June 2022

Influences of Sentence Context and Individual Differences in Lexical Quality on Early Phonological Processing during Silent Reading

Sara Milligan
University of South Florida

Influences of Sentence Context and Individual Differences in Lexical Quality on Early Phonological Processing during Silent Reading

by

Sara Milligan

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters in Psychology
with a concentration in Cognition
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Elizabeth R. Schotter, Ph.D.
Chad Dubé, Ph.D.
Geoffrey Potts, Ph.D.

Date of Approval:
April 27, 2022

Keywords: Eye Tracking, Reading, Phonology, Prediction

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

The current thesis investigates the role of sentence context and individual differences in the quality of sub-lexical representations of words in activation of phonological forms during silent reading. More specifically, this study aims to determine how these situational and participant-level factors influence the use of phonology to aid word recognition during parafoveal processing, before a reader directly fixates the word. Therefore, I manipulated sentence constraint in two eye tracking during reading experiments (one using real-word and one using pseudoword parafoveal previews) that utilized the gaze-contingent boundary paradigm (Rayner 1975) and measured individual's scores on assessments of spelling ability, phonological decoding ability and semantic knowledge. Additionally, I performed follow-up exploratory analyses to investigate the role of lexical frequency in the interaction between sentence constraint and parafoveal processing of phonology. The results of these experiments suggest that the phonological preview benefit (PPB) is not strongly influenced by individual differences, but that the magnitude of the PPB appears to depend on stimulus characteristics, namely sentence constraint and lexical frequency. Additionally, a comparison of the effect sizes and pattern of significant results between the two experiments demonstrates that the PPB benefits from the preview having a holistic lexical representation in memory

Chapter 1: Introduction

Language processing involves the retrieval of component representations from memory in order to arrive at meaningful semantic-level representations. An interesting aspect of language is that it is represented at a number of different levels that have direct correspondences to one another. Bottom-up information is taken in from the environment, either visually or auditorily, and mapped onto orthographic or phonological codes that have internal representations generated through language experience. These codes are then used to home in on the correct holistic lexical representation that carries semantic meaning.

When it comes to learning language, infants begin by developing meaningful connections between auditory stimuli in the environment and objects or abstract ideas to which they refer. For years, they strengthen these connections before learning anything about the visual symbolic representations used in text. Therefore, the primary representational connections are between phonological forms and semantic representations. Research on reading education suggests that learning the correspondence between phonological representations and the orthographic representations of letters is crucial for reading success (Castles, Rastle & Nation, 2018; Rayner, Foorman Perfetti, Pesetsky & Seidenberg, 2001). Therefore, in the process of learning to read, phonological codes serve as intermediaries between orthography and semantics.

Children spend years learning these correspondences, which are continually reinforced by interacting with both spoken and written language. At a certain point, the learned relationships between orthographic representations and semantics may be strong enough for a printed word's semantic meaning to be retrieved without the need for a phonological intermediary. However, it is still possible that having robust, redundant connections between all levels of lexical representation makes word recognition easier and more efficient.

The current project addresses this question of whether, and under what circumstances, skilled adult readers use phonological codes to facilitate visual word recognition in silent reading. The use of phonological codes in reading, particularly during early stages of word recognition (i.e., parafoveal processing), may depend on a number of variables, including (1) the extent to which expectations about the upcoming word are generated by context, (2) the quality of a given reader's component lexical representations (and the strength of the connections between them), and (3) the lexical frequency of the word being recognized.

Sentence context generates expectations that preactivate information about upcoming words in a text and boost the efficiency of word recognition. Evidence from eye tracking shows that highly constraining sentence contexts reduce reading times on a predictable word (e.g., increase the probability that readers will skip it; Rayner & Well, 1996). Expectations are generated by high level conceptual representations rather than bottom-up perceptual information, so if these semantic expectations automatically feed down to pre-activate lower levels of lexical representation, it is possible that they would activate orthography, phonology, or both. Therefore, when context generates strong expectations, phonological codes may be activated to a greater degree and play a larger role in word recognition than when the reader is relying primarily on bottom-up visual information.

Individual differences in the quality of lexical representations may also be a driving factor in the extent to which phonological representations are used during reading. The lexical quality hypothesis proposes that a primary difference between skilled and unskilled readers lies in the quality of their lexical representations. These "common core representations" (Perfetti & Hart, 2002, p. 190) consist of unified orthographic, phonological, and semantic codes. Therefore, individuals with good lexical quality (ie., strong connections between orthographic, phonological, and semantic codes) may be more capable of rapidly activating phonological information from the visual orthographic information that could feed forward to semantics in a redundant fashion. If they are able to activate phonological codes at an early stage of lexical access, they would then be capable of using the connections between orthography and semantics as well as phonology and semantics to create a more resonant route to semantic meaning.

Phonological Processing in Single-Word Reading Tasks

Many previous studies have manipulated phonological information to study phonological processing in single word recognition (ie., in the absence of a sentence context). Although findings from single word reading tasks cannot necessarily be generalized to more natural scenarios like sentence or paragraph reading, they play a fundamental role in understanding the word recognition component of reading. Reaction time tasks like lexical decision and semantic categorization have been extremely useful in studying how various factors influence word processing speed, reflecting processing difficulty and facilitation. Some of the benefits of these tasks are that they do not rely on naming words aloud, which would necessitate phonological processing. Asking participants to respond whether a particular string of letters is a real word or whether a word belongs to a certain semantic category also requires a certain level of lexical or semantic processing to be achieved.

Phonological manipulations in a lexical decision task showed that people take longer to decide that a pseudohomophone (e.g., ROZE) is not a real word compared to a non-homophone nonword (e.g., ROFE; Coltheart, Besner, Jonasson & Davelaar, 1979), indicating that a word's phonology is sufficient to activate lexical representations. Using a semantic categorization task, Van Orden (1987) also showed that phonology alone is sufficient to activate semantic information. People made more categorization errors when the word was a homophone of a word in the probed category (e.g., Is a ROZE/ROBE a flower?). In a semantic relation task where participants responded whether or not two words were related (e.g, TABLE – CHARE vs. NOVEL – CHARE), homophones also activated information about semantic relationships between words. People made more errors and had longer reaction times when the second word was a homophone of a related word (Lesch and Pollatsek, 1998). These studies provide compelling evidence that word recognition involves at least some activation of phonological codes during visual word recognition even though doing so is not required to perform the task.

Phonological Processing during Sentence Reading

As we might expect based on evidence from single word reading tasks, phonological forms are also activated for sentence-embedded words during normal reading. Again, much of this evidence comes from studies that used homophone manipulations. For example, an acceptability judgment task of whole sentences revealed that both adults and children exhibit higher false-positive rates to both anomalous real-word orthographically irregular homophones (e.g., The girl through the ball.) and non-word pseudohomophones (e.g., Her bloo dress was new.) compared to words that are both phonologically and orthographically erroneous (Coltheart, Laxon, Rickard, & Elton, 1988). Similarly, using a proofreading task that had homophone errors in sentences, Daneman and Stainton (1991) found that participants were less sensitive to the errors that were homophones of a contextually plausible word compared to non-homophone errors. In a follow-up study using eye tracking during reading for comprehension (i.e., no proofreading task) Daneman and Reingold (1993) found differences between homophone and spelling control errors, with shorter fixation durations for homophones, but only in measures that incorporated rereading and not in first-pass reading. This later effect went away, however, if the homophones in a pair were of different lengths. However, in a very similar design that also manipulated the constraint of the sentences, Rayner, Pollatsek, and Binder (1998) found significant differences in fixation durations for the homophone error and spelling matched control error in first fixation, single fixation, and gaze durations. They found different effects of orthographic similarity between the correct word and the errors for high and low constraint conditions, but overall the patterns in both high and low constraint were consistent with early activation of phonological codes during silent reading. The results from these sentence reading studies demonstrate that phonological processing in silent word recognition is not isolated to tasks that present words in isolation and involve the participant making an explicit judgment about the word.

Parafoveal Preview in Reading

Phonological codes do appear to be activated at some point during visual word recognition, but an important question is whether phonology actually aids visual word recognition or if it is an epiphenomenal byproduct of the process; the time course of phonological activation can be telling. One issue with generalizing single word reading tasks to natural reading scenarios is that reading involves integration of semantic and syntactic information across multiple words and, therefore, the meanings of the preceding words in the sentence context can influence the process of identifying subsequent words. Furthermore, the earliest stages of visual word recognition during normal reading can occur before a word is even directly fixated. According to the E-Z Reader model (Rayener, Reichle & Pollatsek, 1998), when a certain criterion of word recognition is reached (not necessarily completed) on the currently fixated word, attention shifts to the upcoming word and certain features of that word begin to be processed while eye movements are being programmed to move towards it. During this in-between time, the reader gets a head-start on word recognition of the upcoming word, which lies outside of the fovea in lower-acuity parafoveal vision (see Figure 1). This head-start is referred to as a *parafoveal preview*. Because the parafoveal preview is the earliest possible bottom-up information that can be extracted from a word, it is an ideal place to look when asking how early phonological codes come online.

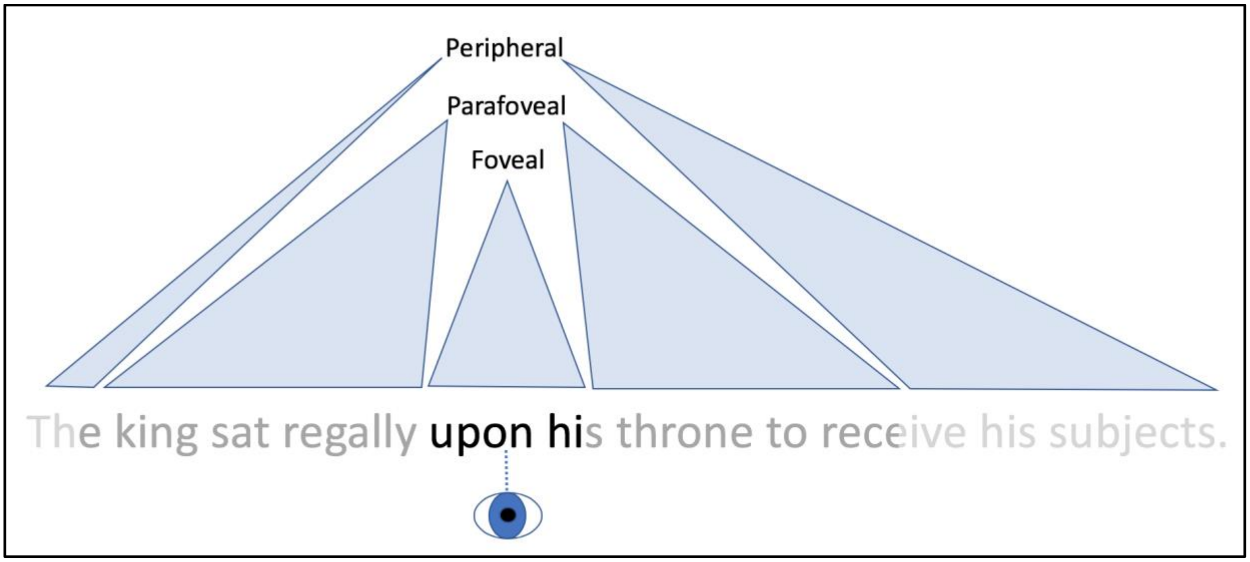

Figure 1. *Illustration of Foveal, Parafoveal, and Peripheral Visual Acuity in Reading*

Eye tracking can provide a useful tool to study online language processing and word recognition during natural reading. Eye movement patterns and gaze durations reflect the time course and mechanisms of cognitive processing. Experimental manipulations of language characteristics can reveal the influence of various properties of words and sentences on cognitive processes involved in word recognition. When it comes to parafoveal processing, the *gaze-contingent boundary paradigm* (Rayner, 1975) has been used to investigate what types of information, and how much of it, readers are capable of extracting from the parafoveal preview. In this paradigm, an invisible boundary is triggered when a person's eyes cross it, causing the word on the opposite side of the boundary to change from a parafoveal *preview* to a foveal *target*. The increased speed with which a word is processed when a facilitative parafoveal preview was available compared to when it was denied is the *preview benefit* and has been estimated to be a reduction in gaze durations of about 20 – 50 ms (Rayner, White, Kambe, Miller, & Liversedge, 2003; for a Meta-analysis see Vasilev & Angele, 2017). Many studies have manipulated the type of visual and linguistic information available in the parafoveal preview and have found benefits of varying sizes for word length and orthography (see Schotter, Angele, & Rayner, 2012), and even semantics (see Schotter, 2018; Andrews & Veldre, 2019).

A *phonological preview benefit (PPB)* was originally reported by Pollatsek, Lesch, Morris, and Rayner (1992), who compared reading speeds on the target (i.e., “cent”) when the preview was a homophone (i.e., “sent”) versus an orthographically matched word (i.e., “rent”) they also included identical and unrelated preview conditions (i.e., “rack” condition; e.g., “The generous man gave every cent/sent/rent/rack to charity.”). They reported mean *first fixation durations* (FFD; duration in milliseconds of the first fixation on the target word) of 275 ms for the homophone preview and 295 ms for the orthographic control preview, which translates to a phonological preview benefit for the homophone compared to the visually similar word of ~20 ms. This reduction in reading time for targets with homophone previews compared to orthographically matched previews reflects the extent to which parafoveal phonological processing facilitates word recognition above and beyond the benefit afforded from orthographic similarity in the parafovea.

Inconsistencies in the Phonological Preview Benefit

A number of phonological preview studies have followed up on these findings (see Leininger, 2014 for a comprehensive review) and recently a Bayesian meta-analysis (Vasilev, Yates, & Slattery, 2019), which included a number of unpublished studies, was conducted in an attempt to estimate the effect size of the PPB. Among the studies included in this meta-analysis, some replicated the original parafoveal preview benefit effect (Miellet & Sparrow, 2004; Blythe, Dickins, Kennedy & Liversedge, 2018; Jouravlev & Jared, 2016; Leininger, 2018, Experiment 4), some replicated after segmenting the data by reading ability or age (Chace, Rayner & Well, 2005; Tiffin-Richards & Schroeder, 2015), and some failed to replicate altogether (Choi & Gordon, 2014; Leininger, 2018, Experiment 3). Ultimately, the meta-analysis concluded that there is a high probability that the PPB effect exists (> 92%) but that the size of the effect was relatively small (4.5 ms) compared to the 20 ms effect originally reported by Pollatsek et al. (1992).

The picture becomes a bit more complicated, however, with the consideration that these studies all had some variation in the types of homophones, the participant populations, and the languages used. This meta-analysis highlights the inconsistencies in the phonological preview benefit literature (e.g., varying effect sizes and some null results), but conflates across the individual studies, assuming that the variation is due to measurement error or a lack of reliability of the PPB. As previously discussed, recognizing words in a sentence is a multifaceted process that involves the integration of perceptual information, sentence context, and an individual's lexical representations to solve the complex problem of extracting meaning from printed symbols. Therefore, it is possible that the variation in these studies is due to systematic differences in the sampled participants and the characteristics of the experimental sentence stimuli.

