University of Southampton

Faculty of Environmental and Health Sciences

School of Psychology

Individual Differences in Eye Movements During Skilled Reading

by

Charlotte Elizabeth Lee

https://orcid.org/0000-0003-0319-5635

Thesis for the degree of Doctor of Philosophy

November 2023

University of Southampton

Abstract

Faculty of Environmental and Life Sciences

School of Psychology

Doctor of Philosophy

Individual Differences In Eye Movements During Skilled Reading

by

Charlotte Elizabeth Lee

The body of work presented here explores the assumption of uniformity in skilled reading which assumes that all skilled readers use comparable cognitive processes and display similar eye movement patterns during reading. This thesis investigates individual differences in skilled readers' eye movements across three experiments. We consider multiple cognitive processes that are active during reading, ranging from fine-grained letter and word level processing to higher-level comprehension tasks. Consistent findings suggest that the most stable predictors of individual differences in skilled readers' word reading processes are skills related to lexical proficiency (such as vocabulary). Readers with greater lexical proficiency generally read more quickly than readers with low quality lexical representations. This difference is facilitated by faster word identification processes, especially when encountering less frequent words. In terms of encoding the position and identity of specific letters within words, consistent with findings from previous literature, skilled readers generally have a strict mechanism for encoding letter identities but are more flexible when encoding letter positions. Our results indicate that the flexibility of letter position encoding varies very little within groups of skilled readers, and as such appears to reach (near) maturation once reading skills are fully developed. A final experiment determined that skilled readers are able to adapt their reading strategies to different comprehension demands, and generally aim to read more quickly over time. Less skilled comprehenders appear to reach a threshold for how much they can speed up, even when comprehension demands are low. An important finding that arose within this thesis was that two often used offline comprehension tests, the NDRT and WIAT-II comprehension tests, never loaded together in principal components analyses, and as such appeared to measure distinct underlying constructs. As a result, we reveal a *Jingle fallacy,* an issue that occurs when two instruments are falsely assumed to measure identical constructs because they share a name. Findings and implications for future research practice are discussed.

Table of Contents

Table of Tables

Table of Figures

List of Accompanying Materials

All data and materials for Chapter 2 are available online at:

https://osf.io/fetw4/?view_only=166f33ba5c584b3aa10efd30972840eb

All data and materials for Chapter 3 including supplementary analyses are available online at:

https://osf.io/b2rdm/?view_only=cf37e55f5c804a98bf7801a8c903d5f3

All data and materials for Chapter 4 are available online at:

https://osf.io/dk82u/?view_only=88d54bdc1f9b4793ba4b2dfebf106085

Research Thesis: Declaration of Authorship

Print name: Charlotte Elizabeth Lee

Title of thesis: Individual Differences in Eye Movements during Skilled Reading

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been submitted for publication as:-

Lee, C. E., Godwin., H. J., Blythe, H. I., & Drieghe, D. (2023). *Individual differences in skilled reading and the word frequency effect*. Manuscript Submitted for Publication.

Lee, C. E., Pagán, A., Godwin., H. J., & Drieghe, D. (2023). *Individual differences and the transposed letter effect during reading.* Manuscript Submitted for Publication.

Lee, C. E., Godwin., H. J., & Drieghe, D. (2023). *The jingle fallacy in comprehension Tests for reading.* Manuscript Submitted for Publication.

Signature: ....Date: 27/06/2023

Acknowledgements

First and foremost, I'd like to thank my Supervisor, Denis for his expertise, support and guidance on this project and without whom this would not have been possible. It has been a great privilege to work with you. Thank you for having an open-door policy for my endless questions and for getting me through 'unprecedented times' when we were faced with what turned out to be a very *long* global pandemic. To Hayward, Ascen and Hazel, thank you for all your helpful comments on these papers. To Hayward in particular, for being in my corner along the way. To the many colleagues in the School of Psychology who helped me to develop as a teacher, thank you for your faith in me. To my officemates and other PhD students (past and present), thank you for the much needed debriefs over coffee and cake. To João in particular, thank you for distracting me, making me laugh, and most of all for making me go outside once in a while in the last year. To Karolina, your help collecting data for this project was invaluable and I will forever be grateful. To Diana, for your kind words at the right time, I hope you know that even the smallest kindness can have the biggest impact. To my wonderful parents, who have given me so much unconditional love (and financial support) throughout my studies to get here (don't worry, this is the last degree). To my incredible friends, old and new, thank you for being patient with me when I prioritised this document over you for the last 6 months. To Sarah and Millie in particular, for being the best friends I could hope to have, for almost 20 years. Finally, Owain, I don't know where to begin. Thank you for parting the clouds on my darkest days, for cheering me on to finish this thesis, for being my personal chef, and for holding my hand through times when we had to stay 6 feet away from everyone else.

The research for this thesis was funded by the UKRI Economic and Social Research Council South Coast Doctoral Training Partnership (Grant Number ES/P000673/1).

Definitions and Abbreviations

SFD	Single Fixation Duration, the duration of the first fixation on a target word during first pass reading, when the word was fixated exactly once.
FFD	First Fixation Duration, the duration of the first fixation on a target word during first pass reading.
GD	Gaze Durations, the sum of all fixation durations on a target word before moving forwards or backwards from it in first pass reading.
GOPAST	Go past times, the sum of all fixations made on a target word before moving on to a later portion of the sentence, this measure includes regressions to previous portions of the text.
PCA	Principal Components Analysis, a statistical method used to reduce the dimensionality of large datasets.
LMM	Linear Mixed Effects Model, a statistical model that accounts for both fixed and random effects.
GLMM	Generalized Lineal Mixed Effects Model, a statistical model that accounts for both fixed and random effects and allows modelling of non-normally distributed data.
NDRT	Nelson Denny Reading Test, a test of reading ability that includes vocabulary and reading comprehension subtests.
WIAT-II	Wechsler Individual Achievement Test 2nd Edition, a test of reading ability that includes word naming, pseudoword decoding and reading comprehension subtests.
ART	Author Recognition Test, a test of print exposure.
LexTALE	Lexical Test for Advanced Learners of English, a test of vocabulary knowledge.
RAN	Rapid Automatized Naming, a test that measures how quickly an individual can name a series of simple visually presented stimuli e.g., letters, digits, colours.

DS	Digit Span, a test of working memory which requires individuals to memorise and recall lists of numbers in forwards and backwards sequential order.
WMC	Working Memory Capacity, the number of items that can be held in working memory.
Zipf	A measure of how frequent a given word is in a language.
LF	Low Frequency, a low frequency word is on that is uncommon in a language.
HF	High Frequency, a high frequency word is one that is commonly occurring in a language.
TL	Transposed Letters, a transposed letter pseudoword is created by switching letter positions within a word.
SL	Substituted Letters, a substituted letter pseudoword is created by replacing letter identities in a word.
Multicollinearity	When independent variables in a model are correlated with each other.

Chapter 1 Introduction

Reading is a fundamental skill that shapes an individual's ability to gain knowledge, communicate with others and participate in society. Reading enables individuals to learn independently and to understand information presented around them and on the web. In contrast, illiteracy highly contributes to societal inequality. People with poor literacy skills are more likely to experience poor health and are more than twice as likely to be unemployed (World Literacy Foundation, 2022). Developing effective reading instruction is clearly something that deserves great attention and is estimated to cost $1.19 trillion USD globally each year (World Literacy Foundation, 2022). Systematic investigations of reading therefore play a vital role in informing educators and policy makers about the best means of reading skill tuition and assessment. To do this effectively, research must first understand the cognitive processes involved during reading and the differences that arise within these processes when readers have different skill levels.

The body of work presented here contributes to the understanding of what makes a skilled reader successful, and how differences in reading behaviours arise within groups of adult readers. The aim of the project was to explore the assumptions of uniformity in skilled reading, which assume that all skilled readers use similar cognitive processes and display similar behaviours (Andrews, 2012). This thesis explores individual differences in skilled reading behaviours across three experiments. The three experiments cover a broad range of levels of cognitive processing during reading ranging from fine-grained word level processing to higher-level comprehension tasks and allow for a broad view of what it means to be a skilled reader and how individual cognitive skills contribute to different aspects of reading.

For the first part of this thesis, we look closely at how common and uncommon words are processed by skilled readers and explore whether specific skills influence readers' word identification speeds. We then delve deeper into word identification processes, to explore how skilled readers use information about differences in letter positions and letter identifies in upcoming words. We assess whether cognitive skills influence this very fine-grained encoding of letter positions within words during this process. For the final empirical section of this thesis, we take a step back to consider overall reading strategies for longer texts and how readers respond to the demands of a reading task. In this we consider offline reading comprehension measures and compare the changes in reading behaviours related to skills assessed by them. However, to understand the influence of individual differences in skilled reading, we must first understand the reading process and the typical behaviour of the average skilled reader.

1.1 Understanding the Reading Process

According to the Simple View of Reading (Gough & Tunmer, 1986), the two main elements of reading are the ability to decode words and the ability to understand oral language information. Word decoding involves both phonological processes and the understanding of alphabetic codes in writing systems such as English. Most words in English can be identified by mapping graphemes (units of one or more letters) onto phonemes (corresponding sounds). It is important to note here that the specific writing system matters, such that languages with non-alphabetic writing systems such as Chinese differ from alphabetic systems in the way that they map onto oral language. In Chinese, each character usually corresponds to one syllable (Hoosain, 1992). In a similar vein, some alphabetic writing systems, for example Norwegian or Swedish, have more transparent orthographies where graphemes (written letter units) consistently map directly onto phonemes (sounds that correspond to them) (Furnes & Samuelsson, 2010). English has a particularly opaque orthography, where oftentimes the same letter combinations can produce multiple sounds (e.g., *ea* in *head* and *hear*) and different letters can produce the same sounds (e.g., *their/there/they're*), depending on surrounding letters and context. From this example we can see that the direct mapping of orthography to phonology is not always sufficient, and for some exception words it is completely ineffective (e.g., yacht). The process of word recognition in English is therefore, in comparison, more complex and also requires some direct mapping of orthography to semantic information for exception words. Computational models of visual word recognition, for example the DRC model (Coltheart et al., 2001); the Triangle model (Harm & Seidenberg, 2004; Plaut et al., 1996); and the CDP+ model (Perry, Ziegler, & Zorzi, 2007; 2010), agree that two cognitive processes are involved in word reading, one that translates orthography to phonology and then to meaning, and one that is able to translate orthography straight to meaning.

While it is clear that word recognition is an integral part of reading success (a text simply cannot be understood if words cannot be identified), additional processes are required for effective understanding of a text. For a broader view of whole text reading and comprehension, the Construction-Integration (CI) model (Kintsch, 1998) explains that a text can be represented in three ways: a surface structure, a textbase and a situation model. The surface structure holds semantic representations of singular words from the text. The textbase integrates the meaning of each word into an explicit semantic representation of the text as a whole (Kintsch & Rawson, 2005). The situation model integrates the textbase with the readers' world knowledge, creating a representation of the situation. A textbase is sufficient for shallow comprehension, but a situation model is necessary for deeper understanding.

To investigate moment-to-moment processes during reading for this project, eye movement patterns are measured and then modelled by both text and reader-skill predictors. Eye movement measures enable researchers to explain complex cognitive processes that are active during specific stages of word recognition and reading (Liversedge & Findlay, 2000; Rayner; 1998).

1.2 Eye Movements During Reading

Due to the anatomy of the human eye, there are limitations on the amount of information that can be obtained in a single moment from a visual stimulus. The retina contains two types of photoreceptors, rods and cones. Rods work in low light levels and have low spatial acuity, allowing only for the perception of coarse-grained information from the visual field. Cones, on the other hand allow for the perception of more fine-grained information from the visual field, including colour. Whereas rods are more evenly distributed across the retina, the majority of the cones are densely located around the central 2 degrees of the retina, called the fovea, which makes the fovea tuned to detect highly detailed visual information. The parafoveal region extends an additional 3 degrees from the fovea and features a much lower density of cones. The parafoveal region has therefore a considerable reduction in visual acuity, which means that information from the visual field may only be coarsely encoded. This results in limited information processing of texts presented in this region. The periphery extends beyond the parafovea, and the number of cones is further dramatically reduced, which means that visual acuity is further degraded, and detailed visual information cannot be gathered.

A reader must, therefore, usually direct the fovea at parts of a text where highly detailed visual information can be acquired. Readers make a series of saccades (ballistic movements) and fixations (moments of relative stillness). During a saccade vision is blurred, which means the reader is functionally blind, and cannot extract new information from the visual field (Matin, 1974). Fixations allow a reader to encode visual information from locations within the text. Research has shown that eye movements are closely linked to cognitive processes during reading. For instance, longer fixation durations are often associated with slower processing of difficult words (Liversedge & Findlay, 2000; Rayner, 1998; Rayner, 2009). A reader may also make regressions, where a backwards saccade is made to re-fixate a previous point in the text, to either correct mislocated fixations (Clifton & Staub, 2011; Mitchell et al., 2008) or to parts of the text that were difficult to understand during the first pass (e.g., where there is ambiguous language) (Frazier & Rayner, 1982; Rayner & Frazier, 1987). Regression rates are therefore interpreted as good indicators of late cognitive processes (after a word or portion of a text has been read for the first time) (Liversedge & Findlay, 2000).

Decisions about when and where to move the gaze are crucial to the reading process and are influenced by several factors originating from both the text and the reader. On average, fixations typically last 200 - 250 *ms* and saccades move across 7-9 letters on average (Rayner, 1998). However, variability in fixation durations can arise depending on the difficulty associated with processing a particular feature of the text (for a review see Rayner, 2009). To read efficiently, an individual must carefully plan eye movements to direct the gaze to parts of the text that will allow them to gain the most useful information. Generally, as the processing demands or difficulty of a text increases, readers need to make more fixations (with an accompanying increase in the number of saccades), their fixations are longer and saccades are shorter, and there is an increase in the rate of regressions (for a review see Rayner, 2009). Readers must use as much information as possible from the text during a fixation to programme a saccade that will land on the most appropriate location for the next fixation.

1.2.1 Eye-trackers and Early versus Late Eye Tracking Measures

Eye-trackers enable researchers to record participants' eye movements during tasks presented in the visual field. Current stationary eye-trackers have a high degree of spatial accuracy and temporal resolution and record a larger number of measurements related to fixations, saccades and regressions made by the participant. The SR Research Eyelink 1000, which is the most commonly used, state-of-the-art eye tracker in reading research, is capable of recording eye movements at rates of 1000 frames per second. A good calibration prior to experimental procedures paired with an appropriate viewing distance in proportion to the size of the presented texts will allow the samples to be spatially accurate to about half a character. A series of algorithms collate the samples into measures that are assigned as fixations or saccades depending on the degree of movement detected. Since eye tracking methods do not require secondary cognitive tasks like lexical decisions (where a participant decide whether a single word presented is a real word or a non-word), or word naming tasks (where participants must also vocalise the word they have read), this methodology is ideal for studying natural reading behaviours.

Using eye tracking equipment, we can separate eye movement patterns into early and late eye measures to understand the time course of reading. Measures that relate to first pass reading of a word (the first time a word is read, before rereading) such as first fixation durations, single fixation durations, and gaze durations are described as capturing "early" stages of word processing, whereas go past times (the sum of all fixations on a target word before moving towards later regions, this measure includes regressions to earlier regions) and total word reading times, are described as capturing "late" stages of word processing (Rayner, Sereno, et al., 1989).

Some key eye movement measures discussed in this thesis are defined in Table 1.1 below with some indication of the stage of processing they are related to.

Table 1.1

Definitions of Eye Movement Measures

Measure	Definition
Skipping	An indication of whether a target word was skipped (not fixated) during first pass reading. This is a very early measure of word processing based on parafoveal processing.
First Fixation Duration	The duration (*ms*) of the first fixation on a target word during first pass reading. This is an early measure of word processing.
Single Fixation Duration	The duration (*ms*) of the first fixation on a target word when the word was fixated exactly once during first pass reading.
Gaze Duration	The sum of all fixation durations (*ms*) on a target word before moving away from it in first pass reading.
Go Past Time	The sum of all fixations (*ms*) made on a target word before moving on to a later portion of the sentence, including regressions to previous parts of the text. A late measure of word processing.
Regression Rate	The percentage of regressive saccades made either in or out of a target word or portion of the text (the former is called the regressions-in rate, the latter the regressions-out rate).
Number of Fixations	A global measure of reading behaviour on a word or region that gives the total number of fixations made during reading. Includes but not limited to fixations made after the first pass.
Average Fixation Duration	A global measure of reading behaviour on an entire text that gives the mean duration of fixations made when reading. Includes fixations made after the first pass.
Average Forward Saccade Length	A global measure of reading behaviour on an entire text that gives the mean length of forward saccades made during reading. Includes saccades made after the first pass.
Total Reading Time	The total duration of time spent reading. It can be calculated for target words, phrases, sentences, passages or whole texts.

Since modern eye-trackers have such a high degree of temporal and spatial accuracy, they are capable of utilising the position of a participant's gaze to trigger components within an experiment. For instance, a paradigm used in the second experiment in this thesis is the *gaze contingent boundary paradigm* where the location of the participant's gaze triggers a change in the presented display when it crosses an invisible line embedded within the text (McConkie & Rayner, 1975; Rayner, 1975). This display change occurs during the saccade that triggers the mechanism and is therefore undetected (Matin, 1974). Information presented in the parafovea, beyond the invisible boundary, is more coarsely encoded compared to information in the fovea, and this paradigm makes it possible to determine the extent to which a word was processed prior to the saccade that triggered the display change. The level of parafoveal processing influences the length of the fixation that lands on the word that has now replaced the original string that had been presented in the parafovea. As a result, it is possible to explore what kinds of information can be extracted from the parafovea to guide subsequent eye movements and to what extent.

1.2.2 Word Skipping

Occasionally, parafoveal processing gives the reader enough information to decide to skip an upcoming word, that is, to not fixate upon it directly, instead planning a saccade to the next word. During reading approximately one third of words are skipped on average (Brysbaert et al., 2005). Many factors influence the decision to skip a word, including text-related factors, individual differences in reading skill (which we will discuss in more detail later) as well as limits of the oculomotor system. The strongest known text-based predictor of word skipping is word length. Short words (e.g., 3 letter words) are skipped more frequently than longer words (Vitu et al., 1995; Drieghe et al., 2004; Just & Carpenter, 1980; Rayner, Sereno, & Raney, 1996). This effect of word length has been found to be more closely related to the spatial width of the word rather than the number of letters directly, as little difference in skipping is observed for words that have a different number of letters but spread across a similar amount of space due to wider fonts or letter spacing (Hautala et al., 2011). As such in the context of word skipping, word length can be considered to be more a visual factor than a linguistic factor.

In terms of the oculomotor system, the likelihood that the upcoming word is skipped also depends on the distance from the current fixation location. A saccade that is launched from a nearby location will be more likely to skip over the subsequent word than a saccade that is launched from a farther distance. Rayner, Sereno and Raney (1996) found that a five-letter word was more likely to be skipped (25 % of the time) when the preceding fixation was located three characters from the beginning of the word, than when it was seven characters away (10 % of the time).

Linguistic properties of the text, such as word frequency (how often a word occurs within a language) and predictability from preceding context also have robust, but smaller effects on word skipping than word length and the saccade launch site. A meta-analysis by Brysbaert et al. (2005) concluded that a high frequency word is skipped more often (16 % of the time) than a low frequency word (11 % of the time). Word length interacts with word frequency for skipping behaviours, and as a result, for longer words (7 letters +), the effect of word frequency is much smaller than for 5 letter words (Rayner & Fischer, 1996). Brysbaert et al. (2005) also concluded that words that are predictable from the preceding context are skipped more often (8 % more) than words that are unpredictable. To control for predictability in experimental investigations of other effects, stimuli are usually pretested using a cloze task. A group of participants independent from the main study are given the materials up until the pre-target word and are asked to complete the sentence. If a word is predictable from the preceding text, participants will produce it more often than if it is unpredictable. These words will then be skipped more often than their less predictable counterparts (e.g., Balota et al., 1985).

1.2.3 Fixation Durations

The oculomotor and text-based features discussed above, also impact upon fixations on words when they are not skipped. If a saccade is launched farther from a target word that is not skipped, fixations land closer to the beginning of words (McConkie et al., 1988) and fixation durations on the target word increase (Pollatsek et al, 1986). The latter effect is probably due to reduced parafoveal processing when the word was in a region of comparatively low visual acuity on the previous fixation. Research has concluded that long words (compared to short words) (Rayner et al., 1996; Kliegl et al., 2004; Pollatsek et al., 2008), low frequency words (compared to high frequency words) (Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Juhasz & Rayner, 2003; 2006; Pollatsek et al., 2008; Reingold et al., 2010; Staub et al., 2010) or words that are unpredictable from the preceding context (compared to those that are predictable) (Erlich & Rayner, 1981; Balota et al., 1985; Binder et al., 1999; Drieghe et al., 2004; Rayner & Well, 1996; Rayner et al., 2001; 2011) are fixated for longer durations.

1.2.4 Perceptual Span

On average, skilled readers can extract information such as word spacing from the parafoveal region up to 14 – 15 character spaces away from the current fixation to the right in English word (towards the upcoming word) and 3 – 4 character spaces to the left in alphabetic languages read from left to right (McConkie & Rayner, 1975). This asymmetric visual field facilitates parafoveal processing of information in upcoming words. For word identification, higher visual acuity is necessary, and the perceptual span is smaller, allowing 3 – 4 letters to the left and 6 – 7 letters to the right (Rayner & Bertera, 1979; Rayner et al., 1981).

1.2.5 Models of Eye Movement Control during Reading

We have presented an overview of what is known about the cognitive processes active during reading as established by experiments that recorded eye movements (for a review, see

Rayner, 2009). Several computational models of eye movements during reading have been formulated based on evidence from eye tracking research, such as the E-Z Reader model (Reichle et al., 1998; Reichle, 2011), the SWIFT model (Saccade-generation With Inhibition by Foveal Targets, Engbert et al., 2002; Richter et al., 2006), OB1-Reader (Snell et al., 2018) and Über-Reader (Reichle, 2020). These models simulate the way a readers' eyes move across a text and adapt to simulate the eye movement patterns associated with difficulties in processing related to oculomotor control and features of a text.

The E-Z Reader (Reichle et al., 1998; Reichle, 2011) assumes that lexical processing is serial, whereby attention moves from one word to the next and that processing of an upcoming word does not begin until the currently fixated word has been fully identified. According to the E-Z reader, the oculomotor system begins planning a saccade from a word to the next when an early stage of word processing called the familiarity check is complete. Attention shifts from the currently fixated word to the upcoming word in the parafovea when the second stage of word identification (for the current word) is complete (the completion of lexical access). Because attention moves to the upcoming word faster than the saccade can be programmed, some processing of the upcoming word begins before the eyes have moved (Rayner et al., 2006). This is how the model accounts for parafoveal processing. The E-Z Reader model assumes that the default target for the next saccade is always the following word, but if the upcoming word can be processed quickly in the parafovea, it will be skipped. Therefore, lexical processes are the driving forces that move the eyes through a text in the E-Z Reader model (Reichle et al., 1998; Reichle, 2011).

The SWIFT model (Engbert et al., 2002; Richter et al., 2006) instead assumes that word processing occurs in parallel. The SWIFT model describes a four-word window of attention within which lexical processing takes place (with a processing gradient reflecting the decreasing visual acuity from the centre of the visual field). The model suggests that saccades occur automatically, via a timing mechanism. If the central word is difficult to process, the current fixation duration can be extended via inhibition of the next saccade. SWIFT (Engbert et al., 2002; Richter et al., 2006) therefore puts less emphasis on cognitive influences compared to the E-Z Reader model (Reichle et al., 1998; Reichle, 2011), though both models are able to account for text-level influences like word frequency and predictability on fixation durations.

More recent models have also been formulated to account for key findings from both reading studies and isolated word paradigms (OB1-Reader, Snell et al., 2018; Über-reader, Reichle, 2020). OB1-Reader (Snell et al., 2018), like SWIFT (Engbert et al., 2002; Richter et al., 2006), assumes that word identification processes occur in parallel, whereas Über-reader

(Reichle, 2020) extends the ideas of serial processing within the E-Z Reader model to a model that incorporates influences other than word processing such as discourse processing (Reichle et al., 1998; Reichle, 2011).

A crucial aspect for this thesis, is that computational models of eye movements generally assume that all readers use similar processes when reading but allow for some variation in the timing and efficiency of these processes (e.g., E-Z Reader, Reichle et al., 1998; 1999; SWIFT, Engbert et al., 2005; Über-Reader, Reichle, 2020; OB1-Reader; Snell et al., 2018). For example, the E-Z Reader model successfully accommodates for the different eye movement patterns of older readers (Rayner et al., 2006) and children (Reichle et al., 2013). However, differences are largely considered by groups of readers (e.g., children vs adults, Reichle et al., 2013) rather than by individuals within these groups, and as such, current computational models of eye movement control in reading do not make specific predictions based on individual differences. As a result, we will not discuss them in much more detail. However, if qualitative differences between individuals' eye movements in relation to their cognitive skills occur (they differ in the processes that are used to recognise and integrate words into sentences for comprehension), computational models would face problems accounting for them. The following section explains how readers continue to differ once reading is fully developed.

1.3 Individual Differences in Eye Movements during Reading

As mentioned earlier, there is an often-implicit assumption that skilled readers all read in a similar manner (Andrews, 2012). However, even within groups of skilled readers, considerable variation in eye movement measures have been observed (see Andrews, 2012; 2015 for reviews). Fixation durations can vary between approximately 50 - 600 *ms* and saccade lengths can be between approximately 1 - 20 letter spaces in English (see Rayner, 2009). In eye movement research, the term *skilled reader* is often used interchangeably with *average reader* and most often refers to an individual who does not struggle to read. While in this thesis we focus our attention on adult readers who do not experience reading difficulties, a number of studies have assessed individual differences in developing readers and those who have reading difficulties such as dyslexia (for a review of the literature on children's eye movements during reading see Blythe & Joseph, 2011). In comparison to the substantial literature concerning children's reading behaviours, relatively little attention has been given to understanding how individual differences can account for variation observed *within* the skilled adult reader population. We highlight here what is currently known about differences observed in this population. We then consider approaches taken for investigations using large test batteries.

Highly skilled readers are typically associated with faster reading speeds, shorter fixations, longer saccades and fewer regressions compared to those with poor or average reading abilities (Rayner, 1998; Ashby et al., 2005). However, the relationship between reading skill and reading speed is not as straightforward (Underwood et al., 1990). Highly skilled readers do not always demonstrate faster overall reading rates and there is a speed-accuracy trade-off where increasing reading speed eventually results in poor comprehension (Rayner et al., 2016). In early studies of individual differences between skilled adult readers, participants were most often grouped based on measures of overall reading ability (e.g., Ashby et al., 2005; Jared et al., 1999; Haenggi & Perfetti, 1994). Ashby et al. (2005) found that average readers read less frequent words more slowly than skilled readers, and made greater use of wider sentence contexts to facilitate identification of less frequent words. We return to discuss this study by Ashby et al. (2005) in more detail when describing tests selected for the current body of work.

1.3.1 Approaches taken in Test Battery Investigations

Human cognitive skills form a multilevel hierarchy, with an overall general factor (Spearman's g) and more specific levels relating to particular tasks (Deary, 2000). Reading relies on multiple specific cognitive skills that enable an individual to make sense of written information, such as letter and word decoding, and language comprehension. Reading also requires the perception of visual information, and more general cognitive abilities such as working memory processes and complex thinking processes, to extract and integrate text meaning with reference to wider knowledge (Kintsch, 1998). An individuals own cognitive skills are often correlated (Deary, 2000), therefore, to fully understand which cognitive skills drive differences observed in reading behaviours and eye` movements during reading, the influence of multiple skills must be accounted for. Some research has assessed multiple skills as predictors of eye movement behaviours using large test batteries to address this issue (e.g., Kuperman, et al., 2018; Kuperman & Van Dyke, 2011; Long & Freed, 2021). These studies have usually examined readers’ eye movements during natural reading using a corpus of uncontrolled reading materials. To account for the influence of text-level characteristics such as word frequency and length within these data, text related variables are then entered as predictors in statistical models alongside skills measured by the test battery (e.g., Kuperman & Van Dyke, 2011). Issues of multicollinearity often arise when including multiple skills that correlate with one another in statistical models. Various statistical approaches have been used to overcome problems with multicollinearity in previous research. For example, principal components analyses are frequently used to group skills based on shared variance to reduce the number of predictors within a model (Veldre & Andrews, 2021; Hasenäcker & Schroeder, 2021). Alternative approaches include using Random Forests non-parametric regression techniques to assess the relative importance of reader skill and text-level predictors (Kuperman et al., 2018) and Structural Equation Modelling (Long & Freed, 2021).

There are pros and cons for each approach discussed here and whilst smaller investigations of individual skills and those directly testing theoretical concepts have the advantage of being easily interpretable, larger investigations allow for more skills to be assessed simultaneously. Larger investigations can examine the influence of one skill whilst controlling for the influence of another but often have to overcome issues with multicollinearity.

The approach for the current body of work focussed at first on investigations using a large test battery during reading experiments where specific word-level characteristics were experimentally manipulated. This approach enabled us to control for other text characteristics that are known to influence eye movements, for example word length and predictability. Thus, we were able to reduce the number of text predictors that were necessary to include in the models, allowing us to include more individual differences predictors without the risk of model overfitting. Next, we describe how tests were selected for the current test battery and describe theoretical explanations and relevant findings that have been associated with them.

1.4 Test Battery Selection

When selecting tests for the current test battery, a high priority was to include tests that would not reach ceiling levels in adult populations. In other words, we required tests with enough sensitivity to differentiate between average and highly skilled individuals in all measures. General English proficiency tests were excluded from the final battery as these are aimed primarily at second language learners and were considered to likely reach ceiling level in native speakers. A large proportion of existing reading comprehension tests are aimed at assessing reading difficulties or reading levels of children, therefore these were also dismissed due to potential ceiling levels for skilled adult populations. Consequently, consideration moved towards reading measures in intelligence quotient tests as these were likely to include normed items which would not reach a ceiling level in adult populations. In the following subsections we consider where theory and previous research link these skills to individual differences in eye movements more generally.

1.4.1 Reading Ability Tests

First, we considered tests of overall reading ability to allow for comparisons with previous studies of individual differences in eye movements that grouped readers based on high and low scores on reading ability tests (e.g., Ashby et al., 2005). Reading ability can be assessed using offline standardised tests that most commonly include measures of reading comprehension and vocabulary size. These scores are then usually combined to give some indication of a reader's overall reading ability.

Ashby et al. (2005) examined the effects of word frequency and predictability on the eye movements of average readers in comparison to more skilled readers. They split a sample of adult participants into two groups based on their scores on the Nelson Denny Reading Test (NDRT; Brown et al., 1993). In the first experiment, participants were given sentences to read whilst their eye movements were recorded that contained a high or low frequency word embedded in the same neutral (unpredictable) sentence frame. Their results showed that overall, the group of high scoring readers generally made shorter fixations than the group of average scoring readers. In addition, highly skilled readers had shorter fixation times on low frequency words in comparison to the group of average readers. For the average group there seemed to be a greater cost to processing speeds for low frequency words than there was for the more skilled group. Ashby et al. concluded that skilled readers had greater familiarity with less commonly occurring words, and were therefore able to identify them more quickly from their mental lexicon.

In a second experiment by Ashby et al. (2005) high and low frequency words (e.g., *house* vs *igloo*) were embedded in two sets of sentence frames, one in which the target word was predictable (HFP*: Wanting children, the newlyweds moved into their first house and were excited.* LFP: *The traditional Eskimo family lived in the igloo built from snow and ice.)* from the preceding context and one in which the target word was unpredictable (HFUP*: The traditional Eskimo family lived in the house built from snow and ice.* LFUP*: Wanting children, the newlyweds moved into their first igloo and were excited*). They found that predictability and frequency effects were additive for skilled readers, with high frequency, predictable target words receiving the shortest fixations and low frequency unpredictable targets received the longest. However, frequency and predictability interacted with reading skill for the less skilled readers who were greatly facilitated by predictability, which reduced the impact of encountering a low frequency word. Ashby et al. (2005) concluded that average readers were using the surrounding context of a sentence to identify words they are less familiar with to a higher extent than the more skilled readers.

Although measures of overall reading ability like the NDRT are standardised and are designed to give a good indication of reading ability, research has found that such measures (and comprehension subtests in particular) are often inconsistent with each other due to differences in underlying constructs (Cutting & Scarborough, 2006; Keenan et al., 2008; Kendeou et al., 2012; Mézière et al., 2023). Therefore, to allow for clear and reliable comparisons with previous studies that have used the NDRT (e.g., Ashby et al, 2005; Veldre & Andrews, 2014), it was selected as one of the more appropriate measures of reading ability for our test battery. The NDRT includes two subtests, one for vocabulary and another for reading comprehension. The vocabulary subtest consists of 80 short descriptions of target words where participants are required to select the most appropriate word or phrase from five options to describe the key word's meaning. The reading comprehension subtest includes 7 texts followed by 5-7 multiple choice comprehension questions (with 5 possible answers) presented below each passage. This test is primarily used as a diagnostic tool for reading difficulties but includes norms for skilled adult readers. Test reliability for the NDRT ranges from .89 to .98 (Brown et al., 1993). In recent studies, a half-timed version of this test (where the standard time given to participants to complete the vocabulary and comprehension tasks is reduced by half) has been found to result in a more normal distribution of scores for skilled adult readers than the full-timed version (Andrews et al., 2020). We therefore opted to use the half-timed version of the NDRT.

With regards to differences in underlying constructs measured by different reading ability tests, and during the process of selecting appropriate measures for our test battery, it became apparent that while most reading ability tests include a measure of comprehension, they often differ in the other subtests that are included. For this reason, and to compare findings to studies of children that utilise this test (e.g., Pagán et al., 2021), the Wechsler Individual Achievement Test (WIAT-II UK; Wechsler, 2005) was included as an additional measure of reading ability. The WIAT-II reading ability test includes three subtests, word reading, pseudoword decoding and reading comprehension. The word reading subtest is comprised of a list of real words that participants must read aloud. The pronunciation difficulty of the words increases further down the list, for example, the test begins with "the" and finishes with "hierarchical". The pseudoword decoding subtest requires participants to read a list of orthographically legal (pronouncable) non-words aloud (e.g., "*flimp*"). The reading comprehension subtest includes passages of varying length and complexity (short fictional narratives, non-fiction text, news articles) that participants read aloud or silently before answering literal and inferential comprehension questions aloud when asked by an experimenter. The internal consistency reliability reported in the WIAT-II user manual is high (α = .98). The WIAT-II is normed with UK and US populations of children from 4 to 16 years, and adults 17 to 85 years. Though the decision to include two measures of overall reading ability may be unusual, the subtests included alongside comprehension for each, were distinct, allowing the opportunity to model individual differences based on word reading and pseudoword decoding measures as well as comprehension scores. Due to the inclusion of separate word reading and pseudoword decoding subtests in the WIAT-II, comparisons with studies would be allowed that included them as predictors of individual differences in children's reading processes (Gómez et al., 2021; Hasenäcker & Schroeder, 2021). Inconsistencies based on which test is selected for use in scientific research (NDRT vs WIAT-II) are discussed throughout this thesis and are the focus of further investigations in Chapter 4. Next, we discuss research that has focussed on investigating the influence of the quality of a readers' lexical representations on reading behaviours.

1.4.2 Lexical Quality

1.4.2.1 The Lexical Quality Hypothesis

The Simple View of Reading (Gough & Tunmer, 1986) suggests that reading difficulties may arise from poor decoding or oral language processing skills. According to the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002), high quality lexical representations facilitate word identification and in turn, enable better reading. A high-quality lexical representation is described as a representation of a word that has strong connections between orthography, phonology, and semantic information. Strong connections mean that when the word is encountered during reading it will trigger rapid retrieval and be identified more quickly than a word that has a low-quality representation. The Lexical Quality Hypothesis emphasises word knowledge and experience, which suggests that adult readers will have more high-quality representations than developing readers, and that better readers can be identified by the number and quality of the lexical representations they hold. In developing readers, who do not yet have much reading experience or word knowledge, word decoding is a strong predictor of reading comprehension, whereas for skilled readers vocabulary size has been found to be a stronger predictor (Braze et al., 2007; Protopapas et al., 2007). Similarly, shorter fixation times on words (indicating faster word identification) have been linked to greater vocabulary knowledge in adult readers (Cop et al., 2015).

The Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002) has guided many investigations into individual differences in skilled readers' word identification during masked priming, lexical decision tasks, and moment-to-moment text processing. Lexical quality has been linked to faster word identification and better use of upcoming information presented in the parafovea (Andrews, 2012; 2015; Andrews & Bond, 2008; Andrews & Lo, 2013; Andrews et al., 2020; Veldre & Andrews, 2014; 2015a; 2015b; 2016; Veldre et al., 2017). Reading comprehension and vocabulary assessments have been included alongside tests of spelling ability designed by Sally Andrews' Lab that have been proven to be sensitive to individual differences in university student samples (Andrews & Hersch, 2010). Spelling ability – at least in a language with opaque orthography to phonology rules such as English - is proposed to be a good assessment of the quality of orthographic and phonological information held in the mental lexicon (Andrews, 2015; Perfetti, 2007).

Veldre and Andrews (2013a) assessed readers' spelling, vocabulary and reading comprehension and assessed readers' perceptual spans using a *gaze-contingent moving window paradigm* (McConkie & Rayner, 1975) whereby the reader is only allowed a fixed window of text around their current fixation, which is programmed to move along with their eyes. The text that falls outside of the window is masked, for example by strings of X. They found that readers who were skilled in both reading comprehension and spelling had larger perceptual spans than those who were less skilled readers and spellers. More skilled readers were able to extract and use information gained in the parafovea more efficiently. Readers with high reading comprehension and spelling abilities were more disrupted when their forward perceptual span was very restricted and benefitted more from larger windows (shown by longer forward saccade lengths) than those with lower reading comprehension and spelling skills. Some research has considered skills related to lexical quality such as vocabulary, reading comprehension using measures that combine scores on tasks (via principal components analyses) (see Veldre & Andrews 2021). However, studies have also found that these skills individually predict distinct eye movement patterns where readers who are good comprehenders are not always good at spelling and vice versa (Andrews & Lo, 2012).

1.4.2.2 Spelling

Spelling has been found to predict patterns in eye movement studies that are distinct from patterns predicted by overall reading ability measures (Andrews & Lo, 2012; Veldre et al., 2017). Veldre et al. (2017) revealed that good spelling abilities predicted a reduced word frequency effect where better spellers were able to identify and read low frequency words more quickly (indexed by smaller gaze durations and total fixation durations on target words) than less skilled spellers. A reduced word frequency effect was not found in relation to better overall reading abilities when spelling ability was accounted for in this study. Veldre et al. (2017) found that proficiency in spelling predicted longer saccades, more word skipping and saccades that landed further into a target word than low spelling abilities. Proficient reading abilities on the other hand predicted fewer fixations, shorter sentence reading times, shorter fixation durations and fewer regressions than low reading abilities. Veldre et al. concluded that spelling and reading abilities contribute to different aspects of eye movement control (see also Slattery & Yates, 2017).

As suggested by the Lexical Quality Hypothesis, spelling appears to be specifically related to orthographic processing and offline measurements of spelling ability, such as the accuracy of written word spellings from a dictated list and the recognition of misspelled words (e.g., Andrews & Hersch, 2010), provide a reliable index of greater orthographic knowledge. It follows that for adult readers, variation in spelling abilities may indicate differences in word identification processes, even where reading abilities are comparable. By including spelling measures in a larger test battery, we may be able to determine how important this skill is when others are accounted for. From the evidence presented here it is reasonable to assume that spelling will be a strong predictor of individual differences in eye movements in relation to word identification processes. Spelling dictation and spelling recognition measures by Andrews and Hersch (2010) were therefore included in our experiments in order to reliably compare findings with other studies on skilled readers that use these measures (Veldre & Andrews, 2014; 2015a; 2015b; 2016; Veldre, Drieghe, & Andrews, 2017). The spelling dictation task consists of 20 single words that are presented from an audio recording. Examples of the words in a sentence and given and participants are required to write down the correct spellings. The spelling dictation task was reported by Andrews and Hersch (2010) to have a test–retest reliability coefficient of .90. The spelling recognition task consists of a list of 44 correct and 44 incorrect word spellings and participants are required to select all the words that are spelled incorrectly. Scoring is such that each correctly spelled word selected scores 1, and each incorrectly spelled word scores -1. The spelling recognition was also reported to have very high test–retest reliability (.93) (Andrews & Hersch, 2010). These tasks were created to assess skilled reader's spelling abilities for research purposes.

1.4.2.3 Vocabulary

Vocabulary size is also closely related to lexical quality as it represents the number of words that are represented in the mental lexicon and their related meanings (Andrews, 2015; Perfetti, 2007). Larger vocabulary sizes have been found to predict generally shorter fixation durations in adult readers (Cop et al., 2015). In addition, readers with a larger vocabulary have also been found to read less common words more quickly than readers with smaller vocabularies (Cop et al., 2015). Kuperman et al. (2018) also found that large vocabularies predicted better reading comprehension in adults, which is in line with previous research that has suggested that for skilled readers, vocabulary size is a stronger predictor of reading comprehension than word decoding (Braze et al., 2007; Protopapas et al., 2007).

The LexTALE test (Lemhöfer & Broersma, 2012), a measure of vocabulary knowledge in English, was therefore included in the current test battery (in addition to the vocabulary subtest included in the NDRT) based on its use in Cop et al. (2015) and also because it had a short testing duration (approximately 3.5 minutes). During the LexTALE test participants are presented with 60 words sequentially and are required to indicate whether each stimulus is a real English word or a pseudoword. The test includes 40 real English words and 20 non-words. To correct for the unequal proportion of real words and non-words, a score is calculated by taking the mean of the percent of correct responses (yes) for real words and the percent of correct responses (no) for nonwords ((number of correct words/40 * 100) + (number of correct nonwords/20 * 100))/2). Lemhöfer and Broersma (2012) report split-half reliability for the LexTALE test ranging from .64 to .81. It was initially created to provide a measure of English proficiency in second language learners, however the test has been used in studies comparing native and non-native adult speakers without performance reaching ceiling levels (e.g., Ernestus, et al., 2017).

