily served to ensure that participants were attending to the content of the passage (i.e., reading for comprehension) instead of focusing solely on reading rate.

Results

Two types of analyses were conducted to address research questions and test hypotheses: global analyses of reading behavior at the passage level (i.e., eye movements averaged across all words within the passage) and target-word analyses examining eye movements averaged across embedded high- and low-frequency target words. For global analyses, mixed analyses of variance (ANOVAs) involving one within-subjects variable (RR, with 4 levels) and one between-subjects variable (age group: adults versus children) were conducted for each eye movement measure. Target-word analyses involving an additional within-subjects variable (word frequency: high versus low) were conducted using 4 (RR) × 2 (age group) × 2 (frequency) mixed ANOVAs. When Mauchly's test indicated a violation of sphericity, degrees of freedom were corrected using Greenhouse-Geisser estimates when epsilon was less than .75 and Huynh-Feldt estimates when epsilon was greater than .75. Simple effects analyses and Bonferroni-corrected comparisons were used to evaluate significant interactions and main effects, respectively. Effect sizes are reported as values of eta squared, which can be interpreted according to Cohen's (1988) benchmark guidelines for small, medium, and large effects (i.e., $\eta^2 = .0099$, .0588, and .1379, respectively).

Prior to analyses, outliers were identified using the outlier labeling rule (Hoaglin & Iglewicz, 1987; Tukey, 1977), which was used to establish cut-off values by multiplying the interquartile range by a factor (*g*) of 2.2, as recommended by Hoaglin and Iglewicz (1987). Specifically, this generated value (*g'*) was subtracted from the score at the 25th percentile and added to the score at the 75th percentile, and scores falling outside of these limits were identified as outliers. Winsorizing (i.e., replacement with the most extreme valid values) was used to adjust outliers; one value for adults and one to two values for children were replaced for variables of gaze duration, total fixation time, number of inter-word regressions, and average number of fixations per word. Following outlier detection and winsorizing, data were analyzed without further transformation in light of research indicating the robustness of ANOVAs to deviations from normality, particularly with adequate sample sizes ($n > 25$–30; Glass, Peckham, & Sanders, 1972; Rasch, Teuscher, & Volker, 2007; Schmider, Ziegler, Danay, Beyer, & Buhner, 2010).

Due to track losses (i.e., loss of eye position recording due to participant head movement or technological issues) encountered for four adult participants and program absences encountered for five adult participants, oversampling was conducted to ensure completion of data collection with 36 adults and 36 children. Within this sample, two adult participants completed only three trials that could be analyzed, and three child participants completed only three trials that could be analyzed; thus, the following analyses are based on a participant sample size of $N = 67$ (adult $n = 34$; child $n = 33$). For all analyses, individual fixations shorter than 120 ms or longer than 1800 ms were omitted, as they were thought to reflect non-reading behavior or momentary track losses; these cutoffs were based on prior research examining changes in reading behavior (e.g., Raney & Rayner, 1995), with the upper bound slightly modified in light of persistent fixations exhibited during reading by adult learners (K. S. Binder, personal communication, November 12, 2014).

Global Analyses

Global measures averaging eye movement parameters across all words in the passage included first fixation duration, gaze duration, total fixation time, number of inter-word regressions, and average number of fixations per word. (See Appendix B for definitions of these variables). Means, test statistics, and effect sizes for each global measure are presented in Table 3.2.

Analyses of all global measures revealed significant improvement in the expected direction across participants' four readings of the passage. Significant main effects for rereading were observed for all measures, including a small effect for first fixation duration ($\eta^2 = .04$) and medium to large effects for all other measures (gaze duration $\eta^2 = .12$; total fixation time $\eta^2 = .18$; number of inter-word regressions $\eta^2 = .11$; average number of fixations per word $\eta^2 = .21$). Pairwise comparisons between the first and second readings (i.e., Trials 1 and 2) indicated immediate significant effects on measures associated with early lexical processing (i.e., first fixation duration, $p = .025$; gaze duration, $p = .031$) and higher-level textual processing (i.e., total fixation time, $p < .001$; average number of fixations per word, $p = .030$). Furthermore, improvements appeared to persist following participants' third and fourth readings (compared to their first reading, as evidenced by pairwise comparisons between Trial 1 and Trials 3 and 4, respectively) and yielded significant aggregate effects after four readings for all measures excluding first fixation duration (i.e., gaze duration, $p < .001$; total fixation time, $p < .001$; number of inter-word regressions, $p = .012$; average number of fixations per word, $p < .001$).

None of the global dependent measures revealed significant interactions between rereading and participant age group, indicating no significant between-group differences in facilitation effects resulting from RR. However, medium to large significant main effects for age group were found for all three measures representing higher-level textual processing (i.e., total fixation time $\eta^2 = .08$; number of inter-word regressions $\eta^2 = .22$; average number of fixations per word $\eta^2 = .10$), indicating increased passage-level sampling behavior and aggregate processing time exhibited by younger readers.

Target-Word Analyses

Target-word measures included first fixation duration, gaze duration, and total fixation time on target words and the average number of fixations per target word. Test statistics are discussed below; visual depictions of means for each target-word measure (separated by group) are presented in Figures 3.1–3.4.

First fixation duration. Analyses of participants' first fixation durations on target words revealed that the three-way interaction for RR × Frequency × Age was not significant, $F(2.88, 187.38) = .46$, $p = .702$, $\eta^2 = .01$. In addition, two-way interactions were not significant for RR and word frequency, $F(2.88, 187.38) = 1.00$, $p = .393$, $\eta^2 = .02$, or RR and participant age group, $F(2.88, 186.90) = .48$, $p = .689$, $\eta^2 = .01$. However, there was a significant two-way interaction effect involving word frequency and participant age group, $F(1, 65) = 4.89$, $p = .031$, $\eta^2 = .07$, suggesting that effects of participant age group differed on high- and low-frequency target words (see Figure 3.1). Specifically, follow-up simple effects analyses revealed that group differences in first fixation duration on low-frequency target words approached statistical significance, $F(1, 65) = 3.69$, $p = .059$, $\eta^2 = .05$, whereas adults' and children's first fixation durations on high-frequency target words did not differ significantly, $F(1, 65) = .027$, $p = .870$, $\eta^2 < .001$. Simple effects of word frequency were not significant for either adults ($p = .156$) or children ($p = .004$).

Participants' first fixation durations on target words did not reveal significant main effects of word frequency, $F(1, 65) = .04$, $p = .843$, $\eta^2 < .001$, or participant age group, $F(1, 65)$

= 1.14, p = .290, η^2 = .02. However, a significant main effect of RR was observed, F(2.88, 186.90) = 2.95, p = .036, η^2 = .04. Follow-up pairwise comparisons revealed a significant decrease in first fixation duration between Trials 3 and 4 (p = .022) but did not indicate significant facilitation compared to Trial 1 after any of the rereading trials.

Gaze duration. Target-word analyses of gaze duration indicated that the three-way interaction for RR × Frequency × Age was not significant, F(3, 195) = .72, p = .541, η^2 = .01. Additionally, the two-way interaction between RR and word frequency was not significant, F(3, 195) = 1.73, p = .162, η^2 = .03. However, participants' gaze durations on target words revealed a significant main effect of RR, F(3, 195) = 7.67, p < .001, η^2 = .10, that took effect immediately after the second reading (p < .001) and a significant, large main effect of word frequency, F(1, 65) = 39.65, p < .001, η^2 = .38.

No significant between-group differences were observed for participants' gaze durations on target words, F(1, 65) = 1.22, p = .274, η^2 = .02. Furthermore, two-way interactions were not significant for RR × Age, F(3, 195) = .62, p = .605, η^2 = .01, or for Frequency × Age, F(1, 65) = .03, p = .858, η^2 < .001, suggesting that participants' gaze durations (and the degree to which they were influenced by RR and word frequency) did not appear to differ significantly for adults and children.