Sentence Context

Some of these PPB studies reused or modified sentences from Pollatsek et al. (1992) and some used unique stimuli that potentially have slightly different properties. In either case, the sentences were not designed to minimize or maximize preactivation of phonological codes from contextually generated expectations (i.e., did not manipulate or control for sentence context). Therefore, it is possible that there is variability in sentence constraint across studies, which could interact with parafoveal processing of phonological codes.

Effects of sentence context on reading and word recognition are often studied by manipulating how constraining a sentence context is for a particular word, and therefore how predictable the word is based on the preceding semantic and syntactic information. A common method of determining the constraint of a sentence context is to use a cloze task (Taylor, 1953), in which participants are asked to produce the word that they most strongly expect to come next in the sentence. The *cloze probability* is then calculated as the probability of a particular word being produced across all participants. Eye tracking studies have demonstrated that the higher the cloze probability of a word in a given sentence context, the more likely a reader is to skip that word (Balota, Pollatsek, & Rayner, 1985; Drieghe, Brysbaert, Desmet, & De Baecke, 2004; White, Rayner & Liversedge, 2005), the shorter their fixation durations are (Inhoff, 1984; Rayner & Well, 1996), and the less likely they are to make regressions to that word (Erlich & Rayner, 1996). If these measures are taken as indicators of how easy a word is to recognize, then a supportive sentence context appears to facilitate recognition.

Converging evidence from event-related brain potentials (ERPs) supports this conclusion. The N400 component, which has been characterized as reflecting the process of accessing semantic representations from long term memory (see Federmeier, 2021), is reduced as cloze probability of a word increases (Wlotko & Federmeier, 2012). It remains an open debate whether this facilitation provided by strong sentence context reflects prediction of a specific lexical item or set of likely candidates (DeLong, Urbach & Kutas, 2005; Hodapp & Rabovsky, 2021; Kuperberg, Brothers, & Wlotko, 2020) or if it reflects easier integration of the encountered word into the sentence-level representation (Ito, Martin, & Nieuwland, 2017). If the parafoveal preview benefit is larger for a predictable word in a high constraint sentence, this would suggest that specific lexical items are predicted and that the prediction feeds down to the level of phonological form.

Lexical Frequency

It has been proposed that phonology may mediate orthographic-to-semantic processing in visual word comprehension via two different mechanisms: *addressed* phonology (relies on stored representations of phonological forms that is tied to holistic lexical items) and *assembled* phonology (relies on knowledge of spelling-to-sound correspondences; Patterson, 1986). The extent to which phonology is assembled appears to depend, in part, on the lexical frequency (i.e., familiarity) of the word, such that lower frequency words that have less precise specification of their component sublexical features require assembled phonological mediation (Walters, 1984, Coltheart, Avons, Masterson, & Laxon, 1991).

When it comes to phonological activation during silent reading, converging evidence also suggests that recognition of lower frequency words tends to depend more on the use of phonological codes for semantic access. For example, Jared and Seidenberg (1991) found that homophone effects on both false-positive error rates and reaction times in a semantic categorization task were exclusive to low frequency category exemplars. Lexical frequency is a robust and reliable predictor of both word skipping rates and fixation durations (see Rayner, 2009); higher frequency words are skipped more often and read faster than lower frequency words. Some models of word recognition (e.g., Dual Route Cascaded model; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) propose that recognizing lower frequency words involves a sublexical route that uses assembled phonology to access semantics. Therefore, it may be that phonological codes are only used during silent reading when a word cannot be easily recognized through a lexical route that proceeds directly from orthography to meaning.

ERPs show different effects of homophone errors for high and low frequency words as well. In the N400 component, spelling control anomalies elicit a more negative amplitude than homophone anomalies only when the sententially appropriate member of the homophone pair is low frequency (Newman, Jared, & Haigh, 2012). In fact, there is virtually no amplitude difference between the contextually correct word and the anomalous homophone, suggesting that the representations accessed during recognition are highly weighted toward phonology over orthography. Therefore, it is quite plausible that the PPB effect might only exist for low frequency words. PPB studies that found small or null effects may have used higher frequency items than those studies that report larger effects, but this variable has not been systematically investigated in the PPB literature.

Individual Differences in Lexical Quality

Another important consideration with respect to Vasilev et al.'s (2019) meta-analysis, is that there is likely variability in the participants' lexical quality both within and between the experiments in the analysis. One of the included studies did look at the effects of general reading ability on the PPB and found that it was present only for 'good readers' (Chace et al., 2005), which lends some evidence to the possibility that lexical quality influences the use of phonological codes in silent reading. Additionally, better spellers extract more information from parafoveal words in general (i.e., larger perceptual spans and increased skipping rates; Slattery & Yates, 2017). Since the other studies did not control for or investigate reading or spelling ability, it is possible that the average PPBs that have been reported are obscuring a more complicated picture in which better readers have larger effects and worse readers have smaller or null effects.

In experimental psychology research, the tradition has been to collect a sample of data from multiple participants and aggregate across individuals to detect the variability associated with a particular manipulation. This approach relies on the assumption that for the given effect of interest, people, in general, are more similar than they are different and that in statistically accounting for variability between people, we can tap into the fundamental mechanisms underlying a particular psychological process. In the last decade or so, however, language cognition researchers have begun to question this assumption and explore how differences between people can systematically change the way in which, or extent to which, they perform a particular cognitive process during language comprehension (e.g., Veldre & Andrews, 2015).

Of particular relevance to the current study, researchers investigating how reading strategies change depending on lexical quality have shown that individuals who are both above average spellers and readers are less reliant on supportive sentence context than individuals on the lower end of the spectrum of reading and spelling ability (Hersch & Andrews, 2011). Nevertheless, good readers and spellers did still show an advantage overall when the sentence context provided support for a particular word. Andrews and Bond (2009) also found evidence that worse spellers are more strongly influenced by sentence context when they have to make a judgment about whether a particular word was in a sentence they just read. Compared to better spellers, they exhibited a larger inhibition effect when the probed word was congruent in the sentence context and related to the meaning of the word they actually saw. Therefore, there is growing evidence that individuals with good lexical quality can perform bottom-up word recognition with minimal reliance on predictions from sentence context, while individuals with less precise lexical quality utilize these top-down sentence level representations to compensate for a reduced capacity for efficient bottom-up word recognition.

A number of studies have also investigated individual differences in phonological activation using homophones in lexical decision and semantic categorization tasks, as well as in sentence reading tasks in which the incorrect homophone is presented. In lexical decision tasks, homophones produce longer response times than non-homophones, which has been taken to suggest that in visual word recognition, homophones produce activation for both members of the pair which causes competition in accessing the correct lexical item (e.g., Pexman, Lupker, & Jared, 2001). Following up on these findings, Unsworth and Pexman (2003) used the same lexical decision task with homophones and also collected offline assessments of the participants' print exposure, reading comprehension, and vocabulary. Using these measures to compute a composite score of 'reader skill', they found opposite effects of homophony on error rates for more and less skilled readers. The less skilled group produced more errors to homophones compared to non-homophone controls, while the more skilled group actually had higher error rates to the controls. They propose that these patterns indicate that increased reading skill results in "more efficient orthographic–phonological mappings and less spurious phonological activation" (p. 77). Therefore, they echo the lexical quality hypothesis, which predicts that better readers have more precise and redundant connections between orthographic and phonological representations.

Yates and Slattery (2019) utilized a different approach to studying individual differences in phonological effects in a lexical decision task by manipulating the *phonological spread* of the words in the task. Phonological spread is a lexical variable that represents how phonologically similar a given word is to other words in the lexicon. It specifically measures how many of the phonemes in a word can be swapped out to produce a real word phonological neighbor (i.e., a word that shares all but one phoneme). They also assessed participants' orthographic quality (measured by spelling production from dictation and spelling error recognition), semantic quality (measured by vocabulary), and reading ability (measured by reading comprehension and reading speed of passages). They found that performance on both spelling tasks interacted with phonological spread when predicting reaction times on the lexical decision task, but in opposite directions. A follow-up principal components analysis revealed that participants who were better at spelling error recognition but worse at spelling production had larger phonological spread effects. They present a framework in which lexical quality has separate, but interacting, subcomponents associated with a *recognition system* that relies on orthography-to-phonology conversion and a *production system* that relies on phonology-to-orthography conversion. Therefore, they propose that people who are good at recognizing words would activate phonology from orthography faster in a lexical decision task, and if they also have weaker phonology-to-orthography connections, this phonological activation would also be slower to feed back to activate orthography. Based on this hypothesis, language expertise not only relies on the connections between orthographic and phonological representations, but also on how these connections vary based on the direction of the activation pathways between them.

Other interesting insights about individual differences in the role of phonology in visual word recognition come from comparisons between typical readers and individuals with dyslexia. In a semantic categorization task study, participants with dyslexia, compared to an age-matched control group, demonstrated larger false-positive error rates for both homophones and pseudohomophones of correct category exemplars (O'Brien, Van Orden, & Pennington, 2013). This pattern was actually in the opposite direction of what the authors hypothesized based on the assumption that dyslexia is characterized by poor phonology-orthography representations, which would reduce the activation of the phonological forms, resulting in fewer errors to homophone foils. A possible explanation of their exaggerated error rates is that the visual word forms do activate phonology but that less precise holistic lexical representations make it more difficult for dyslexic readers to suppress the incorrect homophone.

Homophones vs. Pseudohomophones

Finally, it is possible that the influence of sentence context and lexical quality on the PPB depends on whether the homophone preview is a real word or a pseudoword. The designs of the studies in Vasilev, et al.'s meta-analysis (2019) had some variability and instead of using real-word homophones as the parafoveal preview, some of them used non-word pseudohomophones (e.g., Miellet & Sparrow, 2004; Leininger, 2018, Exp.4). Leininger (2018) used survival analyses to estimate the earliest detectable effect of the PPB in first fixation duration and reported that this effect appeared earlier for pseudohomophones than homophones. One possibility is that real-word homophones cause some inhibition because activation of the semantics of the preview word could potentially result in competition when activating the meaning of the target. Therefore, pseudohomophones may provide a more direct test of the extent to which phonological codes are assembled during parafoveal processing. Furthermore, pseudohomophones do not have holistic lexical representations, so if they elicit a PPB, phonological activation during reading must occur prior to word recognition. When it comes to differences in lexical quality, we might also expect a larger dissociation between individuals with good and bad lexical representations for pseudohomophone previews. Those with stronger connections between orthographic, phonological, and semantic representations might extract more information from the parafovea and experience lexical competition to a larger degree or earlier than individuals with poorer lexical quality. Therefore, the difference in the size of the PPB between those with good and bad lexical quality may be larger for pseudohomophones than homophones.

The Present Study

The goal of this study was to provide a more nuanced and complete picture of the phonological parafoveal preview benefit in silent reading that takes sentence context, preview lexicality, and participant-level differences in lexical quality into account. The research questions specified a priori were:

1. Is phonology activated during parafoveal processing in silent reading to aid recognition efficiency for upcoming words?
2. Does sentence context generate predictions that feed down to the level of phonological form and do these predictions facilitate bottom-up processing of visual word forms in the parafovea?
3. Do individual differences in the precision and redundancy of component lexical representations (i.e., orthography, phonology, and semantics) influence the use of top-

down contextual information and bottom-up phonological information during parafoveal word recognition?

4. Does the phonological parafoveal preview benefit reflect post-lexical whole-word representations or pre-lexical assembled phonological-orthographic correspondences?

To answer these questions, I designed two eye tracking experiments utilizing the boundary paradigm (Rayner, 1975; see Figure 2) with real-word (Experiment 1) and pseudoword (Experiment 2) homophone and orthographic control parafoveal previews that changed to the correct word when the eyes crossed an invisible boundary (see example stimuli below; homophone and orthographic control words are in parentheses and the target words are shown in bold). In addition to the preview manipulation, I manipulated the sentence context so that the target word was either highly predictable (1a & 2a) or unpredictable (1b & 2b). In both experiments, both previews were matched on orthographic overlap with the target but the orthographic control had reduced phonological overlap compared to the homophone, but in Experiment 1 (1a & 1b) the previews were real word homophones, whereas in Experiment (2a & 2b) they were pseudohomophones and pseudoword orthographic controls.

(1a) The boy bought his crush a single red (rows/rods) **rose** for Valentine’s Day.

(1b) The thoughtful man bought his wife a beautiful (rows/rods) **rose** for her birthday.

(2a) The king and queen live in the large (cassul, casmol) **castle** in the countryside.

(2b) They went on a trip to see the ancient (cassul, casmol) **castle** that is in ruins.

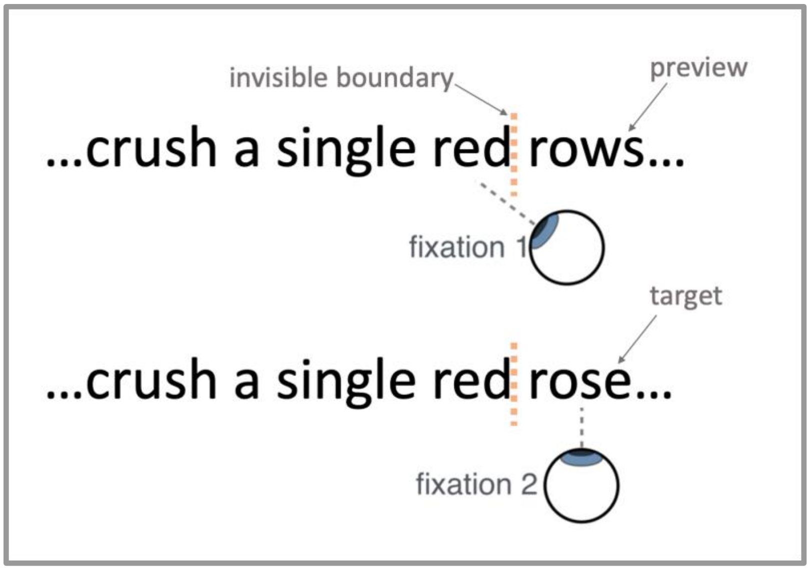

Figure 2. *Example of the Gaze Contingent Boundary Paradigm (Rayner, 1975)*
Note. This example shows a real-word homophone display change trial.

