Spring 2019

# The Word Frequency Effect in English Second Language Students: An Eye Movement Monitoring Study

Chase Alexandria Rathfoot

**Director of Thesis:** Dr. Robin Morris
**Second Reader:** Dr. Alisha Biler

ACKNOLEDGEMENTS

This thesis would not have been possible without the help and support of several individuals throughout my time career at the University of South Carolina. I owe them my sincerest thanks and credit this paper as well as my success to them.

First, I would like to thank my thesis advisor Dr. Robin Morris for her contributions to both this work as well as my education in psychology. Dr. Morris has provided constant support and encouragement for my studies as well as for my professional career. She has helped me with everything from teaching me classes to registering for classes and has allowed me to visit her office more than I have visited home while at USC. Her guidance and insight have allowed me to answer deeper questions, solve bigger problems, and accomplish greater projects than I could've ever done on my own. I am lucky to have her not only as an advisor to this project, but also as a mentor in life.

Second, I would like to thank Dr. Alisha Biler for inspiring this research. During my third year at USC, Alisha allowed me to join her on a research project that would teach me more about this subject matter than any class could have offered. She allowed me work with her directly on a project in which I could learn new theories and software as well as from her and her success – which have found invaluable to this project and my study of neuroscience. I owe Alisha for the initial data and subject matter that have inspired this project, and this work would be impossible without her. I am proud to expand her work into this thesis as well as to call her my friend.

Lastly, I would like to thank the rest of my classmates and colleagues at the University of South Carolina. There are countless individuals who have had a significant on my education that are not mentioned by name, but to whom I owe tremendous thanks for my education and success.

# ABSTRACT

Reading has been the subject of hundreds of thousands of studies. Reading is fundamental skill not only to an individual's education, but to their socialization and an understanding of the behavior is a rapidly expanding area of research. One of the least understood areas of reading behavior can be found is the growing English Second Language (ESL) population – or those who are learning English after demonstrating competence in their native language. As their native language has its own unique set of properties and is presented in its own unique written form, the influence this language has on their reading behavior is significant and only recently examined.

In this study, the work of previous researchers was expanded to demonstrate the word frequency effect and its presentation in two ESL populations: Spanish and Mandarin. Readers of the two languages were contrasted using 16 passages containing target high and low frequency words in hopes to demonstrate that ESL readers rely more heavily on word frequency than their native English reading counterparts. The results of this study affirmed that ESL readers more dramatically express the WFE when compared to native English speakers using both first fixation duration and gaze duration as measures.

Additionally, it was found that the word frequency effect presented differently across language groups with Chinese ESL readers relying on word frequency more heavily than Spanish ESL readers.. These results suggest that the native language of an ESL reader has a significant effect on their reading behavior as it relates to word frequency and possibly more components of the skill. It is recommended that future research should explore the source of these differences and determine whether they are the result of language and or writing system differences.

TABLE OF CONTENTS

# THE WORD FREQUNCY EFFECT IN ENGLISH SECOND LANGAUGE STUDENTS: AN EYE MOVEMENT MONITORING STUDY

Reading is a complicated and well-researched behavior seen in language learners around the world. Through methodologies such as eye movement monitoring, researchers have identified the physical behaviors involved in reading and used them to interpret cognitive processes involved (Rayner, Ashby, Pollatsek, & Reichle, 2004) However, the majority of these studies and interpretations have been based on native language learners – those who have learned a language from birth. In the United States almost ten percent of public-school students identify themselves as English Second Language (ESL), demonstrating a need to expand reading behavior research to second language (L2) readers (U.S. Department of Education, 2018). Previous research has determined that ESL readers demonstrate significant differences when reading in their second language versus their native tongue (Rayner et al, 2012).

ESL readers are defined as those individuals who are learning English after mastering their native language. Their English reading skill is often lower than their skill in their native language or as compared to native English readers. For example, native English readers demonstrate a data pattern known as the Word Frequency effect when they read a text. Because some words are encountered more frequently in written communication than others, these high frequency words are given less time and attention than low frequency words as one reads. The word frequency effect is taken as a marker of basic word recognition processes.

As will be discussed later in this paper, the word frequency effect is thought to be a reflection of one's differences in word recognition – that is, the ability to recognize a word and understand its meaning (Literacy Information and Communication System, 2019). Because word recognition is fundamental to reading comprehension, this study aims to use the word frequency effect to

observe differences between ESL and native English reading behavior as well as explore the origin of those differences. By examining native Spanish and Chinese ESL readers alongside native English readers, this paper hopes to determine the differences in eye movement behavior specifically related to the word frequency effect, a well-researched data pattern in eye tracking studies.

## LITERATURE REVIEW

### L1 Reading Behavior

In the *Psychology of Reading*, researchers defined reading as a highly complex skill, occurring through a multistep process, that is essential to communication in the written form (Rayner, Pollastek, Ashby, Clifton 2012). The complexity of reading lies in the necessity for several types of information processing. Readers must be able to recognize letters and words as representations of sounds and access their definitions. Together, readers then combine words to create and understand sentences through which they can understand and draw conclusions about a given text. This skill is referred to as reading comprehension.

#### *Comprehension*

Rand defines reading comprehension as "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (Snow, 2002). In a review conducted by Cho et al., it was determined that vocabulary plays a direct role in reading comprehension; students who recognized a large set of word families scored higher on reading assessments given in public schools. vocabulary is considered by researchers to be a lexical-level dimension of reading behavior (Cho et al., 2019). When learning new vocabulary words, a reader gains experience with lexical properties of a language. A new vocabulary word

provides them with new access to word families, stems, and conjugations that can then be applied to their preexisting knowledge of the language. Because of this ability to process new information, increases in vocabulary have been linked to increases in word reading. Rather than skimming past unknown words, the reader will instead add it to their personal lexicon and improve their word recognition (Walley, Metsala, & Garlock, 2003).