Total fixation time. Analyses of total fixation time, a measure thought to represent higher-level processing, revealed that there was no significant interaction between RR, word frequency, and participant age group, F(2.78, 180.56) = .83, p = .473, η^2 = .01. However, significant, large effects for RR, F(2.79, 181.14) = 12.04, p < .001, η^2 = .16, and word frequency, F(1, 65) = 41.55, p < .001, η^2 = .39, were observed and qualified by a significant RR × Frequency interaction, F(3, 102) = 10.31, p < .001, η^2 = .23.

Follow-up analyses of RR effects indicated that rereading exerted significant, large effects on participants' total fixation times on both high-frequency target words and low-frequency target words (high-frequency Wilks' Lambda = .83, $F(3, 63) = 4.37$, $p = .007$, $\eta^2 = .17$; low-frequency Wilks' Lambda = .64, $F(3, 63) = 11.91$, $p < .001$, $\eta^2 = .36$). However, effects were notably larger and more persistent on low-frequency targets. Specifically, although participants' total fixation times on both high- and low-frequency target words decreased significantly between Trials 1 and 2 (high-frequency $p = .003$; low-frequency $p = .011$), performance on high-frequency targets then weakened/changed direction, such that total fixation times at Trials 3 and 4 were no longer significantly different than those at Trial 1. In contrast, total fixation times on low-frequency targets indicated continued facilitation at Trials 3 and 4 ($p < .001$ for both). Simple effects analyses examining frequency effects across reading trials revealed that, although rereading yielded significant decreases in total fixation time on low-frequency target words, readers continued to fixate longer on low-frequency targets as compared to high-frequency targets even following rereading. Specifically, word frequency significantly impacted total fixation time during all four readings, including both before (Trial 1 $p < .001$) and after RR (Trial 4 $p = .002$).

No significant between-group differences were observed for participants' total fixation times on target words, $F(1, 65) = 3.65$, $p = .061$, $\eta^2 = .05$. In addition, two-way interactions were not significant for RR × Age, $F(2.79, 181.14) = .06$, $p = .974$, $\eta^2 < .001$, or for Frequency × Age, $F(1, 65) = .04$, $p = .837$, $\eta^2 < .001$. Thus, target-word analyses did not yield any evidence for significant differences between adults' and children's total fixation times on target words or for differential effects of RR and/or frequency on these durations.

Average number of fixations per word. Analyses of the average number of fixations per target word did not indicate a significant three-way interaction involving RR, word frequency, and age group, $F(3, 195) = .92$, $p = .433$, $\eta^2 = .01$. However, as with total fixation time, significant, large effects for RR, $F(2.82, 183.44) = 16.01$, $p < .001$, $\eta^2 = .20$, and word frequency, $F(1, 65) = 68.22$, $p < .001$, $\eta^2 = .51$, were noted, along with a significant RR × Frequency interaction effect, $F(3, 195) = 5.45$, $p < .001$, $\eta^2 = .08$.

Follow-up analyses of RR effects indicated that RR exerted significant, large effects on participants' average fixation counts on both high-frequency target words and low-frequency target words (high-frequency Wilks' Lambda = .80, $F(3, 63) = 5.11$, $p = .003$, $\eta^2 = .20$; low-frequency Wilks' Lambda = .63, $F(3, 63) = 12.56$, $p < .001$, $\eta^2 = .37$). However, effects were notably larger and more immediate on low-frequency targets. Specifically, although participants' average number of fixations on both high- and low-frequency target words reflected significant facilitation (compared to initial reading) during the third and fourth readings (high-frequency $p = .024$ at Trial 3 and $p < .001$ at Trial 4; low-frequency $p < .001$ at Trials 3 and 4), average fixation count decreased significantly between Trials 1 and 2 only on low-frequency targets ($p = .002$). Simple effects analyses examining frequency effects across reading trials revealed that, although rereading yielded significant decreases in the average number of fixations per word on both high- and low-frequency target words, readers continued to fixate more times on low-frequency targets as compared to high-frequency targets even following rereading. Specifically, word frequency significantly impacted average fixation count during all four readings, including initially (Trial 1 $p < .001$) and after RR (Trial 4 $p < .001$).

Consistent with global analyses, a significant main effect of participant age group was observed for participants' average fixation count on target words, $F(1, 65) = 5.57$, $p = .021$, $\eta^2 =$

.08. In contrast, two-way interactions were not significant for RR × Age, $F(2.82, 183.44) = .38$, $p = .758$, $\eta^2 < .01$, or for Frequency × Age, $F(1, 65) = .64$, $p = .426$, $\eta^2 < .01$. That is, the effects of RR and word frequency were not observed to differ significantly between adults and children; however, as evidenced by previously described global analyses, child participants were observed to make significantly more fixations per word than their adult counterparts.

Discussion

Previous eye movement research has demonstrated how consecutive rereading of the same text (such as that involved in instructional procedures like RR) facilitates both adults' and children's reading fluency but in distinct ways. In particular, rereading has been shown to yield consistent improvement in adults' processing of high- and low-frequency words (Raney & Rayner, 1995) but appears primarily to reduce children's need for additional processing time on low-frequency words as opposed to high-frequency words (Ardoin, Binder, et al., 2013; Foster et al., 2013). However, due to previous studies' involvement of only proficient adult readers (i.e., university students or graduates), it is unclear how observed differences between participant samples relate to confounded variables of age, reading skill level, and/or additional characteristics such as linguistic processing abilities. As a result, it is difficult to determine how the effects of rereading may vary as a function of these variables and, thus, how rereading procedures (including RR-based interventions) may differentially impact the reading behavior of different aged and/or differently skilled groups. As both adults and children regularly engage in rereading for various purposes (e.g., to improve understanding, in the context of reading interventions targeting fluency, etc.), it seems important to investigate how their reading performance and behavior change as a function of repeated exposure to a text. Furthermore, given that struggling readers are the most likely to engage in rereading, there is a clear need for investigations of rereading effects to be conducted with unskilled readers (both adults and children) as opposed to proficient ones. Thus, the purpose of the current study was to investigate and compare rereading behavior among two distinct groups of similarly skilled but different aged developing readers (i.e., elementary students and adult literacy learners).

As hypothesized, global analyses of participants' eye movements suggested that rereading (RR) yields significant changes in multiple reading behaviors exhibited by both children and adults with developing reading skills. Specifically, measures associated with early lexical processing (i.e., first fixation duration and gaze duration) and higher-level textual processing (i.e., total fixation time, number of inter-word regressions, and average number of fixations per word) reflected improvement in the expected direction due to rereading. Consistent with previous eye movement studies conducted separately with beginning readers (i.e., elementary students; Ardoin, Binder, et al., 2013; Foster et al., 2013; Zawoyski, Ardoin, & Binder, 2014) and adult readers (Hyönä & Niemi, 1990; Kaakinen & Hyönä, 2007; Raney & Rayner, 1995; Schnitzer & Kowler, 2006), participants in the current study spent significantly less time actively fixating on words, made significantly fewer fixations per word, and exhibited significantly fewer regressions to previously fixated content following four consecutive readings of the same text. Furthermore, as in previous research with similarly skilled elementary student participants (Foster et al., 2013), pairwise comparisons across readings indicated immediate improvements (i.e., significant effects after a single rereading) as well as continued facilitation when reading text a third and fourth time.