1.4.2.4 Word and Pseudoword Decoding

Word and pseudoword decoding skills (which as we mentioned earlier are included as subtests within the WIAT-II) are also linked to the quality of a reader's lexical representations (Perfetti, 1992; 2007; Perfetti & Hart, 2002). Word decoding has been found to play a key role in children's word reading processes (Hasenäcker & Schroeder, 2021), and is a limiting factor for children's comprehension (Perfetti, 1985; Perfetti & Hart, 2001). There is also substantial evidence that for adults, poor decoding skills affect word recognition and comprehension (Rayner, Schotter et al., 2014). Kuperman et al. (2018) found that good word decoding in adults predicted an increase in word skipping and a reduction in the number of fixations, and total reading times observed during passage reading compared to poor word decoding. In a study by Hasenäcker and Schroeder (2021), children's word reading and pseudoword decoding scores, along with vocabulary and spelling measures were grouped together based on analyses of shared variance (via a principal components analysis) and were suggested to contribute to measures of orthographic knowledge. Composite scores based on the variance shared by these measures were found to predict variability in children's eye movements when the orthography of a preview for a target word was altered (for example to create a pseudoword preview by transposing two letters in the target word). Similar findings about individual differences in children's orthographic encoding processes were found related to overall reading abilities (measured by the WIAT-II) (Pagán et al., 2021).

1.4.3 Reading Experience

Whilst reading experience can be considered to be closely related to lexical quality because when individuals gain more reading experience, they have more opportunities to develop a greater number of high-quality lexical representations (Perfetti, 1985), it has also been considered separately in some studies (e.g., Chateau & Jared, 2000; Gordon et al., 2020). Chateau and Jared

(2000) matched university students on their reading comprehension scores (via the NDRT comprehension subtest). Despite this matching of reading comprehension abilities, participants had different levels of print exposure, which is a good indication of reading experience, as measured by the Author Recognition Test (ART; Stanovich & West, 1989). This test requires participants to select the names of real authors from a list of real and fake author names. They found that when reading comprehension scores were controlled for, greater reading experience still predicted better scores on word naming and lexical decision tasks than less reading experience. The authors suggested that reading experience enhances general knowledge and improves word recognition by increasing vocabulary size. Additionally, in a recent study, Gordon et al. (2020) found that when controlling for scores on Rapid Automatized Naming tasks (which will be discussed below), greater reading experience was associated with more word skipping, shorter gaze durations, and modulated the effect of word frequency.

The body of work presented in this thesis also aimed to determine whether differences in eye movements can be seen in relation to differences in print exposure, and whether these differences are due to reading experience alone or via shared variance with other, related skills, for example vocabulary and spelling abilities. The UK Author Recognition Test (Acheson, Wells, & Macdonald, 2008), was therefore selected as an appropriate measure of print exposure for skilled adult readers. This task consists of a list of 65 real author names and 65 foil names. Are required to select the names of real authors and scoring is such that a correct selection of a real author scores 1 and an incorrect selection of foil name scores -1. The ART has been found to have high test reliability (α = .84 in Stanovich & West, 1989; and α = .75 - .89 in a review by Mol & Bus, 2011). In addition, McCarron and Kuperman (2021) found it to be an informative and accurate measure of print exposure for Native English university students.

1.4.4 Other Cognitive Skills

1.4.4.1 Rapid Automatized Naming

As already briefly mentioned, Rapid Automatized Naming (RAN; Denckla & Rudel, 1974) tasks measure how quickly participants can accurately name a sequence of repetitive stimuli (letters, digits, colours or images). Alphanumeric RAN tasks (letters and digits) are often considered to be good predictors of children’s reading skills (Bowey, 2005). However, there has been little consensus about explanations for the relationship between RAN and reading ability (see Kirby et al., 2010 for a review), though it has been suggested that phonological processes are a key component of RAN tasks since they include reading aloud (e.g., Wagner et al., 1994; Torgesen et al., 1997; Clarke et al., 2005; Ziegler et al., 2010). Alternative hypotheses suggest that

RAN tasks measure rapid visual processing (originally suggested by Denckla, 1988), general processing speed (e.g., Cutting & Denkla, 2001; Powell et al., 2007; Kail & Hall, 1994; Savage et al., 2007; Wolf & Bowers, 1999), or automaticity of lexical retrieval (Jones et al., 2016).

Kuperman and Van Dyke (2011) found RAN times (along with word identification tasks) to be reliable predictors of reading behaviours across the time-course of sentence reading for adults who were not college educated. They also found that RAN tasks (alongside word identification and reading ability) modulated the effects of word-level predictors, word-length and word-frequency on readers' eye movements. RAN tasks are also used to indicate reading difficulties and in a study by Kirkby et al. (2022) significant differences in RAN scores were found between dyslexic and non-dyslexic adults. It is important to note that for skilled adult readers who do not have reading difficulties, RAN tasks may be less sensitive to individual differences in reading behaviours. However, Gordon et al. (2020) did find that for university students, high RAN scores were associated with fewer regressions and more rereading than low RAN scores. Reading experience (measured by the ART) was found to be more predictive than RAN scores of early eye movement measures and was more sensitive to changes in eye movement patterns in response to changes in word frequency. Gordon et al. (2020) considered their findings concerning regressions and rereading and suggested that faster RAN times were related to attention and perceptual-motor coordination during reading. To examine this further, we included alphanumeric RAN tasks in the test battery to determine whether it is a strong predictor of eye movement patterns in populations of skilled readers when controlling for the influence of other cognitive skills. For these tasks a series of letters or digits are presented in a grid and the speed at which a participant is able name the characters aloud is recorded. Gordon et al., (2020) reported high test-retest reliability for RAN digits (.89) and RAN letters (.85).

1.4.4.2 Working Memory Capacity

Finally, working memory capacity (WMC) has been suggested to be a 'common cause' variable that predicts performance on both reading and RAN tasks (Papadopoulos et al., 2016). Poor readers may have less efficient processing and memory storage than skilled readers (Daneman & Carpenter, 1980). Indeed, readers with smaller WMC spend more time rereading ambiguous regions of a text than readers with larger WMC when sentences are complex (Clifton et al., 2003). However, many studies have shown that the influence of WMC on eye movements and reading behaviours is likely to be due to overlapping variance from other measures (see Hamilton et al, 2013; Traxler et al.,2012; Van Dyke, et al., 2014). Previous studies have often used sentence span tasks where participants are usually asked to read aloud a number of sentences that are followed by unrelated words (Daneman & Carpenter, 1980; Turner & Engle, 1989; La

Pointe & Engle, 1990; Traxler et al., 2012; Van Dyke et al., 2014). Participants are then required to memorise and recall the words at the end of the session. This measure of WMC shares variance with other reading skills, due to similarities in tasks, for example word identification tasks. For the current test battery, we selected a measure of working memory distinct from reading tasks to minimise shared variance with other measures. A backwards digit span (Gathercole et al., 2004) was selected which requires participants to memorise and recall increasing sequences of digits, first in a forwards order, and subsequently in a backwards order. Scores from the largest successful digit span backwards sequence are used as a measure of working memory capacity. Gathercole et al., (2004) report the mean test–retest reliability coefficient of the backwards digit span task to be .62.

1.5 Thesis Overview

Many previous investigations examine differences in readers' eye movement patterns in response to changes in text-level characteristics using controlled experimental designs. Individual differences in reading skills are either grouped into high and low-skill categories or are modelled on a continuous measure of skill. Text manipulations allow for modelling of individual differences related to specific processes in reading, for example how readers of different skill levels respond to less common (low frequency) words in comparison to more common (high frequency) words (Ashby et al., 2005). However, the aim of this thesis was to explore individual differences based on differences across multiple skills.

Large test battery investigations in previous research have often utilised a corpus approach to study natural reading. In contrast, the studies presented in this thesis used a large test battery to examine the differences in readers' eye movements in response to specific changes in text-level characteristics within controlled experimental designs. This fairly novel approach using highly controlled reading materials when examining multiple cognitive skill predictors meant that fewer text-level variables were required to be included as predictors in statistical models, and therefore reduced the risk of overparameterization.

This thesis investigates individual differences with three experiments. The project involved two rounds of data collection, one for Experiment 1 and another for Experiments 2 and 3 combined. Therefore, the same participants completed Experiments 2 and 3, with the exception of five participants who only contributed to Experiment 2 and five who only contributed to Experiment 3. A large test battery was used across Experiments 1 and 2. The first two experiments focussed on word-level aspects of reading: the first experiment investigated word frequency and the second investigated transposed letter effects. A third experiment compared two offline comprehension measures when assessing individual differences in eye movement patterns during paragraph reading in response to changing comprehension demands.

1.5.1 COVID Statement

The work presented here was impacted by national lockdowns in the UK due to the COVID-19 Pandemic. As a result of national lockdowns, university closures and laboratory restrictions due to social distancing rules, data collection for Experiments 2 and 3 of this thesis was delayed from March 2020 until October 2021. As a result of the delay in data collection and a subsequent experiment programming error with less time available to rectify it, the final experiment was altered to examine individual differences in eye movements in relation to offline reading comprehension tasks instead of in relation to the large test battery. In addition, University guidelines and COVID-19 restrictions that were in place in the UK when lockdown measures were lifted meant that access to university buildings was restricted to an educational context. As such, testing for the second round of data collection (for Experiments 2 and 3) was limited to a university student sample.

1.5.2 Experiment 1: Individual Differences in Skilled Reading and the Word Frequency Effect

The first experiment explored patterns associated with individual differences in the eye movements of average-to-very-skilled readers when presented with words of high and low frequency. High frequency words appear often within a language and most readers are highly familiar with them. In contrast low frequency words are less common, and readers will usually be less familiar with them. For this study, high and low frequency words were matched on length and embedded into identical neutral (unpredictable) sentence frames, and participants read them whilst their eye movements were recorded. A robust finding in the literature is that high frequency words are more easily identified and as a result are read more quickly than low frequency words (Angele et al., 2014; Rayner & Fischer, 1996; Rayner et al., 1996). Some individuals will be more familiar with words considered to be low frequency and will therefore identify them more quickly than others (Ashby et al., 2005; Cop et al., 2015).

In this experiment, participants also completed a test battery of cognitive skill assessments. Previous research has observed that readers with higher reading abilities in general read low frequency words more quickly than less skilled readers (Ashby et al., 2005). However, this pattern has also been associated with other cognitive skills, for example, reading experience (Chateau & Jared, 2000), spelling (Veldre et al., 2014), vocabulary (Cop et al., 2015), rapid automatized naming (Kuperman & Van Dyke, 2011) and working memory capacity (Long & Freed, 2021). This study examined whether a smaller frequency effect in more skilled readers is related specifically to readers' vocabulary knowledge, as identified by Cop et al. (2015), or whether it could also be predicted by more reading-related skills and cognitive assessments.

1.5.3 Experiment 2: Individual differences and the Transposed Letter Effect during Reading

The second experimental investigation takes a closer look at a specific aspect of word identification: the flexibility of letter position encoding. Previous research has shown that skilled readers encode letter positions more flexibly than letter identities, for example, during reading, a word with letters in transposed positions (*mnokey*) is less disruptive than a word where the letter identities are substituted (*markey*) when it is given as a preview for the target word (*monkey*) (Johnson et al., 2007; Johnson & Dunne 2012; Pagán et al., 2016; Kirkby et al., 2022). Letter position encoding has been found to become more flexible as reading skills develop in studies of children (Pagán et al., 2021; Gómez et al., 2021; Hasenäcker & Schroeder, 2021). However, it is not clear whether this process remains variable in skilled readers in relation to individual differences in reading and cognitive skills, or whether it reaches maturation. Experiment 2 assessed the skills of 100 skilled readers and recorded their eye movements. Sentences presented to readers contained an identity preview (identical to the target word) or a pseudoword preview for the target word in which two letters were either transposed (*cmapus*) or substituted (*cnupus*) until the readers' eyes crossed an invisible boundary, which then triggered a display change to reveal the correct target word (*campus*). We were interested in whether readers' performance on offline cognitive skills tasks predicted differences in the magnitude of disruption of a substituted letter preview in comparison to a transposed letter preview. In other words, whether differences between readers in the flexibility of their letter position encoding processes could be predicted by their scores on skills assessed by our test battery.

1.5.4 Experiment 3: The Jingle Fallacy in Comprehension Tests for Reading

The final experimental investigation for this thesis moves away from word-level investigations, to instead focus on how individual differences in offline comprehension tests predict passage reading patterns. Experiment 3 was directly motivated by findings from the first two experiments presented within this thesis. Comprehension subtests from the WIAT-II (Wechsler, 2005) and NDRT (Brown et al., 1993) were not found to load on the same factor within a principal components analysis in either study. This finding suggested that the WIAT-II and NDRT reading comprehension tests described distinct underlying skills. Previous research has shown that comprehension assessments often vary in the constructs that they measure (Cutting & Scarborough, 2006; Keenan et al., 2008; Kendeou et al., 2012; Mézière et al., 2023). The NDRT is widely used to assess reading ability in research investigating individual differences in eye movements (e.g., Ashby et al., 2005; Chateau & Jared, 2000; Veldre & Andrews, 2017; Andrews & Veldre, 2021) and is used primarily in the US and Australia, whereas the WIAT-II is favoured in the UK and is often used when comparing children and adults eye movement behaviours (Pagán et al., 2021). It is therefore important to address how well aligned the NDRT and WIAT-II comprehension subtests are in predicting patterns in the same eye movement data.

Two sets of analyses were conducted on the same eye movement data, each using a different offline measure of reading comprehension (either WIAT-II or NDRT) for the same participants. Comprehension measures differed in format and previous studies suggested that different underlying constructs were assessed by each of them. Individual differences in global eye movement patterns across the reading of a paragraph were examined in relation to an experimental manipulation of comprehension demands. Identical passages were presented with either high or low comprehension demands operationalised by the level of detail needed to answer multiple choice questions that followed them. We assumed that reading comprehension tests would be able to predict differences observed in eye movement patterns as a function of varying comprehension demands.

Previous research has shown that high comprehension demands are associated with more careful reading behaviours (characterised by longer reading times, more fixations, longer fixations, shorter saccades and more rereading) compared to low comprehension demands (Wotschack & Kliegl, 2013). In addition, adults with higher reading skills often read passages more quickly, than those with lower reading skills (Rayner, 1998; Ashby et al., 2005; Andrews & Veldre, 2021). This experiment aimed to, first, determine whether skilled readers were able to adapt their reading strategy to read more carefully when comprehension demands were higher. Next, to examine whether differences in strategy and adaptation over time are predicted by scores on offline comprehension assessments. And lastly, to consider whether the analyses based on two comprehension tests (WIAT-II and NDRT) are qualitatively similar in how they predict modifications in eye movement behaviour to adapt to changing comprehension demands, as would be expected if they measured the same construct, i.e., text comprehension.

Chapter 2 Individual Differences in Skilled Reading and the Word Frequency Effect

Charlotte E. Lee[1], Hayward J. Godwin[1], Hazel I. Blythe[2], and Denis Drieghe[1]

[1] School of Psychology, University of Southampton, Southampton, United Kingdom,

[2] School of Psychology, Northumbria University, Newcastle, United Kingdom.

2.1 Abstract

Variation in eye movement patterns can be considerable even within skilled readers. Here, individual differences and eye movements of 88 average-to-very-skilled readers were assessed to examine the reliability of previous observations of a reduced word frequency effect associated with skilled reading. Shorter fixation durations and higher skipping rates were observed for high frequency compared to low frequency words. High reading ability and vocabulary knowledge predicted reduced frequency effects in gaze duration, demonstrated by faster reading of low frequency words compared to low scorers. A PCA grouped individual differences tests based on shared variance. High 'lexical proficiency' predicted shorter gaze durations, reading times, and increased word skipping. 'Lexical proficiency' and the WIAT-II comprehension test predicted a reduced frequency effect in go past times and all tests apart from the NDRT comprehension test predicted a reduced frequency effect in sentence reading times. Data revealed surprising discrepancies in findings based on two subtests supposedly measuring comprehension (NDRT and WIAT-II), constituting an example of the *Jingle fallacy*: the false assumption that two measures that share a name actually measure the same construct.

Keywords: individual differences, eye movements, word frequency, lexical proficiency, Jingle fallacy

This paper has been submitted for publication and will be referenced in later chapters as:

Lee, C. E., Godwin., H. J., Blythe, H. I., & Drieghe, D. (2023). *Individual differences in skilled reading and the word frequency effect*. Manuscript Submitted for Publication.

Authorship Statement

Charlotte Lee conceptualised and created this experiment, performed all analyses, interpreted data and wrote this manuscript for publication.

Denis Drieghe and Hayward J. Godwin provided project supervision and comments on this manuscript.

Hazel I. Blythe provided comments on this manuscript for publication.

2.2 Introduction

In this study, we investigate individual differences that appear among skilled readers. More specifically, we examine differences in the size of the word frequency effect on eye movements during reading alongside differences in separate cognitive tasks including vocabulary, spelling, print exposure and comprehension. The novelty of the current paper lies in the focus on the frequency effect during reading as a potential indicator of a skilled reader - alongside more established indicators - given that a comparatively small frequency effect during reading has been suggested as indicative of better reading skills (Ashby et al., 2005; Haenggi & Perfetti, 1994).

Though skilled readers may perceive reading as almost effortless, reading does actually involve several complex skills. The Simple View of Reading (Gough & Turner, 1986), states that reading involves decoding meaning from text (word reading) and comprehension. In addition, reading fluency (speed of word recognition which allows automatic interpretation of text meaning, Kirby, 2007) and reading strategies (goal-oriented plans of how to read a text based on prior knowledge, text complexity and intended depth and breadth of processing i.e., Kirby & Woodhouse 1994) contribute to reading success. Moreover, much reading instruction and intervention strategies in the UK focus on phonics instruction, which includes segmenting words into graphemes, mapping them onto phonemes and blending the sounds, to facilitate word reading (Department of Education, 2021). Indeed, it is well established that phonological processing is important when reading alphabetic languages (Rayner et al., 2001; Goswami & Bryant, 2016; Byrne, 1998). The often-implicit assumption across these skills is that there is an optimal way in which reading is achieved, and that skilled readers do so in a uniform way (Andrews, 2012). However, this ignores complex individual differences in skilled reading.

It has been shown that children's word decoding and reading comprehension are predicted by different pre-cursor skills (Castles et al., 2018; Bowyer-Crane et al., 2008; Muter et al., 2004). For example, word recognition was predicted by letter knowledge and phoneme sensitivity, whereas comprehension was predicted by word recognition skills, vocabulary knowledge, and grammatical skills in Muter et al.'s (2004) longitudinal study of children's early reading skills. Evidently, there is scope for varied interventions to help readers who struggle with different aspects of reading.

The gold standard in studying reading is to analyse eye movement patterns. These patterns are considered to reflect cognitive processes active during reading (Liversedge & Findlay, 2000; Rayner, 1998). The two basic types of eye movements made while reading are fixations, which are moments where the eyes remain relatively still to acquire new information, and saccades, which are rapid, ballistic movements during which visual input is mostly suppressed (see Rayner, 2009 for a review). Fixations typically last 225-250 *ms* and saccades on average move across 7-9 letters in English and in similar alphabetic languages (Rayner, 2009). A reader also makes regressions, where a backwards saccade is made to re-fixate a point in the text previously encountered. Variation in these eye movement patterns often reflect the characteristics of the written language or the difficulty of a text. Fixations are longer in duration when processing loads are high, for example when reading low frequency words compared to high frequency words (Inhoff & Rayner, 1986). Generally, as a text increases in difficulty, fixation durations increase, saccades become shorter and more regressions are made (Rayner, 1998).

Importantly, variability in eye movement patterns also corresponds to reader skill. A skilled reader typically reads more quickly, makes shorter fixations, longer saccades and makes fewer regressions than beginner, less skilled or dyslexic readers (Rayner, 1998; Ashby et al., 2005). However, Underwood et al. (1990) demonstrated that there was no direct relationship between speed and reading skill when the latter was defined as the ability to efficiently extract information. Highly skilled readers were not necessarily those who demonstrated faster overall reading rates. A speed-accuracy trade-off exists whereby increasing reading speed eventually results in a decrease in comprehension (Rayner et al., 2016). It is important to note in this context that in eye movement research, a *skilled reader* is usually described as one who does not experience difficulties and the term is often used interchangeably with *average reader.*

Differences in eye movement behaviour can occur even within groups of skilled readers (see Andrews, 2012; 2015 for reviews). Fixation durations vary between 50 - 600 *ms* (sometimes longer), and saccade lengths vary between 1 - 20 letter spaces (or more) (see Rayner, 2009). Indeed, fixation durations, saccade lengths and frequency of regressions all vary widely between readers (Rayner, 1998). Differences are assumed to be quantitative rather than qualitative whereby all skilled readers use similar reading processes with varying efficiency (Ashby et al., 2005). Computational models of eye movement control, such as the E-Z Reader (Reichle et al., 1998; 1999), SWIFT (Engbert et al., 2005), OB1-Reader (Snell et al., 2018) and Über-Reader (see Reichle, 2020), each assume that readers use comparable word recognition processes. If qualitative processing differences were to be found between average and highly skilled readers, models would need to be adapted.

To investigate variability in skilled readers' eye movements, research has focussed on individual differences in related skills, such as spelling, reading comprehension, and working memory capacity (WMC). In the next section we will discuss the cognitive skills that will be examined in this paper starting with overall reading ability measures, followed by vocabulary, spelling, print exposure and finally rapid automatized naming (RAN) and WMC. For these skills we will discuss relevant findings, theory where appropriate, and also mention any research that additionally examined the word frequency effect during reading in combination with testing these skills.

In general, low frequency words (words that are less commonly occurring within a language), are read more slowly and skipped over less often than more common, high frequency words (Angele et al., 2014; Rayner & Fischer, 1996; Rayner et al., 1996; Rayner et al., 2004). This indicates that individuals have greater familiarity with words that are more common within a language and can be explained by implicit learning (e.g., Dijkstra & Van Heuven, 2002; McClelland & Rumelhart, 1981; Morton, 1970). Repeated exposure to common words increases the baseline activation of that item in the lexicon, moving it towards to the activation threshold for lexical selection, which means it will be processed and recognised more quickly (e.g., Monsell,1991). Some individuals will be more familiar with words that are less common within a language and as mentioned previously, differences in the size of this effect have been linked to individual differences in reading skill (Ashby et al, 2005) and are a main focus of this paper.

Firstly, reading ability as a whole can be assessed with standardised reading tests which usually consist of comprehension and vocabulary or other related sub-skills (e.g., word decoding and naming speed). Ashby et al. (2005) grouped readers into high and average proficiency based on the Nelson Denny Reading Test (NDRT; Brown et al., 1993) and examined differences in eye movements during silent reading. They found that highly skilled readers exhibited significantly shorter fixation durations than average readers and found a trend towards smaller word frequency effects for highly skilled readers in first and single fixation durations (the duration of fixations on the target word when the readers fixated the target word exactly once during the first pass) and a marginally significant interaction in the same direction for gaze duration (total time spent fixating a word on the first pass). These researchers did not observe much difference in fixation times on high frequency words when comparing the groups of readers, but there was a much greater difference in the fixation times on low frequency words. Less skilled readers showed a proportionately greater cost to reading times for low frequency words. Ashby et al. (2005) suggested that this may be due to the skilled readers' greater familiarity with less common words. They also found that readers spent more time on the next fixation following a low frequency word (increased spillover effects) compared to a high frequency word when the word is also unpredictable. Though the findings from Ashby et al. (2005) were based on small groups of participants (22 per group) and the interactions reported were marginal, the researcher's claims have been validated by further research in the field. The influence of overall reading ability on eye movement measures (Veldre et al., 2014) and interactions between word frequency and more specific cognitive skills associated with skilled reading have been reported (spelling, Veldre et al., 2014; vocabulary; Cop et al., 2015; print exposure, Chateau & Jared, 2000; rapid automatized naming; Kuperman & Van Dyke, 2011, working memory capacity, Long & Freed, 2021).

When standardised reading ability tests are included for research purposes, it has been noted that different cognitive abilities are addressed depending on which specific reading ability test is selected (Keenan et al., 2008; Kendeou et al., 2012). As a result, inconsistencies in research based on reading ability measures may arise dependent on whether the selected test includes specific sub-skills (e.g., word decoding) in addition to comprehension and vocabulary (see Mézière et al., 2023; Cutting & Scarborough, 2006). This is not an uncommon issue in psychological research, with a large range of different tests available to assess similar constructs, making it difficult to determine what the optimal test is to study these constructs (Flake & Fried, 2020).

We included overall reading ability measures to compare to the trends reported by Ashby et al. (2005). Given the different tests available to measure reading ability, we made the rather unusual decision to include both the NDRT, which is often used in the US and Australia and the Wechsler Individual Achievement Test (WIAT-II UK; Wechsler, 2005), a normed test based on a UK adult sample, in the current study. It is important to remember that these tests are developed to diagnose reading difficulties. There are some notable format differences between these tests. For example, the NDRT features self-paced silent reading comprehension of non-fiction passages and a multiple-choice vocabulary test whereas the WIAT-II features a varied reading comprehension test (featuring single line reading, fiction, and non-fiction paragraph reading, with the option to read aloud) alongside additional subtests (pseudoword decoding and word reading). Given the differences between these two reading ability tests and the potential of these differences impacting our findings, we decided to include both the NDRT and the WIAT-II.

In the context of our second cognitive skill, vocabulary size, and some of the cognitive skills we discuss below, it is important to discuss the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002). Perfetti suggested that skilled readers develop high-quality lexical representations of words that in turn facilitate fast, automatic word recognition. A high-quality representation is defined as one with precision, redundancy and coherence (Perfetti, 2007). This means that it is represented by precise orthographic information that has little interference from similarly spelled words, it consistently triggers phonology of the correct word and is related to specific and redundant semantic information that defines the word. Highly skilled readers have high quality representations of most words (Andrews, 2008; 2012). Good and poor readers may, therefore, be differentiated due to the precision of their lexical representations (Perfetti, 2007; Perfetti & Hart, 2002).

Lexical quality can be divided into orthographic and phonological representations, of which spelling and decoding skills are a proxy, and semantic representations, measured by vocabulary and reading comprehension (Andrews, 2015; Perfetti, 2007). There is an indirect link from reading skills via lexical quality that may lead to a reduced word frequency effect: Individuals with good spelling and decoding skills, a larger vocabulary, better reading comprehension and more reading experience will develop higher quality lexical representations. In turn, they should have greater familiarity with less common words than the general population and require less time to process them than individuals who have less experience or are not as proficient in these cognitive skills.

These predictions about vocabulary size were confirmed in a study of monolingual and bilingual readers reading an entire novel (Cop et al., 2015). For the purpose of this paper, we focus on their examination of monolingual readers. Cop and colleagues found that vocabulary knowledge scores (measured by the LexTALE test; Lemhöfer & Broersma, 2012) indicated that readers with a larger vocabulary exhibited smaller frequency effects compared to those with a smaller vocabulary. This smaller frequency effect was mostly due to comparatively shorter single fixation durations for low frequency words, similar to the pattern based on NDRT scores from Ashby et al. (2005) which includes vocabulary and comprehension.

We now turn to the third cognitive skill of spelling. Good spelling in an opaque language such as English may be a good indicator of precise lexical representations. However, this is not likely to be the case in transparent languages such as Norwegian or Swedish where phonology maps onto orthography more directly, meaning unknown words can be more easily spelled (e.g., Furnes & Samuelsson, 2010). Within a study by Veldre et al. (2017), better spelling, but not reading ability (measured by the NDRT) in native English readers, was associated with a reduced effect of word frequency in gaze duration and total duration of fixations on target words. They found that, in general, higher spelling ability predicted longer saccades, more word skipping and landing positions that were further into the target word than low spelling ability but did not affect fixation times. Skipping probabilities are thought to reflect reaching an advanced stage of word recognition during parafoveal processing through the processing of the upcoming word viewed in the parafovea during a fixation on the previous word (e.g., E-Z Reader; Reichle et al., 1998; 1999). Precise lexical representations would therefore lead to rapid word identification during parafoveal processing, negating the need for the reader to directly fixate an already close-to-being identified word, and result in an increased likelihood of word skipping. Increased word skipping predicted by high spelling scores here is consistent with the idea that good spelling reflects high quality lexical representations which can be processed efficiently in the parafovea.

In the same study, Veldre et al. (2017) found that higher reading ability scores also predicted a reduction in the number of fixations and sentence reading times, smaller fixation durations and fewer regressions compared to lower reading ability scores but did not influence the size of the frequency effect. Even though reading and spelling abilities were moderately correlated in this study, spelling ability predicted differences in eye movements measures that were not predicted by reading ability, indicating that these skills influence different aspects of eye movement control (see also Slattery & Yates, 2017). This is consistent with differential effects of reading comprehension and spelling ability seen in parafoveal processing (see Veldre & Andrews 2015a; 2015b). Spelling ability is a stronger and more reliable index of lexical precision than comprehension tests since readers who have poor lexical representations can compensate with context-based strategies, whereas spelling requires specific word form knowledge (Perfetti, 1992; Andrews, 2012).

Our fourth cognitive skill concerns print exposure (how much experience an individual has of reading novels). Greater reading experience enables readers to develop high-quality lexical representations which facilitate fast and efficient word recognition processes in skilled reading (Perfetti, 1985). Chateau and Jared (2000) studied university students with similar comprehension scores (using the NDRT comprehension subtest) but who had different levels of print exposure (as measured by the Author Recognition Test (ART; Stanovich & West, 1989) and found that students with higher levels of print exposure were more efficient in naming tasks and in lexical decision tasks than those with lower levels. They suggested that since the two groups were matched on comprehension, differences between the groups were due to differences in print exposure rather than general reading ability. Chateau and Jared (2000) suggested that print exposure increases general knowledge and enhances word recognition processes as well as contributing to vocabulary size.

Our fifth cognitive skill is Rapid Automatized Naming (RAN; Denckla & Rudel, 1974). RAN tasks measure an individual’s ability to name a sequence of repetitive and visually familiar stimuli (letters, digits, colours or images) quickly and accurately. They are often considered good measures for predicting reading development. Alphanumeric RAN tasks (letters and digits) in particular, are often found to be predictive of early reading difficulties compared to RAN objects and colours (Bowey, 2005). Performance on a RAN task has been found to predict reduced word frequency effects in single sentence reading for readers who were not college-educated (Kuperman & Van Dyke, 2011). Kuperman and Van Dyke (2011) also found that faster RAN predicted smaller fixation durations and a decreased likelihood of regressions and refixations in a sample of low-to-average-skill readers.

However, when included in models alongside ART scores, Gordon et al. (2020) found that faster RAN did not influence the word frequency effect. Higher ART scores were associated with an increased likelihood of word skipping, shorter gaze durations, and a reduced effect of word frequency on gaze durations, consistent with Chateau and Jared (2000), whereas faster RAN times were instead associated with fewer regressions and rereading times and did not modulate the effect of word frequency in any measures. Gordon and colleagues suggested that the RAN and ART influence different aspects of the word recognition process, with faster RAN indicating efficient perceptual-motor coordination and attentional processes during reading. It has been suggested that the relationship between RAN and reading reflects phonological processes due to the oral reading aspect of the task (e.g., Wagner et al., 1994; Torgesen et al., 1997; Clarke et al., 2005; Ziegler et al., 2010). However, alternative ideas suggest that these observations may be due to rapid visual processing (originally suggested by Denckla, 1988), general processing speed (e.g., Cutting & Denkla, 2001; Powell et al., 2007; Kail & Hall, 1994; Savage et al., 2007; Wolf & Bowers, 1999), or, more recently, automaticity of lexical retrieval (see Jones et al., 2016). Indeed, there is little consensus about why a relationship between RAN and reading ability may appear (see Kirby et al., 2010 for a review). The present study includes RAN tasks to investigate their influence on eye movement patterns when considered alone and alongside other measures.

Finally, a “common cause” variable in the relationship between RAN and reading may be an individual’s working memory capacity (WMC) (Papadopoulos et al., 2016). Daneman and Carpenter (1980) suggested that poor readers are characterised by inefficient processing and storage in WMC. One theory of WMC in reading, the Capacity Theory of Comprehension (Just & Carpenter, 1992), suggests that each reader has limited resources for storage and processing of information, and the capacity differs between individuals. When the resources needed to process information are more than an individual’s capacity allows, processing efficiency is reduced, and information is lost. Therefore, when processing demands are high or capacity is very low, comprehension is diminished. Alternatively, Separate-Sentence-Interpretation-Resource (SSIR) Theory (Waters & Caplan, 1996), suggest that WMC is modular, with one module dedicated to syntactic processing, and the other dedicated to post-parsing sentences. The theory suggests that individuals differ only in the capacity of the second module, not the first. Therefore, individual differences in WMC should appear when looking at late eye movement measures, such as sentence reading times and wrap-up effects, when information is integrated towards the end of a sentence, especially when sentences are complex. Indeed, readers with smaller WMC have been found to spend more time re-reading ambiguous regions of a text than readers with larger WMC when sentences are complex (Clifton et al., 2003). In contrast, Joseph et al. (2015) did not find a direct influence of WMC on reading times and regression probabilities and number of studies have shown that the influence of WMC on reading behaviours is most likely because of shared variance with other cognitive skills (Hamilton et al, 2013; Traxler et al., 2012; Van Dyke, et al., 2014). Interestingly, Long and Freed (2021) found that readers with a greater WMC display smaller differences in gaze durations and sentence reading times for high and low frequency words than readers with a smaller WMC when considered alongside a large test battery of individual differences. Further research is therefore needed to assess these findings when considered alongside individual variation in other cognitive skills.

Our initial approach for the present study aimed to clarify some inconsistencies in research based on differences in analyses and the selected individual differences tests. We first selected the most common tasks that prior research had suggested may influence eye movement patterns; measures of reading ability overall; the NDRT (Brown et al., 1993) and the WIAT-II (Wechsler, 2005), the LexTALE test of vocabulary knowledge (Lemhöfer & Broersma, 2012), spelling (Andrews & Hersch, 2010), print exposure measured by the ART (Acheson et al., 2008), alphanumeric RAN (Denkla & Rudel, 1974), and WMC measured by a backwards digit span test (Gathercole et al., 2004). For simplicity, we will refer to the combination of reading skills and more general cognitive skills tests as cognitive skills.

We next examined separate models in which each test was a single predictor to replicate previous research based on differences in a single predictor. We assessed individual skills tests as predictors of shorter gaze durations, a finding associated with skilled reading. Gaze duration was selected as the dependent eye movement measure in these models to best compare with previous research (Rayner, 1998). We also examined which tests, if any, predicted comparatively smaller differences between reading a high vs low frequency word. In our second approach, tests were grouped into composite scores where variances overlapped using a PCA. This allowed us to examine which skills were closely related. Finally, factors suggested by the PCA were used to predict eye movement measures, allowing an investigation of different cognitive skills and their relationship with eye movements whilst controlling for multicollinearity.

We predicted that, where cognitive tests are considered in separate models, our findings would replicate previous research that linked high overall reading ability (Ashby et al., 2005), better spelling (Veldre et al., 2017), large vocabularies (Cop et al., 2015), high print exposure (Chateau & Jared, 2000; Gordon et al., 2020), faster RAN (Kuperman & Van Dyke, 2011) and larger WMC (Long & Freed, 2021) to shorter gaze durations and a reduced word frequency effect. We also expected skills related to lexical quality to be grouped together by PCA (Andrews, 2015; Perfetti, 2007). We expected lexical quality to be a good predictor of a reduced word frequency effect as seen in previous research about related skills (spelling; Veldre et al., 2017; vocabulary;

Cop et al, 2015; print exposure; Chateau & Jared, 2000; Gordon et al.,2020). Since previous observations of the predictive value of RAN were associated with less skilled populations (e.g. Kuperman & Van Dyke, 2011), where RAN often predicted reading difficulties (Bowey, 2005) and was sometimes not found when included alongside measures of print exposure (Gordon et al., 2020), we hypothesised that the predictive value of RAN times may be secondary to lexical quality in our sample of average-to-very-skilled adult readers. We made similar predictions about WMC, due to suggested associations of WMC and RAN (Papadopoulos et al., 2016).

For completeness, we mention three studies that have examined individual differences within skilled reading in somewhat similar ways. The first was conducted by Kuperman and Van Dyke (2011) who found that individual differences in RAN and word identification tests influenced the magnitude of the word frequency effects om fixation times. They studied a different population of readers, who were non-college-bound participants from low to average reading skill, and took a different methodological approach using a corpus of natural sentence reading as opposed to an experimental design with tightly controlled materials. Kuperman et al. (2018) took an alternative approach to assessing both reader and text-level characteristics in passage reading using a Random Forests non-parametric regression technique to predict eye movement patterns and comprehension accuracy. We also note that different reading tasks were selected for individual differences measures in both of these studies. Most recently, a study was reported by Long and Freed (2021) who used test battery data collected by Freed et al. (2017) to predict eye movement patterns and found that language experience, decoding, and WMC interacted with word frequency to influence eye movement measures. A key difference in their approach is that besides a slightly different choice in the individual differences tests they used, they analysed eye movement data using structural equation modelling whereas this paper uses linear mixed models on individual trial data. Importantly, separate models were created by Long and Freed (2021) for the effects of word length and frequency, and predictability was not controlled for. In our design, words are matched on word length and predictability is controlled, which allows for a more controlled examination of the word frequency effect.

2.3 Method

2.3.1 Transparency and Openness

We report our sample size and how it was selected, all data exclusions based on individual differences and eye movements, all manipulations, and all measures in the study following JARS (Kazak, 2018). Data were analysed in R (version 4.1.1; R Core Team, 2022) using the lme4 package (version 1.1-27.1; Bates et al., 2015). The study design and analysis were not pre-registered. All data, analysis code and materials are available at https://osf.io/fetw4/?view_only=166f33ba5c584b3aa10efd30972840eb

2.3.2 Participants

One hundred participants consisting of students and staff from the University of Southampton as well as members of the wider community (20 Males, *mean age* = 22.54, *SD* = 9.78, *range* = 18-72) took part in the study. Participants were all native English speakers with normal or corrected to normal vision and no known reading difficulties. Student participants were sampled from various courses across the university including psychology, humanities and health sciences. Participants received either course credits (for psychology undergraduate students) or £9 for completing the study. A power analysis was conducted via simulations in R (simR, Green & MacLeod, 2016) based on data from the first 10 participants. This revealed that a sample size of 80 participants would be sufficient to achieve 80 % statistical power for our analyses. Data collection exceeded this target, with a sample of 100 participants. Twelve participants' data were removed due to eye tracking errors and/or extreme low scores in reading tasks that may indicate reading difficulties, meaning that 88 datasets (20 Males, mean *age* = 22.94, *SD* = 10.35, range = 18-68) were therefore included in analyses, still above the target sample size indicated by the power analysis.

2.3.3 Apparatus

Participants used a 14-inch Dell Laptop Computer to complete online tests and questionnaires during the study. These tests were administered using a web browser running Qualtrics. A computerised digit span test was administered using Inquisit on a 19-inch DELL monitor (1024 × 768-pixel resolution). For the WIAT-II (Wechsler, 2005), researchers used the testing flip-pad, scoring sheets and word/pseudoword cards included in the test pack and a stopwatch to record reading times.

Participants' eye movements were recorded during the eye tracking task using an SR Research EyeLink1000 eye tracker (sampling rate = 1000 Hz, max 0.5° calibration error). Sentences were presented on an ASUS HD monitor (1024 × 768-pixel resolution) at a viewing distance of roughly 73 cm to ensure 3 characters equated about 1° of visual angle. Stimuli were presented in Courier New font. Participants' head movements were minimised using a chinrest.

2.3.4 Materials

The study included a battery of tests to assess cognitive skills. Participants were first asked two questions about their reading behaviour: "How often do you read for work?" and "How often do you read for leisure?"

2.3.4.1 Reading Ability Tests

Nelson Denny Reading Test (Brown et al., 1993). The vocabulary subtest presented on the screen was comprised of 80 short descriptions of target words. Participants were asked to select the most appropriate word or phrase from five options to describe the key word's meaning. The reading comprehension subtest required participants to read passages silently before reading and answering multiple choice comprehension questions (with 5 possible answers) that appeared on the same page below the text. This subtest was timed allowing participants 10 minutes to complete it, this is half of the standard time given for this test. This speeded procedure has been used previously and has proven to produce more normally distributed data than the full-timed procedure and can therefore obtain increased discrimination between more proficient readers (Andrews et al., 2020).

Wechsler Individual Achievement Test (WIAT-II UK; Wechsler, 2005). The word reading subtest required participants to read aloud a list of real words that increased in difficulty in pronunciation, the test began with "the" and ended with "hierarchical". Participant's progress was recorded, and testing was discontinued when the participant made six sequential errors. The pseudoword decoding subtest required participants to read aloud a list of orthographically legal nonsense words, for example "*flimp*". Participants' progress was recorded and again testing was discontinued after the participant made six sequential errors. Participants also completed the reading comprehension subtest, where they read passages of increasing length and complexity (types of text included short fictional stories, informational text, advertisements, and how-to passages) aloud or silently before answering literal and inferential comprehension questions aloud when asked by the experimenter. Scores for these three tests were combined, standardised, and normed according to WIAT-II instructions. These normed scores were then used in all analyses.

2.3.4.2 Vocabulary Knowledge

LexTALE task (Lemhöfer & Broersma, 2012). An English word or pseudoword was presented on screen and participants were required to indicate whether the word was a real English word or a pseudoword. Scoring was an average of the percentage of real words correctly identified and the percentage of nonwords correctly identified.

2.3.4.3 Spelling

Spelling Dictation and Spelling Recognition (Andrews & Hersch, 2010). In the spelling dictation task participants listened to recordings of 20 single key words and sentences containing them. Participants were then asked to write down the key words correctly. In the spelling recognition task participants were given a list of 88 correctly and incorrectly spelled words and were asked to select all the words that they identified as being spelled incorrectly.

2.3.4.4 Print Exposure

Author Recognition Test (Acheson et al., 2008). Participants were given a list of real author names and foil names. Participants were asked to mark any real authors they were familiar with. Participants were informed that there were some foil names included in the list.

2.3.4.5 Rapid Automatized Naming

Alphanumeric RAN (Denkla & Rudel, 1974). A series of letters were presented in random order and participants were required to verbally name the characters as quickly as possible while being timed by the experimenter. This was then repeated with a series of numbers.

2.3.4.6 Working Memory Capacity

Digit Span Backward (Gathercole et al., 2004). Number sequences were presented on screen and participants were asked to recall their order in two different tasks: Digit Span Forward (the participant attempts to repeat digit sequences of increasing length forwards), and Digit Span Backward (the participant attempts to repeat digit sequences of increasing length backward). Scores from the largest successful digit span backwards sequence were used in our analyses.