*Word Recognition*

Word recognition refers to the ability of a reader to understand a given word and its meaning. The speed at which a reader is able to identify and comprehend a word is often related to a reader's fluency (Literacy Information and Communication System, 2019). In a study conducted with middle school aged students, automatic word recognition was found to be a significant predictor of reading fluency. Students who recognized a greater number of words in a standardized test, performed significantly higher on reading fluency exams as compared to students who did not recognize the words (Roembke, Hazeltine, Reed, & McMurray, 2019). Similarly, a study conducted throughout the lifespan of adults found that one's reliance on word frequency for word recognition and comprehension increases with aging. This was hypothesized to be the result of years of exposure to high frequency or common words. (Cohen-Shikora and Balota, 2016). Based on the results of these two studies, one can see that different age populations of readers have a varying reliance on word frequency. In the present research, these results will be expanded to college-aged readers as well as to English Second Language readers to explore these differences.

## L2 Reading Behavior

### *Definition*

Native language readers, those who have learned a language from birth, are the most frequently studied reading group. In recent years, bilingualism has been the subject of a growing number of research projects. Bilingualism is broadly defined by researchers Byers-Heinlein and Lew-Williams as "the ability to use two languages in everyday life" and if often expanded to refer to those who speak more than two languages as well (2018). Although some children are raised with two equally dominant languages from birth, it is far more common for a child to learn a second language from ages 5-15 (Byers-Heinlein and Lew-Williams, 2018). In this study, specific attention will be given to a subset of bilingualism referred to as "English Second Language" – those who have learned English after establishing competence in their native language.

### *Native Language and Influence*

It has been suggested that ESL students' speaking, reading, and comprehensive abilities are influenced by their native language and its features (Karim, 2003). For example, the English language generally presents adjectives before the noun they modify as in "The pretty girl walked." However, in the Spanish language, adjectives generally come after the noun they modify. For this reason, Spanish learners will often take longer to process an adjective preceding a noun or mistake it for the noun itself hindering their ability to read effectively (Coe, 1987). In eye movement monitoring studies these processing differences may be reflected in fixation duration and/or location.

Additionally, many languages throughout the word use writing systems differ significantly from the English alphabet. In a study conducted on Chinese and Korean-native ESL students, it was found that the method in which a person's native language is written can physically shape how a new language is processed. When the ESL readers were given a text written in English, they first attempted to read the text as they would in their native language; Chinese ESL readers, for example, relied heavily on lexical networks rather than orthographic networks as their native language writing system was logographic and did not readily relate to English writing systems (Kim et al., 2017). It is important therefore, to also consider orthographic differences and similarities among language groups when studying reading behavior.

**Spanish and Mandarin**

When discussing ESL reading behavior, Spanish and Mandarin provide important and distinct points of comparison. These are two of the most widely spoken languages in the world. In addition, native Spanish and Mandarin Chinese speakers are the largest English Second Language groups in the United States. A 2015 survey of American public schools found the most frequently reported first language spoken by ESL students was Spanish followed by Arabic, Chinese, and Vietnamese in that order (U.S. Department of Education, 2018). Because these language learners are most likely to enroll in English Second Language programs, researchers have focused on the unique characteristics of a language that could potentially make learning English more difficult. These characteristics can include, but are not limited to, their writing system, grammar, structure, and degree of similarity to English.

*The Spanish Language*

Perhaps the language most frequently compared to English is Spanish, due to its similar language properties and shared writing system. In its written form, Spanish, like English, uses the roman alphabet. It is written from left to right, , Letters are combined to make words which are divided by spaces, creating sentences and paragraphs to convey meaning. Like English, Spanish is an alphabetic language in which these written letters represent the sounds made by a speaker. Since letters provide a map of a word's pronunciation, Spanish is considered a phonetic language (Rayner et al., 2012).

Grammatically, Spanish also shares many similarities to English. As in English, Spanish distinguishes function and content words like in English. Function words help facilitate grammar and include conjunctions or articles, where content words contain information and include nouns and verbs. However, unlike English, some of function words are gendered. Spanish is a Romance language, in which nouns and articles are gendered based on predetermined grammar rules (Coe, 1987). For this reason, Spanish readers often struggle with English word endings as they expect different endings based on the gender of a noun. Much as with orthography, these language differences can lead to differences in eye movement behavior and reading comprehension.

*The Chinese Language*

Unlike Spanish, Mandarin provides a dramatically different language and writing system than English. Mandarin is a logographic language, in which individual characters or symbols represent words or larger concepts. These characters can be either pictographic (representing an object) or ideographic (representing an idea). Characters in Mandarin combined thousands of different ways to create words and meanings, with words sometimes represented by a single

character and sometimes by more than one character. Unlike English, the Mandarin writing system also has no spaces (Rayner et al., 2012). Learners are expected to understand conceptually the combination of characters and read them as a single entity rather than separate words or sentences.

Due to the nature of Mandarin and its assumption that learners understand the concepts of words, the language also does not have a need for function words. Where in English, function words are required to indicate possession or supply syntactic context, these concepts are only implied in Mandarin. Another major difference between English and Mandarin is the conjugation of verbs. In English, verbs are conjugated to indicate when they take place in relation to the subject. Ending such as "-ed" indicate a past action whereas "-ing" indicates a present action. However, in Mandarin, verbs are not conjugated at all. Their place in time is assumed based on the context of a sentence and the reader's knowledge (Rayner et al., 2012). Therefore English and Mandarin ESL readers are expected to display dramatically different reading behavior.

*Language Comparison*

As exemplified in the preceding section, an ESL reader's behavior is heavily influenced by their native language. If a reader learned a language in a different writing system, it is likely that they will show dramatic differences or even struggle when reading a text written in English. Additionally, many behavioral differences between readers have identifiable roots in their native language's properties. Because of these writing system and language differences, it is essential to use research methods that are adaptable to all readers and can discern processing differences objectively. Eye movement monitoring allows readers to read without the interruption or added processing burden of a secondary task and provides a continuous record of the reading process as it unfolds over time.