Also in line with hypotheses, global rereading effects (i.e., behavioral changes measured across the entire passage) were not found to differ significantly between children and adults with similar reading skills. That is, none of the global eye movement measures revealed significant interactions between effects of rereading and participant age group. Thus, current results suggest that rereading (the foundation of instructional procedures involving RR) can facilitate improved efficiency of reading at both word and passage levels for unskilled readers, regardless of age. Taken together with previous findings indicating differential effects of rereading based on readers' skill level (i.e., significant interactions between effects of rereading and relative skill level on average number of fixations per word; Zawoyski et al., 2014), current results add to research supporting the idea that eye movements during reading are predominantly influenced by readers' linguistic processing skills rather than age-based oculomotor or perceptual development (Blythe & Joseph, 2011; De Luca et al., 1999; Hutzler & Wimmer, 2004; Reichle et al., 2013; Zang, Liang, Bai, Yan, & Liversedge, 2013).

Consistent with hypotheses based on previous research investigating effects of rereading and word frequency on eye movement parameters (Ardoin, Binder, et al., 2013; Foster et al., 2013; Raney & Rayner, 1995), current participants demonstrated a persistent sensitivity to word frequency even after four consecutive readings of the passage, as evidenced by measures of total fixation time and average number of fixations on target words. Like both adult and child participants in the aforementioned studies, they continued to make more fixations and to fixate longer on low-frequency words as compared to high-frequency words. These results further support past findings that rereading does not eliminate frequency effects in reading (Ardoin, Binder, et al., 2013; Foster et al., 2013; Raney & Rayner, 1995). Thus, it appears that readers along a broad continuum of ages and skill levels—from elementary-age unskilled readers to adults still developing reading fluency to proficient college-age readers—continue to exhibit relative difficulty when reading low-frequency and unfamiliar words as compared to high-frequency and familiar words, even despite improvements related to RR.

In line with previously observed differences between skilled adult readers' and children's behavioral responses to RR (i.e., Ardoin, Binder, et al., 2013; Foster et al., 2013; Raney & Rayner, 1995) and consistent with previous rereading research with similarly skilled elementary-level readers (Foster et al., 2013), target-word analyses revealed significant interactions between rereading and word frequency on multiple measures of reading behavior (i.e., total fixation time and average number of fixations per word). Specifically, although rereading yielded significant effects on both high-frequency and low-frequency target words, effects were notably larger and more prominent (i.e., more immediate for average number of fixations; more persistent for total fixation time) on low-frequency words. Thus, as hypothesized, RR exerted a stronger impact on participants' reading of low-frequency words. However, current findings notably differed from those in past eye movement studies with elementary students (Ardoin, Binder, et al., 2013; Foster et al., 2013) in that rereading in the current study also resulted in significant, albeit lesser, facilitation on high-frequency target words.

The discrepancy between current findings and previous conclusions that RR mainly improves unskilled readers' fluency on low-frequency words (Ardoin, Binder, et al., 2013; Foster et al., 2013; Zawoyski et al., 2014) may relate to the selection of study-specific target words to examine the effects of word frequency. More specifically, whereas the current study attempted to control for additional lexical qualities of target words (i.e., length, position, part of speech, grammatical context), previous research suggesting a lack of facilitation on high-frequency words did not involve matching high- and low-frequency target words for these characteristics. Given that high-frequency words tend to be shorter on average than low-frequency words (Zipf, 1949) and word length has been shown to reliably impact fixation behavior (Just & Carpenter, 1980; Rayner, 1983), it is possible that differences in these characteristics could explain previously observed "floor effects" (i.e., leveling out of facilitation) specific to high-frequency words (e.g., Zawoyski et al., 2014). In contrast, the close matching of target words in the current study may have more evenly allowed for improvement on both high- and low-frequency words. Greater room for improvement on high-frequency words also may explain additional discrepancies between current and past rereading studies with unskilled readers (Foster et al., 2013; Zawoyski et al., 2014), including fewer significant two-way interactions between rereading and frequency and a significant (albeit small) main effect of rereading on first fixation duration in the current study. In addition, differences in the selection and matching of target words could also potentially explain why current results were more similar to past findings with adult readers (Raney & Rayner, 1995) than were those from previous eye movement studies with children. In their study, which likewise indicated significant improvement on both high- and low-frequency words, Raney and Rayner (1995) matched high- and low-frequency target words for length, position, and meaning.

As hypothesized, target-word analyses did not reveal any significant between-group differences in the degree to which observed rereading effects were affected by word frequency; none of the eye movement measures revealed significant three-way interactions between rereading, frequency, and age. However, it bears noting that a few significant between-group differences emerged from global and target-word analyses. Although children and adults both similarly seemed to benefit from rereading of the passage, measures of higher-level textual processing (i.e., total fixation time, number of inter-word regressions, average number of fixations per word) indicated that child participants exhibited significantly more fixations and regressions (which yielded significantly longer summed fixation times) on average compared with their adult counterparts. Thus, children seemed to visually "sample" and refixate around the passage to a greater degree than did adults, who devoted slightly more visual attention to words initially (average global first fixation duration across reading trials = 279.71 to 291.45 ms for adults and 279.72 to 283.37 ms for children) but made significantly fewer discrete fixations and exhibited significantly shorter overall lexical processing times compared to children. This was particularly the case on low-frequency words, as evidenced by group differences in first fixation duration on low-frequency target words (M = 314.64 ms for adults and 279.36 ms for children) approaching statistical significance. Although the origin of these differences is unclear, these data suggest that even unskilled adult readers may exhibit greater efficiency in reading at the passage level (i.e., less "jumping around") than do children with comparable broad reading skills. It may be that adults, even those that struggle with reading, are able to extract textual information within fewer discrete pauses due to improved working memory or age-related differences in reading experience and context usage (Binder & Borecki, 2008; Binder et al., 2007) compared to elementary students whose sight word vocabularies are still developing. However, despite these noted behavioral differences, current findings suggested that unskilled adult and child readers do not seem to significantly differ with regard to the benefits they derive from repeated practice of a given passage.

Limitations and Future Directions

Current results should be carefully interpreted bearing several limitations in mind. First, current participants did not include elementary students receiving resource reading instruction or adults or children receiving English to Speakers of Other Languages (ESOL) instruction. Unfortunately, as a result, this participant sample was not representative of many struggling readers who are likely to receive fluency-based interventions such as RR. However, the National Reading Panel (NICHD, 2000), has suggested that RR is an effective instructional practice for improving the fluency of *all* students through fourth grade and struggling readers through high school. Although current findings have limited generalizability to the specific aforementioned populations, they nonetheless bear relevance to a wider group of students who may benefit from rereading-based interventions as well as adult individuals who are likely to engage in rereading. However, future research is needed to clarify the effects of rereading among individuals of varying ages and skill levels (e.g., elementary-age struggling readers and higher-performing readers, as in Zawoyski et al., 2014).

Another significant limitation is that, although participants were selected and matched based on broad reading achievement scores (i.e., performance commensurate with grade-level performance in Grades 2 through 5), participants likely differed within and between groups along a myriad of dimensions including intellectual ability, level of attention, working memory, task persistence, motivation, learning/educational history, behavioral health concerns, and semi-stable aspects of oculomotor functioning (e.g., binocular coordination, perceptual span). As with all group design studies, current findings should be cautiously interpreted with consideration of individual differences; this seems particularly imperative given the high degree of variability in the adult education population as well as the countless number of ways in which adults and children differ. The current study was designed to focus on the variables of age and skill level due to their established importance based on past research and their ability to be identified and controlled. In addition, the author sought to maximize contrast between subgroups on all variables excepting reading achievement by enlisting readers who were naturally and most unnaturally performing at a given level (i.e., elementary students with grade-appropriate performance and highly impaired older adult readers); it was expected that a lack of significant interactions between rereading effects and participant group would be particularly compelling in the context of maximized differences between subgroups. However, the fact that participants were individuals selected from a particular geographical location and unique educational programs may compromise internal and external validity of specific results. As a result, the utility of current findings should be balanced with careful contemplation of individual variation. It is also the hope of the author that the current study will function as a useful stepping stone in enabling further evaluations of the relationships between eye movements and some of the aforementioned additional group/individual differences.