2.3.4.7 Eye Tracking

Participants were required to read sentences on screen while their eye movements were tracked. Related questions were then presented on the screen and participants were asked to respond using the keyboard to ensure they were reading for comprehension. Sentences featured one word that was either low or high frequency (e.g., “gourd” versus “apple”). There were 78 high and 78 low frequency target words matched on word length. All were nouns with a word length between four and nine characters (*M* = 5.40, *SD* = 0.87). The high frequency words had significantly higher Zipf values (Sublex-UK Zipf word frequencies: Van Heuven et al., 2014), (*M* = 5.28, *SD* = 0.36) than the low frequency words (*M* = 3.22, *SD* = 0.43), *t* (147) = -32.73, *p* < .001. Each pair of target words was embedded in the same neutral sentence frame. Eight participants who did not take part in the main experiment completed a cloze task: they were provided with the experimental sentences up to the target word and were asked to provide single words that could follow this partial sentence. Only 1.28 % of guesses were correct, demonstrating that target words were not predictable from the initial sentence context. Participants read all 78 experimental sentences, with either the high or low frequency target words in each. Stimuli were organised across two lists using a Latin square design. Testing began with 12 practice trials with practice questions. Comprehension questions followed 26 % of the sentences.

2.3.5 Design and Procedure

Participants were given an information sheet and consent form. Participants were first asked to answer some reading background questions. Task order was then randomised apart from eye tracking which always presented in the middle of testing. Participants were allowed breaks where needed. Participants were first asked to complete the NDRT, WIAT-II, LexTALE, Spelling, ART, alphanumeric RAN and backwards digit span tasks.

During the eye tracking task, participants were asked to sit comfortably at the computer, resting their chin on a chinrest and were guided through the set up and calibration of the eye tracker by the researcher. Participants were then required to direct their gaze to a fixation cross presented on the left of the screen. When ready, sentences were presented following the fixation cross. Participants were asked to read and answer questions presented on the screen using the keyboard to respond to ensure they were reading for comprehension. Participants could take breaks when needed.

2.4 Results

2.4.1 Analytic Approach

Our analyses were organised into three stages: First, separate models predicting gaze durations were examined for each individual difference test to assess whether findings replicate previous research which often included only one of these tests. Second, a principal components analysis was performed to determine which tasks load together. This stage identified two components which were extracted and used as composite scores in subsequent models. Third, GLMM models were tested to examine eye movement measures (word skipping probabilities, first fixation durations (FFD), single fixation durations (SFD), gaze durations (GD), go past times and sentence reading times) and whether they may be predicted by the following fixed factors: the two identified components (PC1 and PC2), two reading comprehension subtests (from the NDRT and WIAT-II), and a word reading subtest (from the WIAT-II), word frequency, saccade launch site and trial number. Two-way interactions between a single individual differences score (PC1, PC2, NDRT comprehension, WIAT-II comprehension or WIAT-II word reading) and target word frequency were also considered.

2.4.2 Individual Differences Tests

An overall spelling score was calculated as an average of the two subtests, spelling recognition and spelling dictation (see Andrews & Hersch, 2010). The backward version of the digit span was taken as an overall score (see Gathercole et al., 2004). Overall scores for the NDRT were calculated as an average of comprehension and vocabulary subtest scores. Overall WIAT-II scores were calculated using the age adjusted scoring materials provided which resulted in numerical scores as well as categorical assessments of reading proficiency (Borderline/ Low Average/ Average/ High Average/ Superior). For the purpose of examining these tests as predictors of eye movements, in light of their use in previous research, overall composite test scores were used for models which included only one test as a predictor. However, for the PCA, it is more appropriate to consider the cognitive skills that are included within reading ability tests separately, therefore, for our PCA and subsequent GLMMs, subtests from the WIAT-II and NDRT were included instead. These consisted of reading comprehension (WIAT-II and NDRT), word reading (WIAT-II), pseudoword decoding (WIAT-II), and vocabulary (NDRT). Outliers were examined for each test and as mentioned 12 participants were removed due to very low scores on one or more measures[1]. Of this 12, two were also flagged due to technical difficulties during eye tracking. Descriptive statistics for scores on each test are summarised in Table 2.1.

---

[1] WIAT-II comprehension (2 outliers), NDRT comprehension (1 outlier), NDRT Vocabulary (2 outliers), WIAT-II word reading (2 outliers), WIAT-II pseudoword reading (2 outliers), LexTALE (4 outliers), Spelling (2 outliers), RAN (2 outliers), ART (1 outlier.

Table 2.1

Descriptive Statistics for Tests and Subtests (ms)

	Min	Max	Mean	*SD*
NDRT Total	46.00	149.00	99.04	18.84
NDRT Comprehension	10.00	72.00	33.04	12.91
NDRT Vocabulary	28.00	79.00	66.00	9.11
WIAT-II Total	75.00	127.00	107.04	11.66
WIAT-II Comprehension	65.00	125.00	104.06	15.92
WIAT-II Pseudoword Reading	70.00	118.00	102.49	9.90
WIAT-II Word Reading	85.00	121.00	110.02	6.81
LexTALE	65.00	97.50	88.78	7.10
Spelling	29.98	60.52	41.19	6.01
Author Recognition Test	-6.00	37.00	10.88	7.94
Rapid Automatized Naming	29.98	60.52	41.19	6.01
Digit Span test	3.40	10.83	6.01	1.35

2.4.3 Eye Tracking Analyses

Comprehension accuracy during the eye tracking task was very high (*mean accuracy* = 95.6%). Trials with blinks on the target word or featuring tracking loss were removed prior to analysis. Fixations shorter than 80 *ms* that were made within one character of a previous or subsequent fixation were merged. Afterwards, remaining fixations shorter than 80 *ms* and longer than 800 *ms* were removed, Trials that consisted of fewer than 3 fixations across the sentence were also removed. This stage of data cleaning removed 0.18% of data. Trials where the target word was skipped were removed for analyses based on fixation times. This stage removed 6.26 % of data for the fixation time analyses. Dependent variables were then calculated; first fixation durations (FFD), single fixation durations (SFD), gaze durations (GD), go past times (GOPAST), sentence reading times and word skipping probabilities. When calculating these eye movement measures, data falling outside of 3 standard deviations from the mean for each participant within a condition (high and low frequency) were removed (SFD; 2.18 %, FFD; 2.12 % GD; 2.83 %, go past times; 2.74 %, Sentence reading times; 1.96 %). Descriptive statistics based on participant means for these measures are displayed in Table 2.2 below.

Table 2.2

Descriptive Statistics for Eye Movement Measures (ms)

	Condition	Min	Max	Mean	*SD*
SFD	HF	132.81	277.71	204.65	27.84
	LF	140.21	312.29	232.86	36.76
FFD	HF	131.80	280.28	204.50	27.46
	LF	143.68	324.41	229.66	34.66
GD	HF	135.53	313.12	217.83	33.22
	LF	156.97	420.68	258.01	48.31
GOPAST	HF	135.53	313.12	272.18	36.20
	LF	156.97	442.29	282.34	53.29
Sentence Reading Times	HF	1324.27	4321.74	2587.02	612.67
	LF	1224.90	4779.45	2693.63	657.91
Skipping Probability	HF	0.00	0.49	0.22	0.12
	LF	0.00	0.42	0.14	0.11

Note. Means and *SDs* are calculated based on participant means per condition.

2.4.3.1 Generalized Linear Mixed Models

A gamma distribution was used for all Generalized Linear Mixed Models (GLMM), following guidance for analysing skewed reaction time data without transformation (see Lo & Andrews, 2015) with participants and items as random factors.

2.4.3.2 Models with Single Test Predictors

Separate GLMMs were conducted to model each test focusing exclusively on gaze durations. Fixed factors included the test score measuring the cognitive skill in question, word frequency and the interaction of the test score with word frequency, as well as trial number and launch site. Trial number was never found to be significant in any model and did not contribute to the model fit and was therefore never included in final models. Launch site indicates the distance from which a saccade was launched prior to the target word. Previous research has documented that when a saccade is launched close to a target word, fixation times are often reduced and skipping rates increase (Pollatsek et al., 1986). The following procedure was followed to obtain the final model (see also Dirix & Duyck, 2017). The random effects structure started with intercepts only for subjects and items. The fixed structure was trimmed backwards by removing an interaction or fixed factor and using pairwise chi-square model fit comparisons to determine whether this removal negatively influenced the model fit. However, we always maintained the test itself as fixed factor given that this was the focus of the analysis. Following this, the random effects structure was forwards fitted to find the largest possible structure that converged again using pairwise chi-square model fit comparisons to see if an extra parameter added to the fit of the model. Once the maximal random effects structure possible had been established[2], we finally checked whether parameters in the fixed structure could be removed.

Model outputs are presented in Table 2.3 below. Adjusted *p* values to account a heightened false-discovery rate for multiple comparisons are presented. Gaze Duration models consistently revealed that high frequency words received significantly shorter gaze durations than low frequency words and that saccades launched from a position close to the target word were significantly more likely to result in shorter gaze durations on the target word. Analyses revealed that - when considered in separate models - high scores in the NDRT, WIAT-II and Spelling tests predicted shorter gaze durations than low scores. After correcting for multiple comparisons, the main effect of LexTALE scores on gaze durations was marginal. NDRT, WIAT-II and LexTALE-models revealed that scores on these tests significantly moderated the impact of word frequency on gaze durations. High scorers on these tests were not slowed down as much as low scorers when encountering a low frequency word within a sentence (see Figure 2.1). Note that scores on both reading ability tests are composites of two or more subtests, the NDRT is comprised of comprehension and vocabulary and the WIAT-II includes comprehension, word reading and pseudoword decoding. It was therefore unclear whether comprehension or other skills included in these composites were key in driving the effects found in these models.

---

[2] The final random effects structure for the WIAT-II model was target word frequency and saccade launch site as slopes for subjects, and target word frequency and WIAT-II Scores as slopes for stimuli. For the final NDRT model the random effects structure was target word frequency and saccade launch site as slopes for subjects, and target word frequency and NDRT Scores as slopes for stimuli. For the LexTale model the final random effects structure included target word frequency as a slope for subjects, and target word frequency, LexTale Scores + saccade launch site as slopes for stimuli. For the spelling model the random effects structure was target word frequency and saccade launch site as slopes for subjects, and target word frequency and spelling scores as slopes for stimuli. For the ART model the random effects structure included target word frequency as a slope for subjects, and target word frequency, ART Scores and + saccade launch site as slopes for stimuli. The random effects structure for the final RAN model was target word frequency as a slope for subjects, and target word frequency, RAN scores and saccade launch site for stimuli. Finally, the maximal random effects structure for the backwards digit span model included target word frequency as a slope for subjects and target word frequency, backwards digit span scores and + saccade launch site as slopes for stimuli.

Table 2.3

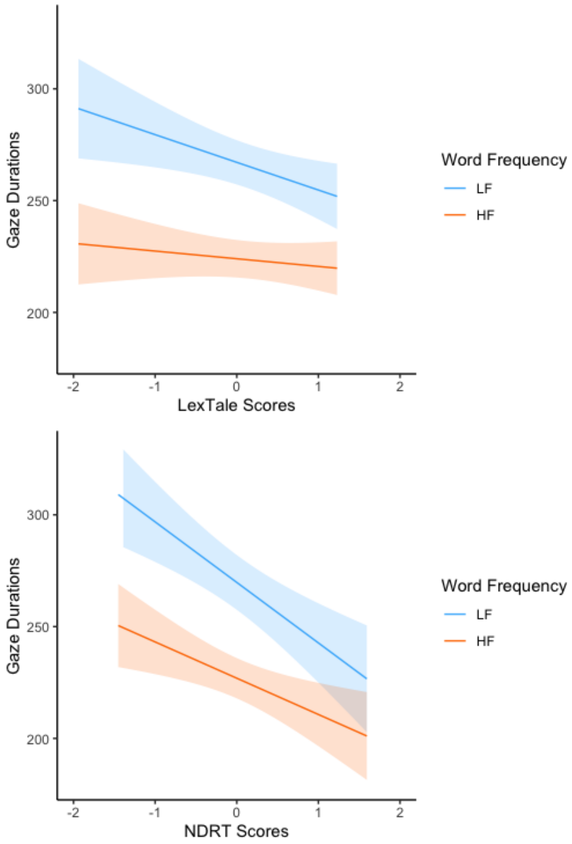
GLMMs to predict Gaze Durations (ms) where Models included a Single Test Predictor

		Est	*SE*	*t*	*p*
WIAT-II	Intercept	248.00	4.53	54.81	<.001 ***
	Word Frequency (LF-HF)	-43.11	3.81	-11.30	<.001 ***
	Launch Site	4.76	0.76	6.30	<.001 ***
	WIAT-II Overall Score	-12.64	4.29	-2.95	.004 **
	WIAT-II Overall Score * Word Frequency	11.61	3.99	2.91	.004 **
NDRT	Intercept	248.47	5.01	49.64	<.001 ***
	Frequency	-42.84	4.46	-9.62	<.001 ***
	Launch Site	4.84	0.76	6.36	<.001 ***
	NDRT Overall Score	-21.64	5.77	-3.75	<.001 ***
	NDRT Overall Score * Word Frequency	10.84	4.17	2.60	.011 *
LexTALE	(Intercept)	245.67	4.25	57.78	<.001 ***
	Word Frequency (LF-HF)	-43.12	3.89	-11.07	<.001 ***
	Launch Site	4.24	0.57	7.42	<.001 ***
	LexTALE	-7.91	3.99	-1.98	.055.
	LexTALE * Word Frequency	8.95	4.15	2.16	.037 *
Spelling	Intercept	247.79	4.53	54.75	<.001 ***
	Word Frequency (LF-HF)	-42.23	4.05	-10.44	<.001 ***
	Launch Site	4.82	0.75	6.43	<.001 ***
	Spelling	-13.06	4.20	-3.11	.003 **
	Spelling * Word Frequency	4.98	3.94	1.27	.212
ART	Intercept	245.34	4.71	52.06	<.001 ***
	Word Frequency (LF-HF)	-41.64	4.21	-9.88	<.001 ***
	Launch Site	4.17	0.59	7.13	<.001 ***
	ART	-7.77	3.98	-1.95	.058.
	ART * Word Frequency	-	-	-	<.001 ***
RAN	Intercept	244.98	4.48	54.73	<.001 ***
	Word Frequency (LF-HF)	-41.34	4.98	-8.30	<.001 ***
	Launch Site	4.18	0.58	7.26	.004 **
	RAN	3.45	3.98	0.87	.004 **
	RAN * Word Frequency	-	-	-	<.001 ***
Backwards Digit Span	Intercept	245.02	4.64	52.77	<.001 ***
	Word Frequency (LF-HF)	-42.12	4.04	-10.42	<.001 ***
	Launch Site	4.11	0.58	7.14	<.001 ***
	Backwards Digit Span	-6.15	3.81	-1.61	.011 *
	Backwards Digit Span * Word Frequency	5.15	3.32	1.55	<.001 ***

Note. Significance is denoted by * < .05, ** < .01, *** < .001. *p* values are adjusted for multiple comparisons via Benjamini and Hochberg's (1995) false-discovery rate correction.

Figure 2.1

The Effect of Word Frequency on Gaze Durations (ms) Moderated by LexTALE, WIAT-II and NDRT Scores



Note. shaded areas represent 95% confidence intervals.

2.4.3.3 Principal Components Analysis

As mentioned previously, WIAT-II and NDRT subtests were included separately when assessing grouping variables via principal components analysis. Scores on all tasks within the test battery were centred to allow comparisons to be made. Correlations between all tests within the test battery are presented in Table 2.4. Tests were moderately positively correlated, except for RAN and backwards digit span tests where smaller correlations were observed. A PCA was conducted on all tests and subtests to identify which tests loaded together. This procedure was used to decompose the correlation matrix and reduce the dimensions for further analysis. Two principal components were identified using parallel analysis which calculates adjusted eigenvalues for the data based on random noise expected via a simulated parallel dataset (Horn, 1965; see Figure 2.2). This method provides guidance about which components should be accepted for interpretation and further analyses. Principal components which fall above the mean of the random eigenvalues generated by the simulated data should be retained. The two components that meet this criterion in our data (PC1 and PC2) collectively contributed 49.23 % of the variance within the data. An individual test loading on a component was considered important if its contribution exceeded 10 % (expected average contribution calculated from 1/number of variables = 1/10).

PC1 was explained by the vocabulary subtest from the NDRT, Spelling, ART, LexTALE and the pseudoword decoding subtest of the WIAT-II (see Figure 2.3). This component uniquely contributed 35.40 % of the variance. We suggest that these tests are related to lexical quality (Perfetti, 2007) though we will refer to them as 'lexical proficiency' to avoid confusion with the theoretical construct. PC2 was explained predominantly by the RAN and to a lesser extent the backwards digit span task, and uniquely contributed 13.83 % of the variance (see Figure 2.4). RAN tasks have been interpreted as an index of general processing speed in cognitive tasks (e.g., Cutting & Denkla, 2001; Powell et al., 2007; Kail & Hall, 1994; Savage et al., 2007; Wolf & Bowers, 1999). Working memory is often suggested to be a "common cause" variable in the relationship between RAN and reading (Papadopoulos et al., 2016). Therefore, the assumption that working memory is related to the RAN when predicting reading ability is supported by the finding that they load together on this component. The remaining tests fell outside of these two components and were considered separately in subsequent analyses. Surprisingly, the two comprehension subtests (NDRT and WIAT-II) were not found to load on the same component. We will return to this issue in the general discussion.

Table 2.4

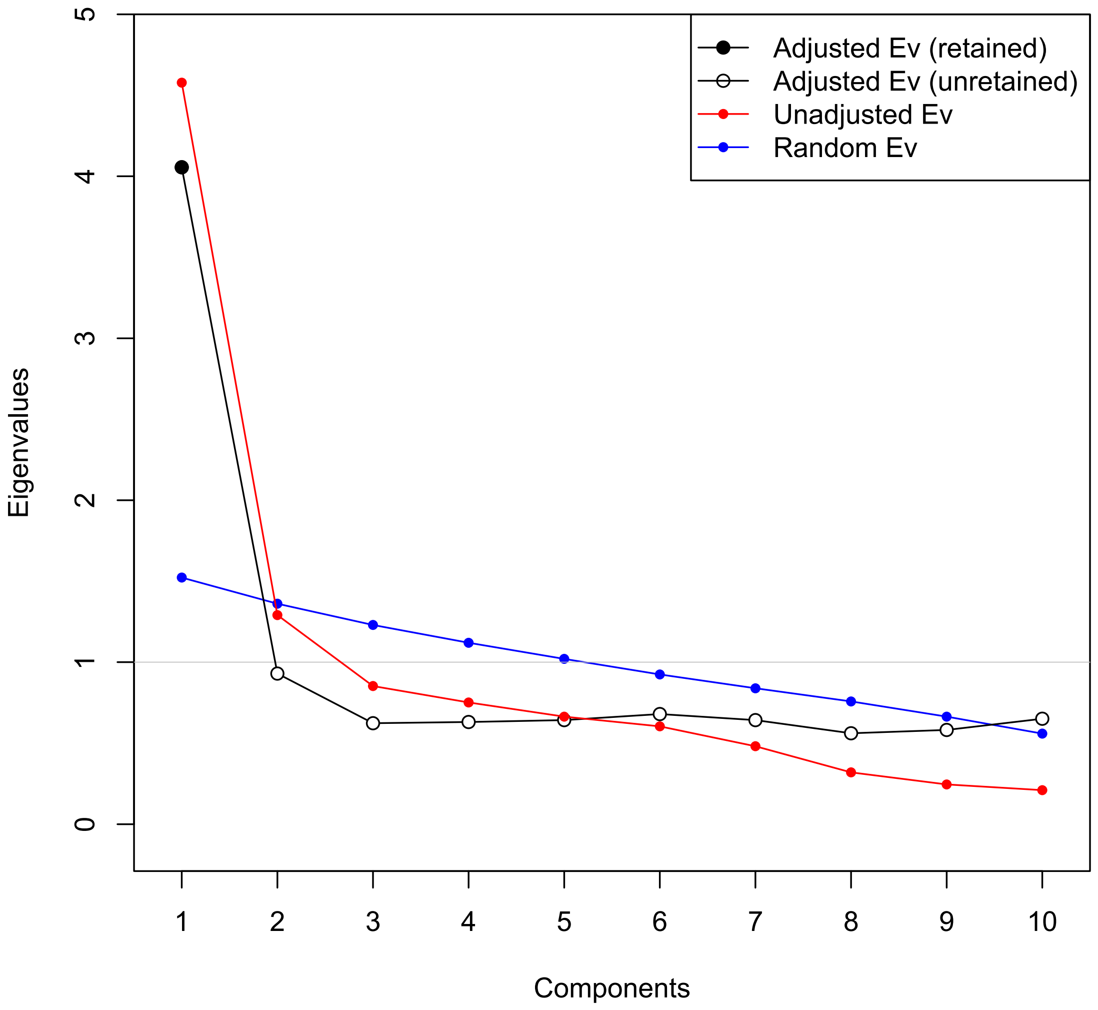
Correlations between Subtests

	NDRT Comprehension	NDRT Vocabulary	WIAT Comprehension	WIAT Pseudoword Decoding	WIAT Word Reading	Spelling	ART	LexTALE	Digit Span
NDRT Vocabulary	0.36***								
WIAT-II Comprehension	0.21	0.51**							
WIAT-II Pseudoword Decoding	0.13	0.30**	0.22**						
WIAT-II Word Reading	0.25*	0.32	0.12	0.44***					
Spelling	0.38***	0.48***	0.27*	0.47***	0.33**				
ART	0.22*	0.58***	0.36***	0.35***	0.24*	0.51***			
LexTALE	0.19	0.48***	0.29**	0.33**	0.43***	0.41***	0.36***		
Digit Span	0.08	0.22*	0.08	0.15	0.14	0.30**	0.06	0.16	
RAN	0.11	0.07	-0.17	-0.27*	0.00	-0.15	0.10	-0.20	-0.24*

Note. Significance is denoted by * < .05, ** < .01, *** < .001.

Figure 2.2

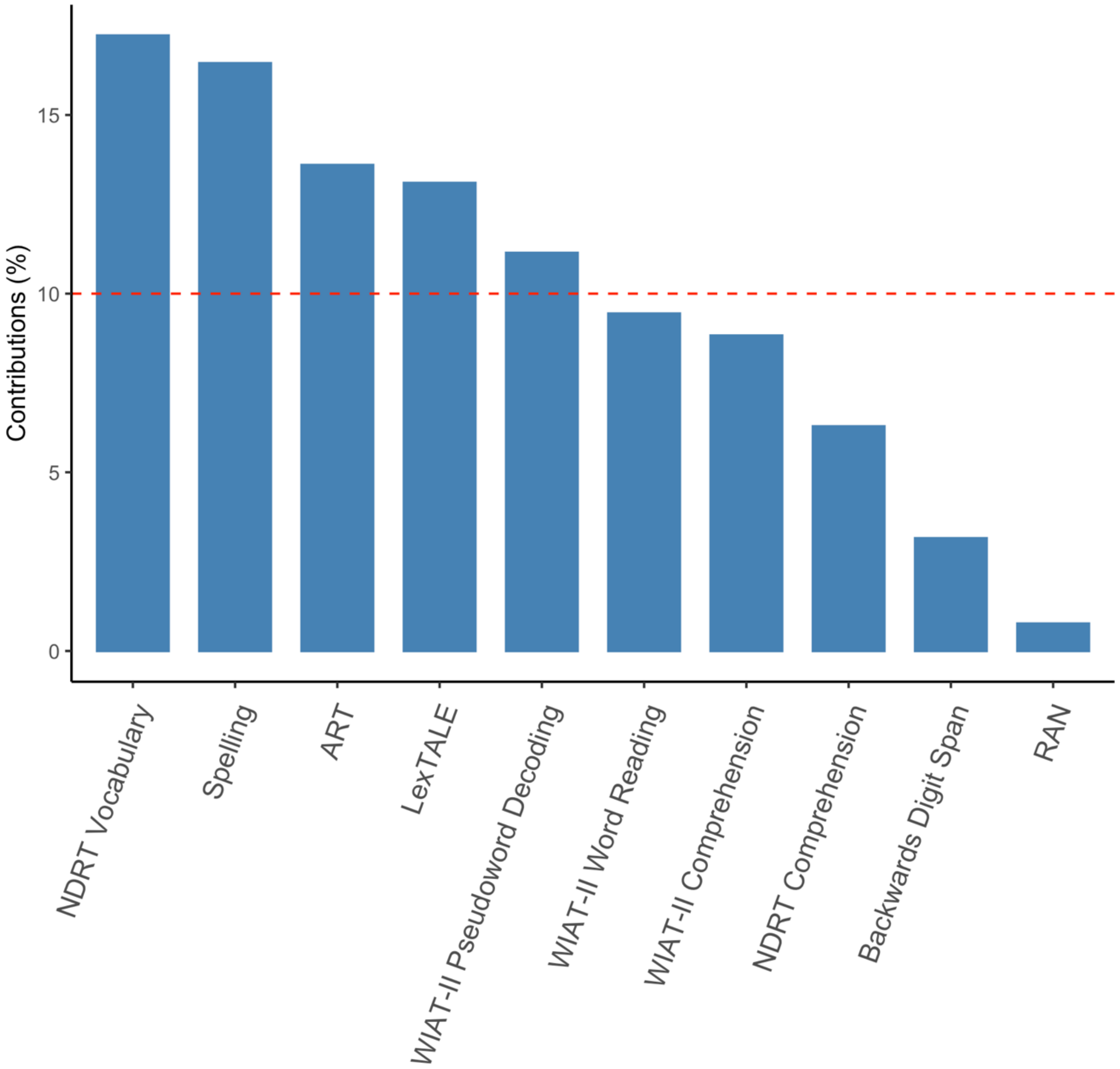
A Plot of Principal Components Identified within the Full Test Battery using Parallel Analysis



Note. Random Ev (blue) refer to randomly generated eigenvalues from a simulated parallel dataset. Unadjusted Ev (red) are the eigenvalues given by the real data. Adjusted Ev (black) show these eigenvalues adjusted to account for expected noise in the data, filled points on this line represent principal components which fall above the mean of the random eigenvalues (1), these components are retained for further analysis.

Figure 2.3

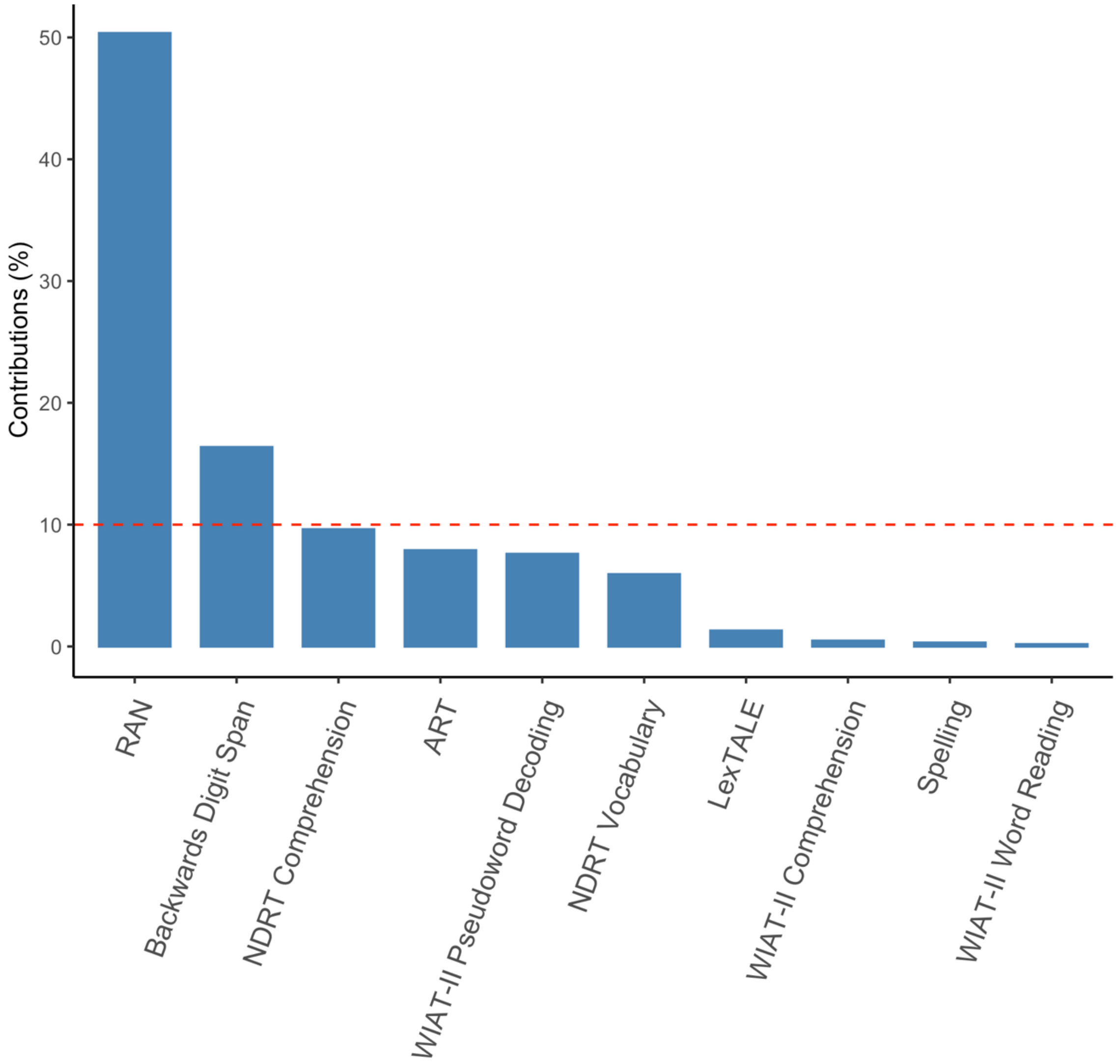
Individual Contributions of Each Individual Differences Measure on PC1



Note. The dotted line represents the expected average contribution (10 %). A contribution above this line is considered important in explaining the component.

Figure 2.4

Individual Contributions of Each Individual Differences Measure on PC2



Note. The dotted line represents the expected average contribution (10 %). A contribution above this line is considered important in explaining the component.

2.4.3.4 Analysis of Grouping variables

GLMMs were conducted for FFD, SFD, GD and go past times using a gamma distribution (as used in the single predictor models as well) with participants and items as random factors. Skipping probabilities were modelled via GLMMs with a binomial distribution. Model trimming followed the same procedure as for the single test models. First, models included all fixed effects: word frequency, trial number, launch site, PC1, PC2, the reading comprehension subtests NDRT and WIAT-II, the word reading subtest of the WIAT-II and two-way interactions between each individual differences test and word frequency (three-way interactions never contributed to the fit). Model comparison Chi-square tests were conducted to investigate whether non-significant fixed effects or interactions could be trimmed without reducing the model fit.

Next, we began to build up the random effects structure to find a converging model closest to the maximal model. Random effects were forward fitted, adding slopes in order of theoretical importance, starting with word frequency, the individual difference test, launch site and trial number. These effects were retained if they contributed to the model fit. Finally, we examined whether any non-significant fixed effects could be trimmed again. Final random effects structures were as follows. The SFD model included frequency as a slope for subjects, and frequency and PC1 scores as slopes for stimuli. The FFD model included frequency and saccade launch site as slopes for subjects, and frequency, PC1 scores and saccade launch site as slopes for stimuli. The GD model included frequency and saccade launch site as slopes for subjects, and frequency and PC1 scores as slopes for stimuli. The GOPAST model included saccade launch site as a slope for subjects, and frequency and WIAT-II comprehension scores as slopes for stimuli. The Skipping model included saccade launch site and trial number as slopes for subjects, and saccade launch site as a slope for stimuli. The sentence reading time model featured an intercept only random effects structure for subjects and stimuli.

Focusing first on the main effects, models revealed that low frequency words received longer first and single fixations, gaze durations and go past times and a lower likelihood of being skipped than high frequency words (Table 2.5 and Table 2.6). In line with previous research (Pollatsek et al., 1986), when saccades were launched nearby the target word, it was subsequently fixated for shorter durations (FFD, SFD, GD and go past times) and skipped more often. Participants who scored highly in tests associated with PC1 had shorter gaze durations and go past times, and more word skipping than participants with lower scores[3]. High scores on the WIAT-II word reading subtest were associated with shorter go past times than low scores.

Turning to the interactions, we found that scores associated with PC1, and WIAT-II comprehension influenced the relationship between word frequency on go past times. Participants who scored highly in the tests associated with PC1 were less negatively impacted by a low frequency word embedded within a sentence than low scorers (see Figure 2.5). High scorers on the WIAT-II comprehension test also exhibited smaller differences in go past times between

---

[3] For clarity, PC1 was negatively associated with the vocabulary subtest from the Nelson Denny Reading Test, Spelling, ART, LexTALE and the pseudoword decoding subtest of the WIAT-II. As such, results in Table 5 and 6 demonstrate that higher PC1 scores are associated with longer gaze durations, go past times, sentence reading times and increased skipping probabilities.

high and low frequency words than those with low scores. However, we observe that go past times for low frequency words were fairly stable across WIAT-II comprehension scores, and instead, we see shorter go past times for high frequency words read by low scorers. The difference here is unexpected and does not fit the pattern we find for WIAT-II overall scores when modelled separately on gaze durations (Figure 2.1) or for WIAT-II comprehension in sentence reading times, as we will discuss next.

Sentence reading times were included in analyses since previous studies have indicated that some individual differences are more closely related to late eye movement measures. Better spellers make longer saccades and better comprehenders read sentences more quickly and make fewer fixations (Veldre et al., 2017), those with a faster RAN score make fewer regressions and refixations (Kuperman & Van Dyke, 2011), and those with a large WMC display shorter sentence reading times (Long & Freed 2021). Sentence reading time models included these fixed effects: word frequency, trial number, PC1, PC2, the reading comprehension subtests NDRT and WIAT-II, the word reading subtest of the WIAT-II and two-way interactions between each individual differences test and word frequency and were built using the procedure described for previous eye movement models. Results revealed that a low frequency word embedded within a sentence significantly slowed the total reading time of that sentence (Table 2.7). Trials occurring later during testing featured shorter sentence reading times than earlier trials. Participants who scored highly in tests associated with PC1 and PC2, and also in NDRT comprehension, WIAT-II comprehension and word reading subtests read sentences more slowly than low scorers on these tests. Participants who scored highly in the tests associated with PC1, PC2, WIAT-II comprehension and WIAT-II word reading tasks were also less slowed down by a low frequency word in sentence reading times (see Figure 2.6).

Table 2.5

GLMMs with Multiple Test Predictors to predict Skipping Probability, First Fixation Durations (ms) and Single Fixation Durations (ms)

	Skipping Probability				FFD				SFD			
	Est	*SE*	*t*	*p*	*Est*	*SE*	*t*	*P*	*Est*	*SE*	*t*	*p*
Intercept	-2.67	0.17	-15.41	<.001***	220.57	3.54	62.31	<.001***	224.69	4.52	49.67	<.001***
Frequency	0.68	0.08	8.70	<.001***	-24.81	3.76	-6.60	<.001***	-29.19	4.17	-7.00	<.001***
Trial Number	0.00	0.00	1.73	.083	-	-	-	-	-	-	-	-
Launch Site	-0.54	0.04	-13.81	<.001***	1.48	0.62	2.38	.017*	1.86	0.42	4.45	<.001***
PC1	-0.17	0.06	-2.85	.004**	1.75	2.04	0.86	.392	-	-	-	-
PC2	-	-	-	-	-	-	-	-	-	-	-	-
NDRT Comprehension	-	-	-	-	-	-	-	-	-	-	-	-
WIAT-II Comprehension	-	-	-	-	-	-	-	-	-	-	-	-
WIAT-II Word Reading	-	-	-	-	-	-	-	-	-	-	-	-
Frequency * PC1	-	-	-	-	-	-	-	-	-	-	-	-
Frequency * PC2	-	-	-	-	-	-	-	-	-	-	-	-
Frequency * NDRT Comprehension	-	-	-	-	-	-	-	-	-	-	-	-
Frequency * WIAT-II Comprehension	-	-	-	-	-	-	-	-	-	-	-	-
Frequency * WIAT-II Word Reading	-	-	-	-	-	-	-	-	-	-	-	-

Note. Significance is denoted by * < .05, ** < .01, *** < .001.

Table 2.6

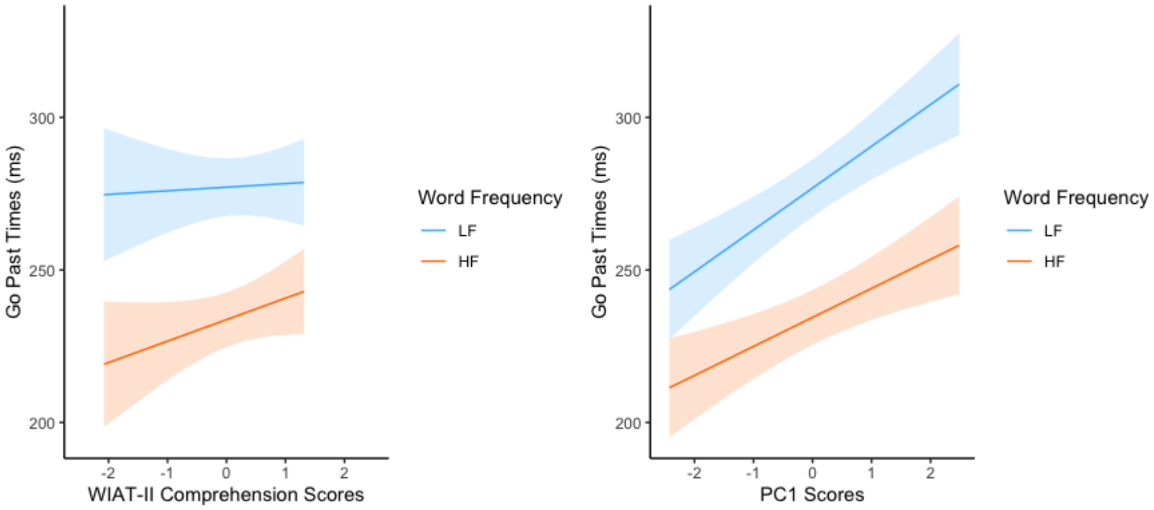
GLMMs with Multiple Test Predictors to predict Gaze Durations (ms) and Go Past Times (ms)

	GD				GOPAST			
	Est	*SE*	*t*	*p*	*Est*	*SE*	*t*	*p*
Intercept	245.87	3.85	63.94	<.001***	253.26	4.51	56.11	<.001***
Frequency	-41.50	3.78	-10.97	<.001***	-43.27	3.35	-12.91	<.001***
Trial Number	-	-	-	-	-	-	-	-
Launch Site	4.82	0.76	6.38	<.001***	5.73	0.73	7.83	<.001***
PC1	5.15	2.23	2.31	.021*	11.61	2.73	4.26	<.001***
PC2	-	-	-	-	-	-	-	-
NDRT Comprehension	-	-	-	-	-	-	-	-
WIAT-II Comprehension	-	-	-	-	4.11	4.28	0.96	.337
WIAT-II Word Reading	-	-	-	-	9.81	4.72	2.08	.038*
Frequency * PC1	-	-	-	-	-4.19	1.24	-3.39	<.001***
Frequency * PC2	-	-	-	-	-	-	-	-
Frequency * NDRT Comprehension	-	-	-	-	-	-	-	-
Frequency * WIAT-II Comprehension	-	-	-	-	5.86	2.63	2.23	.026*
Frequency * WIAT-II Word Reading	-	-	-	-	-	-	-	-

Note. Significance is denoted by * < .05, ** < .01, *** < .001.

Figure 2.5

The Effect of Word Frequency on Go Past Times (ms) as a Function of PC1 and WIAT-II Comprehension Scores



Note. Shaded areas represent 95% confidence intervals.

Table 2.7

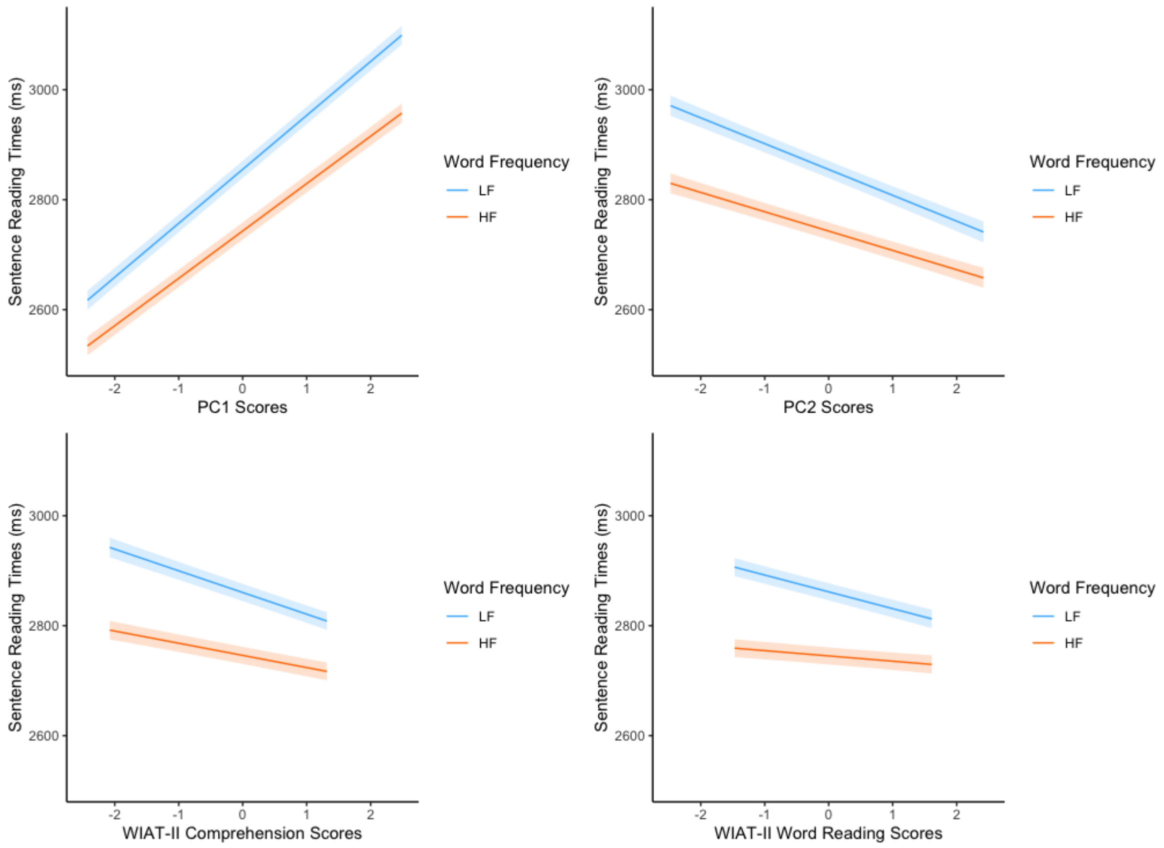
GLMM with Multiple Test Predictors to predict Sentence Reading Times (ms)

	Est	*SE*	*t*	*p*
Intercept	3040.72	1.37	2226.43	<.001***
Frequency	-118.93	1.36	-87.27	<.001***
Trial Number	-4.39	0.15	-29.19	<.001***
PC1	92.14	1.31	70.12	<.001***
PC2	-41.07	1.88	-21.81	<.001***
WIAT-II Comprehension	-30.78	1.23	-24.99	<.001***
WIAT-II Word Reading	-20.07	1.62	-12.38	<.001***
NDRT Comprehension	-113.07	1.42	-79.68	<.001***
Frequency * PC1	-11.94	1.39	-8.58	<.001***
Frequency * PC2	11.83	1.46	8.11	<.001***
Frequency * WIAT-II Comprehension	17.30	2.45	7.07	<.001***
Frequency * WIAT-II Word Reading	20.96	1.47	14.29	<.001***
Frequency * NDRT Comprehension	-	-	-	-

Note. Significance is denoted by * < .05, ** < .01, *** < .001.