## Eye Movement Monitoring

### *A Brief History*

Since the 1900s, researchers have used eye movement monitoring as a method of studying reading behavior. Rayner, in his review of the history of this methodology, argued that eye tracking has been used as a primary tool for understanding the behaviors of the eye that are necessary for reading and the cognitive processes these behaviors could depict (Rayner, 1998). Unlike previous research methods, eye tracking allows a study participant to read fluidly and without interruption, rather than having to ask a reader to perform a specific task or report on their own cognition during a testing session.

Rayner describes modern eye tracking studies as the third era of the research field. Early research into the field of eye tracking focused on the development of an objective measure of reading. Out of several trials came a camera which tracks the readers pupil as it moves across a page and records the length of time in which it performs actions such as moving or fixating (Rayner et al., 2012). This objectivity of the camera allows participants in eye tracking studies to read as they would outside of the research environment and provide a more accurate representation of reading behavior to researchers. Although there are still minor interruptions in the eye tracking data from movements such as blinks, the result of an eye tracking study is a continuous data set which can then be analyzed.

Today, eye tracking research has moved beyond the physical movements of the eye and into the realm of interpretation. Rather than simply recording data, researchers are interpreting data as cognition. As previously discussed, reading is a cognitive process in which a reader must be able to both process and interpret a text; they comprehend the information provided and understand

its value as related to the rest of a text. By observing the specific timing of eye movements, researchers are now using eye tracking to determine how and when this interpretation occurs in the brain. It is hoped that by understand the mechanics of reading, one can understand the cognitive skills involved in reading – and perhaps improve them.

*Eye Tracking in Reading*

Eye movements, as related to reading, are broadly divided into fixations and saccades. Fixations are defined as points in which the eye is not moving and is taking in visual stimuli. In the average native English learner, a fixation lasts between 200 to 300 milliseconds depending on the stimuli presented (Rayner et al., 2012). For example, fixation duration for function words tend to be shorter whereas fixations for adjectives are longer (Rayner, 1998). Saccades are alternatively defined as periods in which the eye is moving, usually covering 7-9 character spaces. Saccades can occur between words or lines on a page and can represent gaps in attention. Generally lasting 25 milliseconds in native English speakers, saccades are interpreted to be times in which the eye takes in little to no visual stimuli (Rayner et al., 2012). During these periods of movement, perception of visual stimuli is suppressed creating a fluid transition between words and rendering the eye's movements unnoticeable to the reader (Krekelberg, 2010). Together, the duration of fixations and saccades are used to interpret the reading behavior of a participant; the differences in the lengths of fixations and saccades can help a researcher interpret the effect of a text's structure or grammar on reading behavior.

*Eye Tracking in Language Comparison*

Since research into English Second Language requires an objective measure that can be used across different writing systems and language groups, eye movement monitoring has become a

staple in the research field. Although they differ in number and length, the fixations and saccades that appear in native English learners' reading also appear in other language groups. For example, a recent research study used eye tracking software to specifically target the eye movements of native Spanish readers. Because Spanish is a rapidly growing language group across the Western world, researchers aimed to discover the natural reading behavior of a Spanish reader while reading in their native language. The results of this study found that participants tended to fixate primarily on content words and mistakes in a given text that was written in Spanish (Betancort, Carreiras, and Sturt, 2009). However in addition to observing natural reading behavior in a given language, eye movement monitoring also allows comparisons to be made between language groups.

By taking data from specifically curated passages written in various languages and comparing it with data collected from native English speakers, it is the hope of researchers to determine the source of differences in eye movement behavior. Research into the field has suggested that the differences in eye movement behavior may arise from differences in a language or writing system, as the eye tracking field was designed to record the movements of English readers (Inhoff et al., 2012). For example, Chinese ESL readers have been known to fixate for longer periods of time on function words when reading English as they are not an essential part of their native language (Rayner et al., 2012).

## Word Frequency Effect

### *Definition*

In modern society, readable text comes in may forms; readers encounter books, magazines, signs, phones, and other sources almost daily. With information coming from several

sources, certain words and phrases become more frequent than others. This variable is known as word frequency and can be objectively measured using complications of written sources that are then analyzed for the appearance of certain words. For example, the Corpus of Contemporary American English (COCA) contains frequency data for over 1 million words in the English language and their related forms and families (Davies, 2008-). Through hundreds of studies it has been found high frequency words take less time process than low frequency words both in oral and written communication (Rayner, 1998). This pattern of having shorter processing time for high frequency words and longer for low frequency words is known as the Word Frequency Effect.

*WFE and Eye Tracking*

In a series of experiments, researchers Inhoff and Rayner (1986) determined that the word frequency effect could be seen in the eye movements of a reader. In their study, Inhoff and Rayner showed that skilled adult readers spend less time reading a high frequency word than a low frequency word. By controlling for word length and complexity, the researchers demonstrated that word frequency affected the fixation durations. Further research by a group of researchers has suggested that the relationship between word frequency and reading time is logarithmic – with words taking exponentially time to read as they increase in frequency (Smith and Levy, 2013). Since these experiments, the word frequency effect has been expanded to other languages and writing systems – some of which have demonstrated similar results to those found in native English speakers.

*WFE in L2 Readers*

In a study conducted in 2018, researchers examined the word frequency effect in bilingual children aged 7-12 years old. Canadian children, whose native language was English, were asked to read passages in both their L1 (English) and their L2 (French). After analysis, researchers determined that although the word frequency effect was apparent in both languages being read, the children displayed a significantly larger effect when reading in their L2 (Whitford and Joanissee, 2018). Similar results have been replicated in adults, as seen in Enkin et al.'s study of English-Spanish readers. In their experiment, adult English readers were asked to read short sentences in their low proficiency L2 (Spanish). The presented sentences varied in difficultly by changing a single high frequency word to a low frequency word, which was found to dramatically increase the difficulty of reading as seen in fixation length (Enkin et al., 2017). The results of these studies lays the foundation for research into the word frequency effect in L2 readers and this study aims to expand these results specifically to adult English Second Language groups.