As is standard in eye movement research examining the impact of word frequency on reading behavior, target words in the current study were selected and dichotomously classified as high- and low-frequency words based on published corpora of frequency norms. However, as discussed above, it is possible that the selection and matching of study-specific target words (e.g., control for lexical variables such as word length, meaning, grammatical context, etc.) may heavily influence results specific to high- and low-frequency words. For example, current target-word results differed from previous findings based on a distinct set of target words that was shared across studies (Ardoin, Binder, et al., 2013; Foster et al., 2013; Zawoyski et al., 2014). Thus, target-word analyses should be interpreted with careful consideration of study-specific procedures/control. To limit the extent to which frequency-related analyses are skewed by differences in study-specific methods, it likely would be beneficial for future researchers to examine the *degree* of frequency for all words in a passage rather than manipulating this variable based upon dichotomous grouping of a small number of high- and low-frequency words. Furthermore, replication of the current and past studies investigating word frequency effects and examination of generalization effects is necessary to determine if findings remain consistent beyond the context of a particular set of methods (i.e., use of a specific set of target words embedded within a specific reading passage).

The current study also was impacted by limitations tied to the use of eye-tracking methodology. Due to technological difficulties and track losses (i.e., instances in which the eye-tracking camera lost sight of the participant's pupil; detailed on p. 93), some eye movements were not recorded, necessitating additional data collection. Unexpected challenges with eye movement recording occurred particularly with adult readers due to older participants' poor vision and use of corrective lenses (especially bifocals with lines, which interfered with tracking), increased squinting behavior, and wearing of dark eye makeup and false eyelashes. In contrast, track losses with children were typically associated with fidgeting or excessive head movement. Researchers conducting eye movement monitoring should carefully consider how use of this technology may pose difficulties with specific participant samples. Furthermore, technological limitations of current eye-tracking systems (e.g., limited tolerance of head and mouth movement) limit generalizability of findings due to the way they dictate/impose specific research methods (e.g., use of silent rereading, which differs from typical implementation of RR by not including monitoring of accuracy, error correction, and performance feedback). As technology progresses, it will be important for future research to utilize experimental methods that more closely approximate typical instructional procedures. This seems particularly important for further examining how adult learners respond (both in terms of effectiveness and acceptability) to strategies typically designed for much younger students.

To avoid overgeneralization of current findings, it seems important to be just as straightforward about the current study's limitations (i.e., what results did *not* reveal) as about its potential utility. Although current results did not indicate significant differences in adult and child participants' responses to rereading, they also did not indicate significant similarities; there simply was not sufficient evidence to suggest significantly different group means. In addition, it is important to recognize that adults and children presumably differ on numerous variables potentially affecting reading behavior (e.g., vocabulary knowledge, purpose in rereading, level of attention, working memory, task persistence, motivation, learning/educational history, and behavioral health concerns) that were not manipulated or investigated in the current study. Thus, although participant subgroups did not respond differently to RR, current findings absolutely do not suggest that adult and child readers are equivalent learners.

Summary and Implications

Despite the aforementioned limitations, results of the current study suggest that repeated practice improves the efficiency of both word- and passage-level reading behavior for unskilled, elementary-level readers at a variety of ages. In light of the high degree of contrast between participant subgroups except with regard to broad reading achievement, it is particularly interesting that readers in the current study did not differ significantly in their response to rereading procedures across multiple measures of behavior. Taken together with previous eye movement research indicating the importance of considering achievement level when evaluating intervention effects (e.g., Ardoin, Binder, Foster, & Zawoyski, in press; Zawoyski et al., 2014), current findings buttress the idea that students' reading behaviors—and the manner in which these change in response to instructional procedures—ultimately reflect and are influenced by their levels of reading skill and achievement. In light of this consideration, but contingent upon further research, it may be that the population with whom rereading-based interventions are effective also includes struggling adult readers whose *reading skill level* (rather than chronological age or grade level) aligns with existing guidelines (e.g., NICHD, 2000) and past empirical evidence.

Although current results indicate that RR appears to facilitate unskilled readers' fluency on both high- and low-frequency words, rereading effects were notably larger and more prominent on low-frequency words. This finding is consistent with those from past eye movement studies of elementary students' rereading (Ardoin, Binder, et al., 2013; Foster et al., 2013; Zawoyski et al., 2014), further supporting the idea that skill level primarily influences reading behavior. In addition, this study aligns with previous research in suggesting that RR may yield variable effects across passages/probes consisting of different vocabulary. Although findings suggest that RR is likely to produce more robust improvement on passages involving low-frequency or unfamiliar words, this potential for gain must be balanced with the added challenges and frustration that this might pose for struggling readers.

Differences between current findings and those from past eye movement studies of rereading have important implications for future reading research. Specifically, as discussed above, different conclusions with regard to readers' improvement on high-frequency target words may potentially have been influenced by use of study-specific reading material. Individuals conducting and consuming reading research should be mindful of how specific results may be tied to particular measures (e.g., experimenter-created passages using specific sets of words). While differing conclusions could appear to relate to salient variables such as participant characteristics (e.g., children versus adults, as in the comparison of Foster et al., 2013 to Raney and Rayner, 1995), they may also be heavily influenced by researchers' use of specific methodological procedures (i.e., controlling/matching for lexical variables such as word length, position, etc.).

Current findings suggest that students with comparable levels of reading achievement may exhibit different patterns of underlying reading behavior. For example, children in the current study generally exhibited an increased tendency to visually "sample" and refixate around the passage compared to similarly skilled adult readers. However, subgroups did *not* significantly differ in the manner in which their reading behavior changed as a function of four consecutive readings of the same text. Ultimately, the current study provides valuable information regarding how RR facilitates the fluency of elementary-level readers of varying ages. By investigating the eye movements of similarly skilled school-age children and adult learners, it bridges an evident gap in the eye movement research literature and expands upon past studies on rereading/RR and age- and skill-based group differences. This study further exemplifies the value of utilizing eye movement recording to examine fine-grained changes in reading fluency (e.g., improvements specific to certain words or behaviors) and behavioral differences between groups of developing readers. Furthermore, it serves as a useful launch pad for continued exploration of variables impacting students' response to instructional procedures and how widely recommended practices may benefit specific groups of students.

References

Ardoin, S. P., Binder, K. S., Foster, T. E., & Zawoyski, A. M. (in press). Repeated versus wide reading: A randomized control design study examining the impact of fluency interventions on underlying reading behavior. *Journal of School Psychology*.

Ardoin, S. P., Binder, K. S., Zawoyski, A. M., Foster, T. E., & Blevins, L.A. (2013). Using eye-tracking procedures to evaluate generalization effects: Practicing target words during repeated readings within versus across texts. *School Psychology Review*, *42*(4), 477-495.

Ardoin, S. P., McCall, M., & Klubnik, C. (2007). Promoting generalization of oral reading fluency: Providing drill versus practice opportunities. *Journal of Behavioral Education, 16*, 55-70. doi: 10.1007/s10864-006-9020-z

Bayram, S., Camnalbur, M., & Esgin, E. (2012). Analysis of dyslexic students' reading disorder with eye movement tracking. *Cypriot Journal of Educational Sciences*, *7*(2), 129-148.