Figure 2.6

The Effect of Word Frequency on Sentence Reading Times (ms) as a Function of PC1, PC2, WIAT-II Comprehension and WIAT-II Word Reading Scores



Note. Shaded areas represent 95% confidence intervals).

2.5 General Discussion

Our approach to assessing the influence of individual differences on the effect of word frequency consisted of first examining separate models in which each test was a single predictor – besides word frequency and the interaction - to allow for a more straightforward comparison with the existing literature. Tests were then grouped based on overlapping variance using a PCA. Finally, eye movement patterns were modelled using these factors alongside individual tests that were not grouped as predictors. Predictions, besides upholding classic main effects related to word frequency and saccade launch site, were first that when cognitive tests are considered in separate models our findings would replicate previous research that linked high NDRT scores (Ashby et al., 2005), large vocabularies (Cop et al., 2015), better spelling (Veldre et al., 2017), high ART scores (Chateau & Jared, 2000; Gordon et al., 2020), faster RAN (Kuperman & Van Dyke,

2011) and larger WMC (Long & Freed, 2021) to smaller gaze durations and a reduced word frequency effect. Secondly, we predicted that skills associated with lexical quality (vocabulary, spelling, print exposure, comprehension and decoding) would be grouped together in the PCA. Finally, in models with multiple cognitive skill predictors, we expected high scores in a factor associated with lexical quality to predict a reduced word frequency effect as seen in previous research (Veldre et al., 2017; Cop et al, 2015; Chateau & Jared, 2000; Gordon et al.,2020). This latter finding would be due to greater familiarity with low frequency words in line with the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002). Other cognitive skills (RAN and WMC) were also included to allow for a comparison with previous studies.

Before delving into these three analytic approaches, we mention that throughout our analyses all models showed the well-documented effect of word frequency on eye movement measures (Rayner & Fischer, 1996; Rayner et al., 1996, Rayner et al., 2004), whereby high frequency words received shorter fixation times and were more likely to be skipped than low frequency words. In addition, all models for word-level measurements (fixation durations, gaze durations, go past times and skipping probabilities) demonstrated that when saccades were launched close to the target word, it was subsequently fixated for shorter durations and skipped more often, again in line with previous research (e.g., Pollatsek et al., 1986).

2.5.1 Models with Single Test Predictors

Simultaneously considering multiple models with single test predictors for a single dependant variable is likely to increase Type I error rates (von der Malsburg & Angele, 2017). For this reason, we opted to follow Schmidtke et al., (2018) in utilising a false-discovery rate correction (Benjamin & Hochberg, 1995) to account for multiple comparisons at this stage of analysis. Following such corrections, fixed effects were subsequently only considered to be significant if the adjusted *p*-value (reported in Table 2.3) was below the *p* <.05 threshold. The main effect of LexTALE scores on gaze durations was found to be marginal following correction for multiple comparisons in these analyses. Future research may instead consider an alternative approach, for example a Random Forests technique (Matsuki et al., 2016), in which multiple comparisons do not present issues with inflated Type I error rates.

Our first prediction was partially supported by models with only one cognitive test as a predictor, in that shorter gaze durations were associated with high scores in both overall reading ability tests (NDRT and WIAT-II) and spelling. These findings are consistent with the idea that shorter gaze durations are associated with better readers compared with average readers (Ashby et al, 2005). However, shorter gaze durations were not predicted by faster RAN or higher scores on the backwards digit span task, suggesting that efficient word identification is facilitated more by top-down processes related to comprehension and word knowledge than bottom-up processes related to WMC and faster processing of single characters.

In addition, we replicated previous research that linked high reading abilities (Ashby et al., 2005) to a reduced word frequency effect and partially replicated research based on large vocabularies (Cop et al., 2015)[4]. Both overall reading ability measures, the NDRT and WIAT-II and the LexTALE task moderated the impact of the low frequency word on reader's gaze durations. High scorers on these tests were not slowed down by a low frequency word as much as low scorers were.

Focusing on vocabulary as measured by the LexTALE, Cop et al. (2015) analysed eye movements of English-speaking monolinguals when reading an entire book (this was a baseline to compare with bilinguals, who read half of the book in their first language and half in their second language). Although patterns in data here and from Cop and colleagues are qualitatively similar, we identified some challenges when comparing them. First, the current study experimentally embedded high and low frequency words into single sentence frames and generated a high degree of statistical power through a large sample of subjects to examine individual differences. In contrast, Cop et al. (2015) examined differences in naturally occurring word frequencies in paragraphs within a novel for the GECO corpus (Cop et al., 2017). They tested just 14 monolingual participants, and in their study, it was the volume of reading material that generated strong statistical power. Previous research has indicated that reading single sentences can produce quite different patterns in reading behaviour compared to reading paragraphs. Radach et al. (2008) demonstrated that frequency effects are more pronounced when reading sentences rather than passages. We exercise caution when comparing these findings directly as discrepancies may be due to differences in reading material format, however we must point out that trends in the data are fairly consistent.

We also considered the finding from Ashby et al. (2005) that readers with high reading proficiency scores based on the NDRT exhibited smaller word frequency effects compared to readers with average proficiency scores. In their study, simple effects revealed that average scorers on the NDRT based on percentile scores (ranging from the 6th to the 70th percentile with a mean at the 40th percentile) showed an effect of word frequency on gaze durations (i.e., LF words

---

4 Cop et al. (2015) observed the interaction between LexTALE and word frequency in SFDs rather than GDs. SFD analyses in the current study did not reveal a significant main effect of the LexTALE (t = - 0.51, p = .612) or any effect of the LexTALE on the relationship between word frequency and SFDs (*t* = 1.45, *p* = .146). There were clear effects of word frequency (*t* = -7.68, *p* < .001) and saccade launch site (*t* = 3.92, *p* < .001).

received longer gaze durations than HF words) whereas high scorers (scoring above the 74th percentile with mean at the 88th percentile) did not. In our data, NDRT scores were found to significantly predict differences in gaze durations, consistent with Ashby et al. (2015). When individuals had higher scores on this test, their gaze durations on target words were shorter. The interaction between word frequency and proficiency as established by the NDRT reported by Ashby et al. (2005) was also observed in our data.

The WIAT-II followed similar patterns, high scores on this reading ability test were associated with shorter gaze durations in general and were less impacted by low frequency words than low scorers were. Since scores on both the NDRT and WIAT-II are composites of two or more skills, it was important to break these down into subtests in the PCA to observe which underlying skills were grouped together and which were key in driving the reduced word frequency effect.

Next, we consider spelling. As part of a study looking at spaced vs unspaced text, Veldre et al. (2017) observed that in normally spaced text, higher spelling ability predicted smaller fixation durations and a reduced frequency effect on all duration measures. In our sample, we found that highly skilled spellers had significantly shorter gaze durations than less skilled spellers, but spelling ability was not found to predict differences in the size of the frequency effect (though trends were in the same direction as Veldre et al., 2017). This discrepancy may be down to a difference in modelling as Veldre and colleagues included NDRT scores within the same models as spelling, whereas we (initially) report models with single predictors, in this case spelling. The influence of spelling ability on gaze durations supports the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002) in that higher quality lexical representations appear to result in more efficient word processing (measured by gaze durations, Rayner, 1998). When considering word frequency, it is not a prerequisite of a good speller to be more familiar with low frequency words, though spelling ability and vocabulary size are often correlated. Therefore, we demonstrate the distinction between high spelling abilities and large vocabularies in this research. Though a good speller may process all words more efficiently than a poor speller, a larger vocabulary may be a better predictor of more efficient processing of less common words in particular.

As predicted by previous findings from Moore and Gordon (2015), individuals with higher ART scores had shorter gaze durations in our eye tracking task, which is consistent with the idea that greater experience leads to more efficient word processing. Print exposure measured by this test has also been previously linked to higher reading ability and a reduced effect of word frequency on gaze durations (Moore & Gordon, 2015). However, an interaction between ART scores and word frequency was not replicated in our data. Kuperman and Van Dyke (2013) offer an explanation for a null effect of print exposure on the size of the word frequency effect. They observed that when print exposure is matched, differences in eye movements in response to high and low frequency words still occurred (Kuperman & Van Dyke, 2013). The authors suggested that it may not be as simple as measuring the amount of exposure to words that determines greater quality lexical representations – but rather that highly skilled readers are better at utilising experience with words to create high-quality lexical representations. This distinction is likely to be subtle in the literature since this ability often correlates with print exposure.

In addition, we note that there are some differences in our sample of readers compared to the sample that the ART was based on for the UK (Acheson et al., 2008). One discrepancy is that ART scores in the current study (*M* = 11.04, *SD* = 7.83) were much lower than scores in the sample from Acheson et al. (2008), when this test was adapted for UK use (*M* = 22.7, *SD* = 10.8). Moore and Gordon (2015) note that the ART is less effective at discriminating differences for lower scores. They also suggested authors appearing on the ART created in 2008 may be somewhat out of date, and therefore may be less useful for determining print exposure for the college-aged samples in 2023[5]. We agree and also propose that in all likelihood it is now increasingly possible to be a voracious reader without reading printed books due to the accessibility of online text. Researchers should consider improvements suggested by Moore and Gordon (2015) when considering the ART for future experiments.

In the RAN and digit span models, there were no significant effects related to the test or interactions of the test with word frequency. RAN tasks have most often been associated with later effects in the eye movement record during reading experiments in previous research, such as refixations on words and second pass reading times of whole sentences (Gordon et al., 2020), therefore this outcome is not totally unexpected (although sentence reading times have the potential to unveil later effects and were predicted by PC2 (which included these measures) in later analyses. We return to this in later sections). However, WMC has previously been linked to a reduction in the word frequency effect in gaze durations (Long & Freed, 2021). WMC in the Long and Freed (2021) experiment was comprised of four measures (Reading span; Unsworth et al., 2005, Alphabet span; Craik, 1986, Minus span; Salthouse 1988, and Visual number span from the Kit of Factor-Referenced Tests; Ekstrom et al., 1976), whereas the current study just uses a

---

[5] Moore and Gordon (2015) conducted a factor analysis on the ART and concluded that two factors were present rather than one clear 'print exposure' factor. They theorized that one of these factors was related to academic or literary reading rather than reading for leisure as the test originally intended to measure (West et al., 1993). They also identified a positive correlation between selecting author names and foil names, suggesting that some participants used a lower criterion for deciding that they know an author, this correlation was moderate in our sample ($r = 0.38$, $p < .001$). Moore and Gordon suggested that stricter penalties for selecting a foil name or including a measure of self-rated confidence could provide more information about a participant's criteria.

Backwards Digit Span task to describe WMC, which may have influenced a different pattern of results. Further research should investigate whether the specific WMC measure selected for analyses is important when considering such differences in the literature.

2.5.2 Principal Components Analysis

Unsurprisingly, we found that scores across most of the test battery were correlated, since individual performances on cognitive tasks often are (Deary, 2000), which make it difficult to interpret individual skills as unique predictors of reading. A principal components analysis revealed two components, allowing related skills to be grouped where variances overlap. The first, lexical proficiency (PC1), was negatively associated with scores on the NDRT vocabulary test, spelling, ART, LexTALE, and WIAT-II pseudoword decoding (higher scores on these tests were associated with lower scores on PC1). We suggest that this variable is related to lexical quality based on the associated skills in line with our second prediction that skills associated with lexical quality would feature overlapping variance. We refer to PC1 as lexical proficiency to avoid confusion with the theoretical construct of lexical quality.

The second factor (PC2) was associated with RAN (faster scores on the RAN were associated with higher scores on PC2) and to a lesser extent the backwards digit span test (more items recalled in the backwards digit span test were associated with higher scores on PC2). We suggest that this variable describes participants' speed of processing. Three tests did not load on these components and were treated as distinct skills in subsequent analyses: NDRT comprehension, WIAT-II comprehension and WIAT-II word reading. It should be noted that word reading was close to the threshold for inclusion in PC1, and had it been included it would not have been surprising since it is related to word knowledge which appears to be a commonality across PC1 measures. Surprisingly, the two comprehension measures included in the test battery did not load together despite supposedly measuring the same construct, this important finding will also be discussed later.

2.5.3 Analyses with Identified Components

Models that included the two identified components and other distinct tests revealed main effects that are consistent with previous research that found more skilled readers identify words more quickly, read sentences faster and skip over words more often than less skilled readers (Rayner, 1998; Ashby et al., 2005). PC1 was predictive of shorter gaze durations, go past times, sentence reading times and higher skipping probabilities when individuals scored highly on the associated tests. Higher skipping probabilities may suggest that participants who score highly on tests associated with PC1 may be advancing faster in the recognition process of the parafoveal word and subsequently make a decision to skip it (Inhoff & Rayner 1986; White, 2008), alternatively they may employ riskier saccade targeting strategies (O'Regan 1990, 1992; Rayner & Fischer, 1996; Rayner, 1998). A risky reading strategy is where some readers (often seen in older adults; Rayner et al., 2006) compensate for less advanced lexical processing by frequently guessing upcoming words, skipping them and often returning when wrong guesses are realised. A follow-up experiment should feature more difficult questions to encourage reading for comprehension, since a risky strategy is likely to have been quite successful in the present study.

Our third prediction was that a factor reflecting the shared variance of skills related to lexical quality would be a strong predictor of a reduced word frequency effect as suggested by previous research about each contributing skill[6] (spelling; Veldre et al., 2017; vocabulary; Cop et al., 2015; print exposure; Chateau & Jared, 2000; Gordon et al., 2020). This prediction was only partially upheld as our lexical proficiency factor PC1 influenced the relationship between word frequency only in some eye movement measures, but not others. In comparison to low scorers, high scorers exhibited shorter go past times and shorter sentence reading times for low frequency words. These findings are in late eye movement measures suggesting that for skilled readers, very precise lexical representations are associated more with faster embedding of meaning into sentence context, rather than faster orthographic decoding of low frequency words. However, an important comment needs to be made regarding not finding an interaction between PC1 and word frequency in earlier measures. Our reported Linear Mixed Models were created using a pruning strategy that aimed to achieve the largest random effects structure possible (see Barr et al., 2013). We suspect pruning techniques used in previous years would often be comparable to intercept only models. If we run intercept only models for the current analyses, consistent interactions between PC1 and word frequency can be found in SFD, FFD and GD, and an WIAT-II comprehension by word frequency interaction in GD can be found. We mention this as it could explain parts of the discrepancies with previous related research. In addition, earlier sections of our own analyses (on single test predictors) found differences in the magnitude of the word frequency effect in gaze durations, which were not found when looking at models based on multiple test predictors. One reason for this may be that PCA groups factors via variance shared amongst the tests included. There may be unique features of the LexTALE and reading ability tests (NDRT & WIAT-II) that are key in predicting this pattern, or perhaps more likely, there is a

---

[6] It should be noted that factors obtained via PCA only reflect shared variance amongst the skills measured, filtering out residual variance that is unique to individual measures. We can therefore only draw conclusions about proxies of latent variables such as lexical quality that ignore the nuance of each cognitive task.

reduction in statistical power to estimate an effect where more parameters are estimated within the model, as in the current analysis.

A previous investigation of individual differences in less skilled readers by Kuperman and Van Dyke (2011) observed strong predictive value of the RAN on eye movement patterns. However, since slower RAN is often associated with reading difficulties (Bowey, 2005), we did not expect it to be a strong predictor of differences in eye movements in our sample of average-to-very-skilled readers. In addition, RAN scores have not always been found to be a useful predictor of a reduced word frequency effect when included in models with measures of print exposure (Gordon et al., 2020). We made similar predictions about WMC, due to the idea that it is a 'common cause" variable in the relationship between RAN and reading (Papadopoulos et al., 2016). PC2 was highly associated with the RAN and to a lesser extent the working memory measure (backwards digit span test). High scores were associated with faster processing of information within whole sentences, but not of single words. In addition, a small reduction in the word frequency effect was found associated with high PC2 scores in sentence reading times which may reflect a small decrease in rereading time following an uncommon word for readers with faster processing speeds.

We first consider the RAN, as it influences PC2 more than any other measure, and again mention that previous research suggests that the RAN is more predictive of later measures in the eye tracking record such as refixations, regressions, foveal-on-parafoveal effects and second pass reading times (e.g., Gordon et al., 2020). The observation that PC2 is predictive of shorter sentence reading times in our data therefore supports the idea that the RAN reflects "efficient coordination of perceptual-motor and attentional processing during reading" (Gordon et al., 2020, p. 553).

PC2 was also somewhat reflective of higher scores on a backwards digit span task (WMC), which supports the idea that working memory is a "common cause" in the relationship between the RAN and reading (Papadopoulos et al., 2016). Overall, results supported the prediction that lexical quality was a stronger predictor of differences in eye movements for average-to-very-skilled readers than RAN and WMC. However, future research should focus on longer passages for reading comprehension and should continue to assess later measures of eye movements in relation to variables associated with WMC and RAN scores to fully assess the literature.

Our test battery featured two comprehension subtests, first we discuss the WIAT-II. This comprehension subtest was not found to influence early eye movement measures in models with multiple test predictors, but was found to predict sentence reading times, a later eye movement measure, and also influenced the relationship between word frequency and eye movements in both sentence reading times and go past times. High scores were associated with shorter sentence reading times when a low frequency word was embedded compared to low scores. However, in go past times, the pattern observed was slightly different, low scorers were found to spend less time reading a high frequency word before moving on than high scorers. No difference was observed for low frequency words. We interpret this pattern with caution as we note that scores on this test are clustered towards high scores, with very few low scores (note that the low scores did not fall outside the 2.5 SD range for outlier justification). We suspect this unusual pattern might be due to strategic differences in whether a processing difficulty is resolved by making a regression or by first finishing the entire sentence. Sentence reading times were in line with the more established pattern that the low frequency word impacts better readers less.

Earlier, we discussed the finding that the overall WIAT-II score (which includes comprehension, word reading and pseudoword decoding) was associated with differences in gaze durations in the model when it was the only predictor entered and interacted with the word frequency effect on gaze durations. It may be that effects seen in early measures were driven by a subtest related to lexical proficiency: pseudoword decoding (which was associated with PC1, and generated similar effects in the models with multiple test predictors) whereas comprehension in the WIAT-II influenced later eye movement measures in the bigger models.

Turning to the NDRT, we found no main effect associated with the NDRT comprehension subtest on the word-based measures (but did on sentence reading times), or interactions between this test and word frequency in any of the models with multiple test predictors. Note that in the earlier reported analyses, when the entire NDRT (including the comprehension subtest) was the only predictor, it led to a main effect and interaction with word frequency on gaze durations. A likely hypothesis is that the reduced word frequency effect seen in associated with NDRT total was driven by the vocabulary subtest (associated with PC1).

We observe here some discrepancies in findings related to comprehension based on which test is selected. High WIAT-II comprehension scores were associated with a reduced word frequency effect in later eye movement measures whereas high NDRT comprehension scores were not in any measures. Since these two comprehension measures were included separately, it could be argued that some of the variance due to NDRT comprehension may have overlapped with variance attributed to WIAT-II comprehension. However, these tests did not load together in the PCA and were not very highly correlated (r = 0.21). We suspect that this demonstrates a case of the *Jingle fallacy* (Thorndike, 1904), – the false assumption that two instruments measure the same construct because they share a name, namely that both tests are measures of a specific "reading comprehension skill" and we think this finding has important repercussions for comparing research that used one of these tests.

Indeed, previous research has indicated that different cognitive abilities are addressed by different tests of reading ability (Keenan et al., 2008; Kendeou et al., 2012). We note some qualitative differences in the way that these tests are conducted that might indicate such differences in the constructs measured. The WIAT-II comprehension test includes a variety of texts including single sentences, fiction and non-fiction passages with the option to read aloud or silently, whereas the NDRT consists of non-fiction passages that participants read silently and the passages remains accessible while individuals answer self-paced questions on screen. Previous research has suggested that differences in the format of reading materials (sentences vs paragraphs; Radach et al., 2008; fiction vs non-fiction; e.g., Graesser et al., 1998; Zwaan, 1994) and in reading strategy (reading aloud or silently; e.g., Hale et al., 2007) can alter the behaviour measured. Additionally, the use of non-fiction passages in the NDRT may mean this test is more related to general knowledge than reading comprehension, as suggested by Coleman et al. (2010) who found greater than chance levels of correct answers after administering a "passageless" version of the test to college students (see also Ready et al., 2013).

Although both tasks were timed, during the NDRT participants were unaware of the time limit until they were asked to stop. We therefore have no reason to expect that participants adopted an unnatural reading pace. The time given to complete the NDRT comprehension subtest in this study was half of the time usually allocated for this test which may produce increased discrimination of scores between more proficient readers (Andrews et al., 2020) but participants were not aware of this. In contrast, participants were aware of the timed element in the WIAT-II comprehension test since the test is administered face-to-face and the timing is explicit in the instructions. However, measures of WIAT-II reading speed (m = 360.84 *ms*, SD = 89.61 *ms*) and NDRT words per minute (m = 272.18, SD = 74.36) were highly correlated (r = -0.65), therefore, we do not expect time pressure differences to be the cause of any differential predictive value between the tests. Instead, we propose that the face-to-face aspect of testing in the WIAT-II comprehension task might have influenced participants' performance. This procedural difference could lead to performance anxiety especially where some sections of the test required participants to read aloud and where questions had to be answered directly to the experimenter, thereby creating qualitative differences in the experience of the participant in the NDRT versus the WIAT-II.

Importantly, the Jingle fallacy found here may explain some inconsistent results across studies that model comprehension on different tests (e.g., Mézière et al., 2023; Cutting &

Scarborough, 2006). However, it is unclear whether such findings are actually in disagreement (failure to conceptually replicate) or whether they are simply measuring separate latent constructs as suggested here when two comprehension tests were not grouped by a single latent construct in our PCA. An analysis of all the conflicting findings in the literature that may be attributed to weak convergent validity is beyond the scope of this paper but should be considered in future research. We advise that these comprehension tests should not be used interchangeably since they appear to tap into independent aspects of reading comprehension. Researchers should proceed with caution when selecting a reading comprehension test in future research.

2.5.4 Summary

This study investigated the patterns of individual differences in skilled reading that have been mentioned in the literature, that is, generally faster reading is associated with higher reading ability (e.g., Ashby et al., 2005; Veldre et al, 2017; Moore & Gordon, 2015;). Our focus was also on the size of the frequency effect given that a comparatively smaller frequency effect during reading has been suggested as indicative of better reading skills (Ashby et al., 2005; Haenggi & Perfetti, 1994).

We sampled average-to-very-skilled college-aged readers, and though differences based on skill may be clearer between less skilled and average readers, it is important to note that differences can still be found within skilled reader populations. We consistently found robust effects of word frequency and saccade launch site across our reader population, but readers varied in their efficiency of word and sentence reading which can be mapped onto differences in cognitive skills.

We first examined separate models in which each test was a single predictor – besides word frequency and the interaction - to allow for a more straightforward comparison with the existing literature. Broadly in agreement with the literature, higher skill was associated with shorter gaze durations in two reading ability tests (WIAT-II and NDRT), and in spelling. A reduced word frequency effect was associated with reading ability tests and vocabulary knowledge but was not found to be associated with other tests included in our test battery.

We grouped shared variance across our test battery into two grouping variables, lexical proficiency and speed of processing. The two comprehension tests (NDRT and WIAT-II) were not grouped into one of the two grouping factors and surprisingly were not grouped in a single factor. We urge researchers to be cautious when selecting one of these two comprehension tests for future research as these tests did not load together and demonstrated an example of the Jingle fallacy. We discussed reasons why these two tests might have qualitatively different ways in measuring comprehension.

When all measures were included in models with multiple test predictors, eye movements often associated with skilled reading (shorter fixation times and higher skipping rates) were most consistently related to a factor we identified as lexical proficiency. Our findings support the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002) in that precise lexical representations support faster word recognition processes as a reduced word frequency effect was associated with higher lexical proficiency (PC1).

2.5.5 Constraints on Generality

This study sampled average-to-very-skilled readers, and it should be noted that differences based on skill may be more clearly observed between less skilled and average readers. Though the sample included a wide age range and included some participants from the local community, a large proportion were undergraduate psychology students.

Chapter 3 Individual Differences and the Transposed Letter Effect during Reading

Charlotte E. Lee[1], Ascensión Pagán[2], Hayward J. Godwin[1], and Denis Drieghe[1]

[1] School of Psychology, University of Southampton, United Kingdom,

[2] School of Psychology, University of Leicester, United Kingdom.

3.1 Abstract

When a preview contains substituted letters (SL; *markey*) word identification is more disrupted for a target word (*monkey*), compared to when the preview contains transposed letters (TL; *mnokey*). The transposed letter effect demonstrates that letter positions are encoded more flexibly than letter identities, and is a robust finding in adults. However, letter position encoding has been shown to gradually become more flexible as reading skills develop. It is unclear whether letter position encoding flexibility reaches maturation in skilled adult readers, or whether some differences in the magnitude of the TL effect remain in relation to individual differences in cognitive skills. We examined 100 skilled adult readers who read sentences containing a correct, TL or SL preview. Previews were replaced by the correct target word when the reader's gaze triggered an invisible boundary. Cognitive skills were assessed and grouped based on overlapping variance via PCA and subsequently used to predict eye movement measures for each condition. Consistent with previous literature, adult readers were found to generally encode letter position more flexibly than letter identity. Very few differences were found in the magnitude of TL effects between adults based on individual differences in cognitive skills. The flexibility of letter position encoding appears to reach maturation (or near maturation) in skilled adult readers.

Keywords: transposed letter effect, individual differences, parafoveal preview benefit

This paper has been submitted for publication and will be referenced in later chapters as:

Lee, C. E., Pagán, A., Godwin., H. J., & Drieghe, D. (2023). *Individual differences and the transposed letter effect during reading.* Manuscript Submitted for Publication.

Authorship Statement

Charlotte Lee conceptualised and created this experiment, performed all analyses, interpreted data and wrote this manuscript for publication.

Denis Drieghe and Hayward J. Godwin provided project supervision and comments on this manuscript.

Ascensión Pagán provided advice on experimental design and comments on this manuscript for publication.

3.2 Introduction

A large body of evidence suggests that eye movements during reading are fundamentally linked to a reader's cognitive processing and reveal processing difficulties related to features of the text (see Rayner, 2009; Liversedge & Findlay, 2000). A number of studies have also looked at the influence of individual differences in reading skills, and how these relate to the patterns of eye movement behaviour (e.g., Kuperman et al., 2018; Kuperman & Van Dyke, 2011; Ashby et al., 2005; Ashby et al., 2012; Chace et al., 2005, Häikiö et al., 2009; Haenggi & Perfetti, 1994; Jared et al., 1999, Luke et al., 2015; Veldre & Andrews, 2014, 2015a, 2015b). Evidence suggests that skilled adult readers process words more quickly than less skilled readers or children, as seen in shorter fixations, shorter gaze durations and fewer refixations (see Rayner, 2009 for a review).

The average skilled adult reader can extract information such as spacing from 14-15 character spaces from the point of fixation in the direction of the upcoming word (rightwards in English readers) and 3-4 character spaces in the direction of the previous word in alphabetic languages (McConkie & Rayner, 1975). This asymmetric visual field is a phenomenon which facilitates the pre-processing of information related to upcoming words. Word identification, which requires higher visual acuity, occurs 3-4 letters to the left and 6-7 letters to the right of fixation in alphabetic languages that are read from left to right (Rayner & Bertera, 1979; Rayner et al., 1981). In practice, the size of the perceptual span varies between readers, and notably increases with age until skilled reading is fully developed (Häikiö et al., 2009; Rayner, 1986; Sperlich et al., 2015; Henderson & Ferreira, 1990). Though differences in the size of the perceptual span can relate to differences in text processing difficulty (as the difficulty of a text increases, the readers' perceptual span shrinks, Rayner, 1986), individual differences are also suggested to influence the size of a reader's perceptual span. Veldre and Andrews (2014) found that adult readers with high spelling and reading abilities had larger perceptual spans during reading than readers with low spelling and reading abilities. A study by Häikiö et al. (2009) found that slower readers identified fewer letters during a fixation than faster readers of the same age group (for Finnish children aged 8, 10 and 12, and for adults). They suggested that slower readers, unlike faster readers, allocate most of their processing resources to words when they are directly fixated on the fovea (2 degrees in the centre of vision). However, readers generally also process some information about an upcoming word parafoveally (in the parafoveal region, approximately 5 degrees to the left or right of fixation) when the eyes are fixating the preceding word.

3.2.1 Parafoveal Processing in Skilled Reading

Parafoveal processing enables the reader to extract information about the upcoming word before it is directly fixated, and when this information is useful, the upcoming word is processed more rapidly. Studies that explore parafoveal processing during reading most often use a *gaze contingent invisible boundary paradigm* (McConkie & Rayner, 1975; Rayner, 1975). This paradigm allows researchers to display a manipulated version of the target word to the right of a reader's gaze until their eyes cross an invisible boundary, whereupon the display is switched to show the correct target. The change occurs during a saccade when the readers' vision is blurred, resulting in this manipulation being usually undetected. Studies have found that reading is facilitated when the preview of the target word is identical to the target word (Rayner, 1975). In addition, when a preview shares orthographic or phonological information with the target word some *preview benefit* is also found*,* where faster processing of the parafoveal word when it is subsequently fixated is facilitated by information gathered from the preview word (Schotter et al., 2012).

3.2.2 Individual Differences in Parafoveal Processing

Individual differences in reading and spelling abilities have been found to modulate the amount of information that can be extracted from the parafovea and used to facilitate word identification (Veldre & Andrews 2015a; 2015b; 2016; Andrews & Veldre, 2019). Skilled readers who are also good spellers extract more parafoveal information about word length and lexical features of a word (Veldre & Andrews, 2015b; 2015a). However, differential effects have been noted for the extraction of semantic information. Good spellers had a reduced preview benefit from semantically related previews (demonstrating competition between semantic and orthographic information), whereas high reading ability has been found to predict a greater benefit from semantically related previews (Veldre & Andrews, 2016).

The current study further investigated individual differences in parafoveal processing of orthographic information in an upcoming word. Our focus was on individual differences in the extraction of letter position and letter identity information in the parafovea. To explore this, we first need to discuss how this information is encoded in isolated word identification.

3.2.3 The Transposed Letter Effect

To 'crack the orthographic code' research has used transposed letter (TL) stimuli, where the position of letters in a real word are swapped to create a nonword, to investigate how precisely letter position information is encoded. Priming paradigms and lexical decision tasks are used to see how word identification is affected by TL pseudowords compared to pseudowords created by substituting letters within a base word. In masked priming studies, the time to identify a real word target (judge) is reduced when a TL prime (jugde) was used compared to a substituted letter (SL) prime (junpe) (e.g., Forster et al., 1987; Lupker, et al., 2008; Perea & Lupker, 2003, 2004; Schoonbaert & Grainger, 2004). When participants must decide whether a letter string is a real word in a lexical decision task, a TL pseudoword increases the response latency and likelihood of errors (misinterpreting the pseudoword as a real word) compared to a SL pseudoword (Colombo, et al., 2017; Grainger et al., 2012; O'Connor & Forster, 1981; Perea & Fraga, 2006). This evidence is consistent with the idea that a TL pseudoword is perceived to be more similar to the base word than a SL pseudoword is, and therefore more difficult to reject in a lexical decision task. These findings suggest that letter positions and letter identities are encoded independently, given that there is a processing advantage when letter identities are preserved despite changes in letter positions.

This flexible letter position encoding mechanism has been incorporated in recent models of word recognition (the SOLAR model (Davis, 1999; 2010); the Open Bigram model (Grainger & Van Heuven, 2003; Grainger et al., 2006; Grainger & Ziegler, 2011; Grainger et al., 2012); the Overlap model (Gómez, Ratcliff & Perea, 2008); the SERIOL model (Whitney, 2001); and the Bayesian Reader (Norris, 2006)).

Reading studies have been consistent with the evidence from isolated word recognition studies. Rayner, White et al. (2006) observed that readers' eye movements were only slightly disrupted when reading a sentence containing a TL pseudoword. Similarly, transposed letter effects have been investigated using the boundary paradigm and have consistently found preview benefits for TL pseudoword previews compared to SL previews (Johnson et al., 2007; Johnson & Dunne 2012; Pagán et al., 2016; Kirkby et al., 2022). There is consensus that letter position information is encoded flexibly in skilled readers (Davis, 2010; Norris & Kinoshita 2012; Whitney, 2001), however, since the extraction and use of parafoveal information is influenced by individual differences in skilled adult readers (Andrews & Veldre, 2019, Veldre & Andrews 2015a; 2015b; 2016), there may be similar modulation of the transposed letter effect in parafoveal preview.

3.2.4 Individual Differences in Children's Letter Position Encoding

The aim of the current study was to assess how individual differences in cognitive skills may influence how letter position information is processed in skilled adult readers. Relatively little research in this field has focussed on individual differences within adult readers, though changes in the magnitude of transposed letter effects have been observed in relation to children's reading abilities (Pagán et al., 2021; Gómez et al., 2021; Hasenäcker & Schroeder, 2022; Colombo et al., 2019; Ziegler et al., 2014). Pagán et al. (2021) investigated the position of a transposition within a word in a reading-like task and noted that the amount of disruption for a misspelled word with a transposition of the 2nd and 3rd characters was smaller for children with higher reading skills with lower reading skills. Using a lexical decision task, Gómez et al. (2021) also found that in 6th grade Catalan children, individual differences in reading ability, specifically in pseudoword reading (measured by a subtest from PROLEC-R; Cuetos et al., 2007) modulated transposed letter effects. Better readers were less likely to confuse TL pseudowords (mohter) with the real base word (mother) than less skilled readers. Negligible differences were associated with word-reading and perceptual processing speed (measured by a symbol search subset of the Wechsler Intelligence Scale for Children; WISC; Wechsler, 2001).

Similarly, Hasenäcker and Schroeder (2022) found that children's orthographic knowledge (a composite score calculated using a principal components analysis of scores on spelling, vocabulary and a word-reading to nonword-reading difference score[7]) modulated transposed letter effects within grades in a longitudinal study of German children from grade 2 to 4. The cost associated with an SL prime was larger for children with higher levels of orthographic knowledge than for those with lower levels of orthographic knowledge, whereas there was no significant cost for a TL prime at any level of orthographic knowledge. Importantly, Hasenäcker and Schroeder (2022) noted that the modulation associated with orthographic knowledge was similar to the modulation observed across grades. They suggested that developmental changes in letter identity encoding are driven by increasing orthographic knowledge in children, for which grade is a good proxy, but that letter position encoding for words is already fairly flexible in early reading development. These investigations demonstrate that letter position encoding becomes more flexible as reading skills improve (Pagán et al., 2021; Gómez et al., 2021; Hasenäcker & Schroeder, 2022). However, it remains unclear whether differences within skilled adult readers remain in relation to individual differences or whether letter position encoding is stable in this population. Next, we consider a model of visual word recognition that discusses letter position encoding in relation to children's reading development, before discussing how it may relate to individual differences in skilled adult readers.

---

[7] Spelling was measured by the Hamburger Schreibprobe (HSP; May, 2002). Nonword-reading and word-reading were measured by the Salzburger Lese- und Rechtschreibtest (SLRT-II; Moll & Landerl, 2010). Nonword-reading scores were subtracted from word-reading scores. Vocabulary was tested using the CFT 20-R (Weiß, 2006).

3.2.5 Orthographic Processing during Reading Development

The Multiple-route model (Grainger & Ziegler, 2011) suggests that the precise positions of letters are important whilst children decode written words phonologically, translating letters to sounds. As children develop reading skills they rely less on this process, and begin to use orthographic processes, bypassing the need to directly convert letters to sounds. The model includes two orthographic routes to achieve this: a fine-grained route where the coding of letter sequences is location-specific, and a coarse-grained route that uses non-continuous-location-invariant bigrams. For example, the word FARM can be coded by the bigrams FA, FR, FM, AR, AM, RM. Therefore, according to this model, an increased reliance on the coarse-grained route as reading skills develop leads to more flexibility in letter position encoding.

Though this model focuses on children's reading development rather than individual differences in adults, some predictions can be adapted for the current study. If skilled readers continue to rely on a coarse-grained route to orthographic decoding, the impact of a transposed letter preview may be stable across skilled adult readers, reflecting a maturation of letter position encoding flexibility. However, given that many cognitive skills remain variable in skilled adults (Martino & Hoffman 2002; Jackson 2005; Welcome et al., 2010), individual differences in cognitive skills may continue to predict differences in the flexibility of letter position encoding once skilled reading is achieved. Similar to differences in the magnitude of the transposed letter effect seen in developing children related to reading ability (Hasenäcker & Schroeder, 2022; Pagán et al., 2021), the effect of a transposed letter may be modulated by individual differences in adults. Extraction and use of this information during parafoveal processing may be greater for adults with better reading and spelling abilities as observed by Veldre and Andrews (2015a; 2015b).

3.2.6 Individual Differences in Skilled Adult Readers

A few studies have investigated a range of individual differences within adult readers' orthographic processing, though, to our knowledge, none have specifically investigated transposed letter effects in this way during reading. Andrews and Lo (2012) used masked priming to investigate individual differences in reading ability, spelling and vocabulary[8] in adult readers. They found that high reading, spelling, and vocabulary skills (calculated as a composite score based on shared variance) was associated with stronger facilitation from an orthographic

---

[8] Measures of written language proficiency were used to assess reading and spelling: Reading Comprehension, Reading Speed, Spelling Dictation, Spelling Recognition and Vocabulary (Andrews & Hersch, 2010).

nonword prime (different from the target in any single letter position). Welcome and Trammel (2017)[9] found comparable patterns associated with phonemic decoding efficiency scores (a subtest of the Test of Word Read Efficiency; Torgesen et al. 1999) where adults with lower scores showed a general benefit of orthographic relatedness (both pronounceable and unpronounceable anagram primes were facilitatory for both word and non-word targets). Adults with higher scores benefitted only in conditions where pronounceable primes were used for word targets. Though these studies did not investigate effects of TL nonword primes specifically, they suggest that differences in orthographic priming may occur in relation to individual differences in these skills. In an investigation of individual differences in masked form priming Adelman et al. (2014) found that those with strong spelling abilities and large vocabularies had faster response times and were less susceptible to priming in general than less skilled spellers and those with smaller vocabularies. This is similar to patterns seen in relation to reading and spelling abilities in parafoveal preview benefit (Andrews & Veldre, 2019, Veldre & Andrews 2015a; 2015b; 2016).

The current study utilised a sentence reading task with an invisible boundary paradigm to explore transposed letter effects in comparison to a larger battery of cognitive tasks. Consistent with previous evidence using masked priming and parafoveal preview paradigms, if the letter position encoding mechanism varies in adult readers, we predicted that spelling and word naming scores would modulate the size of the transposed letter effect during parafoveal processing. There may also be other cognitive skills that play a role, for example, Kuperman and Van Dyke (2011) found that individual differences in rapid automatized naming (RAN) and word identification were the two most reliable measures when predicting eye movements during reading when assessing a large test battery.

The aim of the current study was to investigate whether letter position encoding during parafoveal processing matures in skilled adult readers, or whether individual differences in a range of cognitive skills influence the parafoveal processing of a TL nonword preview in comparison to a SL nonword preview. The following tests were included. First, two commonly used reading ability tests that differ in subtest components were selected; the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005), which features reading comprehension, word reading and pseudoword decoding; and the Nelson Denny Reading Test (NDRT; Brown et al., 1993), which includes a measure of vocabulary and reading comprehension. Since word reading

---

[9] This study measured Verbal IQ (PPVT-4; Dunn and Dunn 2007) ; Sight word efficiency (TOWRE SWE; Torgesen et al. 1999); Phonemic decoding efficiency (nonword reading) (TOWRE PDE; Torgesen et al. 1999); Exposure to print (author recognition test; Acheson et al., 2008); Orthographic choice task (based on Olson et al. 1994; Wordlikeness task (based on Cunningham et al. 2001); Adult Reading History Questionnaire (ARHQ; Lefly and Pennington 2000).

and pseudoword decoding are measures seen in previous investigations of individual differences in children (Hasenäcker & Schroeder, 2022) and adults (Welcome & Trammel, 2017), and other researchers have often used the NDRT as a measure of reading ability (e.g., Veldre & Andrews, 2015a; 2015b) both were included to appropriately assess the literature. In addition, we wanted to further investigate findings from Lee, Godwin et al. (2023) which suggested that the comprehension subtests in these composites measure different aspects of reading ability.

Other tests included spelling (Andrews & Hersch, 2010), print exposure (Author Recognition; Acheson et al., 2008) and vocabulary knowledge (LexTALE; Lemhöfer & Broersma, 2012) which are good proxies for lexical quality (Lexical Quality Hypothesis; Perfetti, 1992; 2007; Perfetti & Hart, 2002). The quality of an individual's lexical representations have been suggested to influence in the amount of information that can be extracted and used in parafoveal preview (Andrews & Veldre, 2019).

Significant differences in RAN scores have previously been found between dyslexic and non-dyslexic adult readers by Kirkby et al. (2022), who also found differences in the size of the transposed letter effect between these groups. Non-dyslexic adults performed significantly more quickly on alphanumeric RAN tasks than dyslexic adults and displayed larger differences between SL and TL previews in a similar reading experiment using an invisible boundary, therefore we included a measure of alphanumeric RAN (Denkla & Rudel, 1974) in the current test battery. Finally, a backwards digit span test was included as a measure of working memory capacity (Gathercole et al., 2004).