*WFE in ESL Readers*

In this paper, previous research into the Word Frequency Effect as observed through eye movement monitoring will be expanded to college-aged English Second Language students. Unlike previous studies conducted on children, the participants in this study have demonstrated advanced proficiency in the English language. Because the study participants will be older and will have had more exposure to high frequency words throughout their lifetime, it is likely that the word frequency effect will be present but less dramatically presented in the data analysis as compared to children. As was discussed in the results of Enkin's longitudinal study of adults, it is also predicted that the results of this study can lead to a greater understanding of reading

comprehension in ESL students and learners. By observing the first fixation duration and gaze duration of target high and low frequency words, this study hopes to directly compare the word frequency effect between both native English and ESL students as well as between two ESL groups. In this way, the results of this study can both affirm previous research in which ESL readers were found to rely more heavily on word frequency than native English readers and new insights can be provided into how a language's properties and a writing system can affect this reliance.

## METHODOLOGY

### Participants

The participants for this study were divided into two experimental groups: L1 (native English learners) and L2 (English Second Language learners). Each of the participants were recruited from the University of South Carolina and were enrolled as either undergraduate or graduate students. Participants in the L2 group were further divided into two sub-groups: native Spanish and native Mandarin learners. In each group (L1, Spanish, and Mandarin) there were a total of 6 participants assessed to be at mid-advanced proficiency level according to a Vocabulary Size Test.

### Vocabulary Size Test

The Vocabulary Size Test (www.vocabsize.com)provided an online assessment of general English language vocabulary knowledge that was used as an assessment of language proficiency in this study. The test was administered on a desktop computer in the laboratory following the reading portion of the experiment. Participants were not allowed to ask for assistance with the test from the researcher nor allowed to access other online resources such as a dictionary or

thesaurus. Participants were asked to complete the 140-question assessment in which they were required to select the proper definition for a presented word. Presented words ranged from very common to very unusual words as based on their frequency in the English lexicon. The assessment was given At the end of the assessment, participants are given a score that represents their approximate vocabulary size. For this experiment, scores lower than 7,000 were categorized as low proficiency and scores above 7,000 were considered high proficiency. All participants for the current study were found to be at high proficiency. Following the competition of the vocabulary test, participants were asked to provide demographic information such as their age and native language as well as affiliation with the University of South Carolina (graduate or undergraduate student). After providing this information, participants were asked if they had any further questions, debriefed on the research intentions, and were compensated for their time.

**Reading Task**

For the reading portion of the experiment, participants were asked to read 16 passages each composed of 5 sentences and between 60-70 words. Passages were constructed to a mid-proficiency level and were created specifically for the testing of L2 readers. Because familiarity and exposure to subject matter is an important factor in L2 reading, passages were controlled for background knowledge and discussed topics that appeal to a general public A complete list of the passages and text are listed in Appendix 1A.

Within each paragraph, two target words were selected: one high frequency and one low frequency. Each target word was a noun between 4-7 letters that did not occur  at the beginning or end of a sentence. proper nouns (names, places, objects) were excluded. The frequency of a word was determined using the Corpus of Contemporary American English, which provides the frequency of words according to a data base of 560 million texts from the years 1990-2017. In

order to ensure that the corpus is unbiased, its sources range from works of fiction and magazines to online newspaper articles and webpages (Davies, 2008-). High frequency was defined in this study as words with a value between 125000 and 450000 according to the COCA. Low frequency was defined as words with a value between 4000 and 40000 . The average word length for the high frequency category was found to be 5.4 letters whereas the average word length for the low frequency words was found to be 6.1 letters. A complete list of the target wordsand their respective word frequency and length can be found in the Appendices 1B and 1C.

**Data Collection**

Each of the vocabulary and reading tests were conducted in a laboratory on the University of South Carolina's campus. Participants were first asked to complete the reading assessment followed by the vocabulary test and were required to provide essential background information (vision status, native language, age, etc.). The experimental process averaged 1 hour in length and all tests were administered by the experimenter individually.

To observe the eye movements of participants, an SR Research Eyelink 1000 eye tracker was used. Participants were asked to sit approximately 0.9 meters from a desktop monitor and were asked to rest their chin in a head stabilizer. During the experiment, the eye tracking software calibrated to the right pupil of the participant after which the participant was asked to read at a comfortable pace. To ensure participants were reading the passages, participants were additionally asked 2 comprehension questions following each paragraph which were not used for assessment.

# RESULTS

## WFE in Native English Readers

To demonstrate the WFE in native English speakers, an initial analysis compared the First Fixation Duration on the target low and high frequency words followed by a comparison using Gaze Duration (Appendices 2A and 2B for individual participant data).As can be seen in Table 1, native English speakers spent less time on high frequency words than on low frequency words. As in previous research, this effect emerges clearly in gaze duration on the target.

**Table 1: Average Fixation Durations in Native English Speakers**

| | High Frequency Words | Low Frequency Words |
|---|---|---|
| Gaze Duration | 579 ms | 682 ms |
| First Fixation Duration | 215 ms | 216 ms |

## WFE in ESL Readers

Using Gaze Duration and First Fixation Duration, the results of native English readers were then compared to those of ESL readers. As discussed in the previous section, native English readers were found to spend less time of high frequency words than low frequency words – thereby demonstrating the word frequency effect. This pattern, however, was not observed in ESL readers. The ESL readers had comparable average  gaze durations, to the native speaker group, suggesting a high level of proficiency. However, they did not demonstrate a word frequency effect suggesting that there is a fundamental difference between ESL and native English readers, even at a relatively high level of ESL proficiency.

