

The Effect of Reading Purpose on Semantic Inferencing About New Words

by

Marielle Christine Snyder

BS Speech Pathology and Audiology, Kent State University, 2019

Submitted to the Graduate Faculty of the
School of Health and Rehabilitation Sciences in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This thesis was presented

by

Marielle Christine Snyder

It was defended on

March 12, 2021

for approval by

Michael Walsh Dickey, PhD, Associate Professor and PhD Program Director, Department of
Communication Science and Disorders

Erin Lundblom, PhD, CCC-SLP, Associate Professor and Clinical Education Coordinator for
Speech Language Pathology, Department of Communication Science and Disorders

Thesis Advisor: Dawna Duff, PhD, Assistant Professor, Department of Communication Science
and Disorders

Copyright © by Marielle Christine Snyder

2021

The Effect of Reading Purpose on Semantic Inferencing About New Words

Marielle Christine Snyder, BS

University of Pittsburgh, 2021

RATIONALE: Vocabulary is critical to academic and social outcomes. Most vocabulary is learned incidentally, without direct instruction. After children become skilled at reading words (by about grade four), this incidental word learning typically happens while reading. However, readers differ in how easily they can infer new word meanings from context. Reading purpose, or the reader's goals for comprehension of the text, affects reading comprehension, and is also thought to affect the quality of this semantic inferencing. Reading instruction to improve fluency, which is common in schools, may shift the reader's focus from text comprehension to speed and accuracy. It is unknown if these instructions to read quickly and accurately have the unintended effect of negatively impacting word learning while reading.

METHODS: This study examines data from a between-subjects study of middle school children who were instructed to read passages with embedded nonwords under one of two conditions: reading for comprehension or reading for speed and accuracy. Eye-tracking data was collected during the task to reflect the readers' online interactions with the text and, particularly, the nonwords within the text. Post-test behavioral measures were also collected to reflect the quality of semantic inferencing.

RESULTS: A linear mixed effects model found no significant differences between the two conditions in performance on the post-test language outcome measures. Similarly, eye-tracking data did not reveal any significant statistical differences between participant groups in measures of either active, reader-initiated reading processes or passive reading processes.

DISCUSSION: Analyses of the present data do not offer support for the theory that a priority on reading speed and accuracy sacrifices the quality of incidental word learning. Further research should investigate individual differences in response to fluency-related instructions.

Table of Contents

Preface.....	x
1.0 Introduction.....	1
1.1 Vocabulary and Reading	2
1.1.1 Vocabulary Acquisition	3
1.1.2 Reading for Vocabulary Learning.....	4
1.2 Semantic Inferencing from Text	5
1.2.1 Eye Movements and Incidental Learning	6
1.2.2 Semantic Inferencing as a Complex, Variable Process	8
1.2.3 Reader Variables Affecting Semantic Inferencing.....	9
1.2.4 External Variables Affecting Semantic Inferencing	10
1.2.5 Reading Purpose and Semantic Inferencing	11
1.3 Fluency and Reading Purpose.....	12
1.3.1 An Example	13
1.3.2 Rationale for Current Study	14
1.4 Precursors to Present Study	14
1.5 Study Purpose	16
2.0 Methods.....	18
2.1 Sample	18
2.2 Procedures.....	20
2.2.1 Nonword Selection	21
2.3 Outcome Measures	22

2.3.1 Quality of Hypotheses About the Semantics of Novel Words	22
2.3.1.1 Definitions of Nonwords	22
2.3.1.2 Context Test Questions.....	23
2.3.2 Eye Movement During Reading.....	24
2.4 Analytic Plan	25
3.0 Results	26
3.1 Primary Research Question.....	26
3.2 Secondary Research Question	30
4.0 Discussion.....	35
4.1 Limitations	38
4.2 Conclusion	38
Bibliography	40

List of Tables

Table 1 Descriptive Statistics for CC and FC Conditions.....	20
Table 2 Group Mean Performance on Language Measures.....	27
Table 3 Language Measures LMM Output, Condition Fixed Effect.....	28
Table 4 Language Measures LMM Output, Condition and PPVT-IV Fixed Effects.....	29
Table 5 First Fixation Duration LMM Output	31
Table 6 Gaze Duration LMM Output.....	32
Table 7 Regression Path Duration LMM Output.....	33
Table 8 Dwell Time LMM Output	33

List of Figures

Figure 1 Representation of Eye Movement Patterns for Encountering New Words (Duff & Davidson, 2019, July).....	7
Figure 2 Recognized Eye Movement Measures (Duff & Davidson, 2019, July)	7
Figure 3 Criteria for Coding Participant Nonword Definitions (Duff, 2015)	23
Figure 4 Between-Group Mean Language Measure Performance	27
Figure 5 Mean Language Measure Performance Differentiated by PPVT-IV Performance	30

Preface

I extend my most sincere gratitude to Dr. Dawna Duff, the advisor for this project and an excellent mentor throughout my time here at the University of Pittsburgh. Additionally, I am incredibly grateful for the contributions of committee members Dr. Michael Walsh Dickey and Dr. Erin Lundblom. The time, effort, and enthusiasm the entire committee has donated to this project is immensely appreciated, especially in light of the multifaceted challenges related to the COVID-19 pandemic.

I would like to also thank the many faculty from the Department of Communication Sciences and Disorders who made a special effort to demonstrate their ongoing support of me and my research endeavors. In addition, I am endlessly grateful to my undergraduate research mentor and ongoing collaborator, Dr. Hayley S. Arnold, whose mentorship has been invaluable and whose words of encouragement continue to propel me in my pursuit of research.

In addition to academic mentors and contributors, successful completion of this project relied on the unfailing support of several close friends, fellow classmates, and family members. A special word of thanks goes to my parents, who continually exemplify hard work and demonstrate an unfailing support for my academic endeavors. I am eternally grateful to all those who have supported my growth as a student and researcher, both for the purposes of this project and beyond.

1.0 Introduction

Literacy development follows a typical developmental timeline in that there is a typical pattern for progression of literacy skill development from birth and continuing throughout the lifetime. One commonly accepted timeline for literacy development is that proposed by Chall (1983), in which stages of typical literacy development are outlined based on grade level. Prior to grade 3, there is a focus on building skills for reading (i.e., learning to read). Around grade 3, the focus switches to a reliance on reading to learn. After this point in literacy development, students must use those initial literacy skills to appropriately access academic information and to continue their literacy development. This research study concerns readers who are in the “reading to learn” phase of literacy.

In accordance with this and similar theories on reading development, the National Reading Panel’s published report (2000) officially recognized five primary components for reading: fluency, vocabulary, phonemic awareness, and phonics, with comprehension as the fifth and most pervasive component of reading. These components interact throughout literacy development, acting together to produce the holistic process and product of reading.

For this study about semantic inferencing for new word learning, there is particular interest in the domains of fluency and vocabulary, particularly in how they combine to create comprehension. Reading fluency, a combination of speed and accuracy in reading (LaBerge & Samuels, 1974), often finds its place at the focus of literacy instruction and, therefore, intervention. Like fluency, vocabulary often finds function at the core of reading instruction. Vocabulary, or the word-level knowledge that contributes to reading, is another facet of literacy that is understood to engage in developmental patterns, particularly as it contributes to and benefits from reading. As

proposed, these facets, along with phonemic awareness and phonics, contribute to produce reading comprehension. These factors are all integral components of reading and, therefore, reading instruction and intervention.

In the next section, I will discuss the intersection between reading and vocabulary, highlighting the importance of reading for vocabulary development. Then, in a brief exploration of semantic inferencing, I explore sources of variation in semantic inferencing that may impact the quality of word-level semantic deduction that occurs while reading a given text. Reading purpose is examined as a variable affecting word inferencing ability, with particular interest in the impact of maintaining fluency as a reading goal. Finally, I will conclude by introducing the methods used for the present study and discussing our research aims and methods in light of the literature discussed.