3.3 Method

3.3.1 Participants

Participants were 100 students and staff from the University of Southampton (88 Females, *mean age*= 19.88 *range* = 18-40). Participants were all native English speakers with normal or corrected to normal vision and no known reading difficulties. Participants received course credits or £25. A power analysis was conducted based on data from the first 10 participants with simulations in R (simR, Green & MacLeod, 2016). A power curve revealed that a sample size of 78 participants would be needed to achieve 80 % statistical power for our analyses. Data collection therefore exceeded this target.

3.3.2 Apparatus

The sentences were presented on a 21" CRT monitor, with a refresh rate of 120 *hz* and a resolution of 1024 x 768, interfaced with a PC at a viewing distance of 60 cm. Sentences were presented in black, size 14, Courier New font on a grey background; three characters equated to approximately 1° of visual angle. Although reading was binocular, eye movements were recorded only from the right eye, using an EyeLink 1000 tracker (S.R. Research Ltd.), with forehead and chin rests in order to minimize head movements. The spatial resolution of the eye tracker was 0.05°, and the sampling rate was 1000 hz.

Participants completed most of the tests and questionnaires during the study on a 14-inch Dell Laptop Computer. Such tests were administered via an online web browser running Qualtrics. Participants were required to respond using a variety of mouse responses and keyboard answers, and response times for timed elements were recorded via a timed mouse click integrated within Qualtrics. A computerised backwards digit span test was administered using Inquisit on a 19-inch DELL monitor (1024 × 768-pixel resolution). During Wechsler Individual Achievement Test (WIAT-II UK; Wechsler, 2005) Reading Subtests researchers used the testing flip pad, scoring sheets and word/pseudoword cards included in the test pack.

3.3.3 Materials

Sixty experimental sentences containing 6-letter target words were partially adapted from Pagán et al. (2016). Target words (nouns or adjectives) were bisyllabic with a CVC structure for the initial trigram, which was always within the same syllabic unit (e.g., *monkey*). Target words were embedded into neutral sentence frames and were rated on a scale of 0 (very unnatural to read) to 100 (very natural) (M = 75.37, SD = 8.48) by 18 participants who did not take part in the main experiment. Three preview conditions were generated for each target word; an identity (ID) condition, in which the preview of the target word was spelled correctly (e.g., *monkey*); a transposed letter (TL, e.g., *mnokey*) condition, where a preview was a nonword with the second and third letters transposed; or a substituted letter (SL, e.g., *mrekey*) condition, where the preview was a nonword with the second and the third letters substituted. It has been noted that word-initial letters are especially important for word identification for both children and adults. White et al. (2008) found that readers were more disrupted by transpositions of external letters (at the beginning e.g., rpoblem or end e.g., problem of a word) than internal letters (e.g., problem/probelm) in a sentence reading study. They found that the greatest disruption to reading was seen in word-initial letter transpositions. For this reason, only internal letters were manipulated in the current experiment. Bigram frequencies for TL previews (M = 109.63, SD =

165.98) and SL previews (*M* = 101.25, *SD* = 150.93) were matched (*t* (118) = 0.77, *p* = .443). None of the target transpositions or substitutions produced real words and all were orthographically illegal.

The three counterbalanced lists were presented within the eye tracking experiment. Participants were randomly assigned to one of these conditions and all read 5 practice sentences followed by 60 experimental sentences (20 per condition). The sentences occupied one line on the screen and the target always appeared in the middle of the sentence. Sentence order was randomised for each participant. Comprehension questions were included following 1/3 of the experimental sentences to encourage reading for comprehension.

Participants were asked two questions about their reading behaviour including "How often do you read for work?" and "How often do you read for leisure?". Next, participants' reading and cognitive skills were assessed by the following tests:

3.3.3.1 Reading Ability Tests

Nelson Denny Reading Test (Brown et al., 1993). Participants completed a vocabulary task, where participants were asked to fill a blank space within a sentence with the most appropriate word. Single words were then presented, and participants were given multiple choices to select appropriate definitions. Next participants completed a reading comprehension task. Participants silently read passages presented on a screen, before answering comprehension questions that were presented below the passages (on the same screen). Participants were asked to record the line they had reached after 1 minute of reading on the first passage. The test was stopped after 10 minutes and answers were recorded for all questions they had answered in this time.

Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005). This included a word reading subtest, where participants were asked to read a list of real words aloud from a sheet of paper which increased in difficulty. The experimenter marked participants' pronunciation accuracy and testing was stopped when the participant made six sequential errors. Next, participants were asked to read a list of orthographically legal nonwords aloud from a sheet of paper (e.g., "flimp") in a pseudoword decoding subtest. The experimenter recorded the participants' pronunciation accuracy and testing was stopped after six sequential errors were made. Participants then completed the reading comprehension subtest, where they read passages (short fictional stories, informational text, advertisements, and how-to passages) aloud or silently before answering literal and inferential comprehension questions orally when asked by the experimenter. All subtests were combined to give an overall score for the whole test. Scores were normed according to test instructions and percentile scores were used in analyses.

3.3.3.2 Vocabulary Knowledge

LexTALE (Lemhöfer & Broersma, 2012). Participants were asked to indicate whether a word presented on screen was a real English word or a pseudoword. There was no time limit for this task.

3.3.3.3 Spelling

Spelling (Andrews & Hersch, 2010). Spelling dictation featured playback of 20 recorded key words. Participants then were asked to write down the correct word spellings. The words were also presented within sentences. Spelling recognition was made up of a list of 88 correctly and incorrectly spelled words. Participants had to select the incorrectly spelled words.

3.3.3.4 Print Exposure

Author Recognition Test (Acheson et al., 2008). A list of real authors and foil names were presented, and participants were asked to identify which were the real ones. Participants were informed that the list featured some foil names.

3.3.3.5 Rapid Automatized Naming

Alphanumeric RAN (Denkla & Rudel, 1974). A randomized series of letters or numbers in a 5x5 grid were presented onscreen. Time taken for participants to name the characters was recorded and the sum of the two conditions was used in analyses.

3.3.3.6 Working Memory Capacity

Digit Span Backwards (Gathercole et al., 2004). Digits were presented in sequences of increasing lengths. Participants were asked to recall them first in the same order as they were presented, and then in backwards order. The length of the longest backwards sequence recalled correctly was recorded for each participant.

3.3.4 Design and Procedure

Testing involved two sessions on different days. During the first session, participants were given an information sheet and consent form and completed the eye tracking task followed by experimenter administered WIAT-II and RAN tasks. For the eye tracking task, participants were asked to sit comfortably at the computer, resting their chin on a chinrest and were then guided through the set up and calibration of the eye tracker by the researcher. Participants were then required to direct their gaze to a fixation cross presented on the left of the screen. When ready, sentences were presented following the fixation cross. Participants were asked to read the sentences and answer questions presented on the screen using the keyboard to respond to ensure they were reading for comprehension. Participants could take breaks when needed.

During a second session participants took part in a separate eye tracking task (unrelated to the current study) and completed the reading comprehension and vocabulary subtests of the NDRT, LexTALE task, spelling dictation and spelling recognition tasks, Author Recognition test, RAN tasks and the backwards digit span task in a randomised order.

3.4 Results

There were two stages of analysis. First, a principal components analysis was conducted to determine which cognitive tests shared variance and loaded together. Subsequently, factors extracted via PCA and tests that fell outside of these identified components were used to model eye movement measures (first fixation durations (FFD), single fixation durations (SFD), gaze durations (GD) and go past times).

3.4.1 Individual Differences Tests

Overall WIAT-II scores were calculated using the age-adjusted scoring materials provided, which resulted in numerical scores as well as categorical assessments of reading proficiency (Borderline/ Low Average/ Average/ High Average/ Superior). Overall scores for the NDRT were calculated as an average of scores on the vocabulary and comprehension subtests. Overall spelling scores were calculated using an average of the spelling recognition and spelling dictation subtests (see Andrews & Hersch, 2010). The highest score on the backward version of the digit span was taken as an overall score (see Gathercole et al., 2004). For examining tests as single predictors of eye movements, in line with previous research, overall composite test scores were used. However, it is more appropriate to consider the cognitive skills that make up reading ability tests separately for PCA, subtests from the NDRT and WIAT-II reading ability tests were included from this point.

Participants met criteria for exclusion if their overall performance on multiple tasks was very poor, or if standardised reading assessments identified a potentially undiagnosed reading disorder. Potential outliers scoring very low on some tasks were assessed in terms of their performance on other tasks and were always found to score above average on standardised reading assessments (WIAT-II and NDRT) and therefore did not meet criteria for exclusion. Very high scores were retained since it is plausible to find individuals who are very high scoring on cognitive tasks within skilled reader populations. No participants were removed from the dataset and all 100 datasets were considered suitable for the current analyses. Descriptive statistics based on scores for each test are summarised in Table 3.1 below. Individual difference test scores were then centred to allow comparisons to be made between measures. Correlations for all test scores are presented in Table 3.2. Tests were moderately positively correlated, except for RAN and Digit span tests which were not (highly) correlated with other measures.

Table 3.1

Descriptive Statistics for Tests and Subtests (ms) (Raw Scores)

	Min	Max	Mean	*SD*
NDRT Total	71.00	147.00	118.66	16.59
NDRT Comprehension	22.00	74.00	57.28	9.94
NDRT Vocabulary	28.00	77.00	61.38	10.35
WAIT Total	84.00	134.00	115.73	9.98
WIAT-II Comprehension	71.00	124.00	109.73	10.00
WIAT-II Pseudoword Decoding	88.00	122.00	108.01	6.58
WIAT-II Word Reading	92.00	121.00	115.18	5.55
Spelling	13.00	63.00	37.54	11.28
Author Recognition Test	-6.00	34.00	6.63	6.15
LexTALE	64.55	100.00	90.10	7.14
Digit Span test	3.27	8.75	5.87	1.20
Rapid Automatized Naming	28.79	63.66	39.87	7.39
NDRT Words Per Minute	80.00	610.00	251.43	85.10

Table 3.2

Correlations between Individual Differences Subtests

	NDRT Comp	NDRT Vocab	WIAT-II Comp	WIAT-II Pseudoword	WIAT-II Word Reading	Spelling	ART	LexTALE	Digit Span
NDRT Vocabulary	0.34**								
WIAT-II Comprehension	0.15	0.60***							
WIAT-II Pseudoword Decoding	0.19	0.39***	0.36***						
WIAT-II Word Reading	0.29**	0.54***	0.47***	0.43***					
Spelling	0.31**	0.60***	0.41***	0.55***	0.40***				
Author Recognition Test	0.19	0.63***	0.36***	0.18	0.25*	0.50***			
LexTALE	0.31**	0.70***	0.47***	0.41***	0.45***	0.57***	0.53***		
Digit Span	0.12	0.14	0.23*	0.13	-0.04	0.25*	0.24*	0.18	
RAN	-0.07	0.09	-0.02	-0.03	0.01	-0.07	-0.15	0.00	-0.13

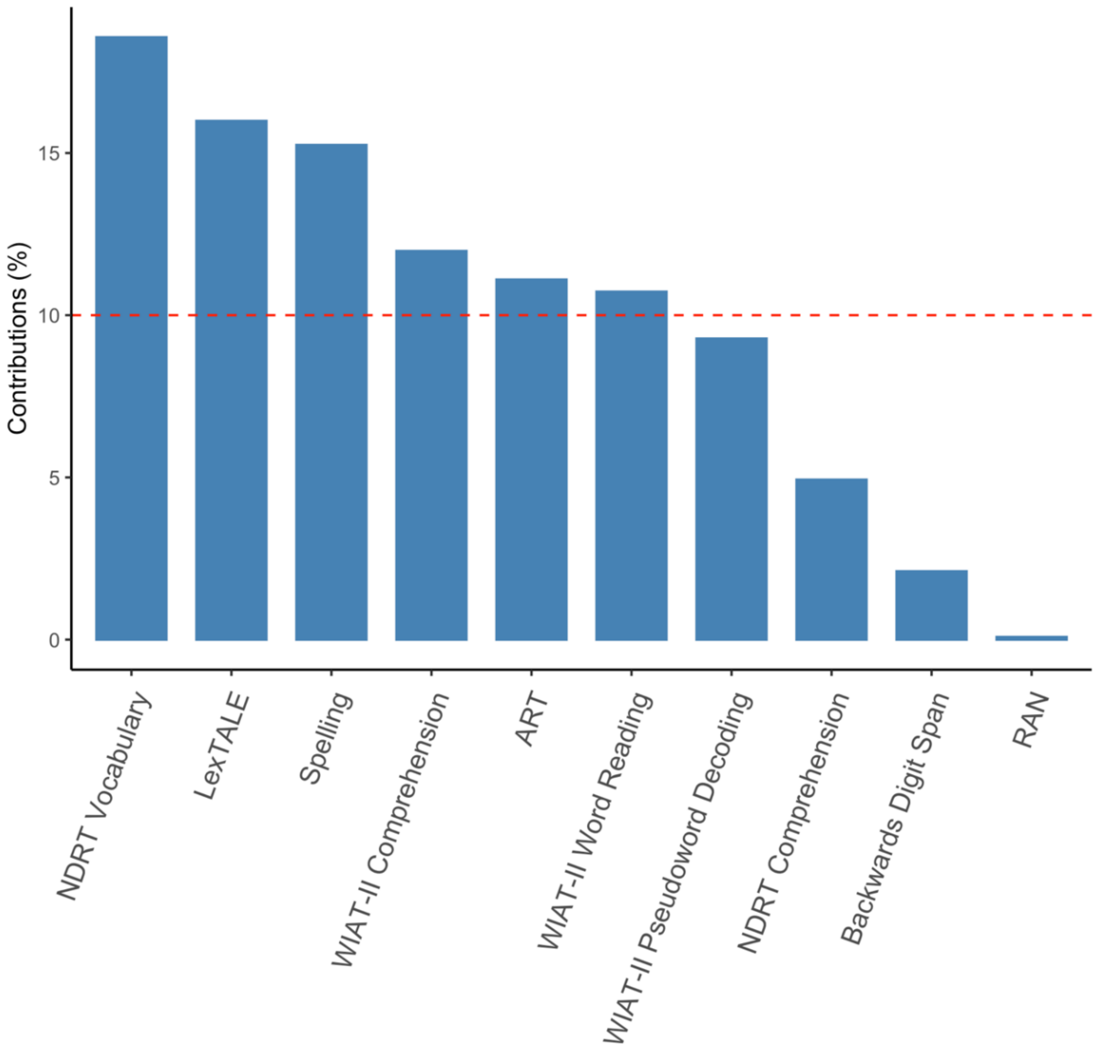
Note. Significance is denoted by * < .05, ** < .01, *** < .001.

3.4.1.1 Principal Components Analysis

A PCA was conducted to identify which tests loaded together to reduce the number of dimensions used in further analysis. NDRT and WIAT-II reading ability test scores are composite scores based on two or more subtests. For the remainder of this paper, subtests that comprise the WIAT-II (comprehension, word reading and pseudoword decoding) and NDRT (comprehension and vocabulary) were considered separately to assess each cognitive skill included within them. A single component was retained after a parallel analysis (Horn, 1965). Parallel analysis calculates adjusted eigenvalues based on random noise expected using a simulated parallel dataset. According to this method, components that fall above the mean of the random eigenvalues should be retained. The component that met this criterion contributed 40.70 % of variance in our data. Tests loadings on a component were considered important if contributions exceeded 10 % (expected average contribution calculated from 1/number of variables = 1/10). In order of magnitude, the extracted component was explained by NDRT Vocabulary, LexTALE, Spelling, WIAT-II Comprehension, ART, and the word reading subtest of the WIAT-II (see Figure 3.1 below). We suggest that these tests are related in their assessment of lexical proficiency. The remaining tests fell outside of this component and were considered separately in subsequent analyses. We note that the comprehension subtests (NDRT and WIAT-II) were not found to load on the same component and will return to this in the Discussion.

Figure 3.1

Individual Contributions of Each Individual Differences Measure on PC1



Note. The dotted line represents the expected average contribution (10%). A contribution above this line is considered important in explaining the component.

3.4.2 Eye Tracking Analyses

Comprehension accuracy across trials was high (M = 98.82%, range = 85 - 100%), indicating that participants were reading for comprehension. Trials with blinks on the target word or featuring tracking loss were removed prior to analysis. Fixations shorter than 80 *ms* made within one character of a previous or subsequent fixation were merged and fixation durations outside of 80 *ms* to 800 *ms* were removed. Trials that involved fewer than 3 fixations across a sentence were also removed. Three trials were removed due to excessive blinking or tracker loss. Data were checked for instances where the boundary change was triggered by a saccade that subsequently landed on the preceding word, where a display changed occurred during a fixation on a pre-target word, or where the display changed had not completed until after 10 *ms* into the fixation on the target word. These trials were then removed (9.82 %). Trials where the target word was skipped (3.06 %) were also removed. Given the low skipping rates, word skipping was not analysed. Outliers were then removed for each dependent variable; first fixation durations (FFD; the duration (*ms*) of the first fixation on a target word), single fixation durations (SFD; the duration (*ms*) of the fixation on a target word when it was only fixated exactly once), gaze durations (GD; the sum of all fixations (*ms*) made on a target word before moving from it in the first pass), and go past times (GOPAST; the sum of all fixations (*ms*) made on a target word before moving on to a later portion of the sentence, this includes regressions to previous portions of the sentence). Data falling outside of 2.5 standard deviations from the mean for each participant within a condition (high and low frequency) were removed (SFD; 1.44 %, FFD; 1.68 % GD; 1.91 %, go past times; 1.72 %). Descriptive statistics based on participant means for these measures are displayed in Table 3.3.

Table 3.3

Descriptive Statistics for Eye Movement Measures (ms)

	Condition	Min	Max	Mean	*SD*
SFD	ID	127.67	310.92	221.79	33.47
	TL	154.67	330.17	229.53	38.05
	SL	160.25	359.80	256.31	38.81
FFD	ID	149.25	306.73	220.46	32.47
	TL	149.14	330.17	227.43	36.08
	SL	160.25	349.15	246.04	33.53
GD	ID	159.67	402.53	250.24	47.60
	TL	162.17	424.50	256.50	54.85
	SL	160.25	448.64	290.20	52.07
GOPAST	ID	164.86	402.53	264.09	54.18
	TL	165.86	454.38	270.76	60.84
	SL	160.25	448.64	303.02	54.31

Note. Means and *SDs* were calculated based on participant means per condition.

3.4.2.1 Generalized Linear Mixed Models

All eye movement measures were analysed using the lme4 package (version 1.1-31; Bates et al., 2015) in R (version 4.2.2; R Core Team, 2022). For all Generalized Linear Mixed Models (GLMM), a Gamma distribution with identity link was used, following guidance for analysing skewed reaction time data without transformation (see Lo & Andrews, 2015) with participants and items as random factors.

Models were trimmed following the procedure described in Dirix and Duyck (2017). Models started with all fixed effects of interest: contrasts for the orthographic preview conditions (ID - TL, and TL - SL), PC1, the NDRT comprehension, WIAT-II pseudoword decoding, RAN and backwards digit span scores. Two-way interactions for each cognitive test and the orthographic preview conditions were also included. In addition, trial number and launch site were included as fixed factors to account for their potential influence on the data. Random factors were intercept only. This was the starting model. Next, non-significant interactions and afterwards fixed effects were sequentially checked to see if they could be removed without reducing model fit, starting with the effect with the largest *p* value. Model comparison Chi-square tests assessed whether removal of such effects impacted model fit. When models reached a point at which no further trimming was possible because all remaining fixed effects were significant, and therefore necessary to be retained, or if the removal of a non-significant effect would reduce model fit, we began building up the random effects structure to reach the maximal model (as is optimal for model analysis according to Barr et al., 2013). Random effects were forward fitted, with slopes for participant and item factors added. The order was as follows: orthographic preview condition, individual differences tests, launch site and finally trial number. Slopes were retained if the model converged and their addition improved the model fit. Finally, when the largest random structure was achieved, any non-significant fixed effects were again checked sequentially to see if trimming them would not reduce model fit. The final random effects structure for each model was as follows. The SFD model included trial number as a slope for subjects, and PC1 scores and trial number as slopes for stimuli. The FFD model included condition (ID/TL/SL) as slopes for both subjects and stimuli. The GD model included condition as a slope for subjects and intercept only for stimuli. The GOPAST model included condition as a slope for subjects and saccade launch site as a slope for stimuli.

Results, displayed in Table 3.4 and Table 3.5, showed that there were small differences between ID and TL previews in FFDs (7.52 *ms*), SFDs (5.24 *ms*), GDs (8.16 *ms*, though this was only marginally significant) and go past times (10.02 *ms*). SL previews resulted in inflated FFD (17.24 *ms*), SFD (27.54 *ms*), GD (33.49 *ms*) and go past times (32.70 *ms*) on the target word compared to TL previews. When saccades were launched from positions close to the target, SFD, GD and go past times were shorter than when saccades were launched from a greater distance. Later trials in the experiment were associated with shorter SFDs, GDs and go past times compared to earlier trials, but no differences were found for FFDs on the target word. Participants who scored highly in tests associated with PC1 generally displayed shorter FFD, SFD, GD and go past times. However, these scores were not found to interact with any orthographic preview condition. No other main effects were found associated with individual differences measures. High scores on the backwards digit span task predicted significantly different SFDs for targets following TL previews compared to ID previews, however as shown in Figure 3.2, this interaction is completely encapsulated by 95 % confidence intervals for ID and TL conditions on these measures, so we will not consider these further. No significant interactions were observed for any other individual differences measures, letter position encoding flexibility remained stable across this population of skilled adult readers.

Table 3.4

GLMMS with Multiple Test Predictors to Predict First Fixation Durations (ms) and Single Fixation Durations (ms)

	First Fixation Duration				Single Fixation Duration			
	Est	*SE*	*t*	*p*	*Est*	*SE*	*t*	*p*
(Intercept)	233.59	3.89	59.99	<.001 ***	245.67	5.06	48.52	<.001 ***
TL-ID	7.52	3.74	2.01	.044 *	5.24	2.47	2.12	.034*
SL-TL	17.24	4.11	4.19	<.001 ***	27.54	2.65	10.38	<.001 ***
Trial Number	-	-	-	-	0.13	0.11	-1.22	.222
Launch Site	-	-	-	-	2.30	0.53	4.32	<.001 ***
PC1	9.09	3.25	2.80	.005 **	8.45	3.66	2.31	.021 *
Backwards Digit Span	-2.61	3.38	-0.77	0.43	-2.14	3.54	-0.60	0.55
TL-ID*Backwards Digit Span	-5.51	3.05	-1.81	.071	-5.75	2.45	2.35	.019 *
SL-TL*Backward Digit Span	2.90	3.48	0.83	.405	0.99	2.54	0.39	.696

Note. Significance is denoted by * < .05, ** < .01, *** < .001.

Table 3.5

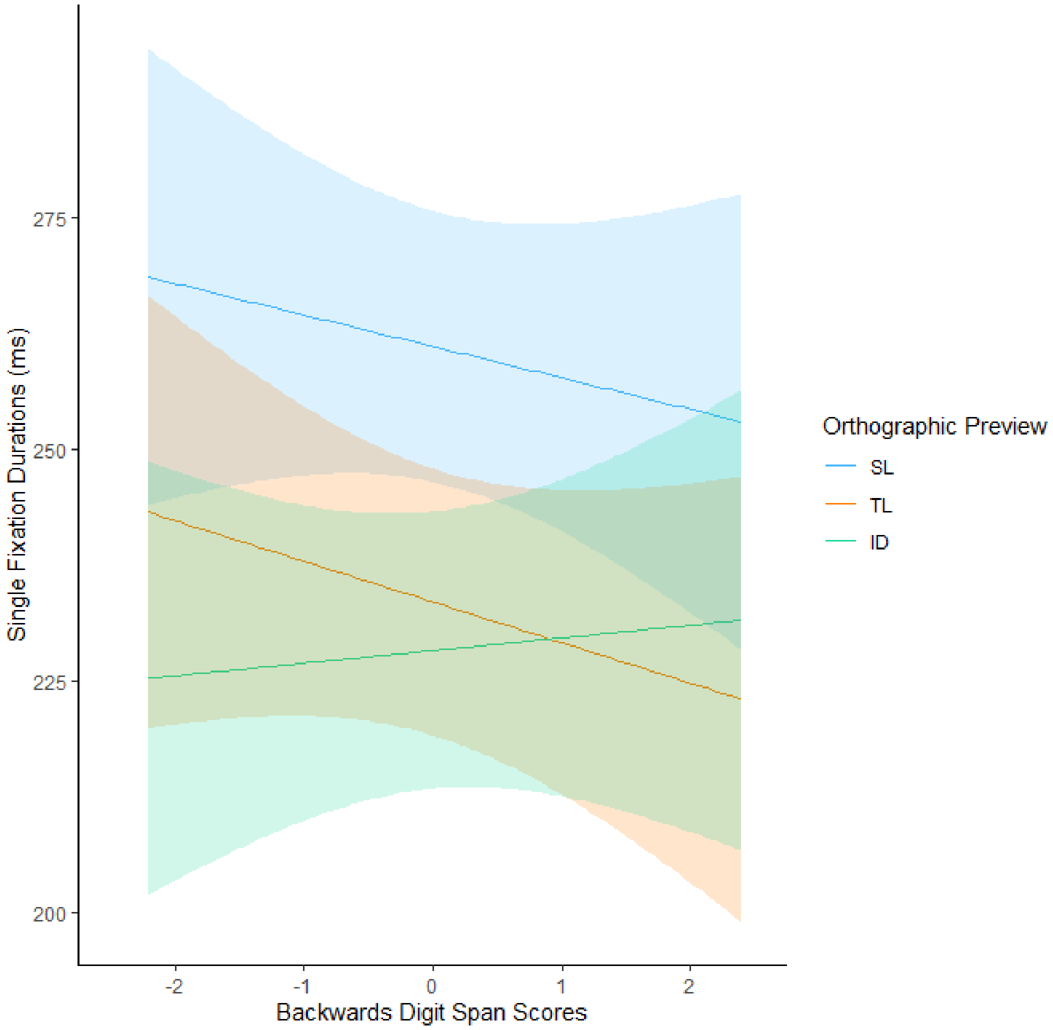
GLMMS with Multiple Test Predictors to Predict Gaze Durations (ms) and Go Past Times (ms)

	Gaze Duration				Go Past Times			
	Est	*SE*	*t*	*p*	*Est*	*SE*	*t*	*p*
(Intercept)	277.38	5.68	48.85	<.001 ***	295.77	5.42	54.52	<.001 ***
TL-ID	8.16	4.33	1.89	.059 .	10.02	4.19	2.39	.017 *
SL-TL	33.49	4.79	6.99	<.001 ***	32.70	4.28	7.64	<.001 ***
Trial Number	-0.17	0.08	-2.24	.025 *	-0.27	0.08	-3.31	.025 *
Launch Site	6.58	0.63	10.49	<.001 ***	7.88	0.67	11.75	<.001 ***
PC1	12.05	4.14	2.91	.004 **	16.35	4.85	3.37	<.001 ***

Note. Significance is denoted by * < .05, ** < .01, *** < .001.

Figure 3.2

The Effect of Orthographic Preview on Single Fixation Durations (ms) Moderated by Backwards Digit Span Scores



Note. Shaded areas represent 95% confidence intervals.

3.5 Discussion

The present study examined whether the flexibility of letter position encoding reaches maturation in skilled adult reader populations or whether it varies in relation to differences in cognitive skills. We first grouped our battery of cognitive skills via overlapping variance by means of a PCA to reduce multicollinearity in models with multiple test predictors. We then analysed the transposed letter effect in relation to these cognitive skills on four eye movement measures (SFD, FFD, GD and go past times).

When saccades were launched from positions close to the target, SFDs, GDs and go past times on the target word were smaller than when saccades were launched from a greater distance. This is in line with previous research (e.g., Pollatsek et al., 1986), in that more information can be gathered from a parafoveal preview during fixations close to the upcoming word. As a result, less time is needed during a subsequent fixation on the upcoming word to identify it. Trials that occurred later in the experiment were associated with shorter SFDs, GDs and go past times than earlier trials indicating readers speeding up somewhat during the experiment, but no differences were found for FFDs.

3.5.1 Transposed Letter Effect

Overall, transposed letter effects were found where SL previews resulted in increased fixation times on target words compared to TL previews. In comparison, there was only a small cost associated with a TL preview in comparison to an ID preview in FFDs, SFDs and go past times, with only a marginally significant difference in GDs, in line with previous findings in reading studies (Rayner, White et al., 2006). These findings are consistent with the idea that in general, skilled readers encode letter position more flexibly than letter identity.

3.5.2 Individual Differences

We adapted predictions based on the Multiple-route model (Grainger & Ziegler, 2011) about developmental differences in letter position encoding to make some predictions about differences that may arise within skilled adult readers. We noted that, if skilled readers generally rely on a coarse-grained route for orthographic decoding there may be a maturation of letter position encoding flexibility in skilled adult readers. This would result in all readers showing similar transposed letter effects with faster identification of target words following TL previews than SL previews. The alternative hypothesis was that only a subset of adult readers may rely more heavily on a coarse-grained route to orthographic decoding leading to larger differences between TL and SL preview conditions. If so, any differences in the transposed letter effect in adults might be observed in relation to individual differences in cognitive skills, as seen in developing readers (Pagán et al., 2021; Hasenäcker & Schroeder, 2022).

We found no systematic differences in letter position encoding related to individual differences in cognitive skills in our main analyses, suggesting that the flexibility of letter position encoding in average-to-skilled adult readers remains fairly stable once reading has developed. At least this is the case for the adults in our sample who were classed as “average” to “superior” on the standardised WIAT-II reading ability. To fully investigate the compatibility of our results with previous literature about individual differences in parafoveal preview benefit (Andrews & Veldre, 2019, Veldre & Andrews 2015a; 2015b; 2016) and differences in transposed letter effects in parafoveal preview in children's reading ability (Pagán et al., 2021), we also ran analyses in which models only included single test predictors for spelling and reading ability. These analyses more closely reflect the analyses reported in these previous studies. The results of these analyses (available online at https://osf.io/b2rdm/?view_only=cf37e55f5c804a98bf7801a8c903d5f3) indicate that the only significant predictor of more flexible letter position encoding was the Nelson Denny reading test (reading comprehension and vocabulary composite score), where higher scores were associated with larger differences between SL and TL preview conditions. However, this observation was not stable across other eye movement measures as it was only observed in single fixation durations. Since this finding was restricted to a single measure and was not significantly predicted by another reading ability measure (WIAT-II reading test), our conclusion remains that the flexibility of letter position encoding in adult skilled readers reaches maturation (or near maturation) and varies very little in relation to cognitive skills.

We suggest that differences in the flexibility of letter position encoding in adults related to individual differences may appear where differences in cognitive skills are larger, for example where samples include struggling or developing readers. Hasenäcker and Schroeder (2022) demonstrated that in a longitudinal study of children, the size of TL effect (in a lexical decision task) was modulated by the readers' orthographic knowledge. We suspect that if there are important differences in the magnitude of the TL effect that occur in relation to reading skill in adults, these differences will be better seen across the entire population of poor to very skilled readers, as opposed to in our sample of average-to-very-skilled. Additionally, the target words in our sentences were not complex, and all skilled readers are likely to have been very successful at identifying them using a coarse-grained word reading strategy (Grainger & Ziegler, 2011). Future research may consider using more complex words to increase the power to discriminate between skilled readers in their use of phonological or orthographic word identification processes of less familiar words.

Similar to our previous investigations of individual differences in skilled adult readers (Lee, Godwin et al., 2023), a PCA in the current analyses grouped together skills that have previously been linked to lexical quality (vocabulary, spelling ability and reading experience; Perfetti, 2007; Perfetti & Hart, 2002; Andrews, 2015). We found that these skills were related to general sentence processing with faster fixations and shorter reading times associated with high levels of reading skill, in line with previous research (Lee, Godwin et al., 2023). This is consistent with the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002) in that words were identified more quickly by participants with higher quality lexical representations indexed by high scores on this component, and is in line with previous research about skills associated with lexical quality (Veldre et al., 2017; Rayner, 1998).

Importantly, we note that in our PCA, the comprehension subtest of the WIAT-II was associated with lexical proficiency (PC1) whereas the comprehension subtest of the NDRT was distinct and that these tests were weakly correlated (r = 0.15) (Dancey & Reidy, 2007). Although both are standardised measures of comprehension, it appears that they measure distinct underlying constructs, (as also noted in a previous investigation by Lee, Godwin et al., 2023). Differences in constructs measured by tests that share a name provide an example of the Jingle fallacy (Thorndike, 1904). Therefore, reading comprehension measures should be selected carefully for future research.

3.5.3 Conclusion

We conclude that in general skilled adults encode letter position more flexibly than letter identity, with a greater processing cost associated with changes in letter identity than changes in letter position. We observed very few individual differences in the flexibility of skilled readers' letter position encoding, suggesting that letter position encoding reaches maturation (or near maturation) and is fairly stable for skilled adult readers.

Chapter 4 The Jingle fallacy in Comprehension Tests for Reading

Charlotte E. Lee, Hayward J. Godwin, and Denis Drieghe

School of Psychology, University of Southampton, United Kingdom.

4.1 Abstract

The *Jingle fallacy* is the false assumption that instruments which share the same name measure the same underlying construct. In this experiment, we focus on the comprehension subtests of the Nelson Denny Reading Test (NDRT) and the Wechsler Individual Achievement Test (WIAT-II). 91 university students read passages for comprehension whilst their eye movements were recorded. Participants took part in two experimental blocks of which the order was counterbalanced, one with high comprehension demands and one with low comprehension demands. We assumed that tests measuring comprehension would be able to predict differences observed in eye movement patterns as a function of varying comprehension demands. Overall, readers were able to adapt their reading strategy to read more slowly, making more and longer fixations, coupled with shorter saccades when comprehension demands were higher. Within an experimental block, high scorers on the NDRT were able to consistently increase their pace of reading over time for both high and low comprehension demands, whereas low scorers approached a threshold where they could not continue to increase their reading speed or further reduce the number of fixations to read a text, even when comprehension demands were low. Individual differences based on the WIAT-II did not explain similar patterns. The NDRT comprehension test was therefore more predictive of differences in the reading patterns of skilled adult readers in response to comprehension demands than the WIAT-II. Our results revealed that these different comprehension measures should not be used interchangeably, and researchers should be cautious when choosing reading comprehension tests for research.

Keywords: Jingle fallacy, Comprehension Demands, Individual Differences

Authorship Statement

Charlotte Lee conceptualised and created this experiment, performed all analyses, interpreted data and wrote this manuscript for publication.

Denis Drieghe and Hayward J. Godwin provided project supervision and comments on this manuscript.

4.2 Introduction

Reading comprehension is a complex task made up of interactions between the features of a text and the skill and strategies of the reader (Kintsch, 1988; Perfetti & Stafura, 2014; Van Den Broek et al., 1999). The Simple View of Reading (Gough & Tunmer, 1986) describes the basic requirements for reading as the ability to decode and identify words in text by converting graphemes into phonemes combined with the ability to understand information presented orally (language comprehension). However, in the more complex Construction-Integration (CI) model of reading comprehension (Kintsch, 1998), text is represented by a surface structure (semantic representations of words within a text), a textbase (a representation of the explicit meaning of the whole text, coherently integrating each word meaning) (Kintsch & Rawson, 2005), and a situation model (where a reader creates a model of the situation, integrating the explicit meaning of the text with their own world knowledge). For shallow comprehension, a textbase is sufficient, however for deeper understanding a situation model is required. Differences in theoretical conceptualisation of comprehension can result in differences in the underlying mechanisms measured by comprehension tests based upon them. Indeed, inconsistencies in research where skills measured by cognitive tasks are used to predict readers' performance on reading comprehension measures have been suggested to reflect differences in underlying cognitive mechanisms (Cutting & Scarborough, 2006; Keenan et al., 2008; Kendeou et al., 2012; Mézière et al., 2023). The current paper strives to shed some light on the problems that researchers may face when selecting reading comprehension tests, and the direct impact that test selection can have on conclusions based upon them in eye tracking investigations.

In some of our previous eye movement investigations of average-to-very-skilled readers (Lee, Godwin et al., 2023; Lee, Pagán et al., 2023) we found that two often-used reading comprehension subtests from standardised reading ability measures failed to load together in a principal components analysis and were only weakly correlated (r = 0.21, (Lee, Godwin et al., 2023), r = 0.15 (Lee, Pagán et al., 2023)). These subtests were from the Wechsler Individual Achievement Test (WIAT-II UK; Wechsler, 2005) and the Nelson Denny Reading Test (NDRT; Brown et al, 1993). We concluded that these comprehension tests might be assessing different underlying skills. Since these measures are both named 'reading comprehension', this would present a clear example of Thorndike's *Jingle fallacy*: that is, the misleading assumption that two measures assess a single underlying construct because they share the same name (Thorndike, 1904). Although not uncommon in psychological research, where a variety of tests are available to assess common constructs, problems when selecting and reporting appropriate measures can lead to questionable research practices when used for scientific purposes (Flake & Fried, 2020).

The aim of the current paper is to extend our previous investigations to directly test the differences between the two tests by using them to predict differences in eye movement patterns reflective of different comprehension demands and to further highlight the pitfalls of comparing research that uses either test for this area of research.

We start by discussing some qualitative differences in the format of the two comprehension tests that may provide some insight into the underlying constructs that are being tapped into by each one. First, the NDRT exclusively features non-fiction passages whereas the WIAT-II features more varied text formats, with some fiction and non-fiction passages as well as single sentences. Previous studies have noted that differences in the format of reading materials (sentences vs paragraphs, Radach et al., 2008; fiction vs non-fiction, e.g., Best et al., 2008; McNamara et al., 2011; Zwaan, 1994) can impact reading behaviour as reflected in eye movement measures. Reading times are longer and rereading is more common for sentences presented within paragraphs than for sentences presented alone, which suggests that text format influences the reading strategy used to comprehend the text (Radach et al., 2008). Best et al. (2008) also showed that comprehension accuracy was higher for narrative texts than expository texts (non-fiction/scientific) and performance on each was predicted by different individual skills. Decoding skills were a key element for successful narrative text comprehension, whereas world knowledge was more important for successful expository text comprehension. While this may suggest that narrative texts included in the WIAT-II where comprehension is suggested to be higher, might be ‘easier’ for skilled readers, it also suggests that a reliance on non-fiction passages in the NDRT may result in greater overlap with general knowledge. This was also suggested by Ready et al. (2013) following work by Coleman et al. (2010) who found that college students could answer the questions on NDRT comprehension tests and achieve a greater-than-chance level of accuracy without actually reading the associated passages. In addition, Eason et al. (2012) concluded that higher-order cognitive skills such as inferencing, planning and organising information were the best predictors of better comprehension when texts and questions were complex, which may indicate cognitive skills that are more likely to be underlying the NDRT comprehension test scores than the WIAT-II scores.

Both tests also feature explicit differences in reading instruction since the WIAT-II includes a combination of silent and oral reading, whereas the NDRT only features silent reading. Reading aloud involves articulating the text as well as the standard process of reading, and evidence from the eye-voice span (the distance between the location of a fixation and the articulated word) demonstrates that oral reading involves additional working memory processes (Laubrock & Kliegl, 2015). Hale et al. (2007) investigated differences in reading aloud and silently and found that for children across grades 4-12 reading comprehension was higher when reading aloud than when reading silently. In addition, some prior research suggests that changing oral and silent reading tasks in comprehension tests may lead to different outcomes, though this has been noted specifically in relation to differences between children with reading difficulties and average readers (García & Cain, 2014). Much less is known about how comprehension changes when adults read aloud. A survey by Duncan and Freeman (2020) reported that, of 529 respondents, 67.5 % said that they read aloud to understand difficult text (though they noted that this was usually only brief). Gambrell and Heathington (1981) reported that 36 % of poor adult readers said that they could read more quickly when reading aloud compared to just 4 % of good readers. It may be that an oral reading task to assess comprehension is less informative for adult readers due to individual differences, though more research is needed to investigate this.

Another notable difference is that testing in the WIAT-II is administered by an experimenter who asks questions aloud to the participant and records their spoken responses on paper, whereas the NDRT is administered independently. This procedural difference could lead to performance anxiety for participants when taking the WIAT-II and may introduce noise into data collected under these conditions. This may be especially important where participants are sometimes asked to read aloud. In contrast it may mean that the NDRT has comparatively less control to determine whether a participant is properly engaging with the task. This aspect highlights another qualitative difference in the administration of the WIAT-II in comparison to the NDRT.

A good comprehension test should be able to predict differences in behaviour between tasks that vary in comprehension demands. Eye movement measures reflect complex cognitive processes active during reading (Liversedge & Findlay, 2000; Rayner; 1998). The current paper therefore investigated global reading strategies for paragraph reading and aims to examine whether the differences we described between the comprehension subtests of the WIAT-II and the NDRT impact their ability to predict eye movement patterns reflecting changes in comprehension demands.

We turn our focus now to individual differences in adult readers' eye movements. Skilled adult readers typically read more quickly, make fewer and shorter fixations, longer saccades and fewer regressions than less skilled readers (Rayner, 1998; Ashby et al., 2005). However, reading skill is not directly related to reading rate, and a speed-accuracy trade-off means that faster reading eventually leads to lower levels of comprehension (Rayner et al., 2016). There is much variability within groups of skilled adult readers with fixation durations varying between approximately 50-600 *ms* and saccade lengths between 1-20 letter spaces (Rayner, 2009; Andrews, 2012). Skilled readers also vary in how they respond to features of a text. Ashby et al.

(2005) found that poor adult readers (identified using NDRT reading comprehension and vocabulary tests) benefitted more from highly constraining sentential contexts compared to skilled readers. Similarly, Bisanz et al. (1992) reported a complex relationship between reading ability and reading times in line with Stanovich's (1980) interactive-compensatory model which stated that poor readers, who had below average bottom-up processing skills, would rely more heavily on contextual cues when they were available. Bisanz et al. (1992) showed that poor readers actually read some sentences more quickly than skilled readers. It has been suggested that some readers might use a 'risky' reading strategy where they read more quickly and make fewer refixations than other readers (O'Regan, 1992).

In addition to the differences observed between readers, intra-individual differences (variability within the same reader) can also influence reading behaviours. It has been well established that task demands can influence the way that readers process a text: skilled readers are able to adjust their reading behaviours (and pace) to the demands of the task (Tinker, 1958) and are able to read thoroughly or superficially when needed (Heller, 1982). Aaronson and Ferres (1984; 1986) noted that skilled readers are more likely to use a 'recall strategy' focussed on structural aspects of a text when a reading task involves direct recall of words/sentences, but when the task involves true/false questions, their focus is driven by the meaning of the text using a 'comprehension strategy'. This research was influential as it gave clear evidence that skilled readers had some autonomy over how deeply they processed a text.