**Table 2: A Comparison of Fixation Duration in ESL and Native English Readers**

| **Gaze Duration** | High Frequency Words | Low Frequency Words |
|---|---|---|
| Native English | 579 ms | 682 ms |
| English Second Language (ESL) | 697 ms | 708 ms |
| **First Fixation Duration** | | |
| Native English | 215 ms | 216 ms |
| English Second Language (ESL) | 237 ms | 243 ms |

## WFE in Chinese and Spanish Readers

Because the results of the analysis showed that ESL readers did not demonstrate the word frequency effect like native English readers, the source of this difference was analyzed. Specific attention was given to the differences between Chinese and Spanish ESL readers - predicting that Chinese readers would demonstrate the effect due to differences between the English and Chinese language and writing systems. By comparing First Fixation as well as Gaze Durations, this prediction was found to be inaccurate.

**Table C: Comparison of ESL Spanish and Mandarin Chinese Fixation Durations**

| **Gaze Duration** | High Frequency Words | Low Frequency Words |
|---|---|---|
| ESL Spanish | 693 ms | 694 ms |
| ESL Mandarin Chinese | 701 ms | 722 ms |
| **First Fixation Duration** | | |
| ESL Spanish | 232 ms | 236 ms |
| ESL Mandarin Chinese | 242 ms | 249 ms |

The results of this study suggest the ESL readers, despite demonstrating comparable reading time and proficiency, do not demonstrate the WFE like native English readers. The further analysis of Chinese and Spanish ESL readers suggests that this lack of pattern is consistent across language groups – despite predictions that they arise due to differences in language and writing systems. Although there is a possibility that with a larger sample size the Chinese ESL readers may demonstrate the effect, it is a more reasonable conclusion that there is an alternative reason for the lack of effect word frequency has on ESL readers. Further suggestions as to what these conclusions may be, are discussed in the conclusion this paper.

## DISCUSSION

The findings of this paper are two-fold; it both confirms previous research into the Word Frequency Effect and how it presents in eye movements, as well as expands the well-known data pattern to English Second Language Students. In an increasingly bilingual society, speaking and reading in a second language is an important but under-researched field. Past research has focused on specific elements of reading – such as comprehension or eye movements – or has strictly compared a language system to another. Seldom has research branched outside of native language speakers into the realm of bilingualism, much less into the ESL community.

In this study, the eye movements of two different English Second Language populations were examined along passages containing both a high frequency and low frequency word. The fixation duration on these target words were then compared to one another and across the three language groups. Previous research has suggested that the fixation duration on these target words can effectively demonstrate the Word Frequency Effect – a comprehensive tool used by readers in

which attention is given primarily to low frequency rather than high frequency words. Although this study does confirm the validity of this measure by demonstrating the WFE data pattern, there was virtually no evidence of the WFE in the ESL reading groups.

In studies conducted on Canadian and Spanish ESL children, it was suggested that second language learners may rely more heavily on high frequency words than native readers (Enkin et al., 2017 and Whitford & Jonaisse, 2018). This presented in their results as a wider gap between the high frequency and low frequency fixation durations for the second language readers as compared to previous research on native English speakers. In the present study, this pattern was examined in both ESL readers in general as well as in specific language groups (Chinese and Spanish). In both cases, the WFE did not appear in the ESL populations suggesting that there is a fundamental difference between the reading behavior of ESL and native English readers regardless of proficiency.

In the analysis of this paper, the native English and ESL readers had comparable reading times (as reported in Gaze Duration). This suggests that the participants selected for this study were indeed advanced proficiency and that the selection criteria for the study was effective. Interestingly, the high proficiency seen in the participants and their similarity to native English readers in this regard was contrasted in the lack of WFE revealed by the study. Whereas native English readers showed a significant reliance on word frequency, ESL readers showed no reliance at all suggesting further research is needed to determine the source of this difference.

Originally, the overall difference between native English and ESL readers was thought to be the result of complications that arise from language and writing systems. Further analysis was conducted to determine whether this was the case, comparing Chinese ESL and Spanish ESL readers as they were previously described to have language and writing system differences from

English. After examining the two language groups and comparing them, it was revealed that neither showed a WFE and were therefore not the source of the lack of WFE in ESL reading. However, this study did not specifically account for differences between the language groups.

By conducting more specific studies, the differences between the language and writing systems of ESL reading groups could be specifically targeted. For example, a study could be done focusing the attention of ESL Spanish readers on modifiers such as adjectives as they differ between the English and Spanish language. Similarly an additional study could observe the reading behavior of Chinese ESL speakers when focusing on function words – since functions words do not exist in the Chinese language. By looking at these specific differences between the languages, it may be possible to determine the source of the overall difference in ESL and native English behavior. Although these additional studies would not reveal the reason for the lack of WFE in ESL readers, they could provide insight into ESL reading behavior as a whole.

In regards to WFE in ESL readers, there are a number of factors that could have contributed to the lack of effect observed in this study – the first of which being the fundamental design of the study. As a method of control, all passages read by participants in this study were written to what was considered a "beginner" level. Passages were controlled for complicated sentence structure, but also for material. Because the texts were extremely readable, it is possible that the participants did not focus their attention to the texts as a whole- much less the target words within the text. Future studies could use materials written for the age and proficiency of all participants to eliminate these questions of readability.

Additionally, the lack of WFE in the ESL readers could have arose from familiarity effect. Although the words selected were considered "low frequency" according to the Corpus of Contemporary American English, it is possible that these words were still highly familiar to the

participants of this study. For example, the words “college” and “grades” may be infrequent when analyzing books, magazines, and popular media, but were perhaps extremely familiar to the college-aged participants in this study. Similarly, to control for cultural bias, passages in this study -and therefore the words contained in them - were intended to be highly familiar to the participants. Topics such as the Olympics and tourist destination allowed for culturally unbiased subject matter, but an increased familiarity of the words associated with them. In future studies, it is recommended that passages be created with this additional criterion in mind.

Although it was previously predicted that English Second Language readers would be more reliant on word frequency than native English speakers, the results of the study reveal that this is not the case. Despite their similarity in proficiencies and reading time to native English readers. ESL students and speakers demonstrate fundamentally different reading behavior than previously thought. This revelation confirms that further research is required in the field of ESL reading research and that there is an opportunity to learn more about native English reading behavior as well. is the hope of this study to begin this research process so that it may continue to grow alongside the ESL population.