1.1 Vocabulary and Reading

To further examine the relationship between vocabulary acquisition and reading, we can look at the reading systems framework by Perfetti and Stafura (2014). The framework is a model of systems contributing to reading comprehension, which is recognized for the complexity it holds. Specifically, lexicon (i.e., word-level knowledge) is posited as a pivotal component for reading comprehension. Additionally, Perfetti and Stafura suggest that not only is lexicon important for informing comprehension processes, but also that these comprehension processes and inferencing are essential in building lexicon. The bidirectional relationship between lexicon and higher-level processes, such as inferencing, can be further explained by the Reciprocal Hypothesis theory, as referenced in Elleman, Lindo, Morphy, and Compton (2009). The theory suggests that reading

comprehension and vocabulary have a reciprocal causal relationship that propagates the recurrent negative loop that has been reported between atypical vocabulary acquisition and developing reading comprehension skills. Vocabulary knowledge supports reading comprehension, but reading comprehension is key for incidental vocabulary learning. This bidirectional relationship is important because it could create a 'vicious cycle' for some learners and a 'virtuous cycle' for others, a pattern notably described by “Matthew Effects” (Stanovich, 1986). An increasing reliance on written modalities compounds any related deficits (i.e., reading comprehension, vocabulary acquisition) as the child moves throughout their school-age years. Academic instruction and assessment become increasingly reliant on written modalities, making the effect of these impairments increasingly poignant (Catts, Hogan, & Adlof, 2005). Good readers get better, while poorer readers continue to fall (relatively) behind. Such important implications vitalize interest in word knowledge in relation to comprehension of a text.

1.1.1 Vocabulary Acquisition

Consideration of vocabulary acquisition is vital to an understanding of the implications of semantic inferencing for literacy development. Acquiring a new vocabulary term is not equivalent to simple rote learning of a unidimensional definition, but rather amassing a conceptual understanding of the meaning (semantics) of the word, the form (phonology) of the word, and the association between the two. We know that vocabulary acquisition starts well before a child begins reading, but for the purpose of this study we will be looking particularly at vocabulary acquisition after that developmental point, when reading is believed to play a central role in word learning.

In the model described in Perfetti and Hart (2002), vocabulary knowledge is depicted as the confluence of the orthographic, phonologic, and semantic knowledge of a word. In this

theoretical framework, orthography and phonology would constitute the word form, with semantics accounting for the word meaning. This semantic content involved in conceptualization of new words is of particular interest to the present study of semantic inferencing for new word learning.

Acquisition of new vocabulary can occur through either direct or incidental learning routes. Direct instruction of vocabulary implies explicit instruction about word form and meaning, such as would take place in an English, science, or social studies class. However, vocabulary learning is often done incidentally, or without direct instruction. In fact, vocabulary acquisition is believed to occur primarily through incidental presentations of novel words (Christ & Chiu, 2018; Clark, 2009; Fernald, Marchman, & Weisleder, 2013). In incidental learning, the semantic properties of the word must be inferred through repeated exposures and interactions with that word in context. This can happen through either auditory exposure (i.e., hearing the word used) or written exposures (i.e., reading the word in context). With many words, and especially those which are more common, both auditory and written exposures will contribute to this inferencing. However, many lower frequency words are only, or nearly always, encountered while reading text (Cunningham & Stanovich, 1998). In this case, the individual must rely on text-based inferencing skills to appropriately infer word meaning and, therefore, improve comprehension of the text. For the purpose of this study, the focus is on incidental vocabulary acquisition in written contexts.

1.1.2 Reading for Vocabulary Learning

The Chall (1983) timeline, referenced previously in this paper, has implications as important for the domain of vocabulary development as it does the broader literacy developmental timeline it describes. The progression outlined by this timeline highlights a pivotal shift in reading

instruction, in which children become more reliant on reading for continued literacy development and academic success. As mentioned, this transition is colloquially referred to as the shift from *learning to read* into *reading to learn*. A study of semantic inferencing of novel words from text, such as the current study, assumes a focus on reading processes after this instructional shift, when semantic inferencing plays a newly prominent role in effective text comprehension and ongoing vocabulary development. In the current study, participants were in 6th grade.

With a shift into reading to learn, academic texts typically begin to incorporate more low-frequency vocabulary. Therefore, written exposure to new words becomes crucial in that it offers a unique depth and breadth of vocabulary not found at the same level through conversation and other spoken contexts (Cunningham & Stanovich, 1998). Both for success in continued literacy development and, likewise, success in the academic setting, effective inferencing for new words in written contexts is essential.

1.2 Semantic Inferencing from Text

Exposures to new words in text-based settings become valuable to a reader's vocabulary when *semantic inferencing* is enacted to deduce that word's meaning through its surrounding context. In other words, to effectively process novel words in text, a reader must engage in *semantic inferencing* about that word. For this to happen, the reader must engage in a variety of complex processing skills. For instance, before semantic inferencing can take place, the word must be read and recognized as an unknown word. Further, if the child uses cognitive resources trying to infer the word's meaning, the child may spend additional time processing the word, or return to the word and its surrounding context to make inferences about the missing information.

To more effectively broach the processes involved in semantic inferencing from text, models proposed by Van den Broek and Helder (2017) may offer valuable insight. The authors' diagrammatic depiction of this model (Van den Broek & Helder, 2017, p. 363) illustrates that both passive and active or reader-initiated processes combine to form the mental representation that is reading comprehension. Activation of these reader-initiated processes (such as inferencing) is moderated by reader's *standards of coherence*. The article refers to standards of coherence as the level at which a reader aims to understand the text, usually on a subconscious level. These standards are subject to change, both across readers (based on individual differences in reading ability) and reading tasks (based reading goals). Different reading goals for a text will affect how well the reader is trying to understand that text.

In this model, passive processes drive reading. As illustrated in the Van den Broek and Helder model (2017, p. 363), these passive processes (i.e., subconscious, associative processes to enact text and semantic memory for text comprehension) combine with active, or reader-initiated, processes. When comprehension is not sufficient to satisfy the standards of coherence for the text using just passive processes (e.g., if comprehension is impeded by the presence of unknown words in the text), the reader engages active processes to improve comprehension until that standard is met. Thus, reader-initiated processes do not always occur, but when they do, they work to improve text comprehension beyond what is achieved with passive processes alone.

1.2.1 Eye Movements and Incidental Learning

One challenge with the constructs of active and passive cognitive processes in the Van den Broek and Helder (2017) model is that they are challenging to operationalize. In this study, we do so by examining eye movements while reading novel words. Eye movements are thought to be an

advantageous reflection of online processes for reading when compared to more behavioral measures (e.g., post-text comprehension questions; Godfroid et al., 2018; Rayner, Chace, Slattery, & Ashby, 2009). That is, examining eye movements offers a unique advantage for the purposes of this investigation because they can give insight into online reading processes.

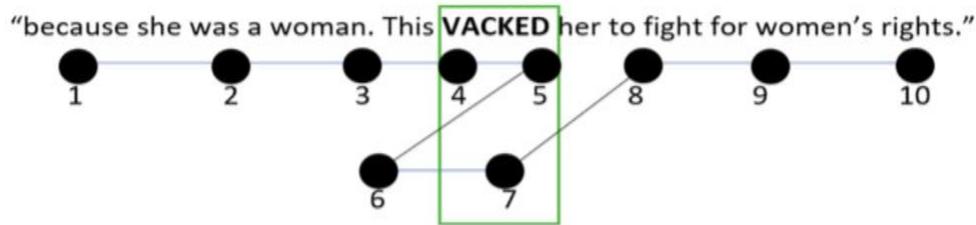


Figure 1 Representation of Eye Movement Patterns for Encountering New Words (Duff & Davidson, 2019, July)

Processing Type	Eye Movement Measure	Fixation Numbers
Passive	First fixation duration	4
	Gaze duration	4 + 5
Reader-initiated	Regressions-in (count)	7
	Regressions-out (count)	6
	Regression path duration (go past time)	4 + 5 + 6 + 7
	Dwell time (total reading time)	4 + 5 + 7

Figure 2 Recognized Eye Movement Measures (Duff & Davidson, 2019, July)

This approach takes advantage of an extensive body of previous work about the types of cognitive processes which map onto different eye movements. There is a substantial body of literature showing that different eye movement measures are associated with different types of processing. For example, fixation durations for a word are longer for less frequent, polysemous, or morphologically complex words (for a review, see Holmquist et al., 2011), which presumably

require more time for lexical access (Rayner, 1998). In contrast, dwell time (fixation numbers 4+5+7; see Figure 2) is sensitive to higher level (and slower) processes, such as difficulty integrating meaning across several words (Rayner, 1998).