It has been noted that when texts are more difficult, a more 'careful' strategy might be used where, in comparison to a risky reading strategy (O'Regan, 1992), readers tend to make more refixations, have smaller average fixation durations and smaller saccade amplitudes (Inhoff & Radach, 1998). Researchers have investigated whether these strategies can be observed for identical sentences when different comprehension demands are placed upon them. Radach et al. (2008) investigated differences in eye movement behaviours related to the specific reading task as well as different text formats. Participants took part in one of two tasks: comprehension, where participants were asked detailed questions about the text; and a word verification task where participants had to indicate which word had appeared in the sentence from some given options. Radach et al. (2008) also compared eye movement measures within these groups for identical sentences that were either embedded within a passage or were presented alone. Researchers concluded that top-down processes influenced by the task (comprehension vs word identification) and format of the text (sentences vs paragraphs) clearly impacted the eye movement record. Word-viewing times were significantly longer on comprehension tasks and more fixations were made on a word in this task than in the verification task, indicating more careful reading when reading for comprehension. Passages were read more quickly on the first-pass but featured more rereading than sentences.

Similarly, Wotschack and Kliegl (2013) investigated the effect of easy 'verification' questions (after 27 % of sentences) compared to 'hard' comprehension questions about sentence meaning (following 100 % of sentences) and found that the more difficult questions were associated with more careful reading as indicated by more rereading and more regressions. However, they found that accuracy was high in both conditions and questioned the strength of their manipulation. In response, Weiss et al. (2018) aimed for a stronger difficulty manipulation and investigated 'easy' lexical verification questions versus 'difficult' comprehension questions that required resolving some syntactic ambiguity. For example, a sentence containing a subjective relative clause such as *The chef that distracted the waiter sifted the flour onto the counter,* was followed by an easy question: *Did a chef do something?* Or a difficult question: *Did the waiter distract the chef?* They did see differences in accuracy between the difficult (83 %) and easy (97 %) conditions, and also found that participants made more regressions and spent more time rereading texts in the difficult condition but that no disruptions were seen in first pass fixation times. Weiss et al. (2018) concluded that inflated differences happened at the end of passages even when the ambiguity occurred earlier in the sentence. Accuracy was not predicted by the magnitude of the disruption, suggesting that the increased processing time was a 'checking mechanism' rather than additional information processing.

Christianson et al. (2017) reached a conclusion similar to Weiss et al. (2018) in a study that investigated rereading behaviours in garden-path sentences (where an ambiguity in the sentence meaning is revealed fairly late in the sentence e.g. *The babysitter who was purchased a gift card thanked the parents*) vs. local coherence structures (where ambiguities were resolved earlier, e.g. *The parents thanked the babysitter who was purchased a gift card).* They found that rereading behaviours were more consistent with confirmatory rereading (checking) than revisionary rereading (for understanding) because rereading was not consistently predicted by critical regions in the sentence structure, and rereading behaviours were not predictors of offline comprehension accuracy.

Recent investigations have looked more closely at rereading behaviours and have started to examine individual differences in rereading. A study by Andrews and Veldre (2021) investigated 'wrap-up' effects in tasks with different comprehension loads in relation to individual differences in reading proficiency (measured by vocabulary, reading comprehension reading rate (NDRT), spelling dictation and spelling recognition (Andrews & Hersch, 2010)). Wrap-up effects (Just & Carpenter, 1980) are where longer reading times are observed at clause and sentence boundaries, where readers integrate information before moving forward in a text (Aaronson & Scarborough, 1976; Rayner et al., 2000). Wrap-up times have been associated with the goals of the reading task, for example in a study by Stine-Morrow et al. (2001) where differences in wrap-up predicted recall but not comprehension success. Importantly, Andrews and Veldre assessed readers' individual differences in spelling, reading comprehension (NDRT), vocabulary and reading rate alongside manipulating how often comprehension questions occurred (after 25 % of passages or 100 % of passages). They found that comprehension load had little effect on wrap-up, however it did lead to shallower (more risky) reading strategies when comprehension demands were low, with longer passage reading times, more refixations and regressions, but no differences in average fixation times or forward saccade lengths. Andrews and Veldre (2021) found that the better readers (as identified via a composite score of the individual differences measures that have been shown to provide a good assessment of lexical quality; Andrews et al., 2020; Andrews; 2015; Veldre & Andrews, 2015a; 2015b; Perfetti, 2007) generally read passages more quickly, made fewer and shorter fixations, longer forward saccades and marginally fewer regressions than poorer readers. They did not find that reading proficiency composite scores interacted with the effect of comprehension load on eye movement measures but they noted that accurate comprehension was associated with more consistent reading behaviour, where readers did not adjust their reading strategy much in response to comprehension load.

Reading strategies may of course be adapted over time during an experiment. For example, readers may read through early trials more slowly when they have high comprehension demands, until they are familiar with the format of the questions in the experimental block, after which they may adjust their reading rate to speed up processing time. This rate of adaptation may be modulated by individual differences, whereby better comprehenders might be able to increase their reading rate to one that is optimal/preferred more quickly over trials than less skilled comprehenders. Therefore, besides examining the differences between predictions based on two comprehension tests, a second goal of current study was to determine whether individuals alter their reading strategies in response to comprehension demands gradually as trials progress. We were interested to see if individual differences in reading ability predicted differences in the rate of adaptation to different comprehension demands as well as whether discrepancies occurred between the two measures of reading comprehension that we included. Following Radach et al. (2008), identical reading materials were used between conditions in the current study to directly compare the influence of comprehension demands placed on the reader via differences in the difficulty of questions that followed.

We expected high scores on the comprehension tests to predict faster passage reading times as faster sentence reading times were associated with higher scores on the comprehension subtests from the WIAT-II (Wechsler, 2005) and the NDRT (Brown et al., 1993) in Lee, Godwin et al. (2023) and Lee, Pagán et al. (2023). Note however that the format of our experimental materials in the current study (paragraphs) was different in comparison to our previous investigations (sentences). Longer reading times and more rereading have been observed for passages compared to sentences (Radach et al., 2008). Similarly, since comprehension was included as part of the composite measure of reading proficiency in Andrews and Veldre (2021), who found that higher reading proficiency predicted faster passage reading times, shorter average fixation durations, longer forward saccades and a greater number of regressions than low proficiency, we expected similar patterns to emerge for our comprehension scores.

We expected that higher offline comprehension scores would predict faster passage reading times, shorter average fixation durations, longer forward saccades and fewer regressions. Higher comprehension demands were expected to increase the number of fixations and the time that participants spent reading the passages. We anticipated that all readers would adapt their reading strategy to become more efficient (they would make fewer fixations, longer saccades, shorter fixations and read passages more quickly), but that there might be individual differences observed in the rate of adaptation or ceiling levels in saccade lengths that poorer readers could reach, since poorer readers have been shown to have shorter rightwards perceptual spans (in languages read left to right) than better readers (Veldre & Andrews, 2014). Similarly, as poorer readers usually exhibit slower reading times and longer fixations than skilled readers (Rayner, 1998; Ashby et al., 2005; Andrews & Veldre, 2021; Lee, Godwin et al., 2023; Lee, Pagán et al., 2023) we anticipated floor effects for poor readers' minimum passage reading times, fixation durations and the number of fixations. Since the intended population was skilled adult readers, it was likely that accuracy would be high across tasks (as was observed in Andrews & Veldre, 2021, and Wotschack & Kliegl, 2013). Therefore, because comprehension accuracy is often higher for narratives than expository texts (Best et al., 2008), expository passages were used in the current study to maximise the likelihood of variability in accuracy scores[10].

---

[10] We note that the NDRT exclusively uses expository texts to measure comprehension, therefore it may be more similar in format to the passages used in this experiment. As noted by Ready et al. (2013) and Coleman et al. (2010), the NDRT may also feature a high degree of overlap with general knowledge or world knowledge, which has been found to be associated with expository text comprehension. Therefore, it will not be surprising if the NDRT predicts higher comprehension accuracy across conditions, than the WIAT-II, which features more varied reading formats. We also anticipate that the WIAT-II may be more noisy in its predictions due to some performance anxiety induced by the experimenter's presence.

4.3 Method

4.3.1 Participants

Participants were 91 students and staff from the University of Southampton (11 Males, *M* = 20.27 years, range = 18 – 45 years). An additional 9 participants took part in the study, but their data were removed from the final dataset due poor overall accuracy (below 60 % where chance level was 50 %). Participants were all native English speakers with normal or corrected to normal vision and no known reading difficulties. Participants received course credits or £25 for completing the study. A power analysis using simulations in R (simR, Green & MacLeod, 2016) based on data from the first 10 participants revealed that a sample size of at least 35 participants would be necessary to generate 80 % statistical power for our analyses. The sample size collected for this study exceeded this target.

4.3.2 Apparatus

Paragraphs and questions were presented on a 21-inch CRT monitor, with a refresh rate of 120 Hz and a resolution of 1024 x 768 at a viewing distance of 60 cm. Passages were presented in Courier New, size 14 font on a grey background; three characters equated to about 1° of visual angle. Although reading was binocular, eye movements were recorded from the right eye only using an EyeLink 1000 tracker (S.R. Research Ltd.). Forehead and chin rests were used to minimize head movements. The spatial resolution of the eye tracker was 0.05°, and the sampling rate was 1000 hz.

Participants used a 14-inch Dell Laptop Computer to complete the NDRT comprehension test administered using an online web browser running Qualtrics. Participants were required to select answers using a mouse. During WIAT-II comprehension test researchers used the testing flip pad and scoring sheets included in the test pack.

4.3.3 Materials

Forty experimental paragraphs (*M* = 138.33 words, *SD* = 19. 28) were adapted from freely available online practice comprehension tests ("Determine the main idea", 2021). Two conditions were created for each paragraph, one with low comprehension demands where participants were asked *"What is the passage about?"* and were given two short options that consisted of a word or phrase (e.g., *Synaesthesia/Claustrophobia*). One option was directly related to the passage and the other was unrelated. In the high comprehension demands condition participants were asked, *"What is the main idea of the passage?"* and two longer and more detailed options were presented from which participants were asked to select an answer (e.g., *People with synaesthesia experience a fusing of different senses/People with synaesthesia may hear a sound when they touch an object*). In this condition, both answers were related to the passage, but one provided a better evaluation of the passage meaning. Questions were similarly phrased but differences were presented by the type of options available, and level of detail needed to select a correct answer. Paragraph naturalness and comprehension question difficulties were independently rated by participants who did not take part in subsequent testing. Passages were rated on a scale from 0 (very unnatural) to 100 (very natural) ($M = 63.04$, $SD = 5.31$) and questions were rated on a scale of 0 (very easy) to 100 (very difficult). High comprehension demand questions were rated as more difficult ($M = 23.71$, $SD = 4.45$) than low comprehension demand questions ($M = 19.97$, $SD = 3.67$), $t(49) = -8.57$, $p < .001$. Two counterbalanced lists were then created so that each participant viewed 20 of each question type but did not view the same paragraph twice. The paragraphs occupied 10 - 13 lines on the screen ($M = 859.95$ characters including spaces, $Max = 1159$ characters).

4.3.4 Design and Procedure

Testing took place over two sessions with a minimum of two days in between them. During the first session participants were given an information sheet and consent form and completed two eye tracking tasks (the first eye tracking task was for a separate experiment), followed by the experimenter administered WIAT-II comprehension test and some other cognitive tasks belonging to an unrelated experiment. The same experimenter administered this task to all participants to control for as much experimental variation between participants as possible. A script was read from the test materials to ensure that instructions were identical for all participants. Participants read passages (short narratives and information texts) aloud or silently and were asked literal and inferential comprehension questions by the experimenter, participants gave spoken responses which the experimenter transcribed.

For the eye tracking task, participants were asked to sit comfortably at the computer, resting their chin on a chinrest and were then guided through the set up and 9-point calibration of the eye tracker by the researcher. Participants were then required to direct their gaze to a fixation cross presented in the upper left portion of the screen. Once participants fixated upon the cross sentences were presented and always began at the location marked by the fixation cross. Participants were asked to read the paragraphs and answer questions presented on the screen using the keyboard to respond. Participants either answered complex (high comprehension demands) or simple questions (low comprehension demands) depending on the condition. The same participants completed both conditions over two sessions. Eye tracking sessions took place on two separate days. At the start of the blocks, participants read five practice paragraphs with questions matching the type for the current condition. Practice questions were followed by 20 experimental paragraphs that were each followed by a comprehension question. Block order was counterbalanced so that participants who read paragraphs and answered questions with high comprehension demands in session 1, then read paragraphs and answered questions with low comprehension demands in session 2 and vice versa. Block order was randomly assigned and within each block trial order was randomised. Participants could take breaks when needed.

During the second session participants took part in the second part of the eye tracking task (featuring the condition that they had not yet completed). Participants were then asked to complete the NDRT comprehension test, and some other online tasks for a separate study. During the NDRT participants silently read up to 7 passages and answered 5 - 8 MCQ questions about them with 4 available options. Questions appeared below the passages on the same screen. On the first passage participants were stopped after 1 minute and were asked to record the number corresponding to the line that they had been reading to measure their reading rate. Testing automatically stopped after 10 minutes and answers were recorded. Testing for the NDRT comprehension test followed a half-timed procedure, in which the standard time limit for completing this test was reduced by half. This procedure has been shown to generate a more normal distribution for university student readers (like those who took part in the current study) than the standard time limits in an investigation by Andrews et al. (2020).

4.4 Results

4.4.1 Data Cleaning

Overall accuracy on comprehension questions was high but not at ceiling level (*M* = 79.26 %, *SD* = 6.38). Reading comprehension scores on WIAT-II were calculated and normed following guidance from the experimenter manual (*M* = 110.47, *SD* = 9.37, range = 71 – 124). NDRT comprehension scores were calculated based on raw scores due to the half-timed aspect of the task. NDRT (*M* = 57.64, *SD* = 11.02, range = 20 - 74), and were weakly correlated with the WIAT-II comprehension scores (*r* = 0.22, *p* = .039) based on criteria from Dancey and Reidy (2007) where a correlation coefficient between 0.10 and 0.30 is considered weak. Scores on both tests were standardised for further analyses.

Eye tracking trials identified by the experimenter as having issues with tracker loss or featuring excessive blinking were removed prior to the analysis. Fixations shorter than 80 *ms* that landed within one character of the previous or next fixation were merged. Then, of the remaining fixations, those shorter than 80 *ms* and longer than 800 *ms* were removed. Practice trials were also removed. Due to an error in the programming of the experiment, texts were presented with a justified alignment which meant that word level data would have confounds between word length and visual extent. For this reason, word level measures such as regressions and refixations were not included in these analyses.

The following global eye tracking measures were calculated for each trial; Number of Fixations (total number of fixations made on a trial); Average Fixation Duration (mean duration in *ms* of all fixations in a trial); Forward Saccade Length (the distance in degrees of visual angle between one fixation and the next); and Total Passage Reading Time (total time in *ms* spent reading the passage in a trial). Trials where total passage reading times fell outside of 2.5 standard deviations from the mean for each participant were removed as outliers (1.31 % of data removed). Data were then removed for each eye movement measure per participant that fell outside of 2.5 standard deviations from the mean (Number of Fixations (0.59 % data removed); Average Fixation Duration (0.88 % data removed); Forward Saccade Length (1.16 % data removed). Descriptive statistics per condition for these measures were calculated across participants and are displayed in Table 4.1.

Table 4.1

Descriptive Statistics for Eye Movement Measures (ms)

	Condition	Min	Max	Mean	*SD*
Number of Fixations	Low	42.00	285.00	138.38	33.43
	High	65.00	269.00	143.64	33.95
Average Fixation Duration (ms)	Low	138.53	285.94	206.14	24.06
	High	144.30	279.17	206.79	23.80
Average Forward Saccade Length (visual degrees)	Low	3.38	9.66	6.07	0.98
	High	3.41	9.25	5.99	0.99
Total Passage Reading Time (ms)	Low	8388.00	58115.00	28854.00	8484.70
	High	11232.00	63869.00	30138.17	8868.98.00

Note. Descriptive statistics are based on participant means per condition.

4.4.2 Linear Mixed Models

Eye movement measures were analysed using the lme4 package (version 1.1-31; Bates et al., 2015) in R (version 4.2.2; R Core Team, 2022). Data were checked for normality and were not transformed for modelling as their distribution closely resembled a normal distribution. Binomial Generalized Linear Mixed Models were used to model accuracy data. The following model building strategy was followed. Models started with fixed effects of interest: the main effect of experimental condition (low vs high comprehension demands), either the NDRT or WIAT-II comprehension test scores and the trial number and all the interactions. A full random structure was included for subjects and items. The maximal model (Barr et al., 2013) was achieved by starting with the full random structure and stepwise trimming of the structure until the model converged. Slopes were first trimmed from the random effects structure where perfect correlations were indicated and subsequently starting with the factors that explained the smallest amount of variance until the model converged.

4.4.2.1 Number of Fixations

Models shown in Table 4.2 and Table 4.3 indicated that overall, more fixations were made on paragraphs where comprehension demands of the questions were high compared to when they were low. The number of fixations decreased slightly over trials, however, a significant three-way interaction between trial number, condition and scores on the NDRT comprehension test revealed a more complex pattern based on individual differences (Table 4.2). Figure 4.1 shows that high scorers on the NDRT comprehension test reduced the number of fixations further into the experimental session. They also made more fixations in the high comprehension demands condition and these two factors did not interact. A different pattern emerged for the low scorers on the NDRT comprehension test. Low scorers did make more fixations on a paragraph at the beginning of the experiment than on trials nearing the end of the experiment, but when comprehension demands were low, this decrease was not as steep. This pattern may indicate that less skilled comprehenders were nearing floor effects where they were close to the minimum number of fixations that they could accommodate whilst still reading for comprehension when comprehension demands were low.

No significant interactions or individual differences were observed for scores on the WIAT-II comprehension test, and in this model a trial by condition interaction was marginally significant (Table 4.3).

Table 4.2

LMM for Number of Fixations predicted by NDRT Comprehension and Interactions with Trial Number and Condition.

	β	95 % *CI*	*t*	*df*	*p*
Intercept	153.47	[147.01, 159.92]	46.60	140.91	< .001 ***
Trial Number	-0.75	[-0.85, -0.64]	-14.38	3300.23	< .001 ***
Condition	9.64	[5.20, 14.07]	4.26	335.04	< .001 ***
NDRT Comprehension	-1.10	[-6.65, 4.46]	-0.39	104.33	.699
Trial Number × Condition	-0.22	[-0.42, -0.01]	-2.09	3302.67	.037 *
Trial Number × NDRT Comprehension	-0.16	[-0.26, -0.06]	-3.13	3299.89	.002 **
Condition × NDRT Comprehension	-4.23	[-8.60, 0.14]	-1.90	346.14	.059.
Trial Number × Condition × NDRT Comprehension	0.25	[0.05, 0.45]	2.41	3300.46	.016 *

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included condition (high/low comprehension demands) as slopes for subjects and stimuli.

Table 4.3

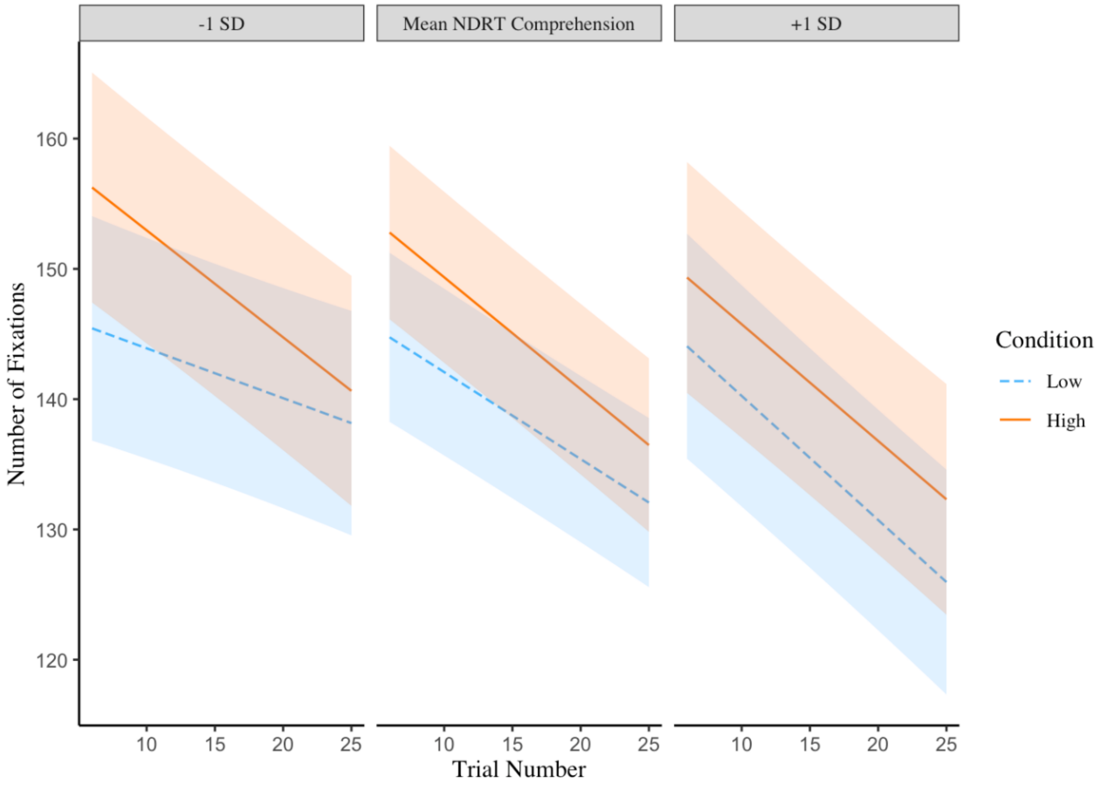
LMM for Number of Fixations predicted by WIAT-II Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*t*	*df*	*p*
Intercept	153.17	[146.69, 159.64]	46.35	140.84	< .001 ***
Trial Number	-0.77	[-0.87, -0.67]	-14.77	3302.63	< .001 ***
Condition	9.22	[4.78, 13.65]	4.07	351.96	< .001 ***
WIAT-II Comprehension	1.90	[-4.39, 8.20]	0.59	104.53	.555
Trial Number × Condition	-0.18	[-0.39, 0.02]	-1.77	3303.08	.076.
Trial Number × WIAT-II Comprehension	0.07	[-0.04, 0.18]	1.21	3300.95	.226
Condition × WIAT-II Comprehension	-0.18	[-5.14, 4.78]	-0.07	351.61	.943
Trial Number × Condition × WIAT-II Comprehension	-0.06	[-0.29, 0.17]	-0.51	3300.98	.611

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included condition as a slope for subjects and intercept only for stimuli.

Figure 4.1

A Three-Way Interaction between NDRT Comprehension Scores, Condition (high vs low Comprehension Demands) and Trial Number on the Number of Fixations made when Reading a Paragraph



Note. Shaded areas represent 95 % confidence intervals.

4.4.2.2 Average Fixation Duration

Table 4.4 and Table 4.5 present models for average fixation durations. These models indicated that overall, average fixation durations increased slightly over trials. A three-way interaction between trial number, condition and scores on the NDRT comprehension test was observed (Table 4.4). Figure 4.2 shows that for low scorers on the NDRT average fixation durations increased from early to late trials in the experiment. For the high scorers a different pattern emerges depending on the comprehension demands with average fixation time going up when comprehension demands are high and going down when they are low.

WIAT-II comprehension scores were not significant predictors of average fixation durations (Table 4.5), though an interaction of trial by condition was marginally significant.

Table 4.4

LMM for Average Fixation Durations (ms) predicted by NDRT Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*t*	*df*	*p*
Intercept	205.48	[200.82, 210.14]	86.42	98.98	< .001 ***
Trial Number	0.10	[0.05, 0.15]	4.00	3300.34	< .001 ***
Condition	-0.41	[-2.71, 1.89]	-0.35	280.96	.729
NDRT Comprehension	-2.18	[-6.75, 2.38]	-0.94	94.13	.351
Trial Number × Condition	0.07	[-0.03, 0.17]	1.43	3301.82	.154
Trial Number × NDRT Comprehension	-0.07	[-0.12, -0.03]	-2.97	3299.18	.003 **
Condition × NDRT Comprehension	-2.62	[-4.89, -0.34]	-2.26	278.34	.025 *
Trial Number × Condition × NDRT Comprehension	0.12	[0.02, 0.21]	2.31	3296.20	.021 *

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included condition as a slope for subjects and intercept only for stimuli.

Table 4.5

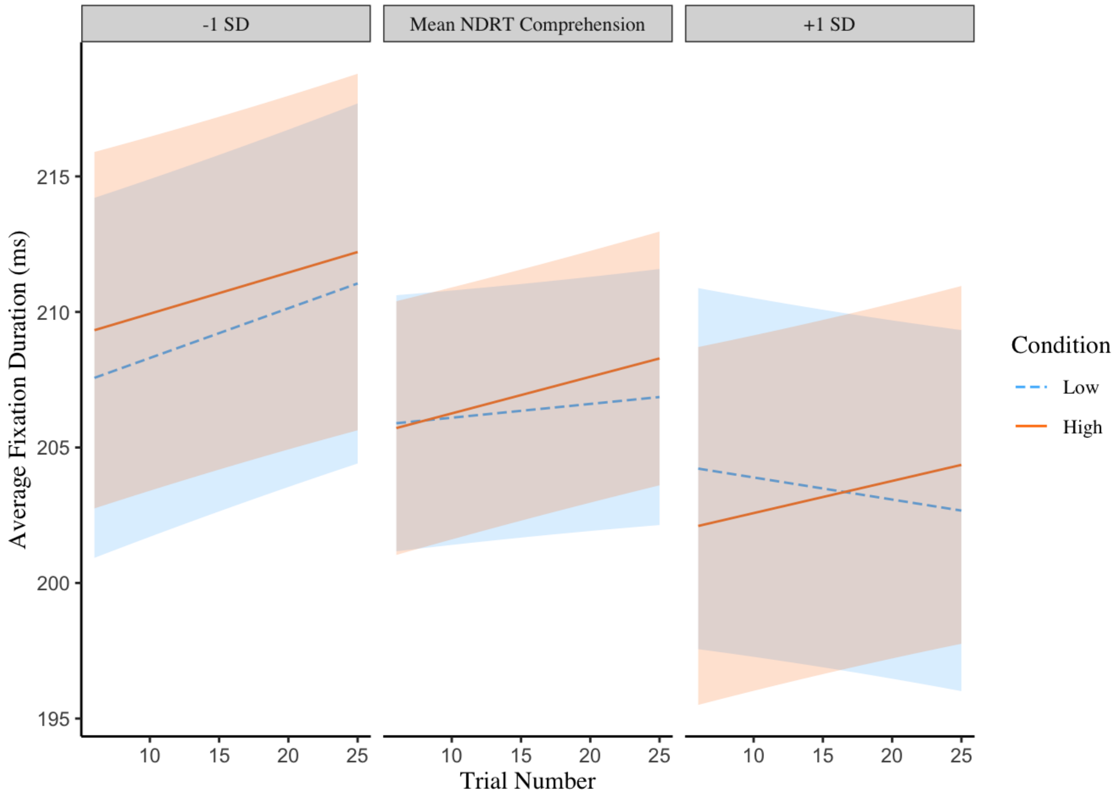
LMM for Average Fixation Durations (ms) predicted by WIAT-II Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*t*	*df*	*p*
Intercept	205.58	[200.90, 210.26]	86.08	98.93	< .001 ***
Trial Number	0.09	[0.04, 0.14]	3.64	3297.93	< .001 ***
Condition	-0.65	[-2.96, 1.66]	-0.55	280.78	.580
WIAT-II Comprehension	-3.27	[-8.44, 1.90]	-1.24	94.18	.218
Trial Number × Condition	0.09	[-0.01, 0.19]	1.75	3298.81	.080.
Trial Number × WIAT-II Comprehension	0.01	[-0.04, 0.07]	0.38	3297.93	.703
Condition × WIAT-II Comprehension	-0.02	[-2.60, 2.56]	-0.01	278.01	.989
Trial Number × Condition × WIAT-II Comprehension	-0.05	[-0.16, 0.06]	-0.87	3298.43	.383

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included condition as a slope for both subjects and stimuli.

Figure 4.2

A Three-Way Interaction between NDRT Comprehension Scores, Condition (high vs low Comprehension Demands) and Trial Number on Average Fixation Durations (ms) when Reading a Paragraph



Note. Shaded areas represent 95 % confidence intervals.

4.4.2.3 Average Forward Saccade Length

Models for average forward saccade lengths are displayed in Table 4.6 and Table 4.7. In both models, longer forward saccades were observed for passages with low comprehension demands than for identical passages with high comprehension demands. A slight increase in forward saccade length over trials was also predicted by both models. Table 4.6 shows that a three-way interaction between trials, conditions and NDRT comprehension scores was significant though numerically small. Differences can be seen in Figure 4.3, where those who scored highly on the NDRT made slightly longer forward saccades when comprehension demands were low compared to when comprehension demands were high, but in both comprehension demand conditions forward saccade lengths became longer further in the experiment. Low scorers also made longer forward saccades when comprehension demands were low than when they were high, however the average length of their forward saccades only increased over time when comprehension demands were high. When comprehension demands were low, these readers did not make longer forward saccades over trials, with comparable average forward saccade lengths across all trials.

No significant effects of individual differences in WIAT-II comprehension test scores were observed for average forward saccade lengths (Table 4.7).

Table 4.6

LMM for Average Forward Saccade Length (degrees of visual angle) predicted by NDRT Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*T*	*df*	*p*
Intercept	5.89	[5.70, 6.08]	60.79	104.80	< .001 ***
Trial Number	0.01	[0.01, 0.01]	7.37	3279.81	< .001 ***
Condition	-0.14	[-0.24, -0.04]	-2.72	340.04	.007 **
NDRT Comprehension	0.02	[-0.16, 0.21]	0.26	96.07	.797
Trial Number × Condition	0.00	[0.00, 0.01]	1.68	3289.38	.092.
Trial Number × NDRT Comprehension	0.01	[0.00, 0.01]	5.19	2762.05	< .001 ***
Condition × NDRT Comprehension	0.10	[0.00, 0.20]	1.94	334.43	.054.
Trial Number × Condition × NDRT Comprehension	-0.01	[-0.01, 0.00]	-2.16	2633.99	.031 *

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included condition as a slope for subjects and NDRT comprehension scores as a slope for stimuli.

Table 4.7

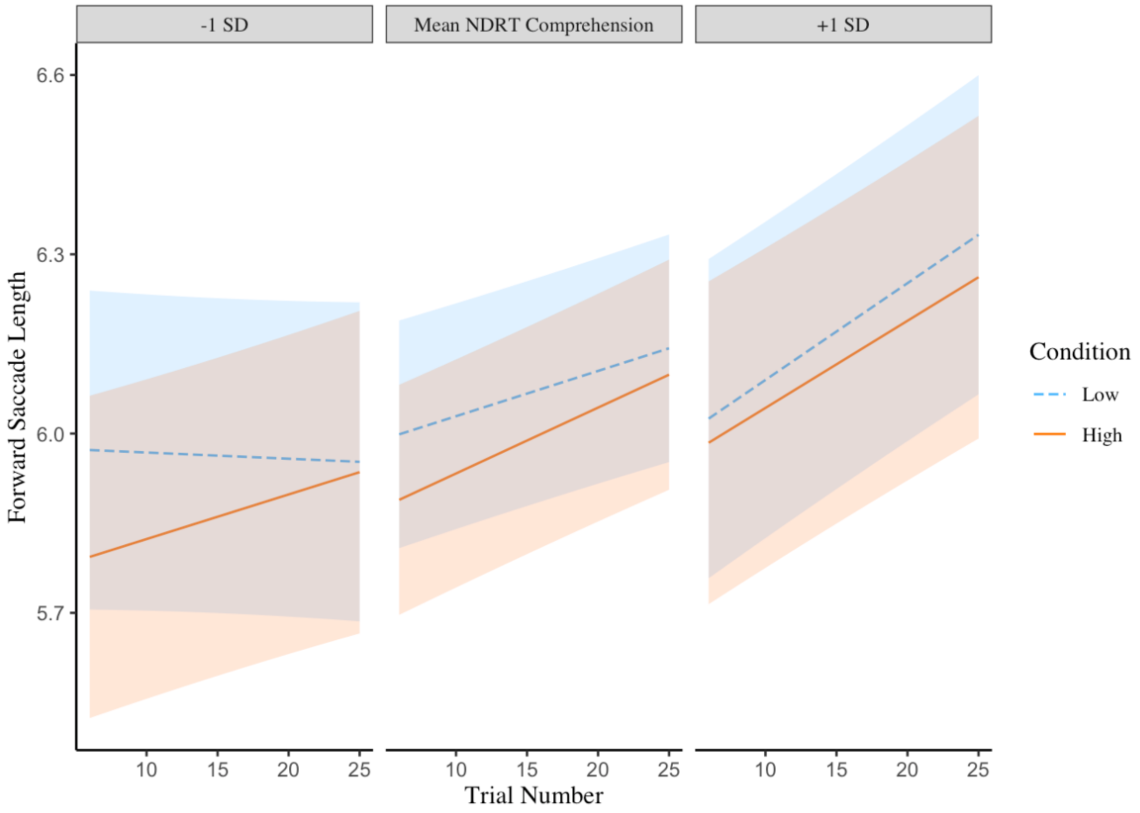
LMM for Average Forward Saccade Length (degrees of visual angle) predicted by WIAT-II Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*t*	*df*	*p*
Intercept	5.88	[5.69, 6.07]	60.24	104.60	< .001 ***
Trial Number	0.01	[0.01, 0.01]	7.99	3284.76	< .001 ***
Condition	-0.14	[-0.24, -0.04]	-2.65	351.51	.009 **
WIAT-II Comprehension	0.06	[-0.15, 0.27]	0.54	96.75	.591
Trial Number × Condition	0.00	[0.00, 0.01]	1.41	3287.86	.159
Trial Number × WIAT-II Comprehension	0.00	[0.00, 0.00]	-0.55	2314.64	.580
Condition × WIAT-II Comprehension	0.01	[-0.10, 0.12]	0.19	353.67	.851
Trial Number × Condition × WIAT-II Comprehension	0.00	[0.00, 0.01]	1.46	2661.39	.144

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included condition as a slope for subjects and WIAT-II comprehension scores as a slope for stimuli.

Figure 4.3

A Three-Way Interaction between NDRT Comprehension Scores, Condition (high vs low Comprehension Demands) and Trial Number on the Average Forward Saccade Length (degrees of visual angle) when Reading a Paragraph



Note. Shaded areas represent 95 % confidence intervals.

4.4.2.4 Total Passage Reading Times

Models for total passage reading times are presented in Table 4.8 and Table 4.9. In both models, passages in trials that occurred later in the experiment for both conditions were read more quickly than earlier passages. Passages were also read more quickly when comprehension demands were low compared to when comprehension demands were high (this was significant in both models). The model presented in Table 4.8 also revealed that total passage reading times were influenced by a three-way interaction between trials, conditions and NDRT comprehension scores. Figure 4.4 shows that high scorers consistently read passages more quickly towards the end of the experimental conditions than at the beginning, and read passages with low comprehension demands more quickly than passages with high comprehension demands. High scorers also read more quickly than low scorers by the end of the experiment in both conditions.

Low scorers on the NDRT comprehension scores displayed a different pattern, where their reading times were longer when comprehension demands were high compared to low at the beginning of the experiment and decreased over trials. However, when comprehension demands were low a potential floor effect was observed for low scorers where only a small decrease in reading times across trials was seen for passages.

No other significant effects were observed in the model including the WAIT-II comprehension scores (Table 4.9).

Table 4.8

LMM for Total Passage Reading Times (ms) predicted by NDRT Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*t*	*df*	*p*
Intercept	31996.48	[30331.38, 33661.57]	37.66	132.60	< .001 ***
Trial Number	-144.55	[-168.41, -120.70]	-11.88	3321.32	< .001 ***
Condition	2152.00	[1081.86, 3222.15]	3.94	304.94	< .001 ***
NDRT Comprehension	-727.71	[-2219.17, 763.74]	-0.96	100.48	.341
Trial Number × Condition	-42.15	[-89.90, 5.59]	-1.73	3323.40	.084.
Trial Number × NDRT Comprehension	-39.11	[-62.68, -15.55]	-3.25	3320.14	.001 **
Condition × NDRT Comprehension	-1270.52	[-2324.79, -216.25]	-2.36	311.73	.019
Trial Number × Condition × NDRT Comprehension	66.07	[19.03, 113.10]	2.75	3321.15	.006 **

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included condition as a slope for both subjects and stimuli.

Table 4.9

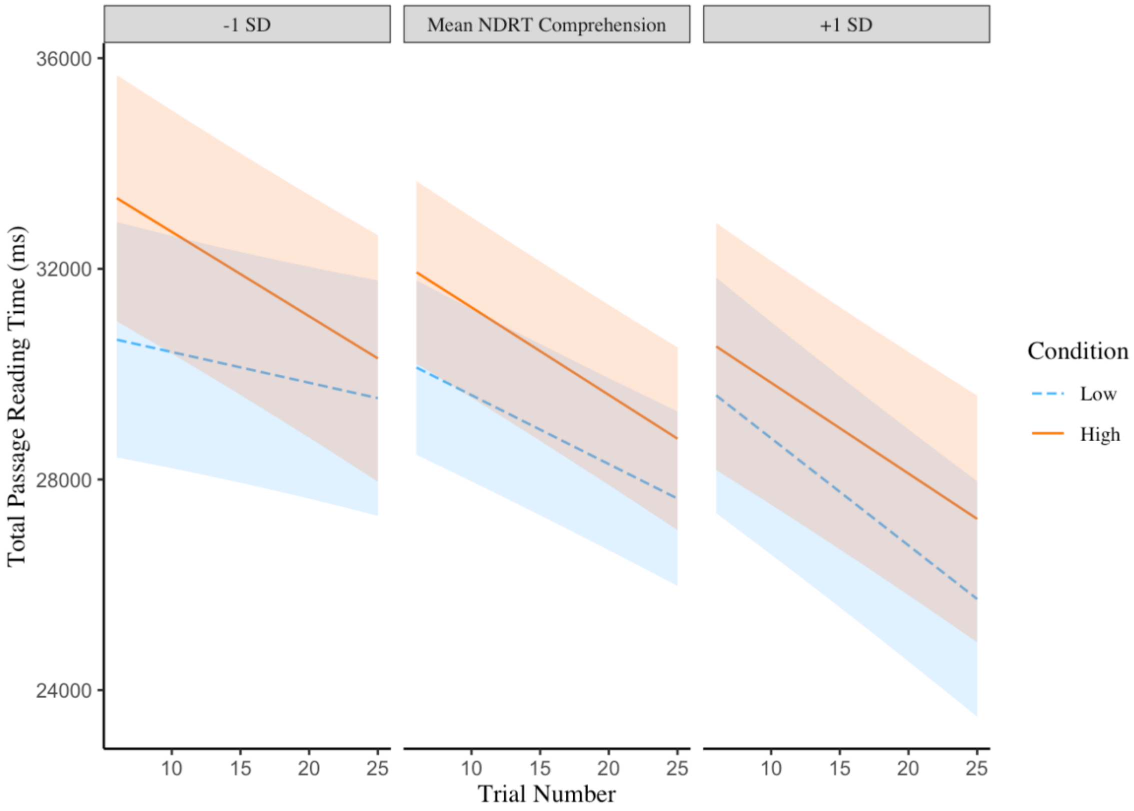
LMM for Total Passage Reading Times (ms) predicted by WIAT-II Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*t*	*df*	*p*
Intercept	31935.52	[30246.02, 33625.02]	37.05	131.55	< .001 ***
Trial Number	-150.52	[-174.47, -126.58]	-12.32	3321.70	< .001 ***
Condition	2059.34	[986.54, 3132.14]	3.76	308.20	< .001 ***
WIAT-II Comprehension	-70.65	[-1782.12, 1640.82]	-0.08	100.25	.936
Trial Number × Condition	-34.52	[-82.43, 13.40]	-1.41	3322.80	.158
Trial Number × WIAT-II Comprehension	18.14	[-8.47, 44.76]	1.34	3320.94	.182
Condition × WIAT-II Comprehension	-291.17	[-1483.74, 901.41]	-0.48	314.56	.633
Trial Number × Condition × WIAT-II Comprehension	-9.23	[-62.48, 44.03]	-0.34	3321.66	.734

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included condition as a slope for both subjects and stimuli.

Figure 4.4

A Three-Way Interaction between NDRT Comprehension Scores, Condition (high vs low Comprehension Demands) and Trial Number on Total Passage Reading Times (ms)



Note. Shaded areas represent 95 % confidence intervals.

4.4.2.5 Accuracy

Neither model showed significant differences in accuracy for high (*M* = 84 %, *SD* = 10.69) compared to low comprehension demands (*M* = 73.78 %, *SD* = 7.65), or across trials (Table 4.10 and Table 4.11). In terms of individual differences in accuracy, one interaction between trials, conditions and WIAT-II comprehension scores was found to be marginally significant (Table 4.11). The pattern observed suggested that when comprehension demands were high, high scorers on this test became more accurate over time, whereas low scorers became less accurate in later trials. However, these trends were marginal.

Table 4.10

Binomial GLMM for Accuracy predicted by NDRT Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*z*	*p*
Intercept	2.65	[1.54, 3.76]	4.66	< .001 ***
Trial Number	0.01	[-0.02, 0.03]	0.56	.576
Condition	-1.57	[-3.97, 0.84]	-1.28	.201
NDRT Comprehension	0.05	[-0.40, 0.49]	0.20	.841
Trial Number × Condition	0.02	[-0.03, 0.07]	0.66	.510
Trial Number × NDRT Comprehension	0.01	[-0.01, 0.04]	0.96	.338
Condition × NDRT Comprehension	-0.09	[-0.95, 0.76]	-0.21	.832
Trial Number × Condition × NDRT Comprehension	0.00	[-0.06, 0.05]	-0.16	.873

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included intercept only for subjects and condition a slope for stimuli.

Table 4.11

Binomial GLMM for Accuracy predicted by WIAT-II Comprehension and Interactions with Trial Number and Condition

	β	95 % *CI*	*z*	*p*
Intercept	2.58	[1.49, 3.67]	4.63	< .001 ***
Trial Number	0.01	[-0.02, 0.04]	0.79	.430
Condition	-1.48	[-3.85, 0.88]	-1.23	.220
WIAT-II Comprehension	0.06	[-0.42, 0.53]	0.23	.815
Trial Number × Condition	0.02	[-0.04, 0.07]	0.59	.558
Trial Number × WIAT-II Comprehension	0.01	[-0.02, 0.04]	0.84	.401
Condition × WIAT-II Comprehension	-0.59	[-1.51, 0.33]	-1.26	.209
Trial Number × Condition × WIAT-II Comprehension	0.05	[-0.01, 0.11]	1.69	.090 .