# APPENDICIES

## APPENDIX 1A: Summary of High Frequency Target Words

| Paragraph | High-Frequency Word | Letter Count | Frequency (COCA) |
|---|---|---|---|
| 1 | father | 7 | 189969 |
| 2 | Years | 5 | 440548 |
| 3 | Country | 7 | 234128 |
| 4 | Money | 5 | 226140 |
| 5 | Team | 4 | 158786 |
| 6 | Time | 4 | 908345 |
| 7 | Water | 5 | 221105 |
| 8 | Problem | 7 | 165777 |
| 9 | Food | 4 | 133863 |
| 10 | Places | 6 | 267939 |
| 11 | Town | 4 | 100389 |
| 12 | History | 7 | 158601 |
| 13 | Life | 4 | 391332 |
| 14 | Student | 7 | 310752 |
| 15 | World | 5 | 408426 |
| 16 | School | 6 | 356731 |
| Average | | 5.4375 | |

APPENDIX 1B: Summary of Low Frequency Target Words

| Paragraph | Low-Frequency Word | Letter | Frequency |
|---|---|---|---|
| 1 | Village | 7 | 36040 |
| 2 | Grades | 6 | 13945 |
| 3 | Tree | 4 | 44687 |
| 4 | Luggage | 7 | 3598 |
| 5 | Sisters | 7 | 17508 |
| 6 | Flights | 7 | 8574 |
| 7 | Fruits | 6 | 8553 |
| 8 | Signs | 5 | 33700 |
| 9 | Beaches | 7 | 4966 |
| 10 | Majors | 6 | 6126 |
| 11 | Clothes | 7 | 36666 |
| 12 | Tourist | 7 | 7784 |
| 13 | Value | 5 | 67738 |
| 14 | College | 7 | 20620 |
| 15 | Cats | 4 | 9293 |
| 16 | Skills | 6 | 55280 |
| Average | | 6.12 | |

APPENDIX 1C: Summary of Passages

| # | Text |
|---|---|
| 1 | The Inuit people in Alaska protect their culture in many ways. They take care of their babies as a connected community. As the kids get older, everyone in the ***village*** works together with other families. For example, all of the **fathers** take their sons to learn how to fish together. The community sees children as their culture's future. |
| 2 | Educational policies have changed a lot over the last 20 **years**. Many colleges give more A's and B's than ever before. People say universities are doing this to keep students happy, like a business. Instead of giving students the scores they deserve, the institutions give the *grades* they want. If this practice continues, people who graduate from the schools won't be good at their jobs. |
| 3 | Scientists have found a 9,550 year-old oak tree in Sweden. The **country** has made a regulation in order to keep it safe. The law makes it illegal to get rid of the *tree.* The order also limits the number of visitors who can see it at one time. The oak will last a lot longer because of the rule from the government. |
| 4 | Discount airlines, like Ryan Air, are finding new ways to make money. The airlines' method is to charge people more **money** for other benefits. Their approach has been to have the lowest prices on flights. After people buy the ticket, the airline's plan has been to have additional fees for ***luggage***. The companies have made a lot of money with their strategy in the last year. |
| 5 | Three identical ***sisters***, or triplets, will participate in the Olympics for the first time. The girls, who are from Estonia, are finishing their training together. Their parents helped them to do the exercises to get ready for the games. The sisters will compete with each other to win first place but do their preparation as a **team**. They enjoy the practice as a family. |
| 6 | It takes a long **time** to go through security lines at an airport now. Airports want to fix this challenge because people keep missing their ***flights***. Passengers are mad because of the issue and want changes. Security thinks they can fix the concern with new machines. They will try the new changes in 2017 and see if the problem gets fixed this year. |
| 7 | In the country of Indonesia, people can buy their groceries on the sea. They travel by boat through the **water** to the shop and look for the food they want. The name of the market means "floating." They have all of the normal ***fruits*** and vegetables, like apples and potatoes, at the stand. They make a lot of money at this store because of the tourists. |
| 8 | In 2015, 56% of employees thought their office was too cold. Some of these workers thought of an answer to the **problem**. Their response was to make a lot of angry ***signs*** for the boss to see. The workers also sent e-mails to their managers as a part of the reaction. Unfortunately, the employees were unsuccessful and are looking for a new solution to warm the workplace. |

| | |
|---|---|
| 9 | The next country you should travel to is the Dominican Republic. There are amazing *beaches* and relaxing areas to go to on your travels. The holiday will provide beautiful pictures and memories forever. You can try new **food** and see different sites on your trip. There are many places to visit, but the best country for your next vacation is the Dominican Republic. |
| 10 | Canadian universities are great **places** for students to study abroad. There are many top universities in this place for several different *majors*. The nation has very cheap housing which is good for students. The bigger cities in the land have good entertainment, like festivals and nightclubs. Students looking to get out of their country should consider the country for a semester. |
| 11 | Do you know who makes the *clothes* of many soldiers in the world? A company called Wyedean makes uniforms for 60 different countries' armies. The firm started in 1852 and is located in a small **town** in England. Many people don't even know the organization is real. With more countries' armies looking for uniforms, the business hopes to grow. |
| 12 | Pamukkale, which is in the country of Turkey, always has a lot of *tourists*. The site was created in 190 B.C. by the Romans. There are more than 50 natural swimming pools in this area. Many visitors go to the spot these days for the healing power of the pools. There is a lot of **history** and culture at the location for tourists to see. |
| 13 | On June 23, 2016 the British people voted to leave the European Union. People around the world reacted to the news in many ways. European leaders were nervous that other countries would want to leave because of the information. British money lost a lot of *value* because businesses were worried by the knowledge. In general, people don't know what **life** will look like now that this report has been released. |
| 14 | Scholarships are very competitive among **students** in the US. Because of the high tuition costs, most people in high school have a wish to get money for *college*. About seven billion dollars are saved each year to help them with their dream. Students spend hours on the Internet looking for opportunities for their hope. Thanks to their hard work, over one million Americans got their desire last year. |
| 15 | Cultures have a variety of superstitions throughout the **world**. One common thought in the USA, Europe, and Asia is that black *cats* are bad luck. Many say the idea started in Egypt over five thousand years ago. People who have this view run away from black cats in the street. Other cultures disagree and say this belief is a crazy idea. |
| 16 | Educational research says children who go to **school** at age four are eventually more successful than those starting at age five. In 2015, 49% of children were in this group in the US. This type does better in reading than children who start school at age five. Researchers also claim kids of this kind will have 23% higher math *skills*. Based on the data, schools encourage this category to all parents. |