The specific eye movements evaluated in this case will be based on previous work by Duff and Davidson (2019, July), as seen illustrated in Figures 1 and 2. Use of these measures is also consistent with more recent work on eye movements in children while reading novel words (Pagán, Bird, NHsiao, & Nation, 2020). Specifically, analysis will focus on measures of first fixation duration, gaze duration, regression path duration (i.e., go past time), and dwell time. Our use of the words “active” and “reader-initiated” processes to relate to these eye movements is novel. However, it is grounded in extensive previous research about the nature of these eye movements.

1.2.2 Semantic Inferencing as a Complex, Variable Process

Semantic inferencing is the result of a complex interaction of processes, and the quality of inferencing is subject to change across individuals, across texts, and across a multitude of other factors. For evaluating the complexity of semantic inferencing processes, *reader variables* can be considered factors about the child that affect the quality of these inferencing efforts. For instance, Daneman (1988) proposed that an individual’s processing capacity may be a key determinant in ability to learn from context, as further explored below. Similarly, children with poor language skills (i.e., poor readers) are thought to be less likely to effectively infer new word meaning from context (Shefelbine, 1990; Swanborn & de Glopper, 2002) . However, in addition to these factors varying from individual-to-individual are those that are not internal to the reader, called *external variables*. These features can include features of the given text (e.g., text difficulty,

number of new words, etc.), or they can be environmental factors that affect how we interact with the text (e.g., expectations set for reading task, supports, cognitive load, etc.).

1.2.3 Reader Variables Affecting Semantic Inferencing

Amid the many factors affecting semantic inferencing of new words in a text, one of the primary considerations would of course be that of the individual's reading comprehension skills. Individual differences in reading comprehension seem to be a stable contributor to a child's ability to derive new word meaning in written contexts (Cain, Oakhill, & Bryant, 2004; Cain, Oakhill, & Elbro, 2003; Cain, Oakhill, & Lemmon, 2004). That is to say, children with poorer reading comprehension also show weaker skills in inferring the meanings of new words.

In addition to comprehension level, other reader skills and qualities are thought to impact quality of semantic inferencing. For instance, low semantic knowledge related to a text's topic has been found to impede acquisition of new concepts, such as new-word learning (Duff, 2015; Graves, 1986; Nagy, Anderson, & Herman, 1987). Several cognitive variables have also been proposed as important to the ability to infer the meaning of new words from text. For instance, Daneman (1988) suggests that processing capacity factors into an individual's inferencing skill. In one of the studies described above (Cain, Oakhill, & Lemmon, 2004), working memory capacity correlated to performance on semantic inferencing measures. This effect was specific to situations where the "clues" about the meaning of the novel word were distal to the word itself, such that the processing demands of the task were higher. Similarly, research suggests effective semantic inferencing may be strongly correlated to the use of meta-cognitive skills (Nagy, 2007). One quasi-experimental study (Baumann, Carr Edwards, Boland, Olejnik, & Kame'enui, 2003) found that students who underwent morphemic and contextual analysis instruction, a form of meta-cognitive

instruction, were more successful at inferring the meanings most novel words than participants who received traditional vocabulary instruction. Whether due to inherent individual differences or differing levels of instruction or preparation, the skills and abilities that the reader brings to a text seem to have definite and complex interaction with their ability to perform semantic inferencing.

1.2.4 External Variables Affecting Semantic Inferencing

One of the most overt factors contributing to ability to process text are those relating to the text itself. Variables such as linguistic difficulty can certainly play a role in the quantity and quality of semantic inferencing that may take place. For instance, findings from Cain, Oakhill, and Lemmon (2004), shed some light on the effect of processing demands on effective semantic inferencing skills. In this study, 9- and 10-year-old participants were assessed on semantic inferencing measures for new words in a provided text, with processing demands (in this case, working memory demands) operationalized by controlling the distance of the referent from the relevant novel word. Results supported a relationship between the processing demands of the task and quality of semantic inferencing, with significantly reduced scores among poor comprehenders for novel words with more distant referents.

Additionally, there may be unrelated environmental factors present at the time of reading that impact complex processing, such as semantic inferencing, during reading. Variables as simple as environmental distractions or low light for reading could plausibly affect a reader's cognitive load when approaching a text and, therefore, might affect the quantity or quality of inferencing.

A final and significant factor to consider is that of reading purpose, or the reader's intentions when approaching a text. A reader's goals when approaching text can have multivariate implications on the process of reading (Cheon & Ma, 2014; Van den Broek & Helder, 2017). In

the framework proposed by Van den Broek and Helder (2017), the standards of coherence (i.e., the reader's situation-dependent criteria for adequate comprehension of a text) for a given text are subject to change depending on the reader's goal for reading. Therefore, higher standards of coherence – as may be enacted when given instructions to read with a focus on text comprehension, for example – would imply increased activation of reader-initiated processes to promote inferencing. Findings from Van den Broek, Lorch, Linderholm, and Gustafson (2001), an investigation of the relationship between inference generation and reader goals, support this theorized implication. In the study, college students were instructed to read either for study or entertainment. Measures reflecting both on-line inference generation and off-line memory revealed patterns of inferential activity associated with reading goal. For instance, participants reading for the purpose of study demonstrated more coherence-building inferences, rather than associations or evaluations, than their counterparts reading for entertainment. The results further supported that participant reading goals influenced the readers' standards of coherence and, therefore, quality of inferencing.

1.2.5 Reading Purpose and Semantic Inferencing

Reading purpose may likewise be thought to affect word-level semantic inferencing for novel word learning. Research on word-level inferencing in relation to reading purpose is limited. In one study, Swanborn and de Glopper (2002) examined the nature of reading purpose, specifically in relation to inferencing at the word level. Grade 6 students were asked to read sample text for varying reasons: for fun, for text comprehension, or for learning about the topic of the text. The article reported differences in the proportions of words learned incidentally between the three groups, but noted that the amount of incidental word learning was correlated to reading ability

across all participants, with low-ability readers hardly learning any words incidentally. Further, there was an interaction between the effect of reading purpose and individual differences; for readers with strong comprehension, reading for learning about the topic produced the most novel word learning. For readers with poor comprehension, the best outcomes were for the “reading for fun” condition. This one study indicated that reading purpose can be important to novel word learning while reading, however it does not speak to the specific reading purposes investigated in the current study.

1.3 Fluency and Reading Purpose

In this study, we investigate how a focus on fluency, specifically, affects how children interact with their text for effective semantic inferencing. When examining reading purpose and standards of coherence, reading fluency must also be considered. Fluency (i.e., the ability to read with appropriate speed and accuracy; LaBerge & Samuels, 1974) often holds a key place in reading instruction. Likewise, school-based assessments and training often include measures of fluency.

For instance, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is a widely used instrument designed for school-age children which has been reported to be useful in over 15,000 schools nationally (Strauss, 2014). However, this set of short reading tasks (e.g., Nonsense Word Fluency subtest, Phoneme Segmentation Fluency subtest) has fallen under substantial criticism as a valid assessment of reading quality (Goodman, 2006; Manzo, 2005, September 28; Riedel, 2007) and, an ancillary criticism, reading fluency (which can be argued to mandate consideration of both speed and comprehension; Samuels, 2006, May). Critics of the widespread adoption of DIBELS voice concern over misalignment of reading needs and intervention services;

students with low DIBELS performance but adequate comprehension abilities may receive ultimately unnecessary intervention, while students with high DIBELS scores but poor comprehension skills may pass unidentified as needing services (Riedel, 2007). Moreover, interventions led by assessment results may result in an inappropriate or ineffective focus on building piecemeal fluency skills (e.g., phoneme segmentation) rather than comprehensively building functional reading skills.

A focus on fluency carries implicit goals (i.e., reading speed and accuracy). Therefore, it is of interest to investigate how reading fluency treatment might affect purpose and, therefore, comprehension and new word learning. It is possible that reading faster may inhibit online processes associated with word learning. Such concern vitalizes interest in the role of fluency in academic instruction and assessment (e.g., use of DIBELS, as described above).