Binder, K., & Borecki, C. (2008). The use of phonological, orthographic, and contextual information during reading: A comparison of adults who are learning to read and skilled adult readers. *Reading and Writing: An Interdisciplinary Journal*, *21*(8), 843-858. doi: 10.1007/s11145-007-9099-1

Binder, K. S., Chace, K. H., & Manning, M. C. (2007). Sentential and discourse context effects: Adults who are learning to read compared with skilled readers. *Journal of Research in Reading*, *30*(4), 360-378. doi: 10.1111/j.1467-9817.2007.00349.x

Blythe, H. I., Häikiö, T., Bertram, R., Liversedge, S. P., & Hyönä, J. (2011). Reading disappearing text: Why do children refixate words? *Vision Research*, *51*, 84-92. doi: 10.1016/j.visres.2010.10.003

Blythe, H. I., & Joseph, H. S. S. L. (2011). Children's eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.). *The Oxford handbook of eye movements* (pp. 801-818). Oxford, Great Britain: Oxford University Press.

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., Findlay, J. M., & Rayner, K. (2006). The binocular coordination of eye movements during reading in children and adults. *Vision Research*, *46*, 3898-3908. doi: 10.1016/j.visres.2006.06.006

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, *49*, 1583-1591. doi: 10.1016/j.visres.2009.03.015

Brysbaert, M., & New, B. (2009). Moving beyond Kŭcera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measures for American English. *Behavior Research Methods*, *41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Bos, C. S. (1982). Getting past decoding: Assisted and repeated readings as remedial methods for learning disabled students. *Topics in Learning & Learning Disabilities, 1*(4), 51-57.

Center on Response to Intervention at American Institutes for Research. (n.d.). *Screening tools chart*. Retrieved from http://www.rti4success.org/resources/tools-charts/screening-tools-chart

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: Effects of reading skill. *Canadian Journal of Experimental Psychology*, *59*(3), 209-217. doi: 10.1037/h0087476

Christ, T. J., Ardoin, S. P., & Eckert, T. L. (2010). *Formative Assessment Instrumentation and Procedures for Reading (FAIP-R): Oral reading passages*. University of Minnesota. Retrieved from http://www.fast.cehd.umn.edu

Christ, T. J., Monaghen, B. D., Van Norman, E. R., Kember, J. M., & White, M. J. (Eds.). (2014). *Formative Assessment System for Teachers technical manual*. Minneapolis, MN: University of Minnesota.

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading* (NCES 2006-469). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. *Brain and Language*, *80*, 617-626.

De Luca, M., Di Pace, E., Judica, A., Spinelli, D. & Zoccolotti, P. (1999). Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologica*, *37*, 1407-1420.

FAST Research and Development. (2014). *FAST aReading*. Retrieved from http://www.fastforteachers.info/#!areading/c1h95

Foster, T. E., Ardoin, S. P., & Binder, K. S. (2013). Underlying changes in repeated reading: An eye movement study. *School Psychology Review*, *42*(2), 140-156.

Glass, G. V., Peckham, P. D. & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*, 237-288.

Greenberg, D., Ehri, L. C., & Perin, D. (1997). Are word-reading processes the same or different in adult literacy students and third – fifth graders matched for reading level? *Journal of Educational Psychology*, *89*(2), 262-275. doi:

Greenberg, D., Ehri, L. C., & Perin, D. (2002). Do adult literacy students make the same word-reading and spelling errors as children matched for word-reading age? *Scientific Studies of Reading*, *6*(3), 221-243. doi: 10.1207/s1532799xssr0603_2

Hoaglin, D. C., & Iglewicz, B. (1987). Fine tuning some resistant rules for outlier labeling. *Journal of American Statistical Association*, *82*, 1147-1149.

Hutzler, F., & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain and Language*, *89*, 235-242.

Hyönä, J. (1995). An eye movement analysis of topic-shift effect during repeated reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1365-1373. doi: 10.1037/0278-7393.21.5.a

Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica, 73*(3), 259-280. doi: 10.1016/0001-6918(90)90026-c

Joseph, H. S. S. L. & Liversedge, S. P. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLoS ONE*, *8*(1): e54141. doi: 10.1371/journal.pone.0054141

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*(5), 708-723. doi: 10.1080/17470210701400657

Joseph, H. S. S. L., Liversedge, S. P., Blythe, H. I., White, S. J., & Rayner, K. (2009). Word length and landing position effects during reading in children and adults. *Vision Research*, *49*, 2078-2086. doi: 10.1016/j.visres.2009.05.015

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354. doi: 10.1037/0033-295x.87.4.329

Kaakinen, J., & Hyönä, J. (2007). Perspective effects in repeated reading: An eye movement study. *Memory & Cognition, 35*(6), 1323-1336. doi: 10.3758/bf03193604

Kirkby, J. A., Webster, L. A. D., Blythe, H. I., & Liversedge, S. P. (2008). Binocular coordination during reading and non-reading tasks. *Psychological Bulletin*, *134*(5), 742-763. doi: 10.1037/a0012979

Kŭcera, H., and Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Kuhn, M. (2004). Helping students become accurate, expressive readers: Fluency instruction for small groups. *The Reading Teacher, 58*(4), 338-344. doi: 10.1598/rt.58.4.3

Kuperman, V., & Van Dyke, J. A. (2011a). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, *65*, 42-73. doi: 10.1016/j.jml.2011.03.002

Kuperman, V., & Van Dyke, J. A. (2011b). Individual differences in visual comprehension of morphological complexity. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the Cognitive Science Society* (pp. 1643-1648). Austin, TX: Cognitive Science Society.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293-323. doi: 10.1016/0010-0285(74)90015-2

Martens, B. K., Eckert, T. L., Begeny, J. C., Lewandowski, L. J., DiGennaro, F. D., Montarello, S. A., . . . Fiese, B. H. (2007). Effects of a fluency-building program on the reading performance of low-achieving second and third grade students. *Journal of Behavioral Education, 16*, 39-54. doi: 10.1007/s10864-006-9022-x

Meyer, M., & Felton, R. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49*(1), 283-306. doi: 10.1007/s11881-999-0027-8

National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Pavlidis, G. T. (1981). Do eye movements hold the key to dyslexia? *Neuropsychologia*, *19*(1), 57-64. doi: 10.1016/0028-3932(81)90044-0chace

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology, 49*(2), 151-173. doi: 10.1037/1196-1961.49.2.151

Rasch, D., Teuscher, F., Volker, G. (2007). How robust are tests for two independent samples? *Journal of Statistical Planning and Inference*, *137*(8), 2706-2720. doi: 10.1016/j.jspi.2006.04.011

Rasinski, T. V., Padak, N. D., McKeon, C. A., Wilfong, L. G., Friedauer, J. A., & Heim, P. (2005). Is reading fluency a key for successful high school reading? *Journal of Adolescent & Adult Literacy*, *49*(1), 22-27. doi: 10.1598/JAAL.49.1.3

Rayner, K. (1983). The perceptual span and eye movement control during reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 97-139). New York: Academic Press.

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology, 41*(2), 211-236. doi: 10.1016/0022-0965(86)90037-8

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422. doi: 10.1037/0033-2909.124.3.372

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

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. doi: 10.1207/s1532799xssr1003_3

Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. F. Lorch, Jr. & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 9-35). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review*, *33*, 110-149. doi: 10.1016/j.dr.2013.03.001

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Buhner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*, 147-151. doi: 10.1027/1614-2241/a000016

Schnitzer, B. S., & Kowler, E. (2006). Eye movements during multiple readings of the same text. *Vision Research, 46*(10), 1611-1632. doi: 10.1016/j.visres.2005.09.023

Shebilske, W. L., & Fisher, D. F. (1980, December). *Eye movements reveal components of flexible reading strategies*. Paper presented at the Annual Meeting of the National Reading Conference, San Diego, CA.