*** **bold** indicates high frequency; *italics* indicates low frequency

APPENDIX 2A: Native English High Frequency Word Fixation Duration Data

| **HIGH FREQUENCY WORDS** | | | | |
|---|---|---|---|---|
| **Participant** | **1st Fixation** | **2nd Fixation** | **Last Fixation** | **Gaze Duration** |
| 1001 | 218 ms | 109 ms | 204 ms | 531 ms |
| 1003 | 222 ms | 174 ms | 178 ms | 575 ms |
| 1004 | 197 ms | 170 ms | 194 ms | 561 ms |
| 1006 | 234 ms | 165 ms | 198 ms | 597 ms |
| 1008 | 207 ms | 185 ms | 184 ms | 576 ms |
| 1010 | 211 ms | 209 ms | 211 ms | 631 ms |
| *TOTAL* | 1289 ms | 1012 ms | 1169 ms | 3471 ms |
| *AVERAGE* | 215 ms | | | 579 ms |

APPENDIX 2B: Native English Low Frequency Word Fixation Duration Data

| **LOW FREQUENCY WORDS** | | | | |
|---|---|---|---|---|
| **Participant** | **1st Fixation** | **2nd Fixation** | **Last Fixation** | **Gaze Duration** |
| 1001 | 220 ms | 151 ms | 186 ms | 557 ms |
| 1003 | 224 ms | 212 ms | 221 ms | 657 ms |
| 1004 | 234 ms | 200 ms | 238 ms | 671 ms |
| 1006 | 182 ms | 177 ms | 173 ms | 531 ms |
| 1008 | 202 ms | 178 ms | 207 ms | 588 ms |
| 1010 | 235 ms | 222 ms | 226 ms | 682 ms |
| *TOTAL* | 1297 ms | 1140 ms | 1251 ms | 3686 ms |
| *AVERAGE* | 216 ms | | | 614 ms |

APPENDIX 2C: English Second Language High Frequency Word Fixation Duration Data

| **HIGH FREQUENCY WORDS** | | | | |
|---|---|---|---|---|
| **Participant** | **1st Fixation** | **2nd Fixation** | **Last Fixation** | **Gaze Duration** |
| 5013 | 210 ms | 271 ms | 216 ms | 697 ms |
| 5016 | 224 ms | 221 ms | 230 ms | 675 ms |
| 5017 | 188 ms | 268 ms | 225 ms | 681 ms |
| 6012 | 234 ms | 160 ms | 207 ms | 602 ms |
| 6025 | 240 ms | 195 ms | 238 ms | 674 ms |
| 7008 | 263 ms | 200 ms | 251 ms | 714 ms |
| 7011 | 200 ms | 232 ms | 241 ms | 673 ms |
| 7019 | 267 ms | 237 ms | 255 ms | 759 ms |
| 7020 | 255 ms | 246 ms | 280 ms | 781 ms |
| 7021 | 289 ms | 179 ms | 270 ms | 738 ms |
| 8022 | 281 ms | 230 ms | 260 ms | 770 ms |
| 8026 | 193 ms | 212 ms | 198 ms | 603 ms |
| *TOTAL* | 2844 ms | 2651ms | 2871 ms | 8366 ms |
| *AVERAGE* | 237 ms | | | 697 ms |

APPENDIX 2D: English Second Language Low Frequency Word Fixation Duration Data

| **LOW FREQUENCY WORDS** | | | | |
|---|---|---|---|---|
| **Participant** | **1st Fixation** | **2nd Fixation** | **Last Fixation** | **Gaze Duration** |
| 5013 | 272 ms | 219 ms | 268 ms | 759 ms |
| 5016 | 228 ms | 167 ms | 221 ms | 615 ms |
| 5017 | 230 ms | 209 ms | 215 ms | 654 ms |
| 6012 | 219 ms | 275 ms | 239 ms | 733 ms |
| 6025 | 250 ms | 279 ms | 246 ms | 774 ms |
| 7008 | 256 ms | 189 ms | 265 ms | 711 ms |
| 7011 | 245 ms | 299 ms | 246 ms | 790 ms |
| 7019 | 252 ms | 194 ms | 232 ms | 678 ms |
| 7020 | 213 ms | 261 ms | 301 ms | 775 ms |
| 7021 | 288 ms | 188 ms | 227 ms | 703 ms |
| 8022 | 239 ms | 223 ms | 234 ms | 695 ms |
| 8026 | 218 ms | 189 ms | 204 ms | 611 ms |
| *TOTAL* | 2910 ms | 2691 ms | 2897 ms | 8498 ms |
| *AVERAGE* | 243 ms | | | 708 ms |