Participants also completed offline assessments of spelling ability, phonological decoding ability, and semantic knowledge to test whether the presence and magnitude of the PPB depends on an individual's lexical quality. My hypotheses and predictions of the pattern of results were as follows:

1. Phonological information is used to facilitate early stages of word recognition, which would be demonstrated by shorter fixation durations and higher skipping rates in the homophone compared to the orthographic control preview condition.
2. Expectations about upcoming words feed down to the level of phonology and these predictions facilitate processing of bottom-up phonological information, which would be demonstrated by larger PPB effects in fixation durations and skipping rates when the target word is predictable given the sentence context (i.e., in the high constraint compared to low constraint condition)

3. Individuals with stronger orthographic and phonological representations of words will generate stronger predictions about the lower-level features of a word and be able to use the preactivation of these features to more efficiently extract phonological form from the bottom-up input in the parafovea. This effect would manifest as a three-way interaction between either spelling ability or phonological decoding ability, sentence constraint, and preview condition in predicting fixation durations and skipping rates. If this interaction is significant for spelling ability, it would suggest that the PPB benefits from more efficient processing of the orthographic forms themselves, which would in turn allow earlier activation of phonological forms. Alternatively, if the significant interaction is with phonological decoding ability, it would suggest that the PPB is augmented by better connections between orthographic and phonological representations in memory, which would speed efficiency only once the orthographic form has been recognized.
4. Parafoveal processing of phonology can occur pre-lexically but may also be inhibited if the phonologically matched parafoveal preview has a lexical/semantic representation that is incongruent in the sentence context. Therefore, I predicted that phonological preview benefit will be observed in both Experiment 1 (real-word homophone previews) and Experiment 2 (pseudoword homophone previews), demonstrating both pre-lexical and post-lexical use of phonological representations for word recognition. However, if whole-word representations also play an important role in phonological parafoveal processing further downstream, I would expect an inhibition effect in Experiment 1 due to competition between the holistic lexical representations of the preview and the target words. Therefore, the PPB, overall, would be larger in Experiment 2 than Experiment 1. However, effects could be larger for real-word homophones if parafoveal phonological

activation only occurs once a lexical item has been recognized (i.e., if activation of phonology is post-lexical).

Additionally, this study involves multiple dependent measures (i.e., skipping rates, SFD, and GZD, as is typical for eye tracking in reading research. Differences in the patterns of effects between these different measures would be informative about the time course of the facilitation that phonological processing provides. Effects in skipping would be the strongest indicator that phonology is preactivated by context and that a matching phonological form perceived parafoveally is sufficient for lexical access even if the orthography is a mismatch. There is evidence that when words are skipped, this is often because they have been sufficiently identified prior to being fixated (Rayner & Duffy, 1988), so word skipping provides a valuable metric for looking at the effects of pre-activation from context and the extent to which phonological processing leads to word recognition in the parafovea. Any effect of the phonological preview manipulation on word skipping would indicate that phonological information is accessed quite early in parafoveal processing. If, however, the PPB only shows up in fixation duration measures, this would suggest that even when the context pre-activates phonological forms, the benefit it affords to word recognition only manifests once the predicted orthographic form is fixated. It is also possible that better phonological decoders would exhibit a PPB in skipping because they are better at converting orthographic forms to phonological forms and may be more efficient at activating phonology parafoveally.

Chapter 2: Experiment 1 - Real-Word Previews

Because Vasilev, et al. (2019) included both homophone preview and pseudohomophone preview studies in their meta-analysis, I conducted two separate experiments that separated these preview types. Furthermore, the use of these two types of previous is informative about whether the PPB reflects a pre-lexical or post-lexical benefit from phonological activation early in the word recognition process. Experiment 1 utilized real-word homophones and orthographic control words as previews.

Method

Participants

One hundred and fourteen native English speakers with normal or corrected-to-normal vision and no history of neurological disorders, reading or learning disabilities participated in Experiment 1. The majority were recruited from the Psychology Department's SONA subject pool at the University of South Florida and compensated with course credit. I also actively recruited higher skilled readers to fill out the high end of the distribution of lexical quality in the sample, so six of the subjects were considered part of this special population and paid $10/hour for their time. Two participants were excluded due to low accuracy in comprehension questions, 3 were excluded due to failure to follow instruction, 1 due to a technical error interrupting the experiment, 1 due to excessive data loss (from a combination of blinks, display change errors, and track loss), and 4 were excluded due to experimenter error in the assessments or eye tracking recording. After these exclusions, 103 participants were retained for the analyses. Based on a power analysis conducted by Vasilev et al. (2019) using the data from their meta-analysis, with ~30 items per condition and ~100 participants, the current study would have a power of around 0.6 - 0.7 (assuming the PPB effect size of 4.5 ms). However, other studies investigating individual differences in the parafoveal preview benefit based on lexical quality have used participant sample sizes of ~100 (e.g., Veldre & Andrews, 2015) and found significant main effects and interactions with lexical quality measures.

Materials and Design

Sentence Stimuli. The experimental design was a fully crossed 2 (sentence constraint: high vs. low) x 2 (preview type: homophone vs. orthographic control) within-subjects, within-items design. To construct the sentence stimuli, 56 homophone word pairs of matched length were selected, with each member of the pair serving as both a target word and a homophone preview in separate stimulus items, yielding 112 total experimental items. Orthographic control words of the same length as the target word were selected for each item and matched as closely as possible to the homophone on lexical frequency, number of letters shared with the target, visual form (e.g., ascenders and descenders), and vowel-consonant pattern, in that order of priority. Four counterbalanced lists were created so that each participant saw each item in only one of the four conditions, resulting in 28 items per condition. Items were presented in a randomized order for each participant. Sentences were constructed around the target words so that each word item had a sentence with high constraint toward the target word and a sentence with neutral constraint that did not lead to predictions of any one word. In addition to the experimental stimuli, frequency manipulated filler sentences with no display change were included to obscure the design of the study. These items were low constraint sentences with target words that were either high or low frequency, taken from Schotter and Leininger (2016).

Norming. Sentences were normed for plausibility and sentence constraint (cloze probability) on Amazon Mechanical Turk on a separate set of participants from the target population. The purpose of the norming was to ensure that the sentences in each constraint condition were equally plausible and that the constraint conditions were functioning as intended to either lead toward the prediction of a particular word or be neutral. For plausibility norming, participants were asked to rate the sentences on how well written they were on a 7-point Likert scale from 1 (i.e., Very Poorly Written) to 7 (i.e., Very Well Written). For cloze norming, participants were given sentence fragments from the beginning of the sentence until the word immediately before the target word and asked to fill in a blank with the word that they would expect to come next in the sentence (See Appendix A for norming task instructions). Norming stimuli were counterbalanced across lists so that each participant only saw one version of each item, balanced across conditions, and 10 responses were collected from separate participants for plausibility rating and cloze response for each sentence. Cloze probability was calculated by finding the proportion of subjects (out of 10) who responded with the target word. Plurals and misspellings of the target word were also accepted. See Table 1 for norming results.

Table 1. *Descriptive Statistics of Plausibility and Cloze Norming Results for Sentence Stimuli by Condition for Experiment 1*

	Cloze Probability	**Plausibility**		
	Target	Target	Homophone	Orthographic
High Constraint	0.75 (0.18)	5.36 (0.51)	3.72 (0.64)	3.22 (0.54)
Low Constraint	0.02 (0.04)	5.29 (0.61)	3.45 (0.70)	3.23 (0.64)

Note. Means are reported with standard deviations in parentheses.

Lexical Quality Assessments. In order to measure individual differences in lexical quality, I administered assessments of spelling, phonological decoding, and semantic knowledge (made up of reading comprehension and vocabulary; see Table 2). In the spelling production task, participants were asked to produce the correct spelling of a word presented auditorily (and used in a sentence) from a recording. The spelling recognition task required participants to identify, from a list of items printed on paper, any words that were misspelled. Half of the items were correct, and half were incorrect. The dependent measure from this task is a d-prime measure of sensitivity to misspellings. All of the phonological decoding tasks involved reading aloud from a printed list. The tasks differed in the types of items read aloud (i.e., real words, nonwords, and words with irregular orthographic-to-phonological correspondences). In the reading comprehension task, participants read a sentence and then selected from a set of 4 pictures which image best represented the meaning of the sentence they had read. Finally, the vocabulary task was a printed multiple-choice test in which participants had to circle the word (4 choices) that was closest in meaning (i.e., a synonym) to a target word.

For the spelling (recognition & production) and vocabulary tasks, subjects wrote their own answers on an answer sheet. In the spelling production task, words were played aloud from a recording and subjects were instructed to write down the correct spelling to the best of their ability. For the reading comprehension task, indicated their answer aloud or by pointing to the correct answer in the printed booklet while the experimenter recorded their responses on an answer sheet. For all three of the phonological decoding tasks, the experimenter used an audio recording device to record the verbal responses of the participant and a stopwatch to keep track of the reading time. After each session, these recorded responses were then coded in a spreadsheet for total number correct and total reading time. Scoring for the assessments followed the scoring standards established for each test. For the novel Irregular Word reading task developed for this study, the final scores were computed by dividing the number of correctly read words by the number of seconds taken to read the entire list, giving a words/second reading rate.

Table 2. *Individual Differences Assessment Tests and Measured Constructs*

Measured Construct	*Test Name*	*Source*
Spelling (recognition)	88-item Spelling Recognition Test	Andrews & Hersch (2010)
Spelling (production)	Woodcock-Johnson III Tests of Cognitive Abilities: Spelling Sub-test	McGrew & Woodcock (2001)
Phonological Decoding (words)	TOWRE-2 Test of Word Reading Efficiency–Second Edition: Phonemic Decoding Sub-test	Torgesen, Wagner, & Rashotte (2012)
Phonological Decoding (nonwords)	TOWRE-2 Test of Word Reading Efficiency–Second Edition: Sight Word Efficiency Sub-test	Torgesen, Wagner, & Rashotte (2012)
Phonological Decoding (irregular words)	Irregular Word Reading test	developed for this study; see Appendix B
Semantic Knowledge (reading comprehension)	Peabody Individual Achievement Test (PIAT) Reading Comprehension Sub-test	Markwardt (1989)
Semantic Knowledge (vocabulary)	Shipley Institute of Living Scale: Vocabulary Sub-test	Shipley, Gruber, Martin, & Klein (2009)

Procedure

The experiment was run on each participant individually (with a single experimenter) in quiet rooms in one testing session. The session consisted of two parts: the administration of assessments, followed by the sentence-reading eye tracking portion.

Assessments. During the assessment portion, the experimenter had the participant sit at an empty table in a testing room and administered each assessment from a booklet on which the materials were printed.

Sentence Reading & Eye Tracking. After the assessment portion of the experiment, subjects were moved to the eye tracking testing room and seated in front of an HP p1230 CRT monitor (screen resolution = 1024 x 768 pixels, refresh rate = 150 Hz), with their eyes 60 cm from the screen. Eye movements were recorded using an SR Research Ltd. Eyelink 1000 eye tracker (sampling rate of 1000 Hz) in tower setup with forehead and chin rests that restrained head movements. To ensure accurate eye tracking, the participants completed a 3-point horizontal calibration procedure that appeared at the vertical midpoint of the screen. The calibration was accepted if it had a maximum error of 0.30 degrees of visual angle at each calibration point. Calibration was re-administered periodically throughout the experiment as needed, determined by a drift check before each trial at a single fixation point in the middle of the screen. Additionally, before each trial, the subject fixated a box on the right-hand side of the screen in the location that the first word of the sentence would appear. Once the eye tracker registered a fixation in the box, the sentence would automatically appear and the trial recording period would begin.

The gaze contingent boundary replicated the procedure described in Experiment 2 in Pollatsek, et al. (1992). The boundary was located immediately after the last letter of the pre-target word and the preview changed to the target word within approximately 5-10 ms of the eye tracker detecting that the boundary had been crossed. In most cases, this display change is fast enough that people are not able to detect the change and do not (consciously) notice anything unusual. Sentences were presented in a randomized order. The sentence text appeared on the screen in black 12-point Courier New (monospaced) font on a white background with 2.66 characters subtending 1 degree of visual angle. Both experimental and filler sentences were followed by yes/no comprehension questions on 25% of trials, with equal numbers of yes and no correct responses, to ensure that participants were reading the sentences for meaning. Participants were given a break halfway through the sentence reading section and at any other time that they requested one. At the end of the sentence reading section, participants were debriefed and asked if they noticed anything unusual about the words or sentences on the screen. If they responded yes, they were asked what they noticed and if they mentioned anything about a display change or words flickering, they were also asked on how many trials they noticed this change occurring. Based on these self-report responses, three participants from Experiment 1 and five from Experiment 2 noticed flickering or letters changing on at least 20% of the trials. Because this study is investigating individual differences, I decided not to exclude these participants from analysis because their ability to detect the changes may be related to their language skills and depth of parafoveal processing.

Data Processing & Analysis

Individual Differences Measures

After all of the data had been collected, assessment scores were converted into z-scores based on data from all participants. To confirm that the tasks were in fact measuring three unique latent variables (i.e., the intended constructs), I conducted a confirmatory factor analysis (Figure 3) with maximum likelihood using the lavaan package (Rosseel, 2012) the R Environment for Statistical Computing in RStudio (version 1.1.456) to test the fit of the model. The analysis was done using the combined assessment data from Experiments 1 and 2.

R Code (lavaan) for the Hypothesized Model:

```
CFA_model1 <-'
phono =~ NA*TOWREwords+TOWREnonwords+Irregular_W
phono ~~ 1*phono
ortho =~ NA*Spelling+Spelling_Rec
ortho ~~ 1*ortho
semantic =~ NA*Vocabulary+Reading_Comp
semantic ~~ 1*semantic
semantic ~~ ortho
semantic ~~ phono
phono ~~ ortho'
```

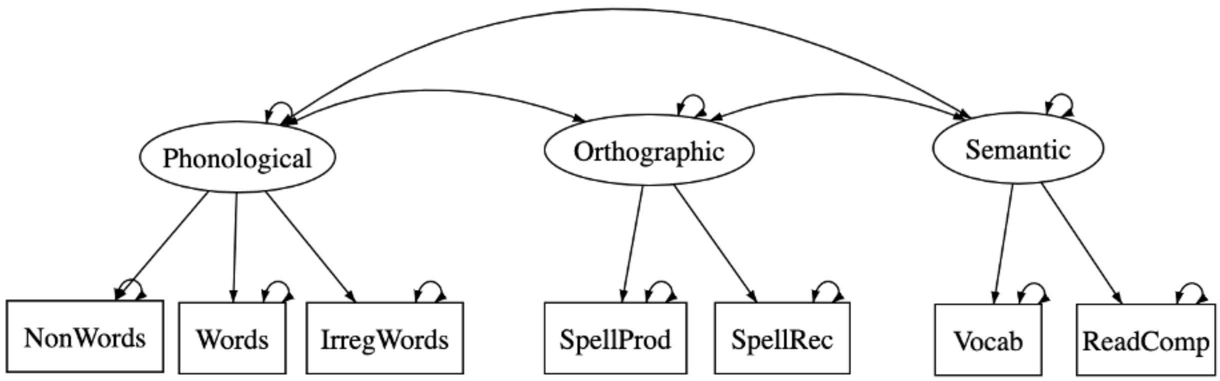

Figure 3. *Lexical Quality Measurement Model Diagram (Original Hypothesized Model)*

The χ^2 was significant (χ^2 (d*f*=11)=72.43, *p*<0.001) for the proposed model and the alternative fit indices also indicated that the fit was not good. The Comparative Fit Index (CFI) was 0.87, Root Mean Square Error of Approximation (RMSEA) was 0.18, and the Standardized Root Mean Square Residual (SRMR) was 0.16. Because the model did not demonstrate good fit, modification indices were computed. Modification indices suggested that model fit could be improved by adding a path from both the Spelling Ability and the Semantic Knowledge latent factors to the Irregular Word Reading measure, which was originally specified as an indicator for Phonological Skill. Because the Irregular Word Reading measure is a novel measure that has not been previously used or validated as a measure of phonological decoding ability, it is conceivable that it is tapping into a variety of skills that overlap between the lexical quality constructs. In fact, the task involves correctly pronouncing somewhat rare and irregular words, so it makes theoretical sense that it would actually be measuring vocabulary knowledge as well as phonological decoding ability (which is more related to the knowledge of spelling-to-sound correspondences). Based on the modification indices and on reconsideration of the nature of the novel task, a second model was constructed in which Irregular Word Reading was removed from the model altogether.