1.3.1 An Example

To understand the implications of an interaction between reading purpose and fluency, including their contribution to reading comprehension and literacy development, one might consider a hypothetical example. Consider a grade 5 student who demonstrates reading fluency skills below typical levels as compared to his peers. As a result, this child has been receiving regular fluency intervention over the years and is continually being assessed on parameters of speed and accuracy for reading. Given this situation, the concern is that if the focus placed on speed and accuracy will come at the cost of those higher-level processes while reading, such as semantic inferencing for new word learning. These time-reliant, complex processes may be sacrificed when the focus of reading is instead placed on speed and accuracy

1.3.2 Rationale for Current Study

The present study examines the relationship between reading purpose and semantic inferencing during text-based comprehension activities. If this relationship exists, it may have significant implications in academic contexts. Instruction across several educational styles and activities may prioritize reading speed and fluency. However, encouraging reading speed and fluency may come at the price of vocabulary acquisition. Deficits in inferencing abilities can have negative implications on a child's vocabulary acquisition during school-age years, affecting their overall language abilities and academic success. More research is necessary to investigate the potential relationship between semantic inferencing from context and reading purpose, and the data presented in this study will attempt to address this gap in the literature.

1.4 Precursors to Present Study

Two studies in particular inform the current investigation: Duff and Davidson (2019, July) and Duff and Brydon (2018). These two studies shared identical procedures which led to, and closely resemble, the methods used for the present study. Both studies used a within-subjects design to compare performance across two reading purpose conditions for the same passage with six nonwords. First, participants were asked to read for comprehension (comprehension condition), then to read the passage a second time and with a goal of learning about new words. After reading the passage the second time, participants were asked to provide definitions for new words and respond to a corresponding series of dichotomous (Y/N) questions about new words in the passage.

In addition to the two posttest language measures, eye tracking was used during passage reading to assess eye movements relative to novel words.

Duff and Brydon (2018) sought to investigate which eye movements are associated with higher scores on language measures. Their data from a sample of 19 children ($M=10;4$, $SD=11$ months) produced interesting findings. For the first reading of the text (comprehension condition), there was an association between higher numbers of regressions out (i.e., the number of times looking at the text leading up to the novel word) and a higher vocabulary measure performance after the second reading (i.e., nonword definitions and context test question scores). For the second reading of the text (learning new word meanings condition), a higher total reading time was associated with an increase in language measures. These findings motivated the suggestion that readers made better hypotheses about the meaning of new words when they spent more time looking at the new word or revisiting the new word. Furthermore, different eye movement measures were associated with novel words in the reading comprehension and novel word learning conditions.

Based on the findings of the latter study, Duff and Davidson (2019, July) posed a new question: does reading purpose affect eye movements associated with reader-initiated (i.e., active) or passive cognitive processes? A sample of 27 children in 5th and 6th grade ($M=10;6$, $SD=1$ month) completed the same procedures, now with the intent of informing whether there is an association between reading purpose condition and types of eye movement (i.e., passive or reader-initiated) used. It was hypothesized that the new word learning condition (where this reading goal was explicitly instructed) would trigger active reading processes for the new words, while the reading comprehension condition would trigger passive reading processes related to the new words. Results indicated that eye movements associated with passive cognitive processes (i.e., first

fixation duration, gaze duration) were not different between the two reading purpose conditions. Meanwhile, some eye movements associated with reader-initiated cognitive processes (i.e., regressions in and total reading time) were different across the two conditions; namely, there were more of these active eye movements when the readers were told to read for the purpose of learning new words. These findings showed that a reading goal of learning words affects eye movements related to reader-initiated cognitive processes. This forms the basis of the current study of the relationship between reading purpose and quality of semantic inferencing about new words.

1.5 Study Purpose

Based on this review of related literature and precursor studies, we have two primary purposes in this investigation. Does a focus on reading fluency (as opposed to comprehension of the text) affect the quality of the hypotheses that the reader makes about new words? Reading for fluency implicates goals of speed and accuracy, thereby altering the reader's purpose for reading the text. Reading purpose is thought to have an effect on many aspects of reading (Cheon & Ma, 2014) and may also affect novel word learning. The question of the effects of reading purpose, such as fluency, on novel word learning was addressed with a between-subjects design, in which readers were either told to read for comprehension or to read for speed and accuracy. We hypothesized that behavioral measures of semantic inferencing (e.g., quality of definitions of novel words, score on content test questions) are expected to be better for a group told to read for comprehension than one instructed to read for speed and accuracy.

This study also aims to explore reader-initiated processing as it contributes to word learning. Specifically, it aims to inform whether reading purpose (fluency vs comprehension)

affects eye movements associated with reader-initiated cognitive processes for new words. The hypothesis is that a focus on reading speed may sacrifice the capacity for other cognitive processes while reading; quality of comprehension, ability to detect nonwords/novel words, and, important to this project, semantic inferencing. Therefore, we predict that eye movements correlated to active reading processes will be higher in the group instructed to read for the purpose of comprehension. Reading faster may inhibit or reduce the online processes needed for components involved in new word learning. In other words, maintaining a goal of reading fluency may inhibit new word learning.

2.0 Methods

This study utilizes a previously collected data set. The procedures used were based on methodology used in previous studies (Duff & Brydon, 2018; Duff & Davidson, 2019, July). In the between-subjects design, participants were randomly assigned to one of two conditions: a *Fluency Condition* (FC) and a *Comprehension Condition* (CC), as described in further detail below. The following sections will describe the sample, procedures, and outcome measures involved.

2.1 Sample

The participants were recruited through *Pitt+Me*, a university-based online research system. Data for this study was originally collected from 31 subjects. However, two subjects had incomplete data due to experimenter error. These two subjects were not included in final analysis, resulting in the reported sample size of 29. All participants ($n=29$) were sixth-grade students (15 males, 14 females) between the ages of 8;5 (years;months) and 12;10 ($M=129.97$ months, $SD=14.37$).

The participant sample is meant to represent a population of readers who do not have dyslexia. A reported history of dyslexia was considered an exclusionary criterion for this study. Additionally, two subtests of the *Woodcock Reading Mastery Tests, Third Edition* (WRMT-III;

Woodcock, 2011) were administered as exclusionary criteria: “Word Identification” (form A)¹, standard scores $n=28$, $M=107.54$, $SD=11.97$, range = 46; “Word Attack” (form A)², standard scores $n=28$, $M=101.68$, $SD=11.96$, range = 49. No participants met the exclusionary criteria of WRMT-III standardized scores outside of 1.5 standard deviations from the mean. Participants in the FC and CC conditions did not differ significantly in performance on these measures according to the independent samples t-tests performed (see Table 2).

In addition to the WRMT-III, the following standardized measures were administered in order to describe participant language skills: the *Peabody Picture Vocabulary Test, Fourth Edition* (PPVT-IV; Dunn & Dunn, 2007), standard scores $n=29$, $M=115.34$, $SD=11.90$, range = 43; the *Gray Oral Reading Test – Third Edition* (GORT-3; Wiederholt & Bryant, 1992)³, scaled scores $n=27$, $M=11.85$, $SD=2.55$, range = 11. This use of a convenience sample may have led to higher performance scores on average for the PPVT-IV and GORT-3.

Participants in the FC and CC conditions were compared on the above parameters using a series of independent samples t-tests (see Table 2 for these and other descriptive measures). Both groups were shown to be comparable for the reported measures, with the exception of PPVT-IV scores, which were significantly higher in the fluency condition. The implications of this finding will be further explored in the “Discussions” section of this paper.

¹ One participant was not administered the WRMT-III “Word Identification” or “Word Attack” subtests due to time limitations. This participant was also not administered the GORT-3, but completed all remaining measures considered in the present data set.

² See footnote 1.

³ The *Gray Oral Reading Test – Third Edition* performance data is missing for two participants. GORT-3 performance data for one participant was discounted due to administration error. The other participant was not administered the GORT-3 due to time constraints during the visit (see footnotes 1, 2).