Spache, G. (1953). A new readability formula for primary-grade reading materials. *Elementary School Journal*, *55*, 410-413.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading. *Remedial and Special Education, 25*(4), 252-261. doi: 10.1177/07419325040250040801

Therrien, W. J., Gormley, S., & Kubina, R. M. (2006). Boosting fluency and comprehension to improve reading achievement. *Teaching Exceptional Children, 38*(3), 22-26.

Thompkins, A. C., & Binder, K. S. (2003). A comparison of the factors affecting reading performance of functionally illiterate adults and children matched by reading level. *Reading Research Quarterly*, *38*(2), 236-258. doi: 10.1598/RRQ.38.2.4

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2003). National assessment of adult literacy. Retrieved from http://nces.ed.gov/naal/kf_demographics.asp

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2014). Table 221.12: Average National Assessment of Educational Progress (NAEP) reading scale score and percentage of students attaining selected NAEP reading achievement levels, by selected school and student characteristics and grade: Selected years, 1992 through 2013. In *Digest of Education Statistics* (2014 ed.). Retrieved from http://nces.ed.gov/programs/digest/d14/tables/dt14_221.12.asp

Zang, C., Liang, F., Bai, X., Yan, G., & Liversedge, S. P. (2013). Interword spacing and landing position effects during Chinese reading in children and adults. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 720-734. doi: 10.1037/a0030097

Zawoyski, A. M., Ardoin, S. P., & Binder, K. S. (2014). Using eye tracking to observe differential effects of repeated readings for second-grade students as a function of achievement level. *Reading Research Quarterly*, *50*(2), 171-184. doi: 10.1002/rrq.91

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Cambridge, MA: Addison-Wesley Press.

Table 3.1
Target Word Frequencies and Lengths

High-Frequency				Low-Frequency			
Word	U (Child)	U (Adult)	Length	Word	U (Child)	U (Adult)	Length
loved	88–122	110.33	5	composed	0–5	1.90	8
enjoyed	22–38	18.16	7	drummed	.5335	.33	7
practice	48–66	45.69	8	excel	.9545	.73	5
planned	33–53	31.84	7	selected	0–9	4.88	8
happening	46–57	90.55	9	chatted	.7341	.37	7
believed	35–98	35.06	8	rehearsing	.4744	3.55	10
thought	810–1130	808.47	7	performed	0–9	7.31	9
entered	26–68	14.65	7	applauded	0–2	.53	9

Note. Child-level frequency norms are statistics specific to Grades 2 through 5 from *The Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995). Adult-level frequency norms were obtained using the $SUBTLEX_{US}$ corpus (Brysbaert & New, 2009).

Table 3.2

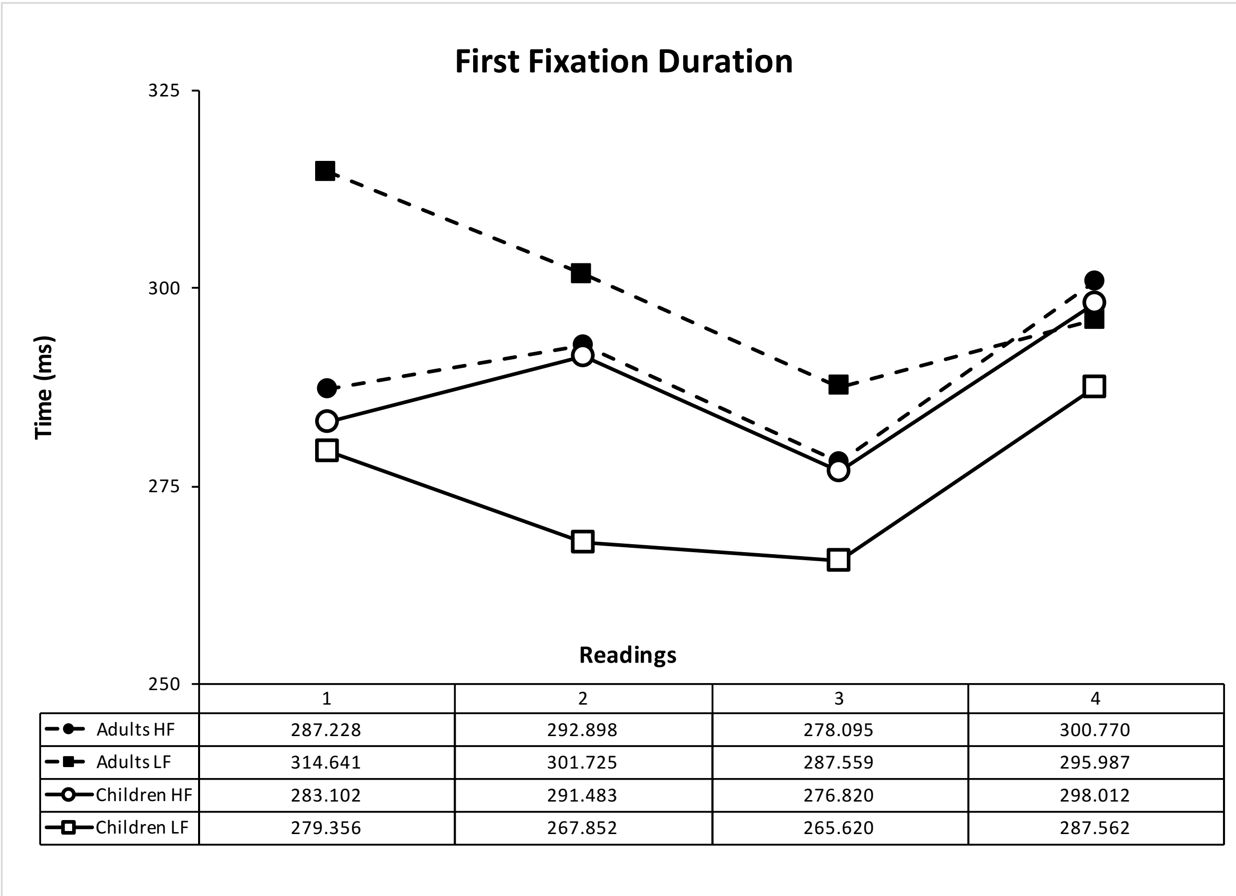
Summary of Global Eye Movement Measures across Readings by Age Group

Measure	Reading 1 *M* (*SD*)	Reading 2 *M* (*SD*)	Reading 3 *M* (*SD*)	Reading 4 *M* (*SD*)
First Fixation Duration (ms) ($N = 67$)	287 (47)	281 (51)*	281 (54)	283 (53)
Adults ($n = 34$)	291 (51)	282 (53)	280 (55)	284 (53)
Children ($n = 33$)	283 (43)	279 (50)	282 (53)	282 (53)
Main Effect (RR): $F(2.46, 159.91) = 3.07$, $p = .039$, $\eta^2 =$				
Between Groups: $F(1, 65) = .04$, $p = .838$, $\eta^2 = .0006$				
Interaction (RR × Age): $F(2.46, 159.91) = 1.64$, $p = .191$, $\eta^2 = .0235$				
Gaze Duration (ms) ($N = 67$)	385 (89)	369 (112)*	360 (92)*	362 (97)*
Adults ($n = 34$)	371 (88)	351 (105)	343 (88)	346 (92)
Children ($n = 33$)	399 (89)	387 (117)	377 (94)	378 (100)
Main Effect (RR): $F(3, 195) = 8.95$, $p < .001$, $\eta^2 = .1207$				
Between Groups: $F(1, 65) = 2.05$, $p = .157$, $\eta^2 = .0305$				
Interaction (RR × Age): $F(3, 195) = .231$, $p = .874$, $\eta^2 = .0031$				
Total Fixation Time (ms) ($N = 67$)	528 (190)	491 (189)*	471 (183)*	470 (174)*
Adults ($n = 34$)	478 (186)	439 (174)	424 (173)	422 (161)
Children ($n = 33$)	579 (182)	544 (192)	519 (184)	519 (175)
Main Effect (RR): $F(2.67, 173.55) = 14.67$, $p < .001$, $\eta^2 =$				
Between Groups: $F(1, 65) = 5.69$, $p = .020$, $\eta^2 = .0805$				
Interaction (RR × Age): $F(2.67, 173.55) = .11$, $p = .942$, $\eta^2 = .0014$				
Number of Inter-word Regressions (#) ($N = 67$)	.304 (.157)	.295 (.143)	.262 (.136)*	.270 (.147)*
Adults ($n = 34$)	.229 (.130)	.234 (.133)	.211 (.131)	.207 (.136)
Children ($n = 33$)	.382 (.144)	.357 (.128)	.314 (.122)	.336 (.129)
Main Effect (RR): $F(2.25, 145.95) = 8.59$, $p < .001$, $\eta^2 =$				
Between Groups: $F(1, 65) = 18.02$, $p < .001$, $\eta^2 = .2170$				
Interaction (RR × Age): $F(2.25, 145.95) = 2.25$, $p = .103$, $\eta^2 = .0293$				