For the second, respecified model, the χ^2 was not significant (χ^2 (d*f* =11) = 7.56, p = 0.75), indicating good model fit. The alternative fit indices also showed excellent fit (CFI = 1.00, RMSEA = 0.00, and SRMR = 0.03). Additionally, the standardized residual correlations were all under 2, indicating good local fit as well. Therefore, the three constructs of interest were better represented as unique factors when the Irregular Word Reading task was excluded. Based on this analysis, I decided to exclude that task from theour composite score predictors. Composite scores were therefore calculated for the constructs of spelling ability, phonological decoding ability, and semantic word knowledge by averaging the z-scores for the tasks associated with that construct (see Table 2), except for Irregular Word Reading, which was not averaged into the phonological decoding scores.

Eye Tracking Measures

Raw eye tracking data was processed via Robodoc and EyeDry software (http://blogs.umass.edu/eyelab/software/), which output trial-level eye movement measures for the target word region that are commonly used in research that uses eye tracking to study reading. The primary dependent measures of interest in this study are skipping rate, single fixation duration (SFD) and gaze duration (GZD). Skipping rate is calculated as the proportion of trials on which the target word is not fixated during first-pass reading. SFD is the duration (ms) of the first fixation on the target word, when the word is only fixated once in first-pass reading (i.e., before the eyes leave to the right or left). GZD is the sum of all fixations on the target word before the eyes leave to fixate another word (includes re-fixation durations)

Results

Analyses were conducted using linear mixed-effects models via the lmer function from the lme4 package (version 1.1-12; Bates et al., 2015) within the R Environment for Statistical

Computing (version 3.3.1). Separate models were used to predict the three dependent eye-tracking measures (skipping rate, SFD, and GZD) by sentence constraint, preview condition, scores for the measured lexical quality constructs, and their interactions. The preview effect was contrast coded (i.e., -0.5, 0.5), while the constraint effect was coded as a treatment contrast (i.e., 0, 1) with high constraint as the baseline because I predicted that the effect of preview would be largest in high constraint sentences. All analyses used the full random effects structure for the experimentally manipulated variables (i.e., preview, sentence constraint, and their interaction as random effects for participants and items).

Skipping Rate

As predicted, there were significant main effects of both parafoveal preview type and sentence constraint on skipping rate (see Table 3). These effects were in the expected direction, with higher rates of skipping when the preview was a homophone of the target word and when the sentence made the target word more predictable (Figure 4A). Because I used a treatment contrast for constraint in which high constraint was the baseline, the main effect of preview reflects the effect of the preview manipulation in high constraint sentences. The analysis also revealed a main effect of spelling ability such that better spellers skipped the target word more often (Figure 5A). However, I found no significant interactions between the experimental variables and individual differences measures.

Table 3. *Results of LMER Model Predicting Eye Tracking Measures by Parafoveal Preview Condition, Sentence Constraint, and Lexical Quality Composite Measures (Experiment 1)*

	Skipping Rate				Single Fixation Duration				Gaze Duration			
Predictors	*Est.*	*SE*	*z/t*	*p*	*Est.*	*SE*	*z/t*	*p*	*Est.*	*SE*	*z/t*	*p*
(Intercept)	0.34	0.05	-8.06	**<0.001**	250.78	3.66	68.50	**<0.001**	267.13	3.36	79.54	**<0.001**
Preview (Homophone vs. Control)	0.74	0.08	-2.94	**0.003**	13.25	3.45	3.84	**<0.001**	16.35	3.77	4.34	**<0.001**
Constraint (High vs. Low)	0.78	0.08	-2.52	**0.012**	14.72	3.20	4.60	**<0.001**	16.10	3.48	4.63	**<0.001**
Phonological Decoding	1.14	0.18	0.82	0.413	-15.33	4.64	-3.30	**0.001**	-15.61	4.87	-3.21	**0.001**
Semantic Knowledge	1.04	0.16	0.24	0.807	-6.49	4.52	-1.44	0.151	-8.84	4.75	-1.86	0.063
Spelling Ability	1.34	0.20	2.00	**0.046**	-0.42	4.31	-0.10	0.922	1.47	4.54	0.32	0.746
Preview x Constraint	1.22	0.17	1.40	0.160	-4.43	4.94	-0.90	0.370	-7.21	5.49	-1.31	0.189
Preview x Phonological Decoding	0.93	0.11	-0.61	0.545	-5.82	4.70	-1.24	0.216	-3.54	5.36	-0.66	0.509
Constraint x Phonological Decoding	1.09	0.09	1.05	0.295	0.67	3.38	0.20	0.844	-2.40	3.86	-0.62	0.534
Preview x Semantic Knowledge	1.05	0.12	0.41	0.683	-2.24	4.46	-0.50	0.616	-3.68	5.13	-0.72	0.473
Constraint x Semantic Knowledge	0.88	0.07	-1.53	0.126	-3.78	3.24	-1.17	0.244	-4.12	3.72	-1.11	0.268
Preview x Spelling Ability	0.99	0.10	-0.14	0.892	8.18	4.26	1.92	0.055	8.28	4.94	1.68	0.094
Constraint x Spelling Ability	1.08	0.08	1.01	0.311	-4.70	3.08	-1.53	0.127	-2.24	3.55	-0.63	0.528
Preview x Constraint x Phonological Decoding	1.29	0.22	1.53	0.125	7.42	7.12	1.04	0.298	4.19	8.02	0.52	0.602
Preview x Constraint x Semantic Knowledge	0.93	0.15	-0.47	0.641	-1.95	6.81	-0.29	0.774	2.57	7.71	0.33	0.739
Preview x Constraint x Spelling Ability	0.96	0.15	-0.27	0.785	-9.62	6.47	-1.49	0.137	-9.70	7.35	-1.32	0.187

Note. All analyses produced singular fits.

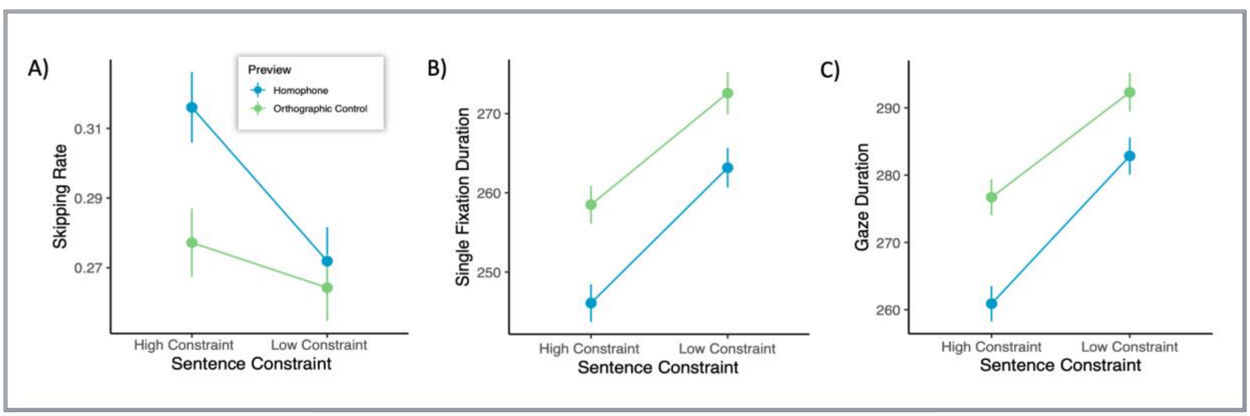

Figure 4. *Overall Effects of Preview Condition and Sentence Constraint on Skipping Rate, SFD and GZD (Experiment 1)*

Note. Error bars represent standard error and points represent means.

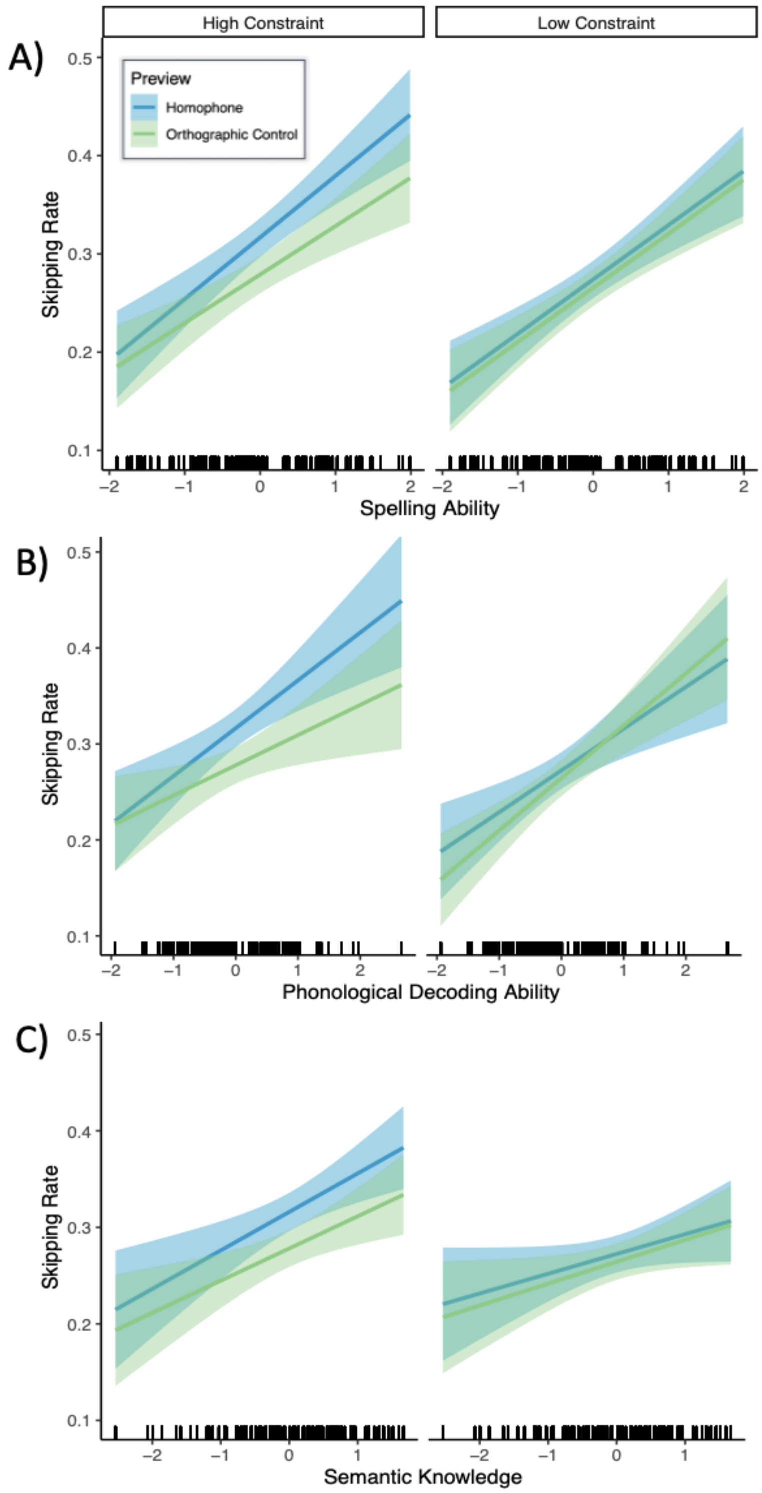

Figure 5. *Skipping Rates as a Function of Constraint and Lexical Quality Measures (Experiment 1)*

Single Fixation Duration

In SFD, I again found significant main effects of both preview and sentence constraint in the expected directions (Table 3; Figure 4B). As demonstrated by the test statistics, these effects appeared to be larger than for skipping rate. However, it is worth noting that the dependent measure of skipping in trial-level data is binomial, which has reduced power compared to the continuous fixation duration measures. Instead of the main effect of spelling ability observed in skipping rates, there was a large main effect (~15 ms) of phonological decoding ability such that better decoders had shorter SFDs. In SFD, there were also no significant interactions between any of the experimental or individual differences variables. However, the interaction between preview and spelling ability approached significance ($p = .055$) and visual inspection of the effect (Figure 6A) suggests that, particularly in high constraint sentences, individuals with better spelling ability had a larger phonological preview benefit in SFD.

Gaze Duration

Statistically, patterns in GZD mirrored those in SFD, with significant main effects of preview, constraint, and phonological decoding ability (Table 3; Figure 4C). Again, the interaction between spelling ability and preview condition approached significance ($p = 0.09$; see Figure 7A) and there was also a marginally significant main effect of semantic knowledge ($p = .06$) on GZD.