Table 1 Descriptive Statistics for CC and FC Conditions

	CC		FC		<i>t</i> (<i>df</i>)	<i>p</i>
	<i>n</i>	<i>M</i>	<i>n</i>	<i>M</i>		
Chronological Age (months)	17	132.71	12	126.08	1.23 (27)	0.23
WRMT-III Word Identification Subtest (standard score)	17	106.24	11	126.08	-0.71 (26)	0.49
WRMT-III Word Attack Subtest (standard score)	17	101.24	11	109.55	-0.24(26)	0.81
PPVT-IV (standard score)	17	111.24	12	121.17	-2.39 (27)	0.02
GORT-3 (scaled score)	17	12.00	10	11.60	0.29 (25)	0.70

2.2 Procedures

For this study, participants were given two texts to read, which were identical between conditions. Each text each included six nonwords. The FC group, meant to represent a reading purpose of speed and accuracy, was instructed to read the texts as quickly and accurately as possible. In this condition, the examiner held a watch to show that the task was being timed. The CC group, representing a reading purpose of comprehension, was instructed to understand the texts as well as possible. Following the reading of text, participants were administered two post-assessment measures of semantic inferencing for the respective novel words: 1) participants were

asked to provide definitions for each of the nonwords, and then 2) participants completed a series of six comprehension questions about the nonwords.

2.2.1 Nonword Selection

Each of the two passages administered contained six embedded nonwords, each appearing only once and taking the place of real-word synonyms (i.e., the target words). Selection of the nonwords was based on methodology described more thoroughly in Duff (2015), and took into account both semantic and orthographic features of the words and the surrounding text.

For each target word represented by a nonword in the text, both the word's semantic complexity and semantic importance for the text's topic were considered. Target words were selected so that they would have adequate difficulty, or semantic complexity. Nonwords were intended to replace real-world vocabulary that had been encountered before by this age group, though not likely compatible with daily use. This evidence-based selection process was guided by principles in "Words Worth Teaching" (Biemiller, 2010) and the "Tier Two" approach from Beck, McKeown, and Kucan (2002). In conjunction with evaluation of word-level semantic complexity, Latent Semantic Analysis (LSA), a computational method that allows for documentation and analysis of the semantic features of a text (Landauer, Foltz, & Laham, 1998), aided in selecting words that had similar semantic importance for comprehending the texts.

Much like the semantic factors, the orthographic composition of nonwords was also vital to prioritizing internal validity. The selected target words represented similar orthographic complexity due to selection of orthographically consistent, single-syllable nonwords (Coltheart & Leahy, 1992). Additionally, nonwords were designed to be easily identified as such based on their phonotactic properties.

2.3 Outcome Measures

This study relies on two types of behavioral data collected during the described procedures: language measures of semantic inferencing and eye movement data, each addressing our primary and secondary research questions respectively.

2.3.1 Quality of Hypotheses About the Semantics of Novel Words

The reader's semantic hypotheses of the novel words embedded in each text were evaluated through two primary outcomes: (1) scored student-provided definitions for each novel word, and (2) context test questions. Language measures were collected via posttest assessment of semantic inferencing.

2.3.1.1 Definitions of Nonwords

Participant definitions of nonwords were obtained in an open-ended format. For each of the twelve nonwords embedded in the texts, participants provided a verbal response to a prompt for a definition (i.e., "What do you think this word means?"). Responses were provided verbally, and audio recorded for later transcription. The transcribed participant definitions were then scored using an established five-category ordinal rating scale previously published in Duff (2019), as represented below in Figure 3. The scale accounts for both semantic and syntactic quality of participant definitions, correlating to a 0-4 assignment on the ordinal scale. Inter-rater reliability was established by having two coders (MS and DD) use the present criteria to score transcribed definitions from a parallel study ($n=240$). These definitions were not part of the current study. Interrater reliability for these definitions, ± 1 was 95.8%. Upon completion of this reliability

measure, scoring for the present data set was completed by a single coder (MS), blind to participant.

0	1	2	3	4
Don't know OR Phonologically-based response OR No semantic relationship OR Very idiosyncratic semantic relationship OR Meets only 1 part of criteria for "1"	Incorrect part of speech OR Semantic content makes some connection to paragraph but not key semantic elements	One key semantic element AND Correct part of speech	Two key semantic elements	Target word OR Exact synonym

Figure 3 Criteria for Coding Participant Nonword Definitions (Duff, 2015)

2.3.1.2 Context Test Questions

The posttest questionnaire involved six dichotomous context test questions for each nonword (e.g., Is drace something you do?). Questions for each nonword were grouped together and ordered to progressively increase in specificity, so as to minimize cues about word meaning that could affect responses. This procedure is based on Killian and colleagues (1995). The combined use of context test questions and the definition scoring system above taps into semantic knowledge at both high levels (definitions) and low levels (context test questions; Duff, 2019)

2.3.2 Eye Movement During Reading

Eye movement data was collected during reading trials using an *Eyelink 1000* to reflect online reading processes. These were the outcome measures for the secondary research question. Eye movements can provide information about participants' encounters with and interactions with each novel word, in real time. This provides a complement to behavioral measures, which may reflect processing that occurred during reading, or that happened as a result of the question posed by the experimenter. To encode the patterns of eye movements, this study will recognize the same eye movement patterns as those used in Duff and Davidson (2019, July), as depicted in Figure 1 and Figure 2. Understanding of these data procedures is described further in the "Eye Movements and Incidental Learning" section of this text.

The recognized eye movement measures reflect both passive and active (i.e., reader-initiated) processes, congruent with the Van den Broek and Helder (2017) representation of online reading comprehension. Passive interactions with the word are characterized through *first fixation duration* (i.e., the amount of time spent first looking at the word) and *gaze duration* (i.e., the total time spent looking at the novel word during the first pass). The first fixation duration and gaze duration can be different when the reader fixates on one part of the novel word, then fixates on a second location within the same word. Meanwhile, reader-initiated processes are represented by those active movements after the reader's first interaction with the word. The eye movements that are believed to relate to more active cognitive processes include *regression path duration* (i.e., go past time) and *dwell time* (i.e., total reading time before moving on in the text).

2.4 Analytic Plan

Participants each underwent two separate passages with respective posttests. Therefore, outcome measures for each participant come from two different texts, introducing the possibility of variance due to text differences or differences between individual nonwords. To account for this complex structure and strengthen the validity of the analysis, a linear mixed effects model (LMM) was used. The LMM builds on traditional linear models of analysis by incorporating a combination of fixed and random effects as predictor variables (Harrison et al., 2018). Commonly fit for biological datasets, this model accounts for non-independent observational units that are hierarchical in nature (e.g., the hierarchical structure created by data that progresses from sequential readings of a first text and second text). *Fixed effects* are the groups or levels at which data is being manipulated (e.g. reading condition). These fixed effects are independent of each other, with common residual variance. Meanwhile, *random effects* are grouping variables that inform the variance existing within and among the hierarchical groups (e.g., story). The primary hypothesis for this study was that posttest measures of semantic inferencing would be higher in the CC than the FC. The secondary research hypothesis posited that eye movements indicative of active reading processes will be longer the CC than for the FC. To address both of these hypotheses, LMMs were used to conduct between-group comparison of both language outcome measures (i.e., coded participant definitions, context test question performance). For the present data set, reading condition (CC vs FC) will be held as a fixed effect, with random effects of participants, story (i.e., each of the two texts read by the participants), item (i.e., each of the twelve nonwords assessed, nested within their respective stories). Among other benefits of using a LMM for fitting data, this method of analysis aims to reduce Type I and Type II error rates and is apt at accounting for any missing data points.

3.0 Results

The data was fit with a linear mixed effects model, as described above. However, the initial model did not converge. Story was therefore removed as a random effect. This effect was selected for removal from the model because item was already nested within the stories. An additional random effect of CTQ item (i.e., each of the 60 CTQ questions administered, nested within item) was included for analysis of the language measures. The model successfully converged with the new parameters. These fixed and random effects were held constant for all remaining LMM analyses.

3.1 Primary Research Question

Our primary research question was interested in whether the difference in instruction between the CC and FC conditions would lead to difference in the quality of semantic inferencing for the nonwords in the texts. It was hypothesized that there would be significant between-group differences in post-test language measures of semantic inferencing. Specifically, scores on language measures were expected to be significantly higher for participants in the CC.

To analyze these between-group differences in language measures, a composite language score was considered. The language score, labeled as TotalScoreAdjusted in the model, represents the summed word definition score (0-4) and the sum of the five corresponding CTQ questions (0-5) for each item. Initially, the analysis of the language score was done with condition as the sole fixed effect, with item, participant, and CTQ as random effects. The R code for this model was as

follows: $\text{lmer}(\text{TotalScoreAdjusted} \sim 1 + \text{Condition} + (1|\text{Item}) + (1|\text{Participant}) + (1|\text{CTQ}))$. This model included 1740 observations (Participant, 29; Item, 12; CTQ, 5). A summary of statistics for the random effects and model output can be found below in Table 3. Contrary to our hypothesis, there was no significant effect of condition in the model ($p=0.883$, $\beta=0.00598$) for the fixed effect of condition on language measures of semantic inferencing.