Table 3.2 Continued

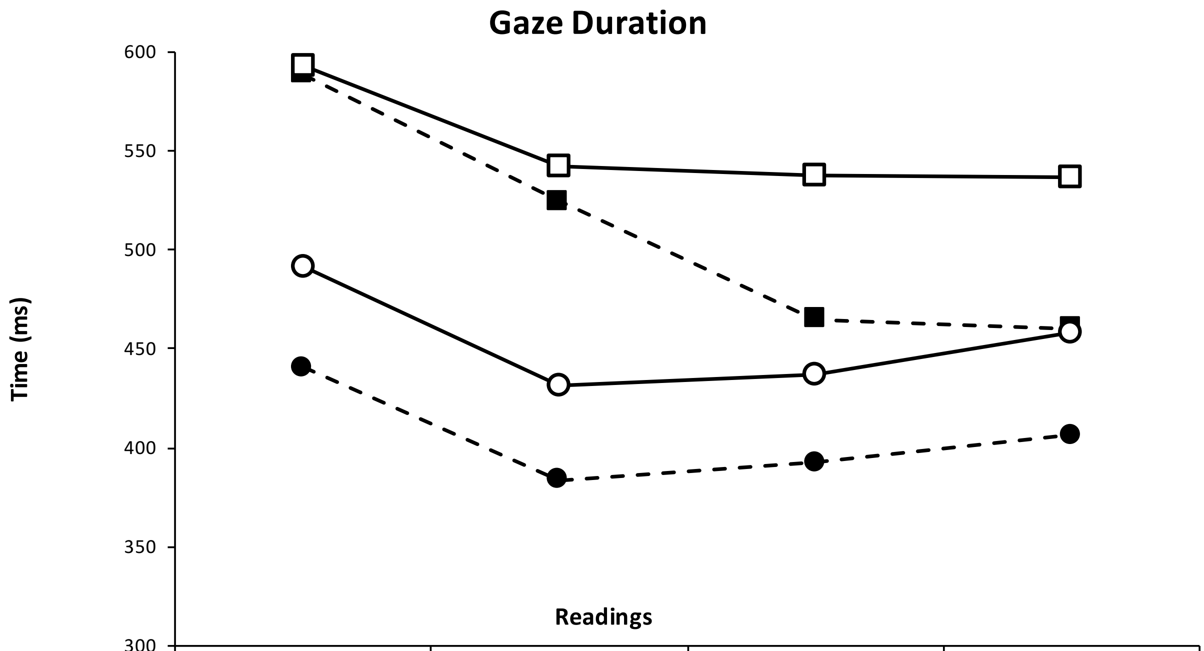
Average Number of Fixations per Word (#) (*N* = 67)	1.514 (.557)	1.384 (.594)*	1.235 (.525)*	1.224 (.478)*
Adults (*n* = 34)	1.349 (.465)	1.214 (.528)	1.112 (.455)	1.082 (.427)
Children (*n* = 33)	1.683 (.598)	1.560 (.614)	1.361 (.569)	1.369 (.489)
Main Effect (RR): $F(2.76, 179.28) = 17.78$, $p < .001$, $\eta^2 =$				
Between Groups: $F(1, 65) = 7.07$, $p = .010$, $\eta^2 = .0981$				
Interaction (RR × Age): $F(2.76, 179.28) = .47$, $p = .689$, $\eta^2 = .0056$				

*Significant pairwise differences between the denoted reading and the first reading, $p < .05$



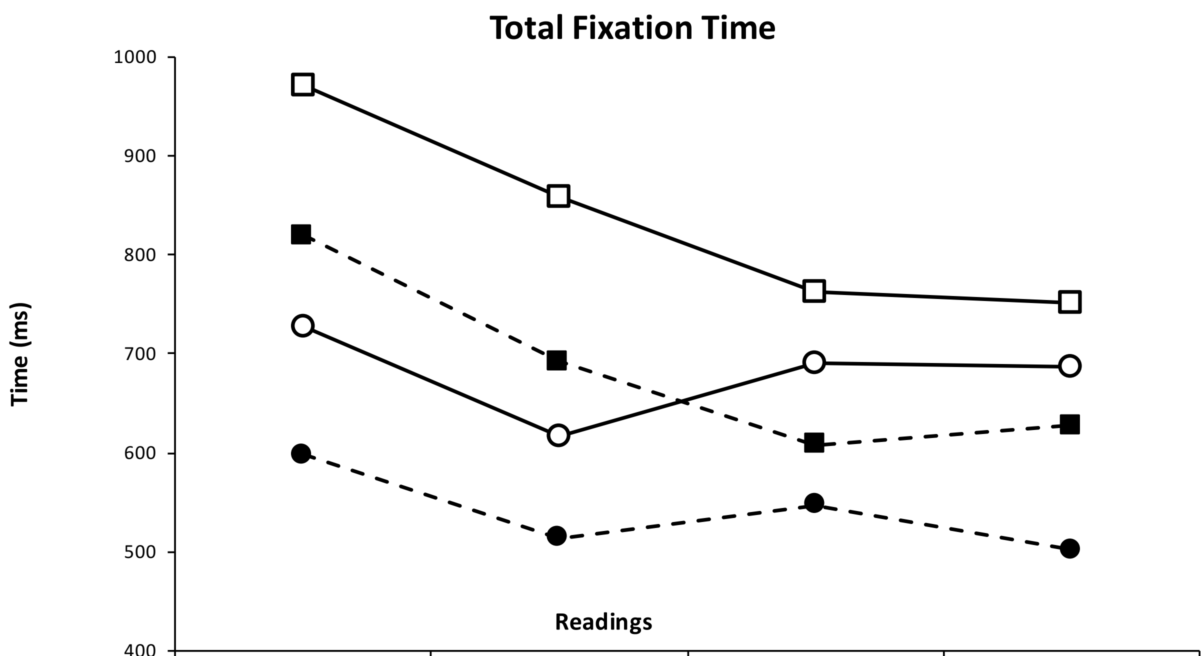
	1	2	3	4
Adults HF	287.228	292.898	278.095	300.770
Adults LF	314.641	301.725	287.559	295.987
Children HF	283.102	291.483	276.820	298.012
Children LF	279.356	267.852	265.620	287.562

Figure 3.1. Average first fixation duration on high- and low-frequency target words across readings, separated by participant age group.