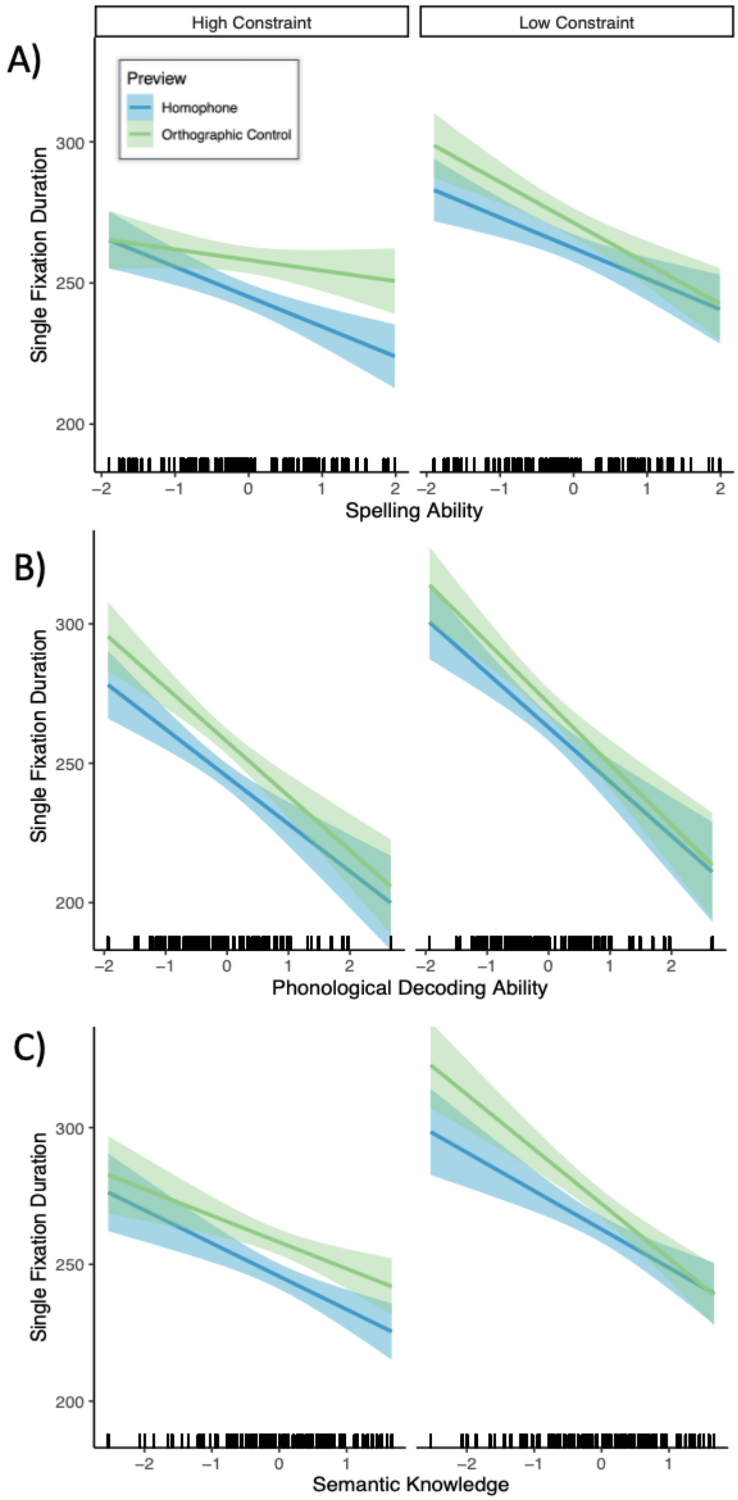

Figure 6. *SFD as a Function of Constraint and Lexical Quality Measures (Experiment 1)*

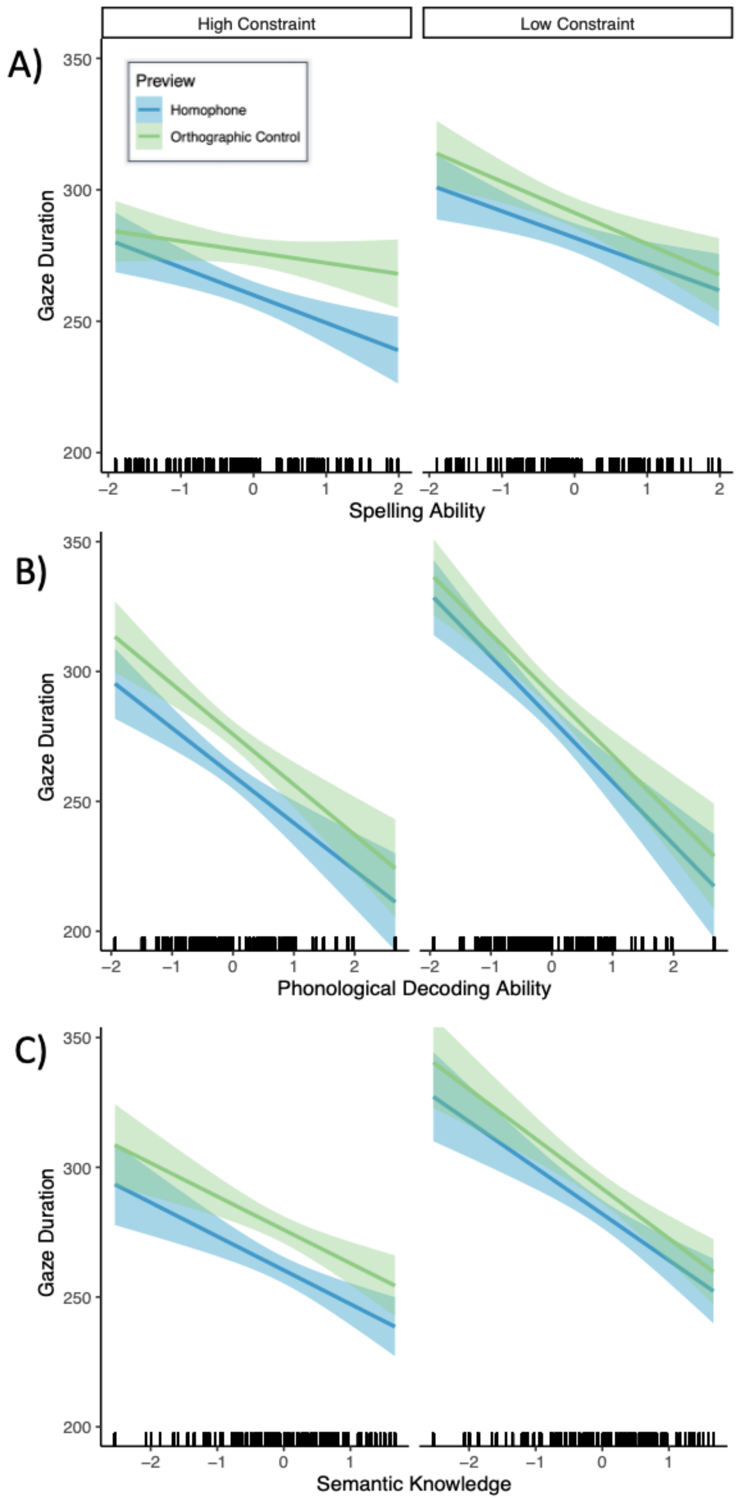

Figure 7. *GZD as a Function of Constraint and Lexical Quality Measures (Experiment 1)*

Exploratory Analysis of Target Word Frequency Effects

Based on prior evidence that lexical frequency may influence the extent to which phonological codes are used during visual word recognition (e.g., Jared and Seidenberg, 1991; Newman, Jared, & Haigh, 2012), I performed exploratory analyses to investigate potential interactions between the PPB, sentence context, and lexical frequency. These findings from semantic categorization error rates and N400 amplitudes suggest that activation of phonology from orthographic forms may be more pronounced when a word is unfamiliar (i.e., low frequency). This makes sense under the framework of a dual route model in which assembled phonology aids semantic access for low frequency, but not necessarily high frequency words.

I was interested in whether the PPB might be dependent on the lexical frequency of the encountered word and the extent to which expectations of low frequency words encourage the pre-activation of their phonological forms. Therefore, I performed a post-hoc analysis predicting each of the dependent eye tracking measures by the experimentally manipulated variables (i.e., preview type and sentence constraint) as well as the lexical frequency of the target word and their interactions.

Table 4. *Results of LMER Model Predicting Dependent Eye Tracking Measures by Preview Condition, Sentence Constraint, and Relative Target Word Lexical Frequency (Experiment 1)*

	Skipping Rate				**Single Fixation Duration**				**Gaze Duration**			
Predictors	*Est.*	*SE*	*z/t*	*p*	*Est.*	*SE*	*z/t*	*p*	*Est.*	*SE*	*z/t*	*p*
(Intercept)	0.20	0.07	-4.95	**<0.001**	237.46	5.57	42.60	**<0.001**	254.21	7.22	35.21	**<0.001**
Preview (Homophone vs. Control)	0.77	0.28	-0.72	0.470	27.35	12.11	2.26	**0.024**	38.56	13.30	2.90	**0.004**
Constraint (High vs. Low)	1.00	0.33	-0.01	0.990	37.03	10.82	3.42	**0.001**	33.97	11.94	2.85	**0.004**
Target Word Frequency	1.19	0.12	1.75	0.080	4.67	2.05	2.28	**0.023**	4.48	2.20	2.04	**0.042**
Preview x Constraint	1.06	0.50	0.12	0.905	-28.92	16.38	-1.77	0.077	-44.36	18.09	-2.45	**0.014**
Preview x Target Word Frequency	0.98	0.11	-0.13	0.894	-4.97	3.94	-1.26	0.207	-7.71	4.31	-1.79	0.074
Constraint x Target Word Frequency	0.92	0.10	-0.78	0.434	-7.42	3.46	-2.14	**0.032**	-5.91	3.82	-1.55	0.122
Preview x Constraint x Target Word Frequency	1.05	0.16	0.31	0.759	8.55	5.26	1.62	0.105	12.71	5.80	2.19	**0.029**

Note. Relative target word frequency was computed by subtracting the log lexical frequency of the homophone from the log lexical frequency of the target word for each individual item.

Because Experiment 1 used real words as previews, I was able to compute the frequency difference between the target word and the preview to investigate the relative role of top-down predictions and bottom-up processing in phonological activation. I did not find any interactions between preview, frequency, and sentence constraint in skipping rate, but there were significant interactive effects of relative target word frequency in SFD and GZD (Table 4). The patterns are very similar between the two fixation measures (see Figures 8B & 8C), but they are most pronounced in GZD, so I will focus my attention on these effects. All three factors (preview type, sentence constraint, and relative target word frequency) had significant effects on GZD such that fixation times were shorter for homophone previews, high frequency sentences, and high frequency target words. Additionally, when controlling for target word frequency, there was a significant interaction between preview condition and sentence constraint. As is clear from Figure 8C, the PPB has the largest effect when the sentence is strongly constraining toward the target word (and therefore the phonological form). Furthermore, this effect interacts significantly with relative target word frequency, such that the PPB is most pronounced when the target word is predictable and when that word is lower frequency.

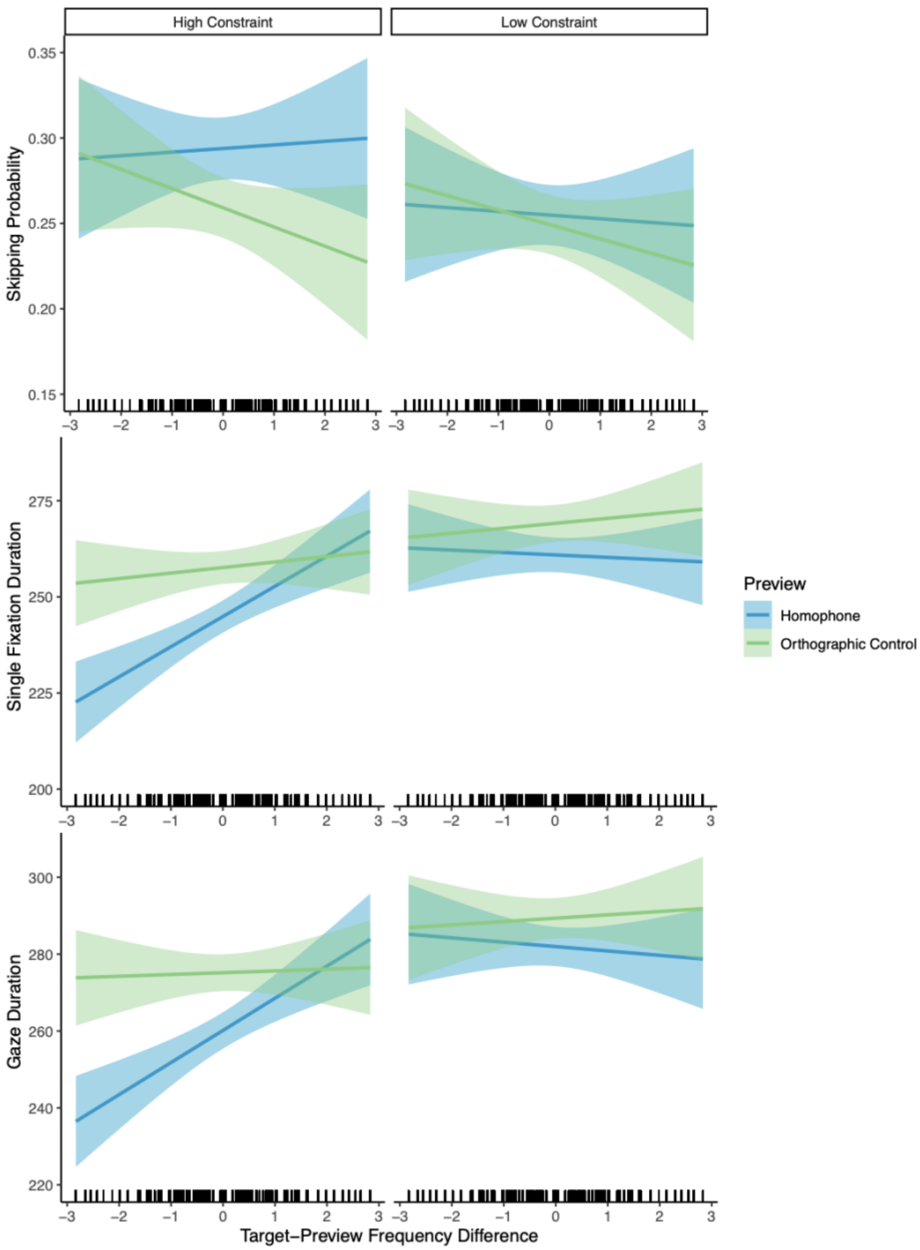

Figure 8. *Skipping Rates, SFD, and GZD by Sentence Constraint and Relative Target Word Frequency (Experiment 1)*

Summary of Experiment 1

Overall, the results of Experiment 1 demonstrated robust main effects of preview condition and sentence constraint in all three eye tracking measures, providing strong evidence that phonological information extracted from the parafovea aids word recognition. Also, phonological decoding was a robust predictor of both fixation duration measures (but not skipping rates). It is worth keeping in mind that all of the target words had homophones. Therefore this main effect of phonological decoding on fixation durations may actually be specific to homophones and not necessarily generalizable to all words.

The follow-up analysis investigating the role of lexical frequency in the PPB demonstrated that, at least for real-word parafoveal previews, the extent to which a word is strongly represented in the lexicon due to repeated encounters with it interacts not only with parafoveal processing of phonology, but also the extent to which preactivation of that word from context feeds down to lower levels of representation (i.e., phonological form).

Chapter 3: Experiment 2 - Pseudoword Previews

The purpose of Experiment 2 was to test whether there were differences in the magnitude of the PPB depending on whether the preview manipulation used pseudowords instead of real words. Pseudowords have no holistic representation in the lexicon, so PPB effects would have to be due to assembled phonology from learned orthography-to-phonology correspondences rather than activation of phonology from a stored lexical item. One advantage to using pseudowords over words in PPB studies is that there are a limited number of same-length homophones in the English language, which can make generating enough stimuli to test an effect (with a decently powered design) challenging, particularly if one is interested in interactions between the PPB and other factors. Therefore, in addition to the theoretical motivation for Experiment 2, there are methodological implications depending on whether the effects differ when using pseudohomophones compared to real-word homophones.

Method

Participants

The inclusion and exclusion criteria for participation were identical to Experiment 1. One hundred and fourteen students from the University of South Florida participated in Experiment 2. Nine of them were recruited as ‘good readers’ and compensated with monetary payment (see Experiment 1) and the remaining participants signed up through the SONA subject pool. Six were excluded due to experimenter error (i.e., failure to save data files or incorrect administration of assessments), one due to data loss from blinks, display change errors, etc., and one because it was determined after participating that they were not a native English speaker. Therefore, 106 participants were included in the analyses.

Materials and Design

The design of Experiment 2 was identical to that of Experiment 1, except for the fact that the parafoveal preview words, in both preview conditions, were made-up pseudowords.