Table 2 Group Mean Performance on Language Measures

Condition	Word Definition + CTQs per Nonword (-1 - 9) Mean (SD)
CC	4.70 (2.71)
FC	4.51 (2.82)

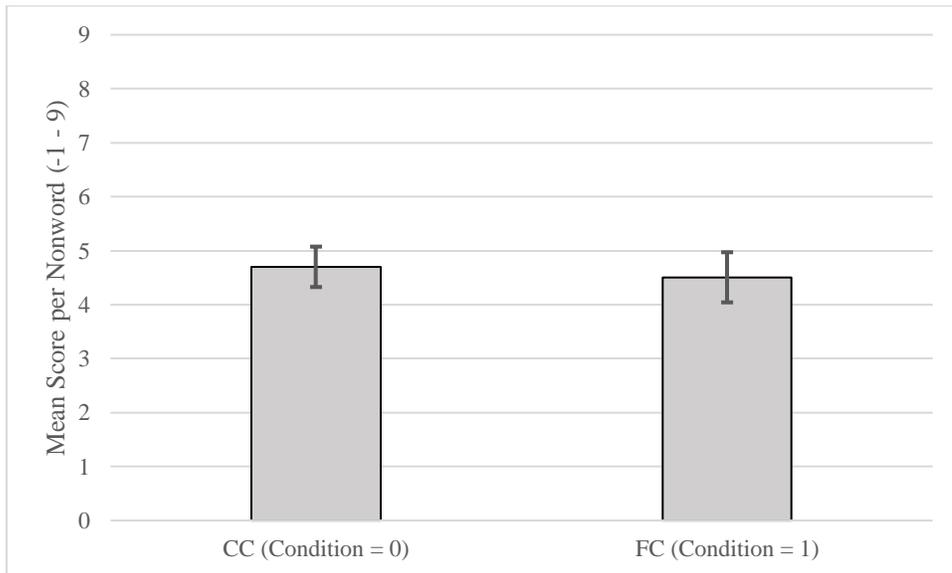


Figure 4 Between-Group Mean Language Measure Performance

Table 3 Language Measures LMM Output, Condition Fixed Effect

Random Effects					
Group Name	Variance	Standard Deviation			
Participant (Intercept)	0.0108	0.1041			
Item (Intercept)	0.0098	0.0988			
CTQ (Intercept)	1.0643	1.0317			
Fixed Effects					
	Estimate	Standard Error	df	t value	Pr(> t)
(Intercept)	0.3924	0.5969	2.0168	0.657	0.578
Condition	0.00598	0.0403	26.9204	0.148	0.883

After running the above model, post hoc analysis was done to test the effects of both the condition (CC vs FC) and PPVT-IV raw scores (i.e., a standardized measure of pre-existing vocabulary knowledge) on language measures of semantic inferencing. In general, one would expect that vocabulary achievement scores would relate to word learning abilities, and there is evidence that this is the case (Gellert & Elbro, 2013; Nash & Donaldson, 2005). Therefore, we would expect that PPVT-IV scores would be significant if they were included in the model. Of interest here was the possibility of an interaction. Specifically, we hypothesized that the effect of reading condition might vary with vocabulary skill, with fluency based instructions showing less of an effect on readers with high vocabulary. The effect of vocabulary could be especially important, given that there were differences in vocabulary achievement (PPVT-IV scores) between the two reading conditions. Random effects for this model remained constant, and the fixed effects were both condition and PPVT-IV raw scores. The R code for the model was as follows: `lmer(TotalScoreAdjusted ~ 1 + Condition*PPVT + (1|Item) + (1|Participant) + (1|CTQ))`. The model similarly included 1740 observations (Participant, 29; Item, 12; CTQ, 5). A summary of

results from the model can be found below in Table 4. This model similarly showed condition to not have a significant simple effect ($p=0.149$, $\beta=-0.5576$). However, PPVT-IV raw score was significant as a fixed effect in this model ($p<0.05$; $p=0.0146$, $\beta=0.0035$). There is no evidence of a significant interaction between reading condition and PPVT-IV raw score ($p=0.1599$, $\beta=0.0030$) in predicting performance on the posttest language measures.

Table 4 Language Measures LMM Output, Condition and PPVT-IV Fixed Effects

Random Effects					
Group Name	Variance	Standard Deviation			
Participant (Intercept)	0.0058	0.0761			
Item (Intercept)	0.0098	0.0988			
CTQ (Intercept)	1.0637	1.0314			
Fixed Effects					
	Estimate	Standard Error	df	t value	Pr(> t)
(Intercept)	-0.2214	0.6407	2.6757	-0.346	0.755
Condition	-0.5576	0.3740	25.0050	-1.491	0.149
PPVT	0.0035	0.0013	25.0459	2.623	0.015
Condition:PPVT	0.0030	0.0021	25.0067	1.449	0.160

Due to the known relationship between performance on the PPVT-IV standardized test and performance on language measures, additional descriptive statistics were run to differentiate between high and low PPVT performers for language measures of semantic inferencing, as represented in Figure 5. High and low PPVT-IV performance was determined in reference to other participants in the present sample, using the whole group mean for PPVT-IV raw score as a cut off between high and low scores within each condition.

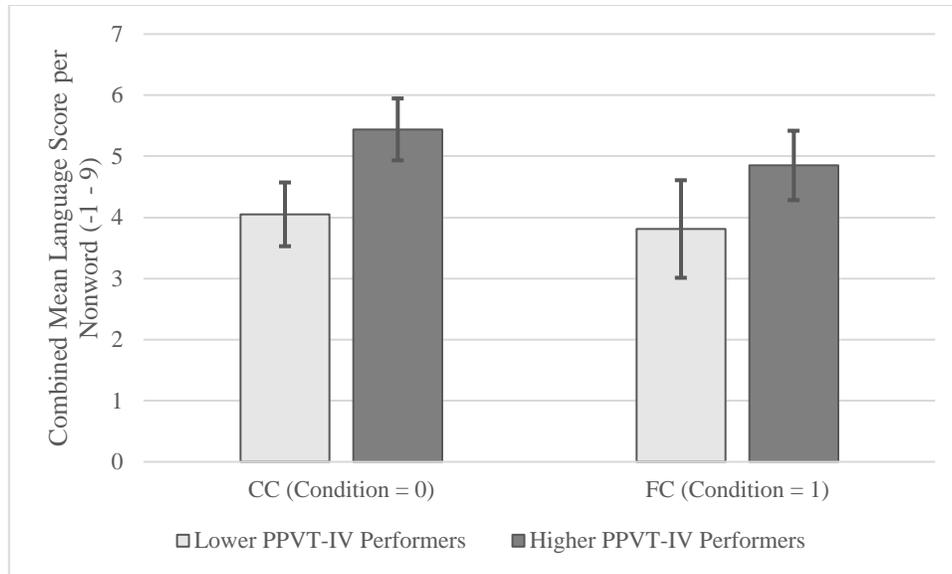


Figure 5 Mean Language Measure Performance Differentiated by PPVT-IV Performance

3.2 Secondary Research Question

The secondary research question concerned differences in online comprehension processes due to differences in reading instruction between the CC and FC groups. For this study, online comprehension processes were operationalized through use of eye tracking data collected during reading. Eye tracking data observed in analyses include two measures associated with passive processes (i.e., first fixation duration, gaze duration) and two measures associated with active, or reader-initiated, processes (dwell time, regression path duration). We hypothesized that there would be significant differences in eye tracking measures dependent upon condition. Specifically, we anticipated higher levels of active reading processes (i.e., dwell time, regression path duration) in the CC than in the FC, with no significant difference between values related to passive processes (i.e., first fixation duration, gaze duration).

To address this research question, the data was fit with LMM models with reading condition as a fixed effect and random effects of item and participant. These fixed and random effects remained stable across all four models run for each of the four parameters of eye tracking data. CTQ item was not included as a random effect, as it had been in the models for the primary research question, due to the nature of the data. Of note, eye tracking data was missing for two participants due to experimenter error. The data for these two participants was included in the study sample for consideration of the primary research question, but excluded from the following analyses. The LMM analysis used was selected in part due to its strength in accounting for such missing data points.