Note. The baseline of the condition term is low comprehension demands. Estimates represent the change when going from low to high comprehension demands. The final random effects structure included intercept only for subjects and condition a slope for stimuli.

4.5 Discussion

The current study investigated two offline reading comprehension tests (the NDRT and the WIAT-II) as predictors of individual differences in skilled readers' eye movements during paragraph reading. Eye movement patterns were investigated under high and low comprehension demands and across trials. Parallel sets of analyses were conducted for each test to determine whether individual differences in offline comprehension tests predicted patterns in eye movement behaviour that was reflective of changes in comprehension demands, and whether readers adapted to comprehension demands over time. The main aim was to determine whether discrepancies arose between the two tests that claim to measure reading comprehension (Lee, Godwin et al., 2023; Lee, Pagán et al., 2023), and a secondary aim was to investigate whether individual differences could be observed in the way that skilled readers adapted their reading strategies over time and in response to comprehension demands. First, we will focus on the overall patterns in the data across global eye movement measures, then on individual differences that were observed, and finally, we will discuss the two offline comprehension tests and differences in the predictive power associated with them for skilled readers.

4.5.1 Overall Patterns

Overall, within an experimental block, paragraphs in later trials were read more quickly than in earlier trials. Reading strategies appeared to become more efficient, or perhaps more 'risky' (O'Regan, 1992) over time, with fewer fixations and increasing saccade lengths. Future investigations would need to include analyses of regressions to determine whether readers do use a riskier reading strategy in later trials since this may be more clearly observed though rereading behaviours. Participants were not more accurate on the comprehension questions in any one condition, or over time in the experiment. Though the increased difficulty in the high comprehension demands condition was confirmed by a pre-test, it may be that for our skilled readers, the higher demands were not enough to reduce their accuracy. Indeed, the pattern observed in the means suggested that participants had higher levels of accuracy when comprehension demands were high, which would be compatible with previous observations by Andrews and Veldre (2021). However, this difference was not significant in the analyses. We did, however, observe differences in eye movement patterns in response to the high and low comprehension demands. Readers were able to adjust their reading behaviours to the comprehension demands (Tinker, 1958) and were able to read more thoroughly when comprehension demands were high and more superficially when comprehension demands were low (Heller, 1982). Passages with higher comprehension demands were read more slowly, and featured more fixations and shorter saccades than passages with low comprehension demands.

4.5.2 Individual Differences

Passage reading has the potential to introduce more variance in eye movement data compared to sentence reading simply due to the increase in processing demands, and the potential for allowing individual differences to be expressed in more varied ways. Slower reading and more rereading is often observed during passages compared to sentences (Radach et al., 2008). Although our data do not echo Andrews and Veldre's (2021) observations of shorter passage reading times, shorter average fixation durations and longer saccades directly related to individual differences in reading proficiency, their findings were based on a composite measure which included vocabulary, reading comprehension, reading rate and spelling, rather than comprehension alone. It may be that the direct effects of individual differences on fixation time measures observed by Andrews and Veldre (2021) are better explained by other measures included in their composite score (e.g., spelling or vocabulary). In our analyses, individual differences as measured by offline comprehension measures seem to predict the response to high versus low comprehension demands in the way that readers adapt over time.

Analysis of eye movements in relation to NDRT comprehension scores presented a clear picture of individual differences in response to comprehension demands. When reading behaviours were measured across trials, there were observable individual differences in the way that readers adapted their behaviour in response to comprehension demands. Differences between readers were smaller at the beginning of the experimental blocks and became larger in later trials where high scorers read more quickly, made fewer fixations and longer saccades than low scorers. High scorers read passages with high comprehension demands more slowly, with more fixations and shorter saccades than passages with low comprehension demands, but the changes over time for high and low comprehension demands were comparable. In contrast, low scorers adapted their reading behaviours at a slower rate and approached a threshold for the fastest reading times, lowest number and duration of fixations, and the largest saccade lengths they were able to accommodate whilst reading for comprehension, even when comprehension demands were low. This evidence that less skilled comprehenders have a lower limit to how quickly they can read for comprehension than highly skilled comprehenders complements the general finding that less skilled readers often read more slowly and make longer fixations than more skilled readers (Rayner, 1998; Ashby et al., 2005; Andrews & Veldre, 2021; Lee, Godwin et al., 2023; Lee, Pagán et al., 2023).

4.5.3 Offline Comprehension Measures

Critically, this pattern of results was highly dependent on which offline measure of comprehension was used to measure comprehension. Analyses of the same participants' eye movement data in relation to their scores on the WIAT-II comprehension test did not predict differences in eye movement patterns for different comprehension demands. Earlier, we described some differences in the format of each test that could indicate differences in the underlying skills measured by them. We return to these now to consider possible reasons why the NDRT revealed patterns in our data that the WIAT-II did not.

Higher comprehension accuracy is often observed for questions following narratives than expository texts (Best et al., 2008). Therefore, expository passages were selected for the current study to ensure that the materials were appropriate for skilled reading and to maximise the likelihood of finding variation in accuracy scores within this population. Potentially as a result of this choice, accuracy scores were not close to ceiling levels in the current study. The NDRT includes expository texts that are more similar to the current study materials than the WIAT-II comprehension test. Therefore, it is reasonable to suggest that comprehension based on similar test materials will account for a comparatively larger proportion of variance in reading behaviour. It has also been suggested that the NDRT is closely related to general knowledge (Ready et al., 2013; Coleman et al., 2010). If the NDRT comprehension measure is highly related to general knowledge, we would expect to see higher levels of comprehension accuracy on our experimental questions for participants who score highly on the NDRT, but this was not observed.

In contrast, since the WIAT-II comprehension test includes some items that must be read aloud, it may feature some overlap with working memory processes (Laubrock & Kliegl, 2015). However, our previous investigations of eye movement behaviours in sentence reading included the WIAT-II and a test of working memory (a backwards digit span task) amongst other reading skill predictors (Lee, Godwin et al., 2023; Lee, Pagán et al., 2023). These investigations did not suggest that there was much overlap between working memory and the WIAT-II comprehension test as they did not load together in principal components analyses (Lee, Godwin et al., 2023; Lee, Pagán et al., 2023). We highlighted some aspects of the WIAT-II comprehension subtest that may mean it also has less power to discriminate between skilled adult readers than the NDRT. First, narrative test comprehension is often higher than expository texts, which may indicate that portions of the WIAT-II comprehension test are not difficult enough to allow much variance within skilled readers. We also noted in the introduction that the reading aloud parts of the test may not be as informative about individual differences in adults as it is for children since adults rely on reading aloud less often (Duncan & Freeman, 2020), though further research would be needed to confirm this. In addition, the face-to-face aspect of the WIAT-II may lead to noisier data for adults where participants might experience performance anxiety.

4.5.4 Conclusion

Overall, it appears that the NDRT comprehension test (notably when following a half-timed procedure) is more sensitive to differences in eye movement behaviours in response to high and low comprehension demands observed between skilled adult readers compared to the WIAT-II comprehension test. Individual differences captured by the half-timed version of the NDRT have been previously shown to be sensitive to individual differences in skilled readers eye movements (Andrews et al., 2020). The current study extends this and suggests it can be used to predict differences in eye movement behaviours across trials in response to varying comprehension demands. We highlight the importance of careful test selection when measuring eye movement behaviour in skilled adult readers and advise that comprehension tests should not be used interchangeably, because they *jingle* (Thorndike, 1904) and that researchers should exercise caution when selecting a reading comprehension test for future research. We echo advice from

Flake and Fried (2020) who call for transparency when reporting test selection processes and urge researchers to select comprehension tests that are clearly based on the theoretical concepts that the researcher wishes to assess.

Chapter 5 Discussion

The overall aim of this thesis was to explore the assumptions of uniformity in skilled reading (Andrews, 2012). Throughout this project, we investigated whether individual differences in skilled readers' cognitive skills predicted changes in reading behaviours in response to differences in features of a text or the demands of the reading task. Across three experimental investigations, this body of work covered a broad range of tasks where individual differences in reading patterns could have arisen.

First, in Chapter 2 (Experiment 1) we examined the effects of word frequency during sentence reading, since it has been found that whilst high frequency words are typically read more quickly than low frequency words (Angele et al., 2014; Rayner & Fischer, 1996; Rayner et al., 1996), research has also found that better adult readers are comparatively less slowed down by low frequency words than less skilled adult readers (e.g., Ashby et al., 2005; Cop et al., 2015). For this experiment we assessed a large test battery of cognitive skill predictors, initially with separate models for each test predictor to allow for clear comparisons with previous research, and then within a single model that accounted for overlapping variance with composite scores for lexical proficiency and general processing speed. In Chapter 3 (Experiment 2), we looked more closely at the process of word identification in parafoveal preview and asked whether letter position encoding is similarly flexible for all skilled readers, or whether differences in response to transposed letter pseudoword previews could be observed within skilled reader populations and could be predicted by differences in cognitive skills. Finally, in Chapter 4 (Experiment 3) we addressed inconsistencies that arose within our findings relating to two offline reading comprehension subtests, from the Nelson Denny Reading Test (NDRT; Brown et al., 1993), and the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005). We assessed the sensitivity of each test when predicting individual differences in skilled reader's passage reading in response to high and low comprehension demands. Experiment 1 was run in one experimental session, the second experimental session comprised of both Experiment 2 and 3.

5.1 Replication of Previous Research

Through the work presented in this thesis we replicated some well-documented effects on eye movement patterns related to features of the text and demands of the task. In addition, we replicated findings related to the constraints of the perceptual span that influence eye movements during reading (i.e., effects of launch site). We also provided evidence that supports the grouping of a number of cognitive skills that have been associated with the theoretical construct of lexical quality in previous research (Perfetti, 2007; Perfetti & Hart, 2002; Andrews, 2015). In this section, we first discuss the replication of text-level influences, then task demands, followed by the effects of saccade launch site, then we discuss the grouping of skills associated with lexical quality, and finally, the replication of findings of individual differences in eye movements based on single test predictors.

5.1.1 Text Influences

Experiments 1 and 2 focussed on well-documented word-level experimental manipulations to investigate individual differences skilled reading. Importantly, these investigations upheld the robust main effects that have been consistently reported in the literature. In Experiment 1, low frequency words were read more slowly and were less likely to be skipped than high frequency words. In general, these findings supported previous research that has found that readers require more time to read less commonly occurring words than words that occur more frequently (Rayner & Fischer, 1996; Rayner, Sereno et al., 1996).

In Experiment 2, we found that, overall, adult readers encode letter positions more flexibly than letter identities, in line with previous findings where transposed letter pseudowords are perceived to be more similar to a base word than a substituted letter pseudoword (Rayner, White et al, 2006; Forster et al., 1987; Lupker, et al., 2008; Perea & Lupker, 2003, 2004; Schoonbaert & Grainger, 2004; Colombo, et al., 2017; Grainger et al., 2012; O'Connor & Forster, 1981; Perea & Fraga, 2006). This finding also supports current models of word recognition that feature a flexible letter position encoding mechanism (the SOLAR model (Davis, 1999; 2010); the Open Bigram model (Grainger & Van Heuven, 2003; Grainger et al., 2006; Grainger & Ziegler, 2011; Grainger et al., 2012); the Overlap model (Gómez, Ratcliff & Perea, 2008); the SERIOL model (Whitney, 2001); and the Bayesian Reader (Norris, 2006)).

5.1.2 Task Demands

Experiment 3 utilised an experimental manipulation of comprehension demands where readers were asked to read identical passages with high or low comprehension demands imposed by the level of detail required to answer comprehension questions. As a result of this experiment, we found that in general, readers could adapt their reading strategies to suit the demands of the task in line with previous research (Radach et al., 2008; Wotschack & Kliegl, 2013; Weiss et al., 2018; O'Regan, 1992; Christianson et al., 2017; Andrews & Veldre, 2021). When comprehension demands were low, readers most often read a text more quickly, using fewer fixations coupled with longer saccades compared to when comprehension demands were high, which may indicate a riskier reading strategy when demands are low (O'Regan, 1992).

5.1.3 Effects of Saccade Launch Site

A consistent finding across Experiments 1 and 2 which considered word-level analyses of skilled reading was that when saccades were launched from prior fixations close to the next word, it subsequently received shorter fixations and was more likely to be skipped than when prior saccades were launched from further away from it, in line with previous research (Pollatsek et al., 1986). This finding demonstrates that information presented in the parafoveal region can be accessed before the word is directly fixated, reducing the time needed to complete lexical identification of an upcoming word, and as a consequence is used to plan subsequent saccade trajectories (Pollatsek et al., 1986). A far away launch site will mean the word is further to the right in the parafovea compared to a nearby launch site, thereby reducing the amount of information that is extracted.

5.1.4 Lexical Quality

Another consistent finding from the first and second testing sessions was that when skills were grouped based on overlapping variance via PCA, there was some consistency in which skills loaded on the same component. Skills that have previously been described as closely related to the quality of readers' lexical representations, vocabulary, spelling and reading experience (Perfetti, 2007; Perfetti & Hart, 2002; Andrews, 2015) were grouped together, which provides evidence that they may share a common underlying construct for skilled adult readers. In addition, the composite measure related to lexical proficiency calculated in Experiment 1 (from the first testing session) included pseudoword decoding whereas word reading was slightly below the criteria for inclusion. In Experiment 2 (from the second testing session) a similar composite measure included word reading but narrowly excluded pseudoword decoding. Therefore, we provided some limited support for the notion that word decoding is also related to lexical quality (Perfetti, 1992; 2007; Perfetti & Hart, 2002). Our findings are in line with previous suggestions that vocabulary is a better predictor of reading comprehension in adults, whereas word decoding skills are better predictors for developing readers (Braze et al., 2007; Protopapas et al., 2007). We found vocabulary to be a more consistent predictor of eye movement patterns during reading than word and pseudoword decoding in Experiments 1 and 2 for our adult participants.

In general, across Experiments 1 and 2, readers with higher lexical proficiency (indicated by composite scores that comprise overlapping variance on skills related to high quality lexical representations: most commonly larger vocabularies, better spelling and more reading experience) had shorter fixation durations on words than readers with lower lexical proficiency scores. This is in clear support of the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002), in that readers with high quality lexical representations stored in their mental lexicon are better equipped to quickly identify words and their meanings. In addition, those with greater lexical proficiency also read sentences more quickly in Experiment 1, suggesting that more efficient word identification enabled readers to access word meanings more quickly which in turn facilitated faster integration of individual words into sentence representations and therefore enabled faster comprehension of sentences than those with lower lexical proficiency (Kintsch, 1998).

5.1.5 Single Test Predictors of Individual Differences

When we considered separate models for test predictors of individual differences within our word frequency investigation, the models replicated some key findings from previous research. We found that faster word identification processes (indicated by shorter gaze durations on words) were predicted by high scores on two offline reading ability measures (NDRT (Brown et al., 1993) and WIAT-II (Wechsler, 2005)), and spelling (Andrews & Hersch, 2010). These findings were consistent with the idea that more proficient readers read more quickly than average readers (Ashby et al, 2005). This finding is also well-aligned with the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002), since more skilled readers are theorised to have more, and better-quality lexical representations than less skilled readers, which in turn allow them to identify words more quickly. Faster gaze durations associated with better spelling is particularly supportive of this theory as it is very strongly linked to readers' lexical representations for readers of English, a language with an especially opaque orthography. In addition, in these analyses we found that higher scores on reading ability tests (NDRT and WIAT-II), and vocabulary knowledge (LexTALE; Lemhöfer & Broersma, 2012) were indicators of reduced word frequency effects in skilled readers. These findings were in line with previous research that found that more proficient readers (indicated by overall reading ability measures) and those with greater vocabulary knowledge are able to identify and read words more quickly than less proficient readers (Ashby et al., 2005; Cop et al., 2015).

In contrast, within these analyses we failed to provide evidence for some previously reported cognitive skill predictors of individual differences in readers' eye movement patterns. In our sample, we found that whilst higher spelling abilities predicted shorter gaze durations compared to lower spelling abilities, we did not observe differences in the size of the frequency effect related to spelling (though trends were in the same direction as Veldre et al., 2017). We noted some differences in statistical modelling in Chapter 2 compared to that of Veldre et al. (2017) that meant this finding was difficult to compare directly. Such limitations of comparing models with different numbers of predictors and model pruning are discussed further in section 5.3.4.

Print exposure measured by the ART (Acheson et al, 2008) has also previously been found to predict a reduced effect of word frequency on gaze durations (Moore & Gordon, 2015), however this was not replicated in our analyses. However, we noted in Chapter 2 that we observed fairly low scores on this test for a skilled reader population and as noted by Moore and Gordon (2015) the ART may be less able to account for variance in reader populations at lower scores.

Similarly, we did not find any individual differences in word reading behaviours related to RAN (Denckla & Rudel, 1974) or WMC (measured by a backwards digit span task, Gathercole et al., 2004) in analyses in Experiment 1 or 2. We did find that a composite score that described shared variance from both RAN and WMC scores was associated with faster sentence reading times which we will discuss in Section 5.2.1 below. Although Kuperman and Van Dyke (2011) found RAN times to be reliable predictors of eye movements across the time-course of sentence reading, we highlighted previously in this thesis that the readers in their study were not college-educated and may represent a larger proportion of less skilled readers than the participants who took part in experiments for the current project (who were primarily university students). In addition, dyslexic adults often perform more slowly on RAN tasks than non-dyslexic adults (Kirkby, 2022), and since we excluded participants with reading difficulties, it may be that that RAN tasks were less sensitive to individual differences in word reading behaviours in our samples.

5.2 Novel Findings

5.2.1 Individual Differences and Word Frequency

In Experiment 1 we included multiple skills within larger models to determine which skills were most predictive of faster low frequency word identification in skilled readers. We found that scores on a composite measure of lexical proficiency (characterised by spelling, vocabulary size, reading experience and pseudoword decoding) were the most consistent indicators of this pattern. This was clear from scores indicating better performance on the associated skills predicting shorter go past times and shorter sentence reading times for low frequency words. Our findings support theoretical explanations of skilled reading where readers with better performance in skills related to high quality lexical representations are better equipped to quickly identify words and are able to read more quickly as a result (Perfetti, 2007; Perfetti & Hart, 2002; Andrews, 2015). In Experiment 1, we observed that lexical proficiency predicted differences in word frequency effects only in late eye movement measures. In Chapter 2 we discussed this finding and suggested that for skilled readers, precise, high quality lexical representations are associated more with faster embedding of meaning into sentence context, rather than faster orthographic decoding of low frequency words. However, in Experiment 2 we did find that a similar composite score related to lexical skills did predict early eye movement measures (shorter first fixation durations and single fixations) as well as late eye movement measures (shorter gaze durations and go past times). Based on this, we may conclude that for skilled adults, high quality lexical representations are also key in facilitating faster word identification processes, which as a result allows readers to process texts more efficiently.

In Experiment 1, a second composite score (PC2) was calculated based on overlapping variance of RAN and (to a lesser extent) WMC (backwards digit span scores), which supported previous ideas that WMC is related to RAN performance (Papadopoulos et al., 2016). We suggested that this composite score represents a reader's general processing speed, though we did not find that a similar composite could be calculated in Experiment 2 and thus we limit somewhat further discussion about the theoretical basis of this grouping. We found that better performance on this measure was associated with faster sentence reading, rather than word reading. We also observed that a small reduction in the word frequency effect was predicted by high PC2 scores in sentence reading times. We suggested that faster sentence reading times in this study associated with higher PC2 scores may reflect less rereading following an uncommon word for readers with faster processing speeds. This explanation is in line with previous research that has found that WMC and RAN task modulate rereading behaviours (Clifton et al., 2003; Kuperman & Van Dyke, 2011; Gordon et al., 2020) which are more likely to occur when readers need to reread difficult or ambiguous portions of a text (Clifton et al., 2003; Frazier & Rayner, 1982; Rayner & Frazier, 1987). Sentences presented in Experiment 1 and 2 in this thesis were neutral, unambiguous and not designed to be difficult for skilled readers, and as a result, we would expect that WMC and RAN tasks would be stronger predictors of individual differences for more complex reading materials. In addition, analyses in Experiments 1 and 2 primarily focussed on word-identification processes, which were appropriate for the word-level experimental manipulations that we wished to investigate (word frequency and transposed letter effects). Therefore, the studies presented here were not very well-suited to assess the influence of WMC and RAN times on rereading behaviours, though we found some evidence in total sentence reading times. We return to discuss our limitations in assessing rereading behaviours in Section 5.3.2.

5.2.2 Individual Differences and Letter Position Encoding

In Experiment 2, we investigated whether individual differences in cognitive skills predicted differences between skilled adult readers in the flexibility of their letter position encoding. As mentioned above, all skilled adults were found to use more flexible mechanisms for encoding letter positions than letter identities, in line with previous research (Rayner, White, et al, 2006; Forster et al., 1987; Lupker, et al., 2008; Perea & Lupker, 2003, 2004; Schoonbaert & Grainger, 2004; Colombo, et al., 2017; Grainger et al., 2012; O'Connor & Forster, 1981; Perea & Fraga, 2006). In terms of individual differences, very little variation in letter position encoding processes was found for skilled readers. Therefore, in general, our findings indicated that letter position encoding flexibility is fairly stable in skilled adult readers, and the development of this processes appears to reach maturation (or near maturation) once reading skills are developed.

A hypothesis for this study that was compatible with the subsequent findings was based upon some key elements described in the Multiple-route model (Grainger & Ziegler, 2011). Whilst this model does not make any explicit predictions about individual differences in skilled readers' letter position encoding, we assumed that skilled readers would rely more on a direct orthographic route to word identification compared to children who primarily use a phonological route, translating letters to sounds before retrieving word meanings. As such we considered it likely that all skilled readers would have a similarly flexible mechanism for letter position encoding that would allow greater facilitation from a transposed letter preview than a substituted letter preview.

5.2.3 Reading Comprehension and the Jingle Fallacy

Within the test battery used to assess readers' cognitive skills in the first two experiments, we included two offline reading comprehension tests (subtests of the WIAT-II, (Wechsler, 2005) and NDRT (Brown et al., 1993)). When tests and subtests were grouped based on shared variance via principal components analysis in both experiments, the WIAT-II and NDRT comprehension subtests were not found to load together. As a result, we suggested that these tests measure distinct underlying constructs, and that they present an example of the *Jingle fallacy* (Thorndike, 1904). The Jingle fallacy is the misleading assumption that two instruments measure the same construct when they share the same name. When different tests are selected by researchers to study the same construct, this assumption leads to differences in findings. As a result of consistent evidence from Experiments 1 and 2 that the NDRT and WIAT-II comprehension tests did not load together, Experiment 3 was formulated to directly compare them as predictors of eye movements during reading for skilled adults. We tested separate models which included either the WIAT-II or NDRT comprehension test as predictor of individual differences in global eye movements for paragraph reading on the same eye movement data from the same participants. We assumed that comprehension tests should be sensitive to differences in eye movement patterns that arise due to changes in comprehension demands. However, we found that the WIAT-II and NDRT comprehension models did not make comparable predictions for global eye movement patterns for skilled adult readers. Our findings demonstrated that researchers should be cautious when comparing studies that include either the NDRT or WIAT-II comprehension measures to predict eye movement patterns during reading. NDRT comprehension test scores were found to be more sensitive to changes in skilled readers eye movement patterns in our analyses than WIAT-II comprehension test scores.

When models considered changes over trials in the experiment, NDRT comprehension scores were found to significantly predict differences in the way that readers adapted their eye movement behaviours to changes in comprehension demands. All readers adapted their reading strategies to the demands of the task in fairly similar ways in early trials, with longer reading times, longer and more frequent fixations, and shorter saccades when comprehension demands were high. However, better comprehenders (as measured by the NDRT) consistently read passages more quickly in later trials (in both high and low comprehension demand conditions). Better comprehenders reduced their average fixation durations and number of fixations, and increased the length of their saccades, towards later trials in the experiment. Less skilled comprehenders also read passages in later trials more quickly than earlier trials, but to a lesser extent than more skilled comprehenders. We concluded that when readers are less skilled comprehenders (as measured by the offline NDRT comprehension test), there is a limit to how quickly they can read whilst maintaining a good level of understanding. This evidence complements previous research of individual differences that suggests that in general more skilled readers read more quickly and make shorter fixations than less skilled readers (Rayner, 1998; Ashby et al., 2005; Andrews & Veldre, 2021).

5.3 Limitations

5.3.1 Skilled Adult Readers

Through the work presented here we can only provide evidence about skilled adult readers (most of whom were university students, with the exception of some participants from the wider community in Experiment 1). Global conclusions about older or younger adult readers, developing readers or adults who experience reading difficulties are not possible to be drawn from this body of work. In this respect, we note that investigations of individual differences in eye movement behaviours that include a more diverse population of readers may produce different results and are worth pursuing.

5.3.2 Consequences of using Justified Text in Experiment 3

A programming error in Experiment 3 meant that passages were presented to participants with a justified alignment. When text is justified, extra whitespace is added between words on a line so that the vertical edges of the text are perfectly aligned. This error meant that word level measures would be confounded by inconsistent differences in the width of the whitespace between words. When the whitespace is larger, the upcoming word will be located further from the currently fixated word and as a consequence, it will be viewed further into the parafovea or periphery, where visual acuity is greatly reduced. It is very likely that a justified text alignment would often lead to less useful parafoveal previews of some words within the passage (Rayner, 1998, 2009; Schotter et al., 2012).

The originally intended analysis for Experiment 3 was to utilise the same test battery used in Experiments 1 and 2 to predict differences in skilled readers' eye movement when reading passages, and in response to different comprehension demands. The motivation for investigating passage reading was that it would be possible to determine whether skills related to lexical proficiency remain strong predictors of word-identification processes during first-pass reading in passages. Skills related to lexical proficiency (associated with PC1 in Experiments 1 and 2) were found to predict faster word identification in sentence reading. In addition, it would be possible to determine whether different skills are more associated with integrative processes, where word meanings are combined to form sentence representations and then integrated with world-knowledge to represent the text as a whole (Kintsch, 1998). When readers find a text difficult to comprehend during the first pass, they often make regressions to reread previous parts of a text that they found difficult to understand or integrate with the rest of the sentence (Frazier & Rayner, 1982; Rayner & Frazier, 1987). Rereading may also be used as a checking mechanism for any text when comprehension demands are high (Weiss et al., 2018). Rereading and regression rates are good indicators of late cognitive processes that occur after first-pass word identification processes (Liversedge & Findlay, 2000). Since previous research has indicated that RAN tasks are strong predictors of regressions and rereading behaviours (Kuperman & Van Dyke, 2011; Gordon et al., 2020), we would expect RAN times to be a stronger predictor of rereading behaviours in the originally intended experiment.

However, due to the justified text alignment, and subsequent confounds related to the width of the word spacing, we opted to focus exclusively on global eye movement measures in our analyses (e.g., average fixation durations, average saccade lengths, total passage reading times). Consequently, the number of observations per paragraph for each participant in our analyses were reduced to a single datapoint per paragraph, which meant that the statistical power to estimate multiple parameters was greatly reduced. As a result, Experiment 3 was adapted and a direct comparison of two reading comprehension tests (subtests from the NDRT and WIAT-II), that had not been found to load together in principal components analyses in Experiments 1 and 2, was selected as a replacement for the large test battery investigation.

The analyses of the test battery presented in this thesis did not focus on the areas of reading that the RAN has previously been suggested to influence, namely rereading behaviours (Kuperman & Van Dyke, 2011; Gordon et al., 2020). As a result, a limitation of this thesis is that we cannot make strong claims about whether the RAN is or is not an important predictor of individual differences in skilled readers' eye movement patterns, other than limited evidence that it predicts later eye movement measures (shorter sentence reading times were associated with greater performance on a composite score related to RAN).

5.3.3 Limitations of Composites based on Principal Components Analysis

In Experiments 1 and 2, we used large test batteries to investigate individual differences in readers' eye movements, and steps were taken to mitigate issues with multicollinearity of test predictors within our test battery investigations. We limited our investigations to controlled experimental manipulations of specific text-level predictors that have been well-documented in the literature to reduce the number of predictors that were necessary in our analyses. In addition, we further reduced the number of predictors in our models by grouping cognitive tests based on shared variance using a principal components analysis.

However, there are limitations associated with modelling individual differences using composite scores based on multiple tests in this way. Composite scores derived from principal components analyses retain only the shared variance between the measures included, and disregard variance that is uniquely associated with each test. When such measures are used as predictors, researchers may only draw conclusions about constructs that are common across included tests. Such models cannot investigate claims that are made about specific tests. To mitigate this in Experiments 1 and 2 we also reported separate models of individual test predictors where it was necessary to compare previous research that made predictions based on specific skills. In doing so, we demonstrated inconsistencies within our own analyses of cognitive skills that arose when comparing findings from models from single test predictors with findings from models that included a composite score. Since PCA preserves the shared variance among predictors and removes unique contributions from each test, it may be that differences between analyses were due to the unique influences of individual tests.

In addition, composite scores are created based on the data provided and therefore are very specific to the current sample and test battery. Whilst well defined theoretical concepts (such as lexical quality) can be useful in defining the underlying constructs that are associated with tests grouped in a PCA, composite scores may only present a proxy measure for such concepts, and we suggest that they should not be named in ways that could cause confusion (or indeed a Jingle fallacy). As a result, we made a distinction between the theoretical construct of lexical quality, and our composite measures of lexical proficiency (PC1) in Chapter 2 and Chapter 3. It is therefore a limitation of this technique that differences in test battery selection and samples mean that clear replication and comparison to the wider literature is difficult and should be considered carefully.

5.3.4 Model Selection and Methods of Analysis

It should also be noted that there are some constraints to be considered when comparing findings from research that include a different number of test predictors, and different methods of analysis or model pruning. In the section above, we noted that when comparing models with one test predictor versus models with multiple test predictors within our analyses in Experiments 1 and 2 we found some differences in findings and suggested that differences may have occurred when comparing single tests with composite scores based on overlapping variance. An additional consideration is that models which include a greater number of parameters have reduced statistical power to estimate effects compared to models with fewer parameters. This presents a challenge for comparing findings based on research that uses test batteries with research that includes fewer test predictors.

In addition, the Linear Mixed Models reported in this thesis aimed for a maximal random effects structure where possible (Barr et al., 2013). Differences in strategies in model pruning can result in different findings, and previous research conducted prior to this guidance may have used model pruning techniques that were more comparable to intercept-only models. As a result, direct comparison of reported findings may reveal inconsistencies. Indeed, we noted in Chapter 2 that when comparing intercept only models in the current analyses to previous research, we can observe a clearer replication of reported interactions between individual differences in cognitive skills and word frequency effects. To limit similar issues that could occur when comparing our own findings across experiments, analyses in Experiments 1 and 2 (which were on the same test battery) followed identical model pruning procedures to obtain the maximal model with a fixed effects structure that best fit the data (see also Dirix & Duyck, 2017). Analyses in Experiment 3 had fewer predictors and were more confirmatory, and therefore used a different model pruning procedure.

In addition, different findings can be observed between studies which use different data cleaning methods (Ezkenazi, 2023), and data transformation techniques (Lo & Andrews, 2015). In Experiments 1 and 2, guidance from Lo and Andrews (2015) was followed, which suggests that GLMMs with gamma distribution are more appropriate for skewed reaction time data as normality is not assumed. It is likely that different results would be observed if LMMs were used with untransformed data or when data were log transformed, a technique used in previous research to fit skewed data to a normal distribution.

For analyses in Experiment 3, global passage reading data were fairly normally distributed and as a result Linear Mixed Models based on the assumptions of normality were considered appropriate (with the exception of Binomial Generalized Linear Mixed Models for modelling accuracy data). In addition, since the aim was to directly compare models based on the NDRT and WIAT-II comprehension tests, it was critical to obtain final models that were as comparable as possible. Since different model pruning techniques can influence the results obtained by final models, it was important to follow identical procedures. As models included fewer predictors and were more confirmatory than Experiments 1 and 2, in both sets of analyses (for the NDRT and WIAT-II comprehension test predictors), models retained all fixed effects of interest and started with full random effects structures for subjects and items. Models were then pruned using stepwise trimming of the random effects structure until they reached convergence, and the maximal model was achieved (Barr et al., 2013).

5.4 Impact on Wider Knowledge and Practice

5.4.1 Computational Models of Eye Movement Control

Whilst current computational models of eye movements during reading (e.g., E-Z Reader, Reichle et al., 1998; 1999; SWIFT, Engbert et al., 2005; Über-Reader, Reichle, 2020; OB1-Reader, Snell et al., 2018), can account for instance for the modulation in eye movements related to differences in word frequency observed in this thesis, in Chapter 1 we mentioned that individual differences relating to cognitive skills are not yet accounted for by computational models of eye movements during reading. E-Z Reader (Reichle et al., 1998; 1999), SWIFT (Engbert et al., 2005), Über-Reader (Reichle, 2020) and OB1-Reader (Snell et al., 2018) all model eye movements during reading under the assumption that, in general, adult readers have similar reading processes, though there is some allowance for differences in timing and efficiency of these processes (e.g., E-Z Reader model can simulate variability in eye movement patterns for older readers (Rayner et al., 2006) and children (Reichle et al., 2013)). However, current models do not account for variation at an individual level. The current body of work has also highlighted that performance on some cognitive tasks can be used to predict modulation of word frequency effects and comprehension demands on eye movement patterns of skilled readers. Findings from Experiments 1 and 2 provide evidence that high lexical proficiency predicts faster word identification processes and facilitates faster recognition of less frequent words. In addition, previous studies have also found quantitative differences in eye movement patterns based on skills that contribute to the quality of readers' lexical representations (Andrews, 2012; 2015; Andrews & Hersch, 2010, Andrews & Lo, 2012; Andrews et al., 2020; Andrews & Veldre, 2021; Veldre & Andrews, 2014; 2015a; 2015b; 2016; Veldre et al., 2017). Therefore, it is likely that future computational models of eye movement control will need to account for these skills. The Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002) provides a good basis for the modulation of reading processes for skilled readers based on the quality of their lexical representations that could be implemented into future models of eye movements in reading.

The evidence gathered throughout this body of work is in agreement with computational models of eye movements whom suggest that readers use similar processes during reading, but do so at different rates (E-Z Reader, Reichle et al., 1998; 1999; SWIFT, Engbert et al., 2005; Über-Reader, Reichle, 2020; OB1-Reader; Snell et al., 2018). According to, for instance, the E-Z Reader model, an early stage of lexical processing, called the familiarity check (L1), must determine whether a word presented is familiar enough to suggest that full lexical access (a subsequent stage called L2) is imminent. It is the L1 stage that is likely to be most influenced by differences in lexical proficiency that we observed in Chapter 2. In the E-Z reader model, L1 is described by the overall rate of lexical processing parameter α_1 with a default value of 104 ms, and parameters for the effect of frequency (with a default value of 3.5 ms) and predictability (with a default value of 39 ms).

$$L1 = \alpha_1 - \alpha_2 \ln(\text{frequency}_n) - \alpha_3 \text{ predictability}_n$$

By increasing the value of α_1, the time needed to complete the familiarity check is also increased, and as such, the average fixation duration becomes longer and often the number of refixations on a word are increased. This in turn causes subsequent forward saccades to be shorter and increases the number of regressive saccades. Together these factors decrease the overall reading rate. In addition, less time is available to parafoveally processes the upcoming word.

Reichle et al., (2013) used the E-Z Reader model to examine and simulate the development of eye movement control and reading skills, comparing children to adults. They concluded that in order to produce the different eye movement patterns observed for these groups (children make longer fixations and shorter saccades than adults), parameters that determine the speed of L1 described above should be adjusted. Indeed, simulations provided evidence to show that by simply increasing the value of the α_1 parameter, as described above, the basic pattern of longer fixations and shorter saccades made by children could be produced. Reichle and colleagues also concluded that adjusting parameters that account for saccade targeting was not necessary to simulate this pattern. However, they did highlight that adjusting parameters for lexical access were not entirely sufficient to explain the patterns of eye movements observed across the development of reading skills. Simulations revealed that the effects of post-lexical variables such as thematic role anomalies (Joseph et al., 2008) required additional adjustments.

In what they refer to as the linguistic-proficiency hypothesis, Reichle and colleagues suggested that the driving force for differences in eye movements based on reading development is driven by increasing proficiency in linguistic processing, that is, the skill involved with first identifying and then processing printed words. Reichle et al. (2013) also noted that the differences between average and skilled readers in Ashby et al.'s (2005) study was similar to the differences observed between children and adults; average readers made longer fixations and shorter saccades than skilled readers. Reichle and colleagues noted that Ashby's findings also aligned with the linguistic-proficiency hypothesis when considering how variation in reading skill modulates eye movement patterns. In Chapter 2 we highlighted that lexical proficiency was the strongest predictor for individual differences in the speed of word identification and in modulating the effects of word frequency. As such, we provided additional strong evidence for the linguistic-proficiency hypothesis in adult readers.

It is likely that by decreasing the value of α_1, we may simulate the eye movement patterns associated with very skilled readers; faster fixation times, fewer refixations and longer saccades. As mentioned above, parafoveal processing is also impacted when the value of α_1, is adjusted. Previous research has also informed us that the extraction and use of parafoveal information is indeed influenced by individual differences in skilled adult readers' lexical representations. Readers who are poor in spelling and reading ability are less able to efficiently use information from the upcoming word compared to better speller and more skilled readers (Andrews & Veldre, 2019, Veldre & Andrews 2015a; 2015b; 2016). Therefore, this adjustment within the model may account for individual differences in basic eye movement patterns related to reading skill in adults. However, to accurately simulate individual differences within skilled readers, it is important to also determine whether significant differences in post-lexical processing seen in relation to reading development are observable within populations of adult readers.

In addition, research must not only determine how the eye movement patterns of skilled adult readers vary in general, but how word frequency and predictability impact these patterns, and specifically to what magnitude. Some of this information has been gathered in the current body of work. In Chapter 2 we observed that overall, skilled adult readers who have high levels of lexical proficiency not only read words more quickly in general, but they are less slowed down by a low frequency word than skilled adult readers who have lower levels of lexical proficiency. Previous research has also determined that less skilled readers rely more on the predictability of a text than more skilled readers do (Ashby et al., 2005; Slattery & Yates, 2017). It is easy to assume that parameter adjustments, for example by adjusting the α_2 parameter, could be made within the E-Z reader model to account for this additional variation within adults, with similar adjustments to α_3 made for the role of predictability (Ashby et al., 2005; Slattery & Yates, 2017). However, similar adjustments were not found to be straightforward when Reichle et al., (2013) examined simulations of children's eye movement patterns. Simulations would therefore be necessary to model differences between skilled and very skilled readers' eye movement patterns.

In Chapter 3, we observed the well-documented effect of transposed letter previews during reading (Rayner, White et al., 2006; Johnson et al., 2007; Johnson & Dunne 2012; Pagán et al., 2016; Kirkby et al., 2022). It is evident that flexible letter position encoding is an important aspect of word recognition processes in skilled readers (as previously found in masked priming (e.g., Forster et al., 1987; Lupker, et al., 2008; Perea & Lupker, 2003, 2004; Schoonbaert & Grainger, 2004) and lexical decision studies (Colombo, et al., 2017; Grainger et al., 2012; O'Connor & Forster, 1981; Perea & Fraga, 2006)). The OB1-Reader (Snell et al., 2018) model of eye movements during reading was formulated to account for the flexibility of letter position encoding in skilled readers and incorporates an architecture equivalent to the Open-Bigram model of word recognition (Grainger & Van Heuven, 2003). The Open-Bigram model is able to account for transposed letter effects identified in isolated word paradigms and reading studies, in which words are decoded based on open-bigrams and their relative positions. As a result, this model is able to explain flexible letter position encoding, however simulations should be used to confirm that it can account for transposed letter effects in the parafovea observed in this thesis. In addition, Über-Reader (Reichle, 2020) includes mechanisms that are consistent with the basic assumptions of the Overlap model (Gomez et al., 2008) such that words are degraded by constraints of visual acuity which reduce the certainty of letter positions and identities. Similarly, a simulation would be necessary to determine whether the model could account for resulting flexible letter position encoding in the parafovea. Results from Experiment 2 demonstrated that the flexibility of letter position encoding is fairly stable once reading skills are developed. If future models of eye movements during reading endeavour to account for individual differences in reading, they should account for this (near) maturation.

With regards to individual differences in eye movements in relation to comprehension demands, some modulation was observed for different offline reading comprehension scores in Chapter 4. Though given the instability of findings across comprehension tests, and lack of other test predictors, we cannot make strong claims that such modulation should be accounted for in computational models of eye movements based on this study. We discuss how further research might investigate these patterns for multiple cognitive skills in Section 5.4.

5.4.2 Assessing Reading Comprehension

Our findings concerning the WIAT-II and NDRT comprehension subtests highlighted a Jingle fallacy (Thorndike, 1904) when they are used in scientific research, in that they are tests which share a name but do not appear to measure the same underlying construct. Evidence was provided in Chapter 4 of this thesis (Experiment 3) that individual differences measured by scores on a half-timed procedure for NDRT comprehension test (Andrews et al., 2020) are sensitive to changes in reading strategies for skilled adult readers, and as such this test may be appropriate for similar investigations in future research. The occurrence of a Jingle fallacy in eye movement research indicates a need for clear guidance on best practices when selecting tests and reporting findings based upon them. We suggest that researchers follow recent guidance from Flake and Fried (2020), to be transparent when reporting how measures were selected for research purposes by answering six questions (*1. What is your construct? 2. Why and how did you select your measure? 3. What measure did you use to operationalize the construct? 4. How did you quantify your measure? 5. Did you modify the scale? And if so, how and why? 6. Did you create a measure on the fly?*). Future research should aim to formulate standardised reading comprehension tests that are designed for the purpose of scientific research, based on theories of reading comprehension. Alternatively, researchers should be guided by theories of reading comprehension to select relevant underlying skills of interest for future investigations of skilled reading.

The Simple View of Reading (Gough & Tunmer, 1986) and Kintsch's 1998 Construction-Integration Model are two prominent theoretical frameworks in the field of reading that can guide researchers in selecting an optimal test battery for examining individual differences in reading skills. The Simple View of Reading theorizes that reading comprehension is the product of two main components: decoding and language comprehension. To assess decoding skills, researchers may include both word and pseudoword decoding from written materials such as those used in the current test battery. Considering our findings in the empirical chapters of this thesis concerning the appropriate use of reading comprehension within test batteries, listening comprehension is likely to be a more appropriate test to gauge language comprehension skills than a test of reading comprehension.