Sentence Stimuli. One hundred and twelve words for which pseudohomophones could be created were selected as target words. Sentences were constructed around each target word that were either highly constraining to the target or neutral, resulting in two sentences per target word (224 total). Pseudohomophones were length-matched non-words that had the same phonology as the target word but different orthography (e.g., *oshin* as the pseudohomophone of *ocean*). Orthographically matched non-word previews were created by swapping out the letters of the pseudohomophone that differed from the target (e.g., *omtin*), so that the orthographic similarity between the two preview conditions and the target were as closely matched as possible.

Norming. Norming procedures and calculations were identical to Experiment 1. See Table 5 below for results.

Table 5. *Descriptive Statistics of Plausibility and Cloze Norming Results for Sentence Stimuli by Condition for Experiment 2*

	Cloze Probability	**Plausibility**		
	Target	Target	Homophone	Orthographic
High Constraint	0.77 (0.17)	5.22 (0.56)	3.12 (0.46)	2.91 (0.46)
Low Constraint	0.03 (0.06)	5.21 (0.53)	3.03 (0.45)	3.90 (0.38)

Note. Means are reported with standard deviations in parentheses.

Lexical Quality Assessments. Individual differences measures and assessment procedures were identical to Experiment 1.

Procedure

The experimental procedure was identical to Experiment 1.

Data Processing & Analysis

Data processing and analysis procedures were identical to Experiment 1.

Results

The analysis approach, including contrast coding and random effects structures, was identical to Experiment 1.

Skipping Rate

In contrast with Experiment 1, the preview manipulation did not have a significant effect on skipping rates (Table 6). Sentence constraint, however, remained a significant predictor of target word skipping (Figure 9A). Interestingly, sentence constraint also interacted with phonological decoding ability such that phonological decoding ability appeared to have a greater effect on skipping rates when a sentence is less constraining toward the target word (Figure 10B).

Table 6. *Results of LMER Model Predicting Dependent Eye Tracking Measures by Parafoveal Preview Condition, Sentence Constraint, and Lexical Quality Composite Measures (Experiment 2)*

	Skipping Rate				Single Fixation Duration				Gaze Duration			
Predictors	*Est.*	*SE*	*z/t*	*p*	*Est.*	*SE*	*z/t*	*p*	*Est.*	*SE*	*z/t*	*p*
(Intercept)	0.21	0.03	-10.89	**<0.001**	234.42	3.26	71.95	**<0.001**	252.63	4.20	60.22	**<0.001**
Preview (Homophone vs. Control)	0.95	0.11	-0.45	0.649	6.19	2.85	2.17	**0.030**	8.00	3.66	2.18	**0.029**
Constraint (High vs. Low)	0.69	0.08	-3.41	**0.001**	18.45	2.51	7.35	**<0.001**	19.43	3.00	6.47	**<0.001**
Phonological Decoding	1.07	0.18	0.42	0.673	-6.58	4.02	-1.64	0.102	-11.34	5.22	-2.17	**0.030**
Semantic Knowledge	1.24	0.19	1.37	0.172	-5.46	3.80	-1.44	0.151	-9.82	4.93	-1.99	**0.047**
Spelling Ability	1.19	0.18	1.10	0.271	-4.50	3.81	-1.18	0.238	3.51	4.92	0.71	0.476
Preview x Constraint	0.96	0.16	-0.22	0.822	-0.68	4.56	-0.15	0.882	-6.16	4.95	-1.24	0.213
Preview x Phonological Decoding	0.91	0.13	-0.67	0.501	0.97	4.10	0.24	0.813	1.70	4.98	0.34	0.733
Constraint x Phonological Decoding	1.25	0.13	2.11	**0.034**	-3.07	3.02	-1.02	0.309	0.00	3.51	0.00	0.999
Preview x Semantic Knowledge	1.17	0.15	1.25	0.210	-1.71	3.88	-0.44	0.660	-1.90	4.68	-0.41	0.684
Constraint x Semantic Knowledge	0.92	0.09	-0.93	0.351	5.06	2.83	1.79	0.074	9.28	3.29	2.82	**0.005**
Preview x Spelling Ability	1.04	0.13	0.29	0.773	-6.90	4.01	-1.72	0.086	-5.23	4.80	-1.09	0.276
Constraint x Spelling Ability	0.99	0.09	-0.09	0.927	0.62	2.93	0.21	0.833	-9.64	3.35	-2.88	**0.004**
Preview x Constraint x Phonological Decoding	1.11	0.21	0.56	0.573	-2.75	6.26	-0.44	0.660	-7.02	7.04	-1.00	0.319
Preview x Constraint x Semantic Knowledge	1.05	0.18	0.30	0.762	1.21	5.88	0.21	0.837	6.53	6.60	0.99	0.323
Preview x Constraint x Spelling Ability	0.77	0.13	-1.51	0.132	9.81	6.06	1.62	0.106	3.87	6.73	0.58	0.565

Note. All analyses produced singular fits.

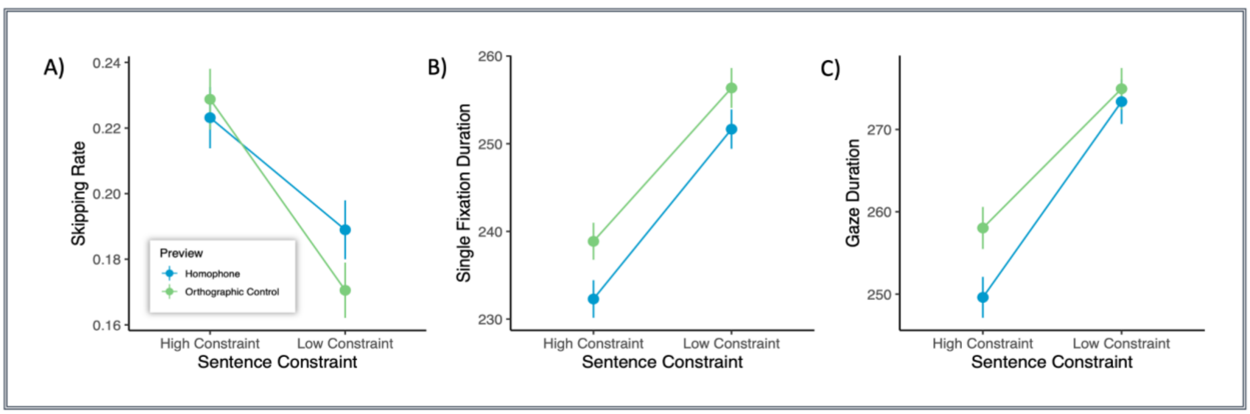

Figure 9. *Overall Effects of Preview Condition and Sentence Constraint on Skipping Rate, SFD and GZD (Experiment 2)*

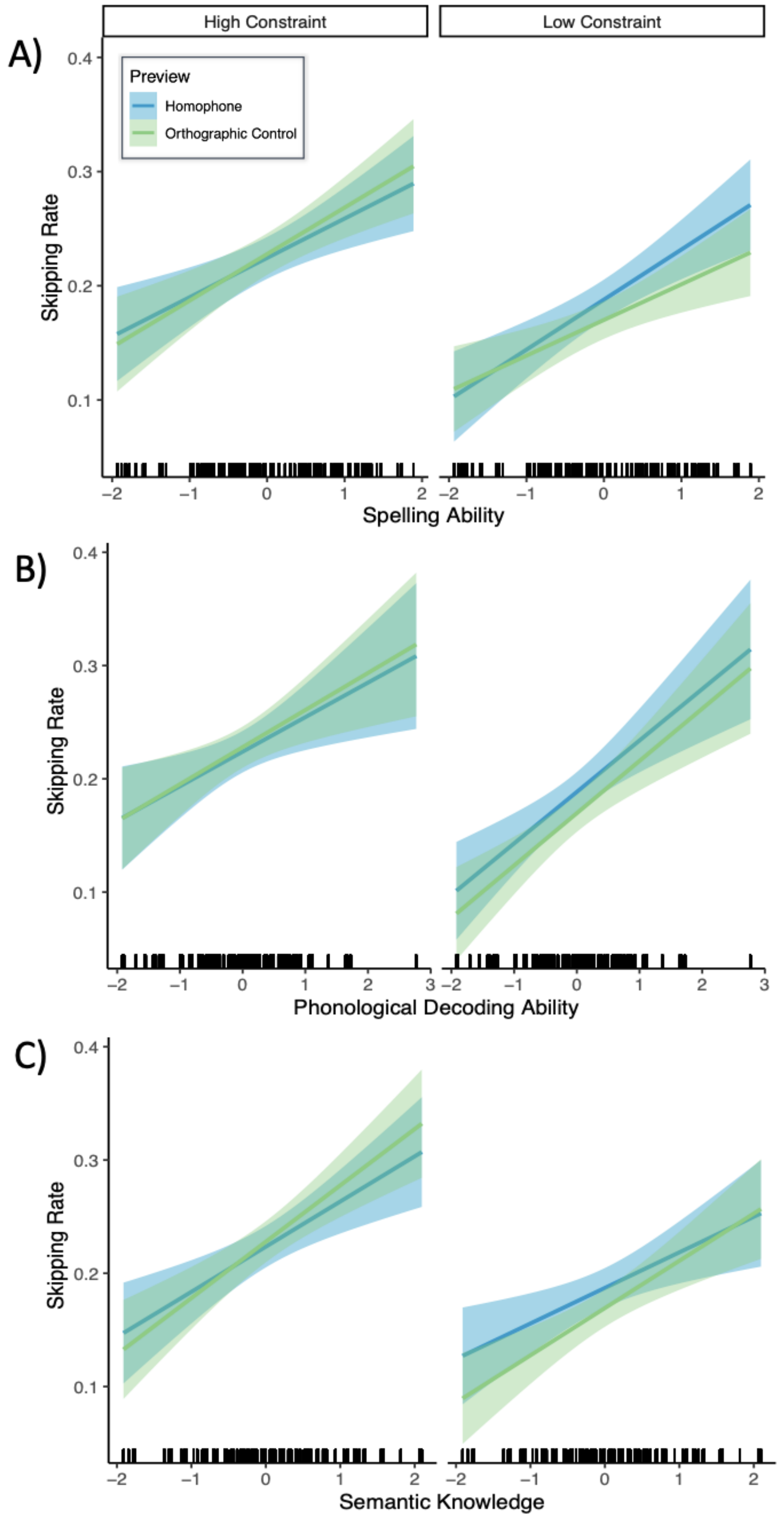

Figure 10. *Skipping Rate as a Function of Constraint and Lexical Quality Measures (Experiment 2)*

Single Fixation Duration

In SFD, the preview manipulation again emerged as a significant predictor of fixation durations (Table 6). Unsurprisingly, sentence constraint was also a significant predictor of SFD (Figure 9B) and relative to Experiment 1, the effect of sentence constraint was much larger than that of preview condition (~18 ms compared to ~ 6 ms for the preview effect). As in Experiment 1, the effect of the preview manipulation had a nonsignificant numerical trend of an interaction with spelling ability ($p = .09$). Based on visual inspection (Figure 11A), this effect also appeared to interact with sentence constraint, such that worse spellers had larger preview effects, but this three-way interaction was not significant ($p = .11$).

Gaze Duration

As in SFD, both preview and constraint were significant predictors of GZD (Figure 9C). Higher scores in both phonological decoding ability and semantic knowledge predicted shorter gaze durations overall (Table 6). Additionally, semantic knowledge and spelling ability interacted with sentence constraint in opposite directions. The effect of spelling ability affected gaze duration more in low constraint sentences, while semantic knowledge mattered more for high constraint sentences (Figure 12A & 12B. As in SFD, sentence constraint was a more robust predictor of GZD than preview condition in Experiment 2 compared to Experiment 1.

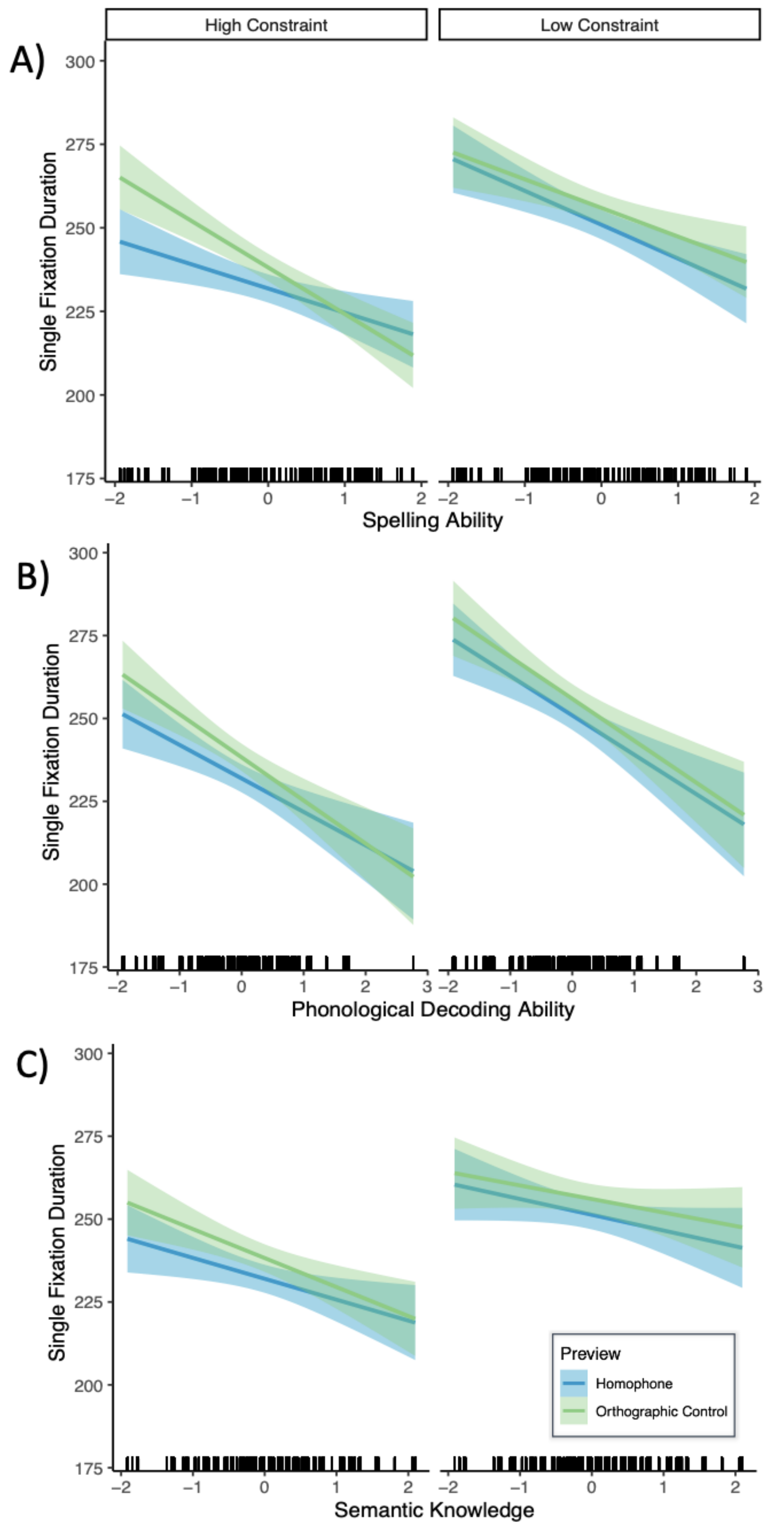

Figure 11. *SFD as a Function of Constraint and Lexical Quality Measures (Experiment 2)*