The measure of first fixation duration, or first fixation time (FFT), quantifies the total time of the reader’s initial first past fixation on the word. A similar model was used for analyzing first fixation time, otherwise known as first fixation duration, the second measure corresponding to active reading processes. The R code for the model was as follows: `lmer(FFT ~ 1 + Condition + (1|Item) + (1|Participant))`. This model had 975 observations (groups: Participant, 26; Item, 12). See Table 5 below for random effects statistics and model output. Consistent with our hypothesis, this model does not show a significant effect of condition on first fixation duration ($p=0.923$, $\beta=9.468$).

Table 5 First Fixation Duration LMM Output

Random Effects					
Group Name	Variance	Standard Deviation			
Participant (Intercept)	56974	238.7			
Item (Intercept)	10083	100.4			
Fixed Effects					
	Estimate	Standard Error	df	t value	Pr(> t)
(Intercept)	434.46	71.89	30.66	6.044	<.001
Condition	9.47	96.70	23.85	0.098	0.923

Gaze duration (GD), or the summed first-pass fixations before leaving the word, was the other measure recognized as indicative of passive processing of the novel words. The R code for analyzing gaze duration was as follows: `lmer(GazeDuration ~ 1 + Condition + (1|Item) + (1|Participant))`. The model included 980 observations (Participant, 26; Item, 12). For a summary of the random effects and model output, see Table 6. Congruent with our hypothesis, the model did not indicate a significant effect of condition for gaze duration ($p=0.900$, $\beta=16.34$).

Table 6 Gaze Duration LMM Output

Random Effects					
Group Name	Variance	Standard Deviation			
Participant (Intercept)	101510	318.6			
Item (Intercept)	12382	111.3			
Fixed Effects					
	Estimate	Standard Error	df	t value	Pr(> t)
(Intercept)	466.81	93.04	28.94	5.017	<.001
Condition	16.34	128.38	23.81	0.127	0.900

Regression path duration, or go past time (GPT), reflects the time from initial fixation on the nonword until the eyes move on to new text. This measure was considered to reflect active interaction with the text (i.e., reader-initiated processing). To analyze between-group differences in GPT, the R code was as follows: `lmer(GoPastTime ~ 1 + Condition + (1|Item) + (1|Participant))`. This model included 980 observations (Participant, 26; Item, 12). See Table 7 below for a summary of the random effects and model output. Contrary to our hypothesis, there was no significant effect of condition on GPT ($p=0.668$, $\beta=118.26$).

Table 7 Regression Path Duration LMM Output

Random Effects					
Group Name	Variance	Standard Deviation			
Participant (Intercept)	420269	648.3			
Item (Intercept)	89971	300.0			
Fixed Effects					
	Estimate	Standard Error	df	t value	Pr(> t)
(Intercept)	847.30	204.97	30.23	4.134	<.001
Condition	118.26	272.69	24.22	0.434	0.668

The measure of dwell time (DT) represents the summation of total fixation time for each nonword, including regressions, and was also considered to represent active reading processes. For analysis of dwell time (DT) between condition groups, the R code was as follows: `lmer(DT ~ 1 + Condition + (1|Item) + (1|Participant))`. The model included 1550 observations (Participant, 26; Item, 12). A summary of statistics for the random effects and model output can be found below in Table 8. Contrary to our hypothesis, this model did not find a significant difference in dwell time between the two conditions ($p=0.478$, $\beta=84.27$).

Table 8 Dwell Time LMM Output

Random Effects					
Group Name	Variance	Standard Deviation			
Participant (Intercept)	83693	289.30			
Item (Intercept)	5459	73.89			
Fixed Effects					
	Estimate	Standard Error	df	t value	Pr(> t)
(Intercept)	378.00	82.30	26.89	4.593	9.17e-05
Condition	84.27	117.02	23.93	0.720	0.478

Analysis of eye-tracking measures related to both passive and active processing found no significant effects. Per our hypothesis, no effect was expected for either of the eye-tracking

measures related to passive reading processes. However, the results were inconsistent with our predictions for the measures relating to active reading processes.

4.0 Discussion

The primary research question in this study was whether differing reading instructions (i.e., read and understand as well as you can (CC) vs. read quickly and accurately (FC)) would affect the quality of semantic inferencing by readers. We hypothesized that instructions promoting reading fluency would predict performance on post-test language measures of semantic inferencing, such that participants in the FC would have lower scores than those in the CC on posttest measures of semantic inferencing. In the secondary research question, we investigated the relationship between differing reading goals and active (i.e., reader-initiated) eye movements, and predicted that reading condition would predict the duration of eye movements associated with reader-initiated processes while it would not predict eye movements associated with passive processes.

Overall, we did not find evidence to support the proposed hypotheses. There was no significant simple effect of reading condition on semantic inferencing outcomes. Similarly, no between-group differences were found in eye tracking measures related to either reader-initiated (dwell time, regression path duration) or passive reading process (first fixation duration, gaze duration). The current data does not support the hypothesis that reading purpose creates a significant difference in active comprehension processes while interacting with the text (as tracked via the aforementioned eye movements) or on the quality of semantic inferencing of new words after reading the text. However, there are several alternative explanations for these findings that deserve further exploration.

Importantly, an independent samples t-test found the CC and FC to differ significantly in their performance on the PPVT-IV assessment (see Table 1), despite random assignment to

condition. Specifically, participants in the FC demonstrated significantly higher scores than their peers in the CC group. The PPVT-IV is a standardized, norm-referenced assessment that measures receptive vocabulary through a series of picture selection prompts given vocabulary words of increasing difficulty. Higher performance on this measure indicates better general vocabulary knowledge, particularly breadth of vocabulary knowledge. In principle, this measure of vocabulary achievement would relate both to vocabulary exposure and word learning ability. As such, it may reflect important individual differences that would directly impact performance on the word learning measure, which is the primary outcome in this study. Indeed, PPVT-IV scores did significantly predict participant performance on the posttest measures of semantic inferencing. A LMM showed PPVT-IV scores to significantly predict participant performance on the posttest language measures of semantic inferencing about nonwords, regardless of condition. Because of these differences in PPVT-IV performance between groups, we cannot rule out the possibility that differences in previous vocabulary attainment may have impacted word learning between the randomly assigned reading conditions. If so, the effect of vocabulary would have worked against the hypothesis, so that effects of previous vocabulary attainment and expected effects of reading condition may have worked in opposite directions.

Also of consideration is whether task procedures elicited the desired simulation of fluency instructions in the school setting. The instructions may not have carried appropriately salient instructions to create similar functional change in the readers' motivation or goals for the task across participants. The current study did not examine data to suggest whether participants read the passages more quickly or accurately in response to fluency instructions. Therefore, it is unknown whether these instructions effectively changed these aspects of reading for the participants. If participants did not change how they read based on the reading instructions, then no

significant effect of condition would be expected for either research question. The stability across conditions for both the eye movements associated with reader-initiated processes (i.e., those expected to change with differing goals and comprehension processes) and with passive processes (i.e., those that are not expected to change across tasks) is consistent with the idea that there was no functional change in participants' approaches to reading based on the instructions.

Finally, individual differences in response to instructions could play influential roles in determining the extent to which reading patterns change with fluency instruction. In other words, children may respond differentially to the same instructions. Because this was a between subjects design, it is difficult to account for this possibility in the statistical analysis. However, one possible source of interaction was considered in a post hoc analysis, namely a possible interaction between reading condition and PPVT-IV scores. This analysis did not reveal any significant interaction. General standards of coherence could also impact readers' responses to reading instructions. For example, some readers may have had a high standard of coherence, expecting full comprehension of the text, which was maintained despite instructions to read quickly. This commitment to comprehension may not have been as robust in children with generally lower standards of coherence. Finally, the salience of reading instructions during the task may have varied with a myriad of related variables, including those related to the experimenter or environmental variability.

4.1 Limitations

The higher PPVT-IV performance in the FC group may partially account for the lack of significant results in respect to our hypotheses. This difference occurred despite random assignment to groups. However, it does limit our ability to interpret the findings in this study.