APPENDIX 2E: ESL Language Comparison High Frequency Word Data

| **HIGH FREQUENCY WORDS** | | | | |
|---|---|---|---|---|
| ***Chinese*** | **1st Fixation** | **2nd Fixation** | **Last Fixation** | **Gaze Duration** |
| 8026 | 193 ms | 212 ms | 198 ms | 603 ms |
| 7020 | 255 ms | 246 ms | 280 ms | 781 ms |
| 7021 | 289 ms | 179 ms | 270 ms | 738 ms |
| 7008 | 263 ms | 200 ms | 251 ms | 714 ms |
| 6025 | 240 ms | 195 ms | 238 ms | 674 ms |
| 5013 | 210 ms | 271 ms | 216 ms | 697 ms |
| *TOTAL* | 1450 ms | 1303 ms | 1453 ms | 4207 ms |
| *AVERAGE* | 242 ms | | | 701 ms |
| | | | | |
| ***Spanish*** | **1st Fixation** | **2nd Fixation** | **Last Fixation** | **Gaze Duration** |
| 5016 | 224 ms | 221 ms | 230 ms | 675 ms |
| 5017 | 188 ms | 268 ms | 225 ms | 681 ms |
| 6012 | 234 ms | 160 ms | 207 ms | 602 ms |
| 7011 | 200 ms | 232 ms | 241 ms | 673 ms |
| 7019 | 267 ms | 237 ms | 255 ms | 759 ms |
| 8022 | 281 ms | 230 ms | 260 ms | 770 ms |
| *TOTAL* | 1394 ms | 1348 ms | 1418 ms | 4160 ms |
| *AVERAGE* | 232 ms | | | 693 ms |

APPENDIX 2G: ESL Language Comparison Low Frequency Word Data

| **LOW FREQUENCY WORDS** | | | | |
|---|---|---|---|---|
| ***Chinese*** | **1st Fixation** | **2nd Fixation** | **Last Fixation** | **Gaze Duration** |
| 8026 | 218 ms | 189 ms | 204 ms | 611 ms |
| 7020 | 213 ms | 261 ms | 301 ms | 775 ms |
| 7021 | 288 ms | 188 ms | 227 ms | 703 ms |
| 7008 | 256 ms | 189 ms | 265 ms | 711 ms |
| 6025 | 250 ms | 279 ms | 246 ms | 774 ms |
| 5013 | 272 ms | 219 ms | 268 ms | 759 ms |
| *TOTAL* | 1497 ms | 1326 ms | 1511 ms | 4333 ms |
| *AVERAGE* | 249 ms | | | 722 ms |
| | | | | |
| ***Spanish*** | **1st Fixation** | **2nd Fixation** | **Last Fixation** | **Gaze Duration** |
| 5016 | 228 ms | 167 ms | 221 ms | 615 ms |
| 5017 | 230 ms | 209 ms | 215 ms | 654 ms |
| 6012 | 219 ms | 275 ms | 239 ms | 733 ms |
| 7011 | 245 ms | 299 ms | 246 ms | 790 ms |
| 7019 | 252 ms | 194 ms | 232 ms | 678 ms |
| 8022 | 239 ms | 223 ms | 234 ms | 695 ms |
| *TOTAL* | 1413 ms | 1366 ms | 1387 ms | 4166 ms |
| *AVERAGE* | 236 ms | | | 694 ms |

REFERENCES

Biler, A.(2018). *The Role Of Cohesion In Second Language Reading Comprehension*. (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/4690

Cho K., & Krashen, S. (1994). Acquisition of Vocabulary from the Sweet Valley Kids Series: Adult ESL Acquisition. *Journal of Reading*, *37*(8), 662-667. Retrieved from http://www.jstor.org/stable/20172388

Coe N. in Swan, M. & Smith, B. *Learner English: A teacher's guide to interference and other problems.* (1987) Cambridge University Press.

Cohen-Shikora, E. R., & Balota, D. A. (2016). Visual word recognition across the adult lifespan. Psychology and Aging, 31(5), 488–502. https://doi-org.pallas2.tcl.sc.edu/10.1037/pag0000100.supp (Supplemental)Byers-Heinlein, K., & Lew-Williams, C. (2013). Bilingualism in the Early Years: What the Science Says. LEARNing landscapes, 7(1), 95-112.

Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available online at https://corpus.byu.edu/coca/

Enkin, E., Nicol, J., Brooks, Z., & Zavaleta, K. (2017). Reading in Spanish as a Second Language: An Eye-Tracking Study. *Reading Matrix: An International Online Journal*,*17*(1), 1-15. Retrieved March 1, 2019, from http://www.readingmatrix.com/files/16-wn327906.pdf

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*,*40*(6), 431-439. doi:10.3758/bf03208203

Karim, K. (2003). First Language (L1) Influence on Second Language Reading (L2): The Role of Transfer. *Working Papers of the Linguistics Circle,17*, 49-54.

Kim, S. Y., Liu, L., & Cao, F. (2017). How does first language (L1) influence second language (L2) reading in the brain? Evidence from Korean-English and Chinese-English bilinguals. *Brain and Language,171*, 1-13. doi:10.1016/j.bandl.2017.04.003

Krekelberg, B. (2010). Saccadic Suppression. *Current Biology,20*(5). doi:https://doi.org/10.1016/j.cub.2009.12.018

Literacy Information and Communication System (LINCS). (n.d.). Assessment Stategies and Reading Profiles: Word Recognititon. Retrieved March 1, 2019, from https://lincs.ed.gov/readingprofiles/MC_Word_Recognition.htm

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The Effects of Frequency and Predictability on Eye Fixations in Reading: Implications for the E-Z Reader Model. *Journal of Experimental Psychology: Human Perception and Performance,30*(4), 720-732. doi:10.1037/0096-1523.30.4.720

Roembke, T. C., Hazeltine, E., Reed, D. K., & McMurray, B. (2019). Automaticity of word recognition is a unique predictor of reading fluency in middle-school students. Journal of Educational Psychology, 111(2), 314–330. https://doi-org.pallas2.tcl.sc.edu/10.1037/edu0000279.supp (Supplemental)

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. Cognition, 128(3),302-319. doi:10.1016/j.cognition.2013.02.013

Snow, C. (2002). Reading for understanding : Towards a research and development program in reading and comprehension. Santa Monica, CA: RAND Reading Study Group. Whitford, V., & Joanisse, M. F. (2018). Do eye movements reveal differences between monolingual and bilingual children's first-language and second-language reading? A focus on word frequency effects. *Journal of Experimental Child Psychology,173*, 318-337. doi:10.1016/j.jecp.2018.03.014

U.S. Department of Education, National Center for Education Statistics. (2018). *The Condition of Education 2018* (2018-144), English Language Learners in Public Schools