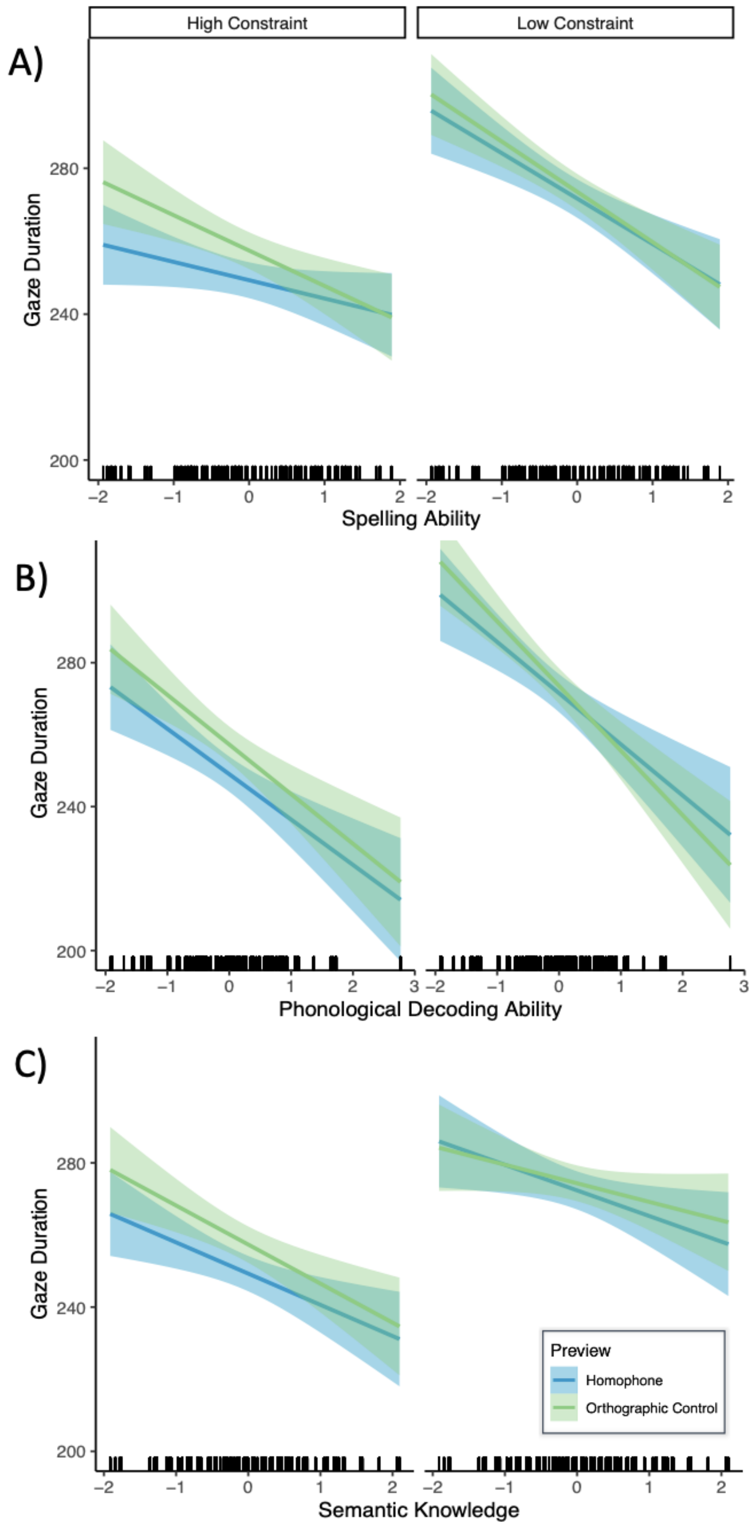

Figure 12. *GZD as a Function of Constraint and Lexical Quality Measures (Experiment 2)*

Exploratory Analysis of Target Word Frequency Effects

We conducted the same exploratory analysis as in Experiment 1 to examine effects of lexical frequency did not replicate these results (Table 7). In Experiment 2, the previews were nonwords, however, so I used the raw target word frequency as a predictor rather than the relative frequency. Because a nonword essentially has a lexical frequency of zero, however, these measures could be conceptualized as equivalent. Interestingly, in Experiment 2, both SFD and GZD appeared to be most sensitive to target word frequency, and although there was a significant effect of parafoveal preview, I did not find any significant interactions, although, in general, the patterns appear to be in the same direction as in Experiment 1 (Figure 13). Potential explanations of these differences between experiments are explored further in the general discussion.

Table 7. *Results of LMER Model Predicting Dependent Eye Tracking Measures by Preview Condition, Sentence Constraint, and Target Word Lexical Frequency (Experiment 2)*

	Skipping Rate				**Single Fixation Duration**				**Gaze Duration**			
Predictors	*Est.*	*SE*	*z/t*	*p*	*Est.*	*SE*	*z/t*	*p*	*Est.*	*SE*	*z/t*	*p*
(Intercept)	0.04	0.01	-8.65	**<0.001**	257.80	9.57	26.93	**<0.001**	280.43	8.34	33.61	**<0.001**
Preview (Homophone vs. Control)	1.39	0.62	0.74	0.457	27.50	12.39	2.22	**0.026**	33.11	15.21	2.18	**0.029**
Constraint (High vs. Low)	1.42	0.61	0.81	0.418	10.44	10.72	0.97	0.330	12.26	12.84	0.95	0.340
Target Word Frequency	1.67	0.17	4.94	**<0.001**	-7.06	2.71	-2.60	**0.009**	-8.30	2.27	-3.66	**<0.001**
Preview x Constraint	0.73	0.46	-0.49	0.621	-5.76	18.89	-0.30	0.761	-7.72	21.01	-0.37	0.713
Preview x Target Word Frequency	0.89	0.11	-0.89	0.372	-6.44	3.74	-1.72	0.085	-7.62	4.58	-1.66	0.096
Constraint x Target Word Frequency	0.81	0.10	-1.74	0.082	2.42	3.20	0.76	0.449	2.35	3.84	0.61	0.542
Preview x Constraint x Target Word Frequency	1.07	0.19	0.40	0.686	1.36	5.65	0.24	0.809	0.36	6.32	0.06	0.954

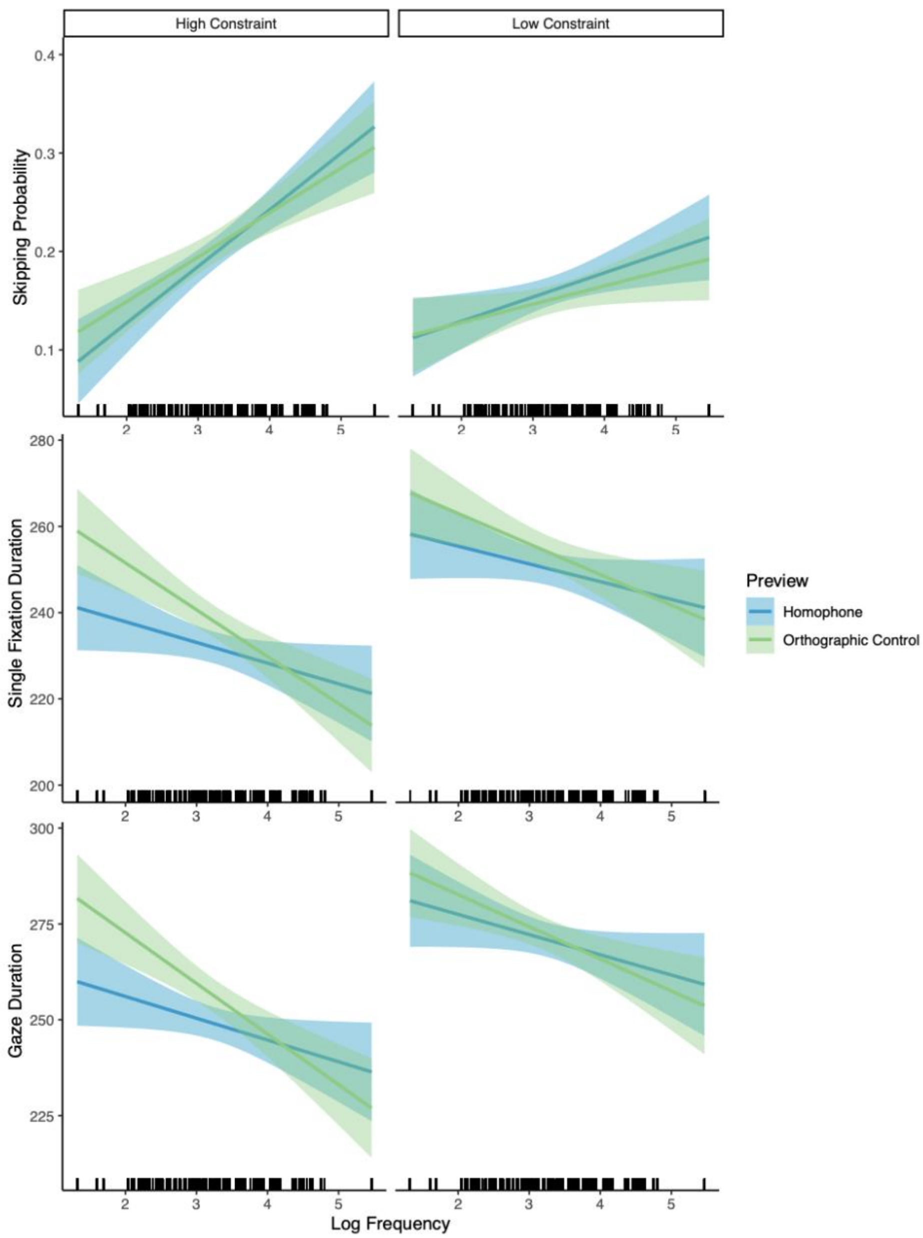

Figure 13. *Skipping Rates, SFD, and GZD by Sentence Constraint and Relative Target Word Frequency (Experiment 2)*

Summary of Experiment 2

In Experiment 2, in which the parafoveal previews were not real words represented in the lexicon, the robust main effects of parafoveal preview and phonological decoding ability were diminished in fixation durations and entirely wiped out in skipping rates. In contrast, the effect of sentence context remained a strong predictor of all three eye tracking measures.

Also, in contrast to Experiment 1, we did not find significant interactions between word frequency and any of the experimentally manipulated variables or any of the lexical quality measures. In fact, much of the variance in skipping rates and fixation durations was attributed to lexical frequency of the target word on its own. Overall, these patterns suggest that activation of phonology during parafoveal processing depends, at least in part, on the familiarity of the preview itself.

Chapter 4: General Discussion

Overall, this study replicates previous findings of the PPB in fixation durations and controlling for the variables of sentence context, lexical frequency, and participant-level lexical quality, I also found larger PPB effects in SFD and GZD than the effect size estimate (~4.5 ms) reported in Vasilev, et al. (2019). This was particularly true in our exploratory analyses controlling for lexical frequency, in which the main effect of PPB (using high constraint as a baseline), was as large as 27.5 ms for SFD (Experiment 2) 38.5 ms for GZD (Experiment 1). These substantial effect sizes alone suggest that the magnitude of the PPB depends on lexical and contextual factors that have not been previously considered. When it comes to skipping rates, which rely on extensive processing of a word during parafoveal preview, our results were more equivocal because I only found significant main effects of the preview manipulation on word skipping in the primary analysis for Experiment 1. Therefore, it may be that the unfamiliarity of the pseudoword previews in Experiment 2 discouraged participants from skipping the target words. Nevertheless, it appears that phonological information extracted parafoveally can promote activation of a lexical item before a word is even fixated during silent reading. The fact that our effects were larger and more consistent in fixation duration measures, however, suggests that phonologically mediated semantic access continues to influence later stages of word recognition as well.

Sentence Context Effects

In both experiments, the sentence constraint manipulation had a significant effect on all three dependent eye tracking measures in the primary analyses. It did not, however, interact significantly with the preview manipulation, suggesting that the size of the PPB is not modulated by sentence context. These results further suggest that the supportive representations or predictions generated by sentence context do not preactivate phonological forms to aid bottom-up processing when a word is ultimately encountered.

Interestingly, in Experiment 2, constraint did interact with phonological decoding ability in predicting skipping rates, such that the effect of phonological decoding ability was larger for low constraint sentences. Looking at Figure 10B, it is clear that these effects are not driven by higher skipping rates for better decoders, but rather a reduction in skipping for poorer decoders in low compared to high constraint sentences. In a low constraint sentence, in which readers have to rely more on bottom-up input, it is possible that people who are not able to rapidly convert orthographic to phonological representations are unable to achieve a recognition threshold that would promote skipping. Relatively higher skipping rates for these individuals in high constraint suggests that they are particularly reliant on sentence context when it comes to deciding whether it is necessary to fixate an upcoming word or not. Furthermore, the fact that this interaction was only significant in the non-word preview experiment suggests that having an unfamiliar preview with no holistic lexical representation further impairs poorer phonological decoder's ability to activate phonological forms via an assembled phonology route. It is worth noting that phonological decoding ability was a significant and robust predictor of overall skipping rates, so poorer phonological decoders had much lower skipping rates overall.

In Experiment 2, I also found significant interactions between constraint and spelling ability and constraint and semantic knowledge in GZD, but these interactions were in opposite directions. Spelling ability had a larger effect on GZD in low constraint, while semantic knowledge had a larger effect in high constraint sentences. These patterns make a lot of sense considering that spelling ability likely reflects sensitivity to fine-grained orthographic representations, while semantic knowledge may reflect better representations of words at the semantic level. Therefore, better spellers would be better at utilizing bottom-up input for word recognition (which would be emphasized when top-down information is sparse), while readers with better semantic representations are likely better at generating semantic expectations from context (which would be amplified when a sentence context activates semantic expectations about upcoming words).

Returning to the role of context in the PPB, our main analyses did not show any interactions between these variables. However, the follow-up analyses that included interactions of these factors with lexical frequency, did reveal evidence that the size of the PPB is modulated by sentence context, but only in Experiment 1. In addition to significant main effects of preview and constraint in both fixation duration measures, I found a significant interaction between them in GZD, such that the PPB was larger in high constraint sentences, aligning with our prediction for this interaction. Furthermore, I found a significant three-way interaction between preview, constraint, and relative target-preview word frequency. As is evident in Figure 13C, the PPB effect appears to be driven by high constraint sentences in which the target word is lower frequency than the preview.