In the discussion, we raised several possible explanations for the lack of a significant result. To disambiguate these hypotheses, we would need data that was not included in the planned study. Specifically, we did not consider data on reading times, or on general standards of coherence in the participants.

Finally, this study examined a relatively small sample size ($n=29$), given that it is a between subjects design. Furthermore, missing data was not equally spread across conditions, so that the two groups were of different sizes. The data set was fit with the statistical model thought best to account for the small size, expected variability, and missing data points. However, further studies should investigate whether the observed non-effects are also found in a larger sample.

4.2 Conclusion

Future research may build on the present findings and procedures to better inform the effect of reading purpose on children's ability to perform complex comprehension processes. Importantly, interpretation of the current results were confounded by unanticipated between-group differences in PPVT-IV performance. Given the implications of these differences on the interpretation of the results for our hypotheses, additional exploration of the same hypotheses is warranted. Future replications of this study or completion of similar studies may better inform our

hypotheses by completing similar analyses between groups with more comparable underlying language performance.

Furthermore, research should further evaluate independent differences in standards of coherence and response to instructions, along with the effect of these considerations on the present topic of the fluency instruction and high level cognitive processes while reading. Accounting for this differential response to instruction may offer better insight as to the true effect of imposing fluency-driven directions for reading tasks and, on a larger scale, prioritizing fluency in reading instruction in the schools. Also, though not possible for the present data set, considering total reading time for each passage may inform participant response to fluency instruction. Similarly, a Reading Motivation Scale was administered to many participants in the current data set among other measures. Data from this measure was not included in the planned analyses for this study. However, further investigating these scale scores as possible predictors of performance on inferencing measures and eye-tracking behaviors would be a promising area of future investigation. We might predict that scores on the Reading Motivation Scale might predict performance on the behavioral measures, and/or the eye movement measures. There may also be an interaction between reading motivation scale scores and reading condition.

This study synthesized relevant precursor studies and prior findings in the literature to offer a valuable analysis of the current data set. Though the analyses completed do not support the proposed hypotheses, questions raised within this study may inform future research of the present hypotheses. Matriculating findings may contribute to a developing understanding of the implications of reading instruction and focus in the academic setting, particularly on learning of novel words.

Bibliography

- Baumann, J. F., Carr Edwards, E., Boland, E. M., Olejnik, S., & Kame'enui, E. J. (2003). Vocabulary tricks: Effects of instruction in morphology and context on fifth-grade students' ability to derive and infer word meanings. *American Educational Research Association, 40*(2), 447-494.
- Beck, I. L., McKeown, M., & Kucan, L. (2002). *Choosing words to teach*. New York: Guilford Press.
- Biemiller, A. (2010). *Words worth teaching: Closing the vocabulary gap*. Columbus: McGraw-Hill SRA.
- Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*(1), 31-42. doi:<https://doi.org/10.1037/0022-0663.96.1.31>
- Cain, K., Oakhill, J., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *J. Child Lang., 30*, 681-694. doi:10.1017/S0305000903005713
- Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671-681. doi:10.1037/0022-0663.96.4.671
- Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25-40): Lawrence Erlbaum Associates Publishers.

- Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Cheon, H. J., & Ma, J. H. (2014). The effects of reading purpose on reading comprehension and perceived difficulty. *English Teaching, 69*(2), 51-69.
- Christ, T., & Chiu, M. (2018). Hearing words, learning words: How different presentations of novel vocabulary words affect children's incidental learning. *Early Education and Development, 1556-6935*. doi:10.1080/10409289.2018.1484648
- Clark, E. V. (2009). *First language acquisition*. New York, NY: Cambridge University Press.
- Coltheart, V., & Leahy, J. (1992). Children's and adult's reading of nonwords: Effects of regularity and consistency. *Journal of experimental Psychology: Learning, Memory, and Cognition 18*, 718-729.
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22*, 8-15.
- Daneman, M. (1988). Word knowledge and reading skill. In M. Daneman, G. E. Mackinnon, & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 6, pp. 145-175): Academic Press.
- Duff, D. (2015). *Lexical semantic richness: Effect on reading comprehension and on readers' hypotheses about the meanings of novel words*. (Doctor of Philosophy Dissertation), University of Iowa, Iowa City, Iowa.
- Duff, D. (2019). Has vocabulary intervention had an effect? A valid, reliable, and (fairly) quick outcome measure for semantic knowledge. *Lang Speech Hear Serv Sch, 50*(4), 506-517. doi:10.1044/2019_LSHSS-VOIA-18-0134
- Duff, D., & Brydon, M. (2018). Reading new words what eye movements relate to successful incidental learning.

- Duff, D., & Davidson, M. M. (2019, July). *Eye movements when reading novel words in text: The effect of reading purpose*. Paper presented at the The Society for the Scientific Study of Reading (SSSR) Annual Meeting, Toronto, Canada.
- Dunn, L. M., & Dunn, L. M. (2007). Peabody picture vocabulary test-iv. In. Circle Pines, MN: American Guidance Service.
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, 2(1), 1-44. doi:10.1080/19345740802539200
- Fernald, A., Marchman, V. A., & Weisleder, A. (2013). Ses differences in language processing skill and vocabulary are evident at 18 months *Developmental Science*, 16(2), 234-248. doi:doi:10.1111/desc.12019
- Gellert, A. S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, 31(1), 16-28.
- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., . . . Yoon, H. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, 21, 563-584. doi:https://doi.org/10.1017/S1366728917000219.
- Goodman, K. S. (2006). A critical review of dibels. In K. S. Goodman (Ed.), *The truth about dibels: What it is, what it does* (pp. 1-39). Portsmouth, NH: Heinemann.
- Graves, M. F. (1986). Vocabulary learning and instruction. In E. Z. Rothkopf & L. C. Ehri (Eds.), *Review of research in education* (Vol. 13, pp. 49-89). Washington, D. C.: American Educational Research Association.

- Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D., Goodwin, C. E. D., . . . Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6(e4794). doi:<https://doi.org/10.7717/peerj.4794>
- Holmquist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK.: Oxford University Press.
- Killian, A. S., Nagy, W. E., Pearson, D. P., Anderson, R. C., & Garcia, G. E. (1995). *Learning vocabulary from contexts: Effects of focusing attention on individual words during reading*. Retrieved from Champaign, Illinois:
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information process in reading. *Cognitive Psychology*, 6, 293-323.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.
- Manzo, K. K. (2005, September 28). National clout of dibels test draws scrutiny: Critics say reading tool's scope fails to justify its braod use. *Education Week*, 25, 1-12.
- Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52-77). New York: Guilford.
- Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24, 237-270.
- Nash, M., & Donaldson, M. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research*, 48(2), 439-458. doi:10.1044/1092-4388(2005/030)

- Pagán, A., Bird, M., NHsiao, Y., & Nation, K. (2020). Both semantic diversity and frequency influence children's sentence reading. *Scientific Studies of Reading, 24*(4), 356-264. doi:10.1080/10888438.2019.1670664
- Panel, N. R. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, D. C.: National Institute of Child Health and Human Development.
- Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamins.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22-37. doi:10.1080/10888438.2013.827687
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2009). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10*(3), 241-255. doi:https://doi.org/10.1207/s1532799xssr1003_3
- Riedel, B. W. (2007). The relation between dibels, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*(4), 546-567. doi:10.1598/RRQ.42.4.5
- Samuels, S. J. (2006, May). *Introduction to reading fluency*. Paper presented at the annual meeting of the International Reading Association, Chicago.
- Shelfbine, J. L. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior, 22*(1), 71-97.

- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-407. doi:10.1598/rrq.21.4.1
- Strauss, V. (2014, June 25, 2014). Is this really how we should test reading development in kids? *The Washington Post*. Retrieved from <https://www.washingtonpost.com/news/answer-sheet/wp/2014/06/25/is-this-really-how-we-should-test-reading-development-in-kids/>
- Swanborn, M. S. L., & de Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning*, 52(1), 95-117.
- Van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54(5-6), 360-372. doi:10.1080/0163853x.2017.1306677
- Van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, 29(8), 1081-1087.
- Wiederholt, J. L., & Bryant, B. R. (1992). Gray oral reading test , third edition. In. Austin, TX: Pro-Ed.
- Woodcock, R. W. (2011). Woodcock reading mastery tests - third edition. In. Oxford: Pearson.