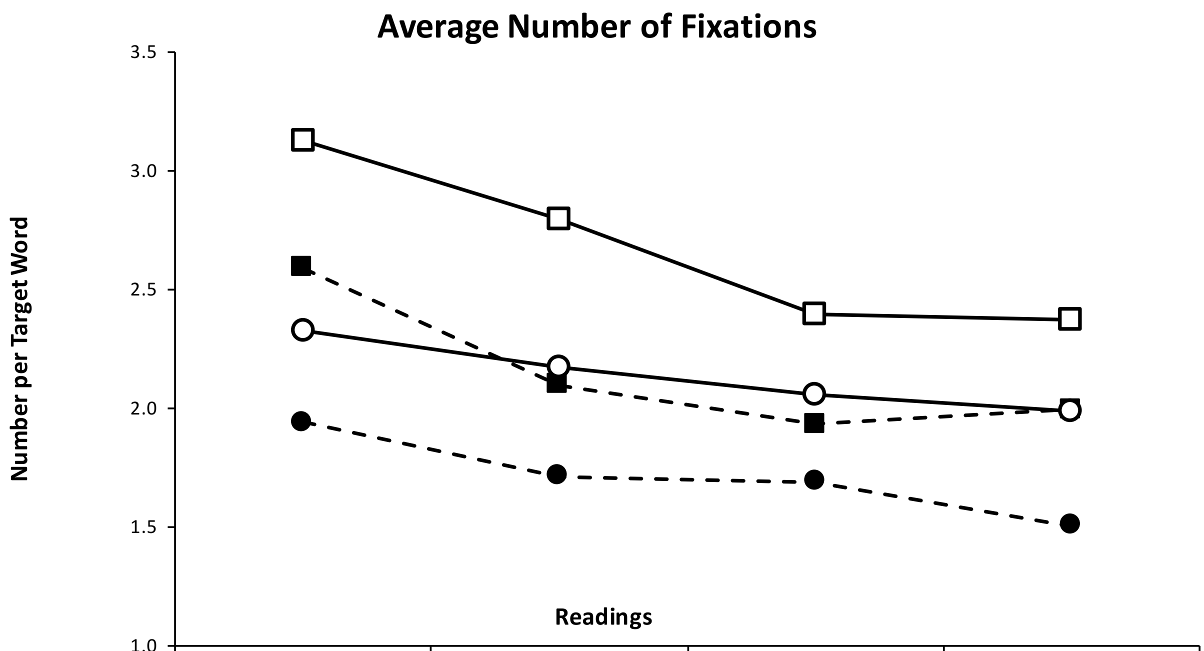
	1	2	3	4
Adults HF	440.588	383.571	392.396	406.189
Adults LF	588.437	524.281	464.934	460.201
Children HF	491.090	431.859	436.594	458.372
Children LF	592.843	542.613	537.627	536.951

Figure 3.2. Average gaze duration on high- and low-frequency target words across readings, separated by participant age group.



	1	2	3	4
Adults HF	598.095	514.277	547.366	501.655
Adults LF	819.504	692.412	608.028	627.212
Children HF	728.111	616.369	690.627	687.050
Children LF	971.963	858.968	763.433	752.233

Figure 3.3. Average total fixation time on high- and low-frequency target words across readings, separated by participant age group.



	1	2	3	4
Adults HF	1.941	1.713	1.691	1.504
Adults LF	2.588	2.099	1.934	1.993
Children HF	2.330	2.170	2.061	1.989
Children LF	3.125	2.799	2.398	2.371

Figure 3.4. Average number of fixations (average fixation count) on high- and low-frequency target words across readings, separated by participant age group.

CHAPTER 4

GENERAL CONCLUSION

The overall purpose of this two-study dissertation was to address substantial gaps in the existing literature and to extend upon and clarify recent eye movement research by investigating (a) the technical adequacy of recently utilized eye movement measures and (b) potential differences in the observed rereading behavior of similarly skilled adult and child readers. Findings from Study 1 suggest that recent eye movement measures of children's reading, particularly passage-level measures of fixation duration, demonstrate stability over time and consistency across passages and capture a considerable degree of variability in elementary students' levels of reading fluency and broad reading achievement. Findings from Study 2 suggest that repeated practice improves the efficiency of both word- and passage-level reading behavior for unskilled, elementary-level readers at a variety of ages. Results from both studies also indicate notable changes in reading behaviors relative to specific target words (especially low-frequency words), such that eye movements on these words are more likely to reflect subtle changes in processing (e.g., facilitation as a function of rereading) and demonstrate lesser reliability and validity compared to global eye movements.

Taken together, results from both studies affirm the link between students' eye movements during reading and their level of reading skills (e.g., broad reading achievement, as measured by more traditional assessment tasks). Estimates of concurrent criterion-related validity from Study 1 do so quite explicitly, indicating significant moderate to high correlations between measures of fixation duration/count and standardized measures of reading fluency, basic reading, comprehension, and overall reading skills. Findings from Study 2, though not involving similar numerical associations, also suggest that students' reading behaviors—and the manner in which these change in response to instructional procedures—ultimately reflect and are influenced by their levels of reading skill and achievement, even alongside variability in age.

Future research with larger and more diverse participant samples is needed to enhance generalizability of current results to a wider population and applied educational settings. However, despite the limitations discussed in the above manuscripts, the current studies inspire a higher degree of confidence in recent eye movement research and suggest potentially greater utility in employing eye-tracking methodology to explore the reading behavior of a wider range of individuals (e.g., adult literacy learners). In sharp contrast with early cognitive eye-tracking research, which simply focused on the "end state" of successful reading among skilled adult readers, current eye movement studies offer the educational community increased potential for better examining and understanding the development of reading skills and how reading behaviors relate to and are impacted by modern-day assessment and instructional procedures. This exciting progress suggests that the field of applied eye movement research, while still young and affected by several flaws and limitations, can offer unique insight into previously unstudied reading behavior.

APPENDIX A

Experimenter-Developed Reading Passage

Martin liked playing music with his band. He played the guitar and was very good at it. He also composed new songs and drummed with his band. He loved making music and enjoyed sharing it with others. He hoped he could get a good job making music.

Martin and his band were going to play a show. They had to practice because it was one week away. Many people would be there to listen to their music. They wanted to excel at playing. First, Martin selected the songs that they would play. Next, he planned the set list with all the people in the band. They needed to do a lot before the show.

The next day, the band members came to Martin's house. They chatted about when the show was happening and who would be there. Then they spent several hours rehearsing their songs. Martin was happy that everyone was doing their best. He believed that they sounded very good. The people coming to the show would like what they would hear.

On the day of the show, Martin was nervous. The band had never played for so many people before. He thought that they would sound good. All the people in the band were ready for the show. When it was their turn, they entered onto the stage. They performed their first song, and the crowd liked it. When it was over, they applauded for Martin and his band. Everyone had a nice time.

APPENDIX B

Common Definitions of Eye Movements and Dependent Measures

(e.g., Rayner et al., 2006)

Average number of fixations per word: fixation count divided by number of words of interest (e.g., number of words in a passage, number of target words)

Average proportion of words initially skipped: number of words on which no fixations were made during the first pass of reading divided by number of words of interest (e.g., number of words in a passage)

First fixation duration: the duration of the first fixation on a word, regardless of the number of fixations made on the word

Fixations: pauses on words that allow readers to extract information from fixated points and surrounding areas of visual acuity

Fixation count: total number of fixations

Gaze duration: the sum of all of the fixations made on a word prior to movement to another word

Regressions: backward saccadic movements that typically reflect additional processing of previously identified text or correction for "overshooting" eye movements

Saccades: rapid movements in which the eyes move from one point to another and vision is suppressed

Total fixation time: the sum of all fixations, including regressions, on a word

Number of inter-word regressions: total number of regressions between words (versus within words)