Two alternative, but not mutually exclusive, explanations of this pattern are plausible. First, when a particular word is predictable given the sentence context, and that word has weaker lexical quality due to being less familiar, it may be that preactivation of the word's features is more strongly weighted toward phonology than orthography. You can know the meaning of an uncommon word and know how it sounds without knowing exactly how it is spelled. They may also actively recruit knowledge of the word's phonology in anticipation of encountering a word in their lexicon with an imprecise orthographic representation. So, by this account the larger PPB for low frequency words in a high constraint sentence is due to the differential strength of preactivation/prediction in these lower levels of lexical form. Alternatively, if a reader is able to pre-activate a lexical item from context but the lower-level forms of that word are imprecisely represented (because they are low frequency), when the higher frequency homophone is presented in the parafovea, the reader may leverage the relative familiarity of that orthographic form to more quickly convert from orthography to phonology. Because the phonological form of the bottom-up input matches that of the contextually preactivated phonology, when the correct orthographic form is encountered in the fovea, reading time may be reduced because the semantic and lexical representations may have already been partially activated. By this account, the driving factor in the PPB interaction with frequency and context is the ease of converting the orthographic form of the higher frequency parafoveal preview to a phonological representation. As stated previously, these accounts are not incompatible and it is possible that both the increased preactivation of phonology and the ease of extracting phonology from the preview contribute to this interactive effect.

Individual Differences in the PPB

In the current study, one of the primary questions was how the use of phonology in silent reading differs systematically across people based on their lexical quality (i.e., the precision and redundancy of their orthographic, phonological, and semantic representations of words in memory). Ultimately, I found no interactions between any of the lexical quality constructs and the PPB in any of the dependent measures. There were marginally significant effects of spelling ability on the PPB in Experiment 1, but to interpret these effects as though they were significant violates the principles of null hypothesis statistical testing. This study was also relatively high powered and had a similar participant sample size to other eye tracking studies that found significant individual differences in parafoveal preview benefit based on lexical quality measures (e.g., Veldre & Andrews, 2015). Therefore, the most parsimonious interpretation of our results is that lexical quality does not influence the use of phonology during parafoveal processing.

However, I did find that lexical frequency interacts with the PPB and influences the role of sentence context in the PPB. Lexical frequency was not a factor in our original analyses that included individual differences measures, however, so it is possible that the effects of lexical quality on the PPB also depend on lexical frequency. Lexical frequency has been shown to influence the activation of phonology during visual word recognition differently for people with different levels of reading skill (e.g., Unsworth and Pexman, 2003) and spelling ability (e.g, Yates & Slattery, 2019), so it is theoretically plausible that it plays a key role in this aspect of visual word recognition as well. Therefore, future directions for this study will involve follow-up analyses that test this potentially explanatory variable in the relationship between lexical quality and the PPB.

Another consideration regarding the null effects of lexical quality in this study is that the lexical quality hypothesis specifically emphasizes the importance of "fully specified and redundant representations" (p. 201; Perfetti & Hart, 2002). In the current study, I only truly represent the *specificity* of the representations by measuring and analyzing each lexical quality construct separately. Because these moderately correlated measures are entered separately into the same model, they are in effect controlling for one another and therefore their *redundancy* is not modeled at all. The activation of phonology during reading necessarily arises from converting orthographic and semantic codes to phonological codes because the phonology is not represented directly in the text. Therefore, when it comes to phonological processing during silent reading, it is possible that the correspondence between semantic knowledge and phonological decoding, and between spelling ability and phonological decoding, is more relevant than any one of these abilities on its own. Future investigations will therefore focus on deriving such correspondences using the current data in an attempt to represent both the specificity and redundancy of these sublexical representations. One potentially promising analytical method is criterion pattern profile analysis, which uses regression and structural equation modeling to identify a *pattern* of activation across latent factors that best predicts a particular criterion of interest (Wiernik, Wilmot, Davison, & Ones, 2021). This method would potentially allow us to identify a particular pattern across the measured lexical quality subcomponents that best predicts the magnitude of the PPB. Such a method would consider the redundancy of these representations because it would take into account the relationships between them.

Lexicality Effects on the PPB

We conducted two experiments that used either real word previews or non-word previews with the aim of comparing the PPB effect sizes to gain further insight into whether phonological activation during silent reading is pre-lexical or post-lexical. Furthermore, I considered the possibility that real word previews with semantic representations that were incompatible with the sentence context may result in competition between the two stored lexical representations and reduce the PPB. To the contrary, overall I found numerically larger PPB effects in the experiment with real word previews, suggesting that phonological activation based on parafoveal processing benefits from the preview having a stored holistic representation in memory. Therefore, it seems as though the PPB relies on a kind of feedback process in which the familiarity of the preview begins to activate that stored representation, which in turn increases the activation of the phonological form, which then facilitates semantic access. Furthermore, it may be that conceptualizing a sub-process in word recognition as pre- or post-lexical is not appropriate because it assumes that there is a single point in time at which all of the sublexical representations are subsumed by a holistic lexical one, but the pattern of results from this study suggest it may be a more iterative and resonant process.

Chapter 5: Conclusion

The aim of this study was to provide a clearer picture of the role of phonology in the cognitive processes involved in silent reading, specifically how these processes that allow us to extract meaning from visual symbols might vary depending on particular characteristics of the stimuli, the individuals being studied, and the context. What I found is that all of these factors interact in complex ways. I replicated the phonological preview benefit, supporting the conclusion that phonology is activated during silent reading and that it promotes word recognition and contributes to efficiency in reading. I did not find any evidence from the present analyses that a reader's lexical quality impacts the extent to which they utilize phonological codes extracted from parafoveal vision. I did, however, find evidence that the magnitude of the PPB depends on whether the sentence context promotes the preactivation/prediction of the target word and that the lexical frequency of the target word influences not only the PPB itself, but also the extent to which the PPB is modulated by context. Therefore, it appears as though the PPB does not vary drastically between individuals based on their lexical quality, but that it is sensitive to stimulus properties and to top-down representations generated by the context in which a word is encountered. As a whole, the findings from this study demonstrate that estimating an effect size of the PPB without taking these factors into account is somewhat uninformative and potentially misleading. This study also demonstrates that the development of more comprehensive theories about language processing, and cognitive processing more generally, can benefit from investigations that take a deeper look at how the seemingly extraneous factors that are present when I study a particular cognitive phenomenon might interact in the process of extracting meaning from the stimuli around us.

References

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139*(2), 299.

Andrews, S., & Veldre, A. (2019). What is the most plausible account of the role of parafoveal processing in reading?. *Language and Linguistics Compass, 13*(7), e12344.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364-390.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1– 48.

Blythe H. I., Dickins J. H., Kennedy C. R., & Liversedge S. P. (2018). Phonological processing during silent reading in teenagers who are deaf/hard of hearing: An eye movement investigation. *Developmental Science, 21*(5), e12643.

Chace, K. H., Rayner, K., & Well, A. D. (2005). Eye movements and phonological parafoveal preview: effects of reading skill. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*(3), 209.

Coltheart, V., Avons, S. E., Masterson, J., & Laxon, V. J. (1991). The role of assembled phonology in reading comprehension. *Memory & Cognition, 19*(4), 387-400.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. (1979). Phonological encoding in the lexical decision task. *The Quarterly Journal of Experimental Psychology, 31*(3), 489-507.

Coltheart, V., Laxon, V., Rickard, M., & Elton, C. (1988). Phonological recoding in reading for meaning by adults and children. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 387.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review, 108*(1), 204.

Daneman, M., & Reingold, E. (1993). What eye fixations tell us about phonological recoding during reading. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 47*(2), 153.

Daneman, M., & Stainton, M. (1991). Phonological recoding in silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 618.

Drieghe, D., Brysbaert, M., Desmet, T., & De Baecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology, 16*(1-2), 79-103.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6), 641-655.

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology, 59*(1), e13940.

Hodapp, A., & Rabovsky, M. (2021). The N400 ERP component reflects an error‐based implicit learning signal during language comprehension. *European Journal of Neuroscience, 54*(9), 7125-7140.

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior, 23*(5), 612-624.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience, 32*(8), 954-965.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning?. *Journal of Experimental Psychology: General, 120*(4), 358.

Jouravlev, O., & Jared, D. (2016). Cross-script orthographic and phonological preview benefits. *The Quarterly Journal of Experimental Psychology*, 1-10.

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience, 32*(1), 12-35.

Lee, H. W., Rayner, K., & Pollatsek, A. (1999). The time course of phonological, semantic, and orthographic coding in reading: Evidence from the fast-priming technique. *Psychonomic Bulletin & Review, 6*(4), 624-634.

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin, 140*(6), 1534.

Leinenger, M. (2018). Survival analyses reveal how early phonological processing affects eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(7), 1316-1344.

Lesch, M. F., & Pollatsek, A. (1998). Evidence for the use of assembled phonology in accessing the meaning of printed words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(3), 573.

Liu, W., Inhoff, A. W., Ye, Y., & Wu, C. (2002). Use of parafoveally visible characters during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1213.

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: I. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123*(2), 107.

McCann, R. S., Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance*, *14*(4), 693.

Miellet, S., & Sparrow, L. (2004). Phonological codes are assembled before word fixation: Evidence from boundary paradigm in sentence reading. *Brain and language, 90*(1-3), 299-310.

Newman, R. L., Jared, D., & Haigh, C. A. (2012). Does phonology play a role when skilled readers read high-frequency words? Evidence from ERPs. *Language and Cognitive Processes, 27*(9), 1361-1384.

O'Brien, B. A., Van Orden, G. C., & Pennington, B. F. (2013). *Do dyslexics misread a ROWS for a ROSE?. Reading and Writing, 26*(3), 381-402.

Patterson, K. (1986). Lexical but nonsemantic spelling?. *Cognitive Neuropsychology, 3*(3), 341-367.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. Precursors of functional literacy, 11, 189-212.

Pexman, P.M., Lupker, S.J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 139–156.

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. Journal of Experimental Psychology: Human Perception and Performance, 18(1), 148.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. Cognitive Psychology, 7(1), 65-81.

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

Rayner, K., & Duffy, S. A. (1988). On-line comprehension processes and eye movements in reading. In M. Daneman, G. E. MacKinnon, & T. G. Waller (Ms.), Reading research: Advances in theory and practice (pp. 13-66). New York: Academic Press

Rayner, K., Pollatsek, A., & Binder, K. S. (1998). Phonological codes and eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 476.

Rayner, K., Reichle, E. D., & Pollatsek, A. (1998). Eye movement control in reading: An overview and model. In *Eye guidance in reading and scene perception* (pp. 243-268). Elsevier Science Ltd.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*(4), 504-509.

Rayner, K., White, S. J., Kambe, G., Miller, B., & Liversedge, S. P. (2003). On the processing of meaning from parafoveal vision during eye fixations in reading. In *The mind's eye* (pp. 213-234). North-Holland.

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Rosseel Y (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1–36.

Schotter, E. R., & Leinenger, M. (2016). Reversed preview benefit effects: Forced fixations emphasize the importance of parafoveal vision for efficient reading. Journal of Experimental Psychology: Human Perception and Performance, 42(12), 2039.

Schrank, F. A. (2011). Woodcock-Johnson III Tests of Cognitive Abilities.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523.

Shipley, W., Gruber, C., Martin, T., & Klein, M. (2009). Shipley Institute of Living Scale-2. Los Angeles, CA Western Psychological Services

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. Journal of Experimental Psychology: Human Perception and Performance, 37(6), 1924.

Slattery, T. J., & Yates, M. (2017). Word skipping: Effects of word length, predictability, spelling and reading skill. *The Quarterly Journal of Experimental Psychology*, (just-accepted), 1-40.

Tiffin-Richards S. P., & Schroeder S. (2015). Children’s and adults’ parafoveal processes in German: Phonological and orthographic effects. *Journal of Cognitive Psychology, 27*(5), 531–548.

Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). Test of Word Reading Efficiency–Second Edition. *Austin: Pro-Ed*.

Unsworth, S. J., & Pexman, P. M. (2003). The impact of reader skill on phonological processing in visual word recognition. *The Quarterly Journal of Experimental Psychology Section A, 56*(1), 63-81.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15*(3), 181-198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 371.

Vasilev, M. R., & Angele, B. (2017). Parafoveal preview effects from word N+ 1 and word N+ 2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic bulletin & review*, *24*(3), 666-689.

Vasilev, M. R., Yates, M., & Slattery, T. J. (2019). Do Readers Integrate Phonological Codes Across Saccades? A Bayesian Meta-Analysis and a Survey of the Unpublished Literature. *Journal of Cognition*, *2*(1).

Veldre, A., & Andrews, S. (2015). Parafoveal preview benefit is modulated by the precision of skilled readers' lexical representations. *Journal of Experimental Psychology: Human Perception and Performance, 41*(1), 219–232.

Veldre A., & Andrews S. (2018). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during reading. *Journal of Memory and Language,* 100, 1–17.

White, S. J., Rayner, K., & Liversedge, S. P. (2005). The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. *Psychonomic Bulletin & Review, 12*(3), 466-471.

Wiernik, B. M., Wilmot, M. P., Davison, M. L., & Ones, D. S. (2021). Meta-analytic criterion profile analysis. *Psychological Methods, 26*(2), 186.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage, 62*(1), 356-366.

Yates, M., & Slattery, T. J. (2019). Individual differences in spelling ability influence phonological processing during visual word recognition. *Cognition, 187*, 139-149.

Appendix A: Norming Instructions

Plausibility Rating Task

INSTRUCTIONS

In this task, you will be presented with sentences and be asked to make judgments about them.

Please read each sentence carefully and indicate how well-written the sentence is using the scale provided.

You will be given the options: Very well written, Well written, Somewhat well written, Neither, Somewhat poorly written, Poorly written, Very poorly written

Please use your best judgement and select the option that you feel best describes the sentence. Grammatical and spelling errors should be considered and receive lower ratings.

Cloze Sentence Completion Task

INSTRUCTIONS

In this task, you will be presented with sentences and be asked to make judgments about them.

Please read each sentence carefully and indicate how well-written the sentence is using the scale provided.

You will be given the options: Very well written, Well written, Somewhat well written, Neither, Somewhat poorly written, Poorly written, Very poorly written

Please use your best judgement and select the option that you feel best describes the sentence. Grammatical and spelling errors should be considered and receive lower ratings.

Appendix B: Irregular Word Reading Task

Irregular word reading

synagogue
phlegm
coyote
rhapsody
calf
gnome
cough
Wednesday
cyst
bureau
brooch
guard
lasagna
gnaw
abyss
castle
beret
subpoena
busy
lingerie
pseudonym
seize
gauge
sieve
laugh
rhinoceros
indict
receipt
doubt
pneumonia

ghost
aisle
adjective
debut
solder
weigh
tongue
sergeant
queue
chute
heir
impugn
island
ghetto
depot
rhyme
debt
weird
knead
quiche
ricochet
buoy
malign
bough
paradigm
sleigh
meringue
guy
subtle
biscuit
algae
dough
sword

debris
choir
colonel
yacht
asthma
mnemonic
foreign
cello

Total time: __________
Total correct:________