Kintsch's model describes the active construction of the meaning of a text that occurs during reading. Within this theory, Kintsch emphasizes the processes of understanding the explicit information within a text and the subsequent integration of this information with existing knowledge. In this model, a text is first represented by single word meanings within a text (surface structure), then by integration of these words to create a meaningful representation of the text as a whole (textbase), and finally a model of the situation that the text lies within, integrating the information in the text with relevant word knowledge (situation model). To assess these processes, researchers must consider each stage of comprehension.

A representation of the surface structure of a text is inherently linked to word decoding and vocabulary. An individual must first decode written information and match it to its corresponding semantic information held in the mental lexicon. Assessments of both word and pseudoword decoding, along with vocabulary would be appropriate measures to assess the processes involved in creating a surface representation of a text.

Next, to assess an individual’s ability to integrate separate word meanings into a representation of a text as a whole (textbase), assessment should focus on whether individuals can summarise the main ideas of a text, infer text meaning and integrate information across different texts or parts of the same text (e.g., the multiple-text integration paradigm by Beker et al., 2016). Finally, to assess integration of information gathered from a text with prior world knowledge, researchers must use assessments that measure background knowledge related to a text and the ability to apply this knowledge to enable better interpretation of a text. For example, the Global Integrated Scenario-based Assessment (GISA; O’Reilly & Sabatini, 2013; Sabatini, O’Reilly, & Deane, 2013) measures a reader’s ability to integrate, evaluate, and synthesize multiple texts and includes a measure of the reader’s ‘domain and topic specific’ background knowledge.

Finally, throughout the experiments included in this thesis we gained insight into cognitive skills that influence faster word recognition processes and passage reading strategies in adults.

This insight may inform future investigations that wish to assess the effectiveness of interventions or training for these skills to improve the reading skills of developing or struggling readers.

5.5 Future Directions

In this thesis we established that the quality of skilled readers' lexical representations are good indicators of individual differences in the moment-to-moment processing of words within sentences, and as a result, the time taken to read whole sentences. Therefore, future investigations should investigate to what extent lexical proficiency influences later processes involved in passage reading for skilled adult readers. Such investigations should consider the relative importance of multiple cognitive skill predictors, as was originally intended for the final experiment in this thesis. In Section 5.3.2. we highlighted a limitation of the current research in identifying important influences of RAN and WMC in later reading processes. However, we anticipate that in an investigation of passage reading across multiple skills, faster RAN times would predict fewer regressions and reduced rereading behaviours in skilled readers, as this measure has been found to be important in predicting later reading processes in previous studies (Kuperman & Van Dyke, 2011; Gordon et al., 2020). Future studies should also investigate readers' responses to changes in high-level cognitive influences, such as comprehension demands investigated in Experiment 3, or text difficulty, in relation to a larger test battery. Such experiments should consider including a measure of intelligence, for example non-verbal reasoning to minimise conceptual overlap with linguistic tasks, that may play an important role in integrating text meanings with world knowledge (Kintsch, 1998).

Future studies could also explore the maturation in flexible letter position encoding with longitudinal studies from developing readers to skilled adults, to bridge the gap between studies of individual differences in children (Pagán et al., 2021; Hasenäcker & Schroeder, 2022) and the investigation with skilled adult readers in Chapter 3. Hasenäcker and Schroeder (2022) found that children's orthographic knowledge (as measured by spelling, vocabulary and a word-reading to nonword-reading difference score) was a good predictor of flexible letter position encoding across grades 2 to 4. The skills that Hasenäcker and Schroeder (2022) associated with orthographic knowledge were similar to the skills we associated with lexical proficiency (PC1) in Chapter 2 and Chapter 3. An extension of this study to follow developing readers to adulthood would allow researchers to understand in what circumstances maturation of flexible letter position encoding occurs, and whether orthographic skills remain the strongest predictors of differences across reading development.

Finally, to overcome the Jingle fallacy for comprehension tests used for scientific research, we suggested in Section 5.4 that there is a need for a new standardised measure of reading comprehension for research purposes. Future studies would first be required to assess the relative contribution of cognitive skills that are integral to successful reading comprehension (measured by performance on experimental tasks rather than offline comprehension tests) across a range of reading materials with varied complexity and both narrative and expository texts for skilled adults. Such studies would provide data driven evidence for the key underlying skills for reading comprehension success that would form a good basis for developing a new standardised measure for future research.

5.6 Concluding Summary

By examining individual differences in skilled adult readers' eye movements, we determined that the most consistent predictor of individual differences in the skilled reader's word identification processes is related to lexical proficiency. In line with the Lexical Quality Hypothesis (Perfetti, 1992; 2007; Perfetti & Hart, 2002), we highlighted that readers with greater lexical proficiency read sentences more quickly than readers with low quality lexical representations, and that this difference is facilitated by faster word identification processes, especially when encountering less frequent words. This research also found evidence that letter position encoding processes approach maturation in skilled adult readers with very little differences related to cognitive skills. In addition, this body of work uncovered individual differences in how readers adapt their behaviours over trials in response to high and low comprehension demands. Skilled comprehenders (as measured by the NDRT comprehension subtest) were fairly consistent in that their reading strategy was to speed up, and read less carefully over time, whereas less skilled comprehenders reached a threshold for how quickly they could read when comprehension demands were low. However, when comparing predictive models based on two distinct offline comprehension measures (NDRT and WIAT-II), we uncovered evidence that a Jingle fallacy exists where the tests were not found to describe performance on the same underlying construct related to reading comprehension. As a result, this research highlighted an important limitation for studies that have used either reading comprehension measure (NDRT and WIAT-II) to compare skilled readers' eye movement patterns across studies and we gave recommendations for future research practice.

List of References

Aaronson, D., & Ferres, S. (1984). Reading strategies for children and adults: Some empirical evidence. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 189–220. https://doi.org/10.1016/S0022-5371(84)90137-3

Aaronson, D., & Ferres, S. (1986). Reading strategies for children and adults: A quantitative model. *Psychological Review, 93*(1), 89–112. https://doi.org/10.1037/0033-295X.93.1.89

Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence*. Journal of Experimental Psychology: Human Perception and Performance, 2*, 56–70. https://doi.org/10.1037/0096-1523.2.1.56

Acheson, D. J., Wells, J. B., & Macdonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students, *Behavioural Research Methods, 40*(1), 278–289. https://doi.org/10.3758/BRM.40.1.278

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., Perry, J. R., Lupker, S. J., Forster, K. I., Cortese, M. J., Scaltritti, M., Aschenbrenner, A. J., Coane, J. H., White, L., Yap, M. J., Davis, C., Kim, J., & Davis, C. J. (2014). A behavioral database for masked form priming. *Behavior Research Methods, 46*(4), 1052–1067. https://doi.org/10.3758/s13428-013-0442-y

Andrews, S. (2008). Lexical expertise and reading skill. *Psychology of Learning and Motivation, 49*, 247-281. https://doi.org/10.1016/S0079-7421(08)00007-8

Andrews S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In Adelman J. (Ed.), *Visual word recognition*, (pp. 151–172). Hove: Psychology Press.

Andrews, S. (2015). Individual differences among skilled readers: The role of lexical quality. In A. Pollatsek & R. Treiman (Eds.), *The Oxford Handbook of Reading* (pp. 129-148): Oxford University Press.

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General, 139*(2), 299–318. https://doi.org/10.1037/a0018366

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 152–163. https://doi.org/10.1037/a0024953

Andrews, S., & Veldre, A. (2019). What is the most plausible account of the role of parafoveal processing in reading? *Language and Linguistics Compass* 46-53. https://doi.org/10.1111/lnc3.12344

Andrews, S., & Veldre, A. (2021). Wrapping up sentence comprehension: The role of task demands and individual differences. *Scientific Studies of Reading, 25*(2), 123–140. https://doi.org/10.1080/10888438.2020.1817028

Andrews, S., Veldre, A. & Clarke, I. E. (2020). Measuring Lexical Quality: The Role of Spelling Ability. *Behavior Research Methods, 52,* 2257-2282. https://doi.org/10.3758/s13428-020-01387-3

Angele, B., Laishley, A. E., Rayner, K., & Liversedge, S. P. (2014). The effect of high and low frequency previews and sentential fit on word skipping during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(4), 1181–1203. https://doi.org/10.1037/a0036396

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 58*(6), 1065–1086. https://doi.org/10.1080/02724980443000476

Ashby, J., Yang, J., Evans, K. H. C., & Rayner, K. (2012). Eye movements and the perceptual span in silent and oral reading. *Attention, Perception, & Psychophysics, 74*(4), 634–640. https://doi.org/10.3758/s13414-012-0277-0

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. https://doi.org/10.1016/0010-0285(85)90013-1

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278. http://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B, 57,* 289--300. http://www.jstor.org/stable/2346101.

Best, R.M., Floyd, R.R., & McNamara, D.S. (2008) Differential competencies Contributing to children's comprehension of narrative and expository texts, *Reading Psychology, 29*(2), 137-164, https://doi.org/10.1080/02702710801963951

Binder, K. S., Pollatsek, A., & Rayner, K. (1999). Extraction of information to the left of the fixated word in reading. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 1162–1172. https://doi.org/10.1037/0096-1523.25.4.1162Drieghe et al., 2004

Bisanz, G. L., Das, J. P., Varnhagen, C. K., & Henderson, H. R. (1992). Structural components of reading time and recall for sentences in narratives: exploring changes with age and reading ability. *Journal of Educational Psychology, 84*(1), 103–114. https://doi.org/10.1037/0022-0663.84.1.103

Blythe, H. I., & Joseph, H. S. (2011). Children's eye movements during reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 644-662). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0036

Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology. The science of reading: A handbook* (pp. 155-172). Malden, Blackwell Publishing. http://doi.org/10.1002/9780470757642.ch9

Bowyer-Crane, C., Snowling, M.J., Duff, F.J., Fieldsend, E., Carroll, J.M., Miles, J., Götz, K. & Hulme, C. (2008), Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry, 49:* 422-432. https://doi.org/10.1111/j.1469-7610.2007.01849.x

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. Journal of learning disabilities, 40(3), 226-243. https://doi.org/10.1177/00222194070400030401

Brown, J. A., Fishco, V. V., & Hanna, G. (1993). *Nelson-denny reading test: Manual for scoring and interpretation, forms G & H.* Rolling Meadows, IL: Riverside Publishing

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Cognitive processes in eye guidance* (pp. 53–77). Oxford: Oxford University Press.

Byrne, B. J. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle.* Psychology Press.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

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

Chateau, D., & Jared, D. (2000) Exposure to print and word recognition processes. *Memory and Cognition 28,* 143–153. https://doi.org/10.3758/BF03211582

Christianson, K., Luke, S. G., Hussey, E. K., & Wochna, K. L. (2017). Why reread? Evidence from garden-path and local coherence structures. *Quarterly Journal of Experimental Psychology, 70*, 1380–1405. https://doi.org/10.1080/ 17470218.2016.1186200

Clarke, P., Hulme, C., & Snowling, M. (2005). Individual differences in RAN and reading: a response timing analysis. *Journal of Research in Reading, 28*, 73–86. https://doi.org/10.1111/j.1467-9817.2005.00255.x

Clifton, C., & Staub, A. (2011). Syntactic influences on eye movements during reading. In Liversedge, S. P., Gilchrist, I. D. and Everling, S. (Eds.), *Oxford handbook on eye movements,* 767–786. Oxford, UK: Oxford University Press. 896–909. https://doi.org/10.1093/oxfordhb/9780199539789.013.0049

Clifton C.., Traxler, M.J., Mohamed, M.T., Williams, R.S., Morris, R.K., & Rayner, K. (2003) The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language. 49,* 317–334. https://doi.org/10.1016/S0749-596X%2803%2900070-6

Coleman, C., Lindstrom, J., Nelson, J., Lindstrom, W., & Gregg K. N. (2010) Passageless comprehension on the Nelson-denny reading test: well above chance for university students. *Journal of Learning Disabilities 43*(3):244-9. https://doi.org/10.1177/0022219409345017.

Colombo, L., Sulpizio, S., & Peressotti, F. (2017). Serial mechanism in transposed letters effects: A developmental study. *Journal of Experimental Child Psychology, 161*, 46–62. https://doi.org/10.1016/j.jecp.2017.04.002

Colombo, L., Sulpizio, S., & Peressotti, F. (2019). The developmental trend of transposed letters effects in masked priming. *Journal of Experimental Child Psychology,186,* 117–130. https://doi.org/10.1016/j.jecp.2019.05.007

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), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Cop, U., Dirix, N., Drieghe, D. & Duyck, W (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods 49*, 602–615 https://doi.org/10.3758/s13428-016-0734-0

Cop, U., Keuleers, E., Drieghe, D. & Duyck, W. (2015) Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review 22,* 1216–1234. https://doi.org/10.3758/s13423-015-0819-2

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities, mechanisms, and performance* (pp. 409–422). Amsterdam, Netherlands: North-Holland.

Cuetos, F., Rodríguez, B., Ruano, E., & Arribas, D. (2007). *PROLEC-R: Català—Bateria d'avaluació dels processos lectors [PROLEC-R: Catalan—Battery of assessment of reading processes].* TEA Ediciciones.

Cunningham, E., Perry, K.E., & Stanovich, K.E. (2001) Converging evidence for the concept of orthographic processing. *Reading and Writing, 14*, 549–568. https://doi.org/10.1023/A:1011100226798

Cutting, L. E., & Denckla, M. B. (2001). The relationship of rapid serial naming and word reading in normally developing readers: An exploratory model. *Reading and Writing, 14,* 673-705. https://doi.org/10.1023/A:1012047622541

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277-299. https://doi.org/10.1207/s1532799xssr1003_5

Dancey, C. P., & Reidy, J. (2007). *Statistics without Maths for Psychology.* Pearson Education.

Daneman, M., & Carpenter, P. A. (1980) Individual differences in working memory and reading, *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. https://doi.org/10.1016/S0022-5371(80)90312-6.

Davis, C. J. (1999). *The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition.* Unpublished doctoral dissertation, University of New South Wales, Australia. https://doi.org/10.26190/unsworks/13769

Davis, C. J. (2010). SOLAR versus SERIOL revisited*. European Journal of Cognitive Psychology, 22,* 695–724. https://doi.org/10.1080/09541440903155682

Deary, I. (2000). *Looking down on human intelligence: From psychometrics to the brain* (Vol. 36). England: Oxford University Press.

Denckla, M. B. (1988). *What does it take to learn (after you've learned to read)? in* Paper Presented at the Research Symposium Conducted at the Meeting of the New York Orton Society, (New York, NY, USA: Literacy Assistance Center).

Denckla, M. B., & Rudel, R. (1974). Rapid "automatized" naming of pictured objects, colors, letters and numbers by normal children. *Cortex, 10*(2), 186-202. https://doi.org/10.1016/S0010-9452(74)80009-2

Department of Education. (2021). *The reading framework: Teaching the foundations of literacy.* United Kingdom Department of Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1102800/Reading_framework_teaching_the_foundations_of_literacy_-_Sept_22.pdf

Determine the main idea of the passage. (2021, October 27). Retrieved from http://www.uk.ixl.com

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dirix, N., & Duyck, W. (2017). An eye movement corpus study of the age-of-acquisition effect. *Psychomomic Bulletin & Review, 24*(6), 1915–1921. https://doi.org/10.3758/s13423-017-1233-8

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. https://doi.org/10.1080/09541440340000141

Duncan, S., & Freeman, M. (2020) Adults reading aloud: A survey of contemporary practices in Britain. *British Journal of Educational Studies, 68*(1), 97-123, https://doi.org/10.1080/00071005.2019.1610555

Dunn, L.M., & Dunn, D.M. (2007) *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader–text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology, 104*(3), 515–528. https://doi.org/10.1037/a0027182

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. https://doi.org/10.1016/S0022-5371(81)90220-6

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42*(5), 621–636. https://doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777-813. https://doi.org/10.1037/0033-295X.112.4.777

Ernestus, M., Kouwenhoven, H., & Van Mulken, M. (2017). The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics, 62,* 50-64. https://doi.org/10.1016/j.wocn.2017.02.003

Eskenazi, M. A. (2023). Best practices for cleaning eye movement data in reading research. *Behavior Research Methods,* 1-11. https://doi.org/10.3758/s13428-023-02137-x

Flake, J.K., & Fried, E.I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science. 3*(4):456-465. https://doi.org/10.1177/2515245920952393

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 39A* (2), 211–251. https://doi.org/10.1080/14640748708401785

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology, 14*(2), 178-210. https://doi.org/10.1016/0010-0285(82)90008-1

Freed, E. M., Hamilton, S. T., & Long, D. L. (2017). Comprehension in proficient readers: The nature of individual variation. *Journal of Memory and Language, 97*, 135–153. https://doi.org/10.1016/j.jml.2017.07.008

Furnes, B., & Samuelsson, S. (2010) Predicting reading and spelling difficulties in transparent and opaque orthographies: a comparison between Scandinavian and US/Australian children. *Dyslexia. 16*(2):119-42. http://doi.org/10.1002/dys.401.

Gambrell, L. B., & Heathington, B. S. (1981). Adult disabled readers' metacognitive awareness about reading tasks and strategies. *Journal of Reading Behavior, 13*(3), 215-222. https://doi.org/10.1080/10862968109547409

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84*(1), 74–111. https://doi.org/10.3102/0034654313499616

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gómez, P., Marcet, A., & Perea, M. (2021). Are better young readers more likely to confuse their mother with their mohter? *Quarterly Journal of Experimental Psychology, 74*(9), 1542–1552. https://doi.org/10.1177/17470218211012960

Gómez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115,* 577–600. http://doi.org/10.1037/a0012667

Gordon, P. C., Moore, M., Choi, W., Hoedemaker, R. S., & Lowder, M. W. (2020). Individual differences in reading: Separable effects of reading experience and processing skill. *Memory and Cognition, 48,* 553-565. https://doi.org/10.3758/s13421-019-00989-3

Goswami, U., & Bryant, P. (2016). *Phonological skills and learning to read.* Psychology Press.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. RASE: Remedial and *Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Graesser, A. C., Kassler, M. A., Kreuz, R. J., & McLain-Allen, B. (1998). Verification of statements about story worlds that deviate from normal conceptions of time: What is true about Einstein's Dreams? *Cognitive Psychology, 35*(3), 246–301. https://doi.org/10.1006/cogp.1998.0680

Grainger, J., Granier, J. P., Farioli, F., Van Assche, E., & van Heuven, W. J. B. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 865–884. http://doi.org/10.1037/0096-1523.32.4.865

Grainger, J., Lété, B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition, 123*(2), 280–292. https://doi.org/10.1016/j.cognition.2012.01.003

Grainger, I. J., & Van Heuven, W. J. B. (2003). Modelling letter position coding in printed word perception. In P. Bonin (Ed.), *Mental lexicon: "Some words to talk about words"* (pp. 1–23). Hauppauge, NY: Nova Science.

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology, 2*, 54. http://doi.org/10.3389/fpsyg.2011.00054

Green P., & MacLeod C.J. (2016). simR: an R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504, https://CRAN.R-project.org/package=simr.

Haenggi, D., & Perfetti, C. A. (1994). Processing components of college-level reading comprehension. *Discourse Processes, 17*(1), 83-104. http://doi.org/10.1080/01638539409544860

Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology, 102,* 167–181. http://doi.org/10.1016/j.jecp.2008.04.002

Hale, A. D., Skinner, C. H., Williams, J., & Hawkins, R., Neddenriep, C. E., & Dizer, J. (2007). Comparing comprehension following silent and aloud reading across elementary and secondary students: Implication for curriculum-based measurement. *The Behavior Analyst Today. 8*(1), 9. https://doi.org/10.1037/h0100101.

Hamilton, S.T., Freed, E.M., & Long, D.L. (2013). Modelling reader and text interactions during narrative comprehension: A test of the lexical quality hypothesis, *Discourse Processes, 50*:2, 139-163, https://doi.org/10.1080/0163853X.2012.742001

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review, 111*(3), 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Hasenäcker, J., & Schroeder, S. (2022). Transposed and substituted letter effects across reading development: A longitudinal study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(8), 1202–1218. https://doi.org/10.1037/xlm000106

Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision research, 51*(15), 1719-1727. https://doi.org/10.1016/j.visres.2011.05.015

Heller, D. (1982). Eye movements in reading. In R. Groner & P. Fraisse (Eds.), *Cognition and eye movements* (pp. 139–154). Amsterdam: North Holland

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 417–429. http://doi.org/10.1037/0278-7393.16.3.417

Hoosain, R. (1992). Psychological reality of the word in Chinese. In H.C. Chen and O. J. L. Tzeng (Eds.) *Advances in psychology* (Vol. 90, pp. 111-130). Amsterdam: North-Holland Publishing Company.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika 30*, 179–185. https://doi.org/10.1007/BF02289447

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.*), Eye guidance in reading and scene perception* (pp. 29–53). Elsevier Science Ltd. https://doi.org/10.1016/B978-008043361-5/50003-1

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics, 40*(6), 431–439. https://doi.org/10.3758/BF03208203

Jackson, N. E. (2005). Are university students' component reading skills related to their text comprehension and academic achievement? *Learning and Individual Differences, 15*(2), 113–139. https://doi.org/10.1016/j.lindif.2004.11.001

Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from proofreading and eye movements. *Journal of Experimental Psychology: General, 128*(3), 219–264. https://doi.org/10.1037/0096-3445.128.3.219

Johnson, R. L., & Dunne, M. D. (2012). Parafoveal processing of transposed-letter words and nonwords: Evidence against parafoveal lexical activation. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 191–212. https://doi.org/10.1037/a0025983

Johnson, R. L., Perea, M., & Rayner, K. (2007). Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview. *Journal of Experimental Psychology: Human Perception and Performance, 33(*1), 209–229. https://doi.org/10.1037/0096-1523.33.1.209

Jones, M. W., Snowling, M. J., & Moll, K. (2016). What automaticity deficit? Activation of lexical information by readers with dyslexia in a rapid automatized naming Stroop-switch task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 465. https://doi.org/10.1037/xlm0000186

Joseph, H. S. S. L., Bremner, G., Liversedge, S. P., & Nation, K. (2015) Working memory, reading ability and the effects of distance and typicality on anaphor resolution in children, *Journal of Cognitive Psychology, 27*:5, 622-639, https://doi.10.1080/20445911.2015.1005095

Juhasz, B. J., & Rayner, K. (2003). Investigating the Effects of a Set of Intercorrelated Variables on Eye Fixation Durations in Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1312–1318. https://doi.org/10.1037/0278-7393.29.6.1312

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition, 13*(7-8), 846-863. https://doi.org/10.1080/13506280544000075

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

Just M. A., & Carpenter P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology, 30*(6), 949. https://doi.org/10.1037/0012-1649.30.6.949.

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist, 73*(1), 1–2. https://doi.org/10.1037/amp0000263

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281–300. https://doi.org/10.1080/10888430802132279

Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction, 22*(5), 354–367. https://doi.org/10.1016/j.learninstruc.2012.02.001

Kendeou, P., Van Den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research and Practice, 29*(1), 10-16. https://doi.org/10.1111/ldrp.12025

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge University Press.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.*), The science of reading: A handbook* (pp. 209–226). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch12

Kirby, J. R. (2007). *Reading comprehension: Its nature and development. Encyclopedia of Language and Literacy Development.* London, ON: Canadian Language and Literacy Research Network, 1-8.

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45,* 341–362. https://doi.org/10.1598/RRQ.45.3.4

Kirby, J.R., & Woodhouse, R.A. (1994). Measuring and Predicting Depth of Processing in Learning. *Alberta Journal of Educational Research, 40*, 147-161.

Kirkby, J.A., Barrington, R.S., Drieghe, D., & Liversedge, S.P. (2022) Parafoveal processing and transposed-letter effects in dyslexic reading. *Dyslexia. 28*(3), 359-374. https://doi.org/10.1002/dys.1721

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*(1-2), 262-284. https://doi.org/10.1080/09541440340000213

Kuperman, V., Matsuki, K., & Van Dyke, J. A. (2018). Contributions of reader- and text-level characteristics to eye-movement patterns during passage reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(11), 1687–1713. https://doi.org/10.1037/xlm0000547

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language, 65*(1), 42–73. https://doi.org/10.1016/j.jml.2011.03.002

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology, Human Perception and Performance, 39*(3), 802–823. https://doi.org/10.1037/a0030859

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16(6), 1118–1133. https://doi.org/10.1037/0278-7393.16.6.1118

Laubrock, J., & Kliegl, R. (2015). The eye-voice span during reading aloud. *Frontiers in Psychology. 6*(1432). https://doi.org/10.3389/fpsyg.2015.01432

Lee, C. E., Godwin., H. J., Blythe, H. I., & Drieghe, D. (2023). *Individual differences in skilled reading and the word frequency effect*. Manuscript Submitted for Publication.

Lee, C. E., Pagán, A., Godwin., H. J., & Drieghe, D. (2023). *Individual differences and the transposed letter effect during reading.* Manuscript Submitted for Publication.

Lefly, D.L., & Pennington, B.F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities, 33,* 286–296. https://doi.org/10.1177/002221940003300306

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods, 44*, 325-343. https://doi.org/10.3758/s13428-011-0146-0

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences. 4(*1):6-14. https://doi.org/10.1016/s1364-6613(99)01418-7.

Lo, S. & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6:*1171. http:/dx.doi.org/10.3389/fpsyg.2015.01171

Long, D. L., & Freed, E. M. (2021). An individual differences examination of the relation between reading processes and comprehension. *Scientific Studies of Reading, 25*:2, 104-122, https://doi.org/10.1080/10888438.2020.1748633

Luke, S. G., Henderson, J. M., & Ferreira, F. (2015). Children's eye-movements during reading reflect the quality of lexical representations: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 41*(6), 1675–1683. https://doi.org/10.1037/xlm0000133

Lupker, S. J., Perea, M., & Davis, C. J. (2008). Transposed-letter effects: Consonants, vowels and letter frequency. *Language and Cognitive Processes, 23*(1), 93–116. https://doi.org/10.1080/01690960701579714

Martino, N. L., & Hoffman. P. R. (2002). An investigation of reading and language abilities of college freshmen. *Journal of Research in Reading, 25*, 310–318. https://doi.org10.1111/1467-9817.00178

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*(12), 899–917. https://doi.org/10.1037/h0037368

Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The Random Forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading, 20*(1), 20–33. https://doi.org/10.1080/10888438.2015.1107073

May, P. (2002). *HSP 1-9. Diagnose orthographischer Kompetenz. Zur Erfassung der grundlegenden Rechtschreibstrategien mit der Hamburger Schreibprobe. Neustandardisierung 2001 [HSP 1-9. Diagnosis of orthographic competencies. Capturing basic orthographic strategies with the Hamburg writing test. New standardization 2001].* Klett Verlag.

McCarron, S.P., & Kuperman, V. (2021). Is the author recognition test a useful metric for native and non-native English speakers? An item response theory analysis. *Behavioural Research Methods, 53,* 2226–2237. https://doi.org/10.3758/s13428-021-01556-y

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception. An account of basic findings*. Psychological Review, 88*, 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision research, 28*(10), 1107-1118. https://doi.org/10.1016/0042-6989(88)90137-X

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics,* 17, 578–586. http://doi.org/10.3758/BF03203972

McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education, 4*(1), 229–257.

Mézière, D. C., Yu, L., Reichle, E. D., Von Der Malsburg, T., & McArthur, G. (2021). Using Eye-Tracking Measures to Predict Reading Comprehension. *Reading Research Quarterly.* https://doi.org/10.1002/rrq.498

Mitchell, D. C., Shen, X., Green, M. J., & Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis. *Journal of Memory and Language, 59*(3), 266-293. https://doi.org/10.1016/j.jml.2008.06.002

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137,* 267–296. https://doi.org/10.1037/a0021890.

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest: Weiterentwicklung der Salzburger Lese- und Rechtschreibtests (SLRT): Manual [SLRT-II: Reading and writing test: Advancement of the Salzburg reading and writing test (SLRT)].* Verlag Hans Huber.

Monsell, S. (1991). The nature and locus of word frequency effects inreading. In D. Besner & G. W. Humphreys (Eds.), *Basic Process in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Erlbaum.

Moore, M., & Gordon, P.C. (2015). Reading ability and print exposure: Item response theory analysis of the author recognition test. *Behavior Research Methods. 47*(4), 1095-1109. https://doi.org/10.3758/s13428-014-0534-3.

Morton, J. (1970). A functional model of human memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203–254). New York: Academic Press

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665–681. https://doi.org/10.1037/0012-1649.40.5.665

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327–357. https://doi.org/10.1037/0033-295X.113.2.327

Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review, 119*(3), 517–545. https://doi.org/10.1037/a0028450

O'Connor R. E. & Forster K. I. (1981). Criterion bias and search sequence bias in word recognition. *Memory and Cognition 9,* 78–92. https://doi.org/10.3758/BF03196953

Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In: Lyon, G. R. (Ed), *Frames of reference for the assessment of learning disabilities: New views on measurement issues.* Paul H Brookes Publishing, Baltimore, pp 243–277

O'Regan J. K. (1990). Eye movements and reading. In Kowler, E. (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 395-453). Amsterdam: Elsevier, North-Holland.

O'Regan, J. K. (1992). Optimal viewing position in words and the strategy-tactics theory of eye movements in reading. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 333-354). New York: Springer-Verlag https://doi.org/10.3758/BF03213829

Pagán, A., Blythe, H. I., & Liversedge, S. P. (2016). Parafoveal preprocessing of word initial trigrams during reading in adults and children. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 411-432. https://doi.org/10.1037/xlm0000175

Pagán, A., Blythe, H. I., & Liversedge, S. P. (2021). The influence of children's reading ability on initial letter position encoding during a reading-like task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(7), 1186–1203. https://doi.org/10.1037/xlm0000989

Papadopoulos, T.C., Spanoudis, G.C., & Georgiou, G.K. (2016). How is RAN related to reading fluency? A comprehensive examination of the prominent theoretical accounts. *Frontiers in Psychology. 7*:1217. https://doi.org/10.3389/fpsyg.2016.01217

Perea, M., & Fraga, I. (2006). Transposed-letter and laterality effects in lexical decision. *Brain and Language, 97*(1), 102–109. https://doi.org/10.1016/j.bandl.2005.08.004

Perea, M., & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art,* (pp. 97–120). Psychology Press.

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51*(2), 231–246. https://doi.org/10.1016/j.jml.2004.05.005

Perfetti, C. A. (1985). *Reading ability.* New York, NY, US: Oxford University Press.

Perfetti, C. A. (1992) The representation problem in reading acquisition. In Gough PB, Ehri LC and Treiman R (eds), *Reading acquisition.* Hillsdale, NJ: Erlbaum, pp. 145–174.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67–86). American Psychological Association. https://doi.org/10.1037/10459-004

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy, 11*, 67-86. https://doi.org/10.1075/swll.11.14per

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive psychology, 61*(2), 106-151. https://doi.org/10.1016/j.cogpsych.2010.04.001

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56

Pollatsek, A., Juhasz, B. J., Reichle, E. D., Machacek, D., & Rayner, K. (2008). Immediate and delayed effects of word frequency and word length on eye movements in reading: a reversed delayed effect of word length. *Journal of Experimental Psychology: Human Perception and Performance, 34(*3), 726. https://doi.org/10.1037/0096-1523.34.3.726

Pollatsek, A., Rayner, K., & Balota, D. A. (1986). Inferences about eye movement control from the perceptual span in reading. *Perception and Psychophysics, 40*, 123–130. https://doi.org/10.3758/BF03208192

Powell, D., Stainthorp, R., Stuart, M., Garwood, H., & Quinlan, P. (2007). An experimental comparison between rival theories of rapid automatized naming performance and its relationship to reading. *Journal of Experimental Child Psychology, 98*(1), 46-68. https://doi.org/10.1016/j.jecp.2007.04.003

Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2007). Development of lexical mediation in the relation between reading comprehension and word reading skills in Greek. *Scientific Studies of Reading, 11*(3), 165-197. https://doi.org/10.1080/10888430701344322

R Core Team (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Radach, R. R., Huestegge, L., & Reilly, R. G. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research, 72,* 675-688. https://doi.org/10.1007/s00426-008-0173-3

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

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

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

Rayner, K. (2009). The Thirty-fifth Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Ashby, J., Pollatsek, A., & Reichte, 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. https://doi.org/10.1037/0096-1523.30.4.720

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science, 206*(4417), 468-469. https://doi.org/10.1126/science.504987

Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research, 41*(7), 943-954. https://doi.org/10.1016/S0042-6989(00)00310-2

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10*(3), 241-255. https://doi.org/10.1207/s1532799xssr1003_3

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition, 14*(3), 191-201. https://doi.org/10.3758/BF03197692

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception and Psychophysics, 58*(5), 734–747. https://doi.org/10.3758/BF03213106

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 31-74. https://doi.org/10.1111/1529-1006.00004

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *The Quarterly Journal of Experimental Psychology, 39*(4), 657-673. https://doi.org/10.1080/14640748708401808

Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental*

Psychology: Human Perception and Performance, 7, 167–179. http://doi.org/10.1037/0096-1523.7.1.167

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A, 53*(4), 1061-1080. https://doi.org/10.1080/713755934

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging, 21,* 448– 465. https://doi.org/10.1037/0882-7974.21.3.448

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So much to read, so little time: How do we read, and can speed reading help*? Psychological Science in the Public Interest, 17*(1), 4-34. https://doi.org/10.1177/1529100615623267

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1188–1200. https://doi.org/10.1037/0096-1523.22.5.1188

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 4*(3-4), SI21–SI49. https://doi.org/10.1080/01690968908406362

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance, 37*(2), 514–528. https://doi.org/10.1037/a0020990

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. https://doi.org/10.3758/BF0321455

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics, 27*(6), 537-544. https://doi.org/10.3758/BF03198682

Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding wrods with jubmled letters: There's a cost. *Psychological Science, 17,* 192–193.https://doi.org/10.1111/j.1467-9280.2006.01684.x

Ready, R. E, Chaudhry, M. F., Schatz, K.C., & Strazzullo, S. (2013). "Passageless" administration of the nelson-denny reading comprehension test: Associations with IQ and reading skills. *Journal of Learning Disabilities 46*(4):377-84. http://doi.org/10.1177/0022219412468160.

Reichle, E.D. (2011). Serial attention models of reading. In Liversedge, S. P., Gilchrist, I. D. and Everling, S. (Eds.), *Oxford handbook on eye movements,* (pp767–786). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0042

Reichle, E. D. (2020). *Computational models of reading: A handbook.* Oxford University Press.

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S., White, S. J., & Rayner, K. (2013). Using EZ Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review, 33*(2), 110-149. https://doi.org/10.1016/j.dr.2013.03.001

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125-157. http://doi.org/10.1037/0033-295X.105.1.125

Reichle, E.D., Rayner, K., & Pollatsek, A. (1999) Eye movement control in reading: Accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research, 39*(26), 4403-4411.

Reingold, E. M., Yang, J., & Rayner, K. (2010). The time course of word frequency and case alternation effects on fixation times in reading: Evidence for lexical control of eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 36*(6), 1677–1683. https://doi.org/10.1037/a0019959

Richter, E. M., Engbert, R., & Kliegl, R. (2006). Current advances in SWIFT. *Cognitive Systems Research, 7*(1), 23-33. https://doi.org/10.1016/j.cogsys.2005.07.003

Salthouse, T. A. (1988). The role of processing resources in cognitive aging. In M. L. Howe & C. J. Brainerd (Eds.), *Cognitive development in adulthood* (pp. 222-242). New York, NY: Springer-Verlag.

Savage, R., Pillay, V., & Melidona, S. (2007). Deconstructing rapid automatized naming: Component processes and the prediction of reading difficulties. *Learning and Individual Differences, 17*(2), 129-146. https://doi.org/10.1016/j.lindif.2007.04.001

Schmidtke, D., Van Dyke, J. A., & Kuperman, V. (2018). Individual variability in the semantic processing of English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(3), 421–439. https://doi.org/10.1037/xlm0000442

Schoonbaert, S., & Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes, 19*(3), 333–367. https://doi.org/10.1080/01690960344000198

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, and Psychophysics, 74*(1), 5–35. https://doi.org/10.3758/s13414-011-0219-2

Slattery, T. J., & Yates, M. (2017). Word skipping: Effects of word length, predictability, spelling and reading skill. *Quarterly Journal of Experimental Psychology, 71*(1), 250–259. https://doi.org/10.1080/17470218.2017.1310264

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review, 125*(6), 969- 984. http://doi.org/10.1037/rev0000119

Sperlich, A., Schad, D. J., & Laubrock, J. (2015). When preview information starts to matter: Development of the perceptual span in German beginning readers. [Advance online publication]. *Journal of Cognitive Psychology, 27*, 511–530. http://doi.org/10.1080/20445911.2014.993990

S.R. Research Ltd. (2017). *Eyelink user manual* 1.0.12. Mississauga, Ontario, Canada

Stanovich, K. E. (1980). Towards an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32–71. https://doi.org/10.2307/747348

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading and Research Quarterly, 24,* 402–429. https://doi.org/10.2307/747605

Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1280–1293. https://doi.org/10.1037/a0016896

Stine-Morrow, E. A. L., Milinder, L., Pullara, O., & Herman, B. (2001). Patterns of resource allocation are reliable among younger and older readers. *Psychology and Aging, 16*, 69–84. https://doi.org/10.1037/0882-7974.16.1.69

Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements.* Teacher's College, Columbia University, Science Press, New York. https://doi.org/10.1037/13283-000

Tinker, M. (1958). Recent studies of eye movements in reading. *Psychological Bulletin, 55*, 215–231.

Torgesen, J.K., Wagner, R., & Rashotte, C. (1999) *TOWRE–2 test of word reading efficiency. Pro-Ed,* Austin.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997) Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children, *Scientific Studies of Reading, 1*(2), 161-185, https://doi.org/10.1207/s1532799xssr0102_4

Traxler, M.J., Long, D.L., Tooley, K.M., Johns, C.L., Zirnstein, M., & Jonathan, E. (2012). Individual Differences in Eye-Movements During Reading: Working Memory and Speed-of-Processing Effects*. Journal of Eye Movement Research, 5*(1), 5. https://doi.org/10.16910/jemr.5.1.5

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language, 28*(2), 127-154. https://doi.org/10.1016/0749-596X(89)90040-5

Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension: The relationship between reading skill, reading speed, and visual inspection. *Language and Speech, 33,* 69–81. https://doi.org/10.1177/002383099003300105

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. https://doi.org/10.3758/BF03192720

Van Den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah: Erlbaum.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition, 131*(3), 373-403. https://doi.org/10.1016/j.cognition.2014.01.007

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67,* 1176-1190.

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *Quarterly Journal of Experimental Psychology, 67*(4), 703–727. https://doi.org/10.1080/17470218.2013.826258

Veldre, A., & Andrews, S. (2015a). Parafoveal lexical activation depends on skilled reading proficiency. *Journal of Experimental Psychology: Learning Memory and Cognition, 41*(2), 586–595. https://doi.org/10.1037/xlm0000039

Veldre, A., & Andrews, S. (2015b). 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. https://doi.org/10.1037/xhp0000017

Veldre, A., & Andrews, S. (2016) Semantic preview benefit in English: Individual differences in the extraction and use of parafoveal semantic information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(6), 837–854. https://doi.org/10.1037/xlm0000212

Veldre, A., Drieghe, D., & Andrews, S. (2017). Spelling ability selectively predicts the magnitude of disruption in unspaced text reading. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1612–1628. https://doi.org/10.1037/xhp0000425

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception and Psychophysics, 57*, 352-364. https://doi.org/10.3758/BF03213060

Von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of Memory and Language, 94,* 119-133. https://doi.org/ 10.1016/j.jml.2016.10.003

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a grouping variable longitudinal study. *Developmental Psychology, 30*(1), 73–87. https://doi.org/10.1037/0012-1649.30.1.73

Waters G.S., & Caplan D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*(4), 761–772. https://doi.org/10.1037/0033-295X.103.4.761

Wechsler, D. (2001). *Wechsler Intelligence Scale for Children–Revised (Spanish adaptation).* TEA Ediciones.

Wechsler, D. (2005). *Wechsler Individual Achievement Test 2nd Edition (WIAT-II II).* London: The Psychological Corp.

Weiss, A. F., Kretzschmar, F., Schlesewsky, M., Bornkessel-Schlesewsky, I., & Staub, A. (2018). Comprehension demands modulate rereading, but not first-pass reading behavior. *Quarterly Journal of Experimental Psychology, 71*, 198–210. https://doi.org/10.1080/17470218.2017.1307862

Weiß, R. H. (2006*). Grundintelligenztest Skala 2 (CFT 20-R) mit Wortschatztest (WS) und Zahlenfolgentest (ZF) [Basic test of intelligence scale 2 (CFT20R) with vocabulary test (WS) and digit sequence test (ZF)].* Hogrefe.

Welcome, S.E., Leonard, C.M, & Chiarello, C. (2010) Alternate reading strategies and variable asymmetry of the planum temporale in adult resilient readers*. Brain and Language 113*, 73–83. https://doi.org/10.1016/j.bandl.2010.01.003

Welcome, S. E., & Trammel, E. R. (2017). Individual differences in orthographic priming relate to phonological decoding skill in adults*. Cognitive Processing, 18*(2), 119–128. https://doi.org/10.1007/s10339-017-0793-x

West, R. F., Stanovich, K. E., & Mitchell, H. R. (1993). Reading in the real world and its correlates. *Reading Research Quarterly,* 35-50. https://doi.org/10.2307/747815

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 205. https://doi.org/10.1037/0096-1523.34.1.205

White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: The importance of word beginning letters. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1261–1276. http://doi.org/10.1037/0096-1523.34.5.1261

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review, 8,* 221–243. http://doi.org/10.3758/BF03196158

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*(3), 415–438. https://doi.org/10.1037/0022-0663.91.3.415

World Literacy Foundation (2022). *Why Literacy?* Retrieved July 7, 2023, from https://worldliteracyfoundation.org/why-literacy-2/

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Quarterly Journal of Experimental Psychology, 66,* 548–562. https://doi.org/10.1080/17470218.2011.625094140S.

Ziegler, J. C., Bertrand, D., Lété, B., & Grainger, J. (2014). Orthographic and phonological contributions to reading development: Tracking developmental trajectories using masked priming. *Developmental Psychology, 50*(4), 1026–1036.https://doi.org/10.1037/a0035187

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., ... & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21*(4), 551-559. https://doi.org/10.1177/0956797610363406

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 920–933. https://doi.org/10.1037/0278-7393.20.4.