UNIVERSITY
OF OSLO

Master Thesis

The Effects of Interword Spacing on Visual Word Recognition

An Eye-tracking Study Using a Flanking-word Visual World Paradigm

Stefania A. Kyriakidou

Special Needs Education

40 Credits

Department of Special Needs Education

Faculty of Educational Sciences

Spring 2023

The Effects of Interword Spacing on Visual Word Recognition

An Eye-tracking Study Using a Flanking-word Visual World Paradigm

2023

The Effects of Interword Spacing on Visual Word Recognition

An Eye-tracking Study Using a Flanking-word Visual World Paradigm

Stefania Antonia Kyriakidou

https://www.duo.uio.no

Print: University of Oslo

Abstract

It has recently been reported that reading fluency goes beyond accuracy and speed. Efficient processing of word sequences has also been found to be a prerequisite for fluent reading. Processing successive words involves to some extent readers' ability to process not only the currently fixated word but also the upcoming words at a single glance. Parafoveal processing facilitates reading efficiency by partially preprocessing upcoming words during the preceding fixation leading to shorter subsequent fixations on these words. However, recent studies on single word recognition, using a flanking design resembling multi-element contexts encountered in sentence reading, have documented interference effects induced by adjacent words. These findings suggest that parafoveal processing comes with an intrinsic cost, at least at a word level.

The purpose of the present study was to determine whether spatial proximity between parafoveal and foveal words further modulate the interference effects induced by adjacent words on single word recognition. To do so, a flanking-word Visual World Paradigm is used. Eye-tracking data from a sample of 54 Norwegian-speaking adult skilled readers were collected and analysed.

Through subject and item Analyses of Variance (ANOVA) spacing was found to have a statistically significant main effect on processing individual, yet non-isolated, words. Closer proximity between foveal and parafoveal words led to slower and more fragile word recognition as compared to normal (default) spacing. Target words were equally divided into two frequency groups, the high-frequency and the low-frequency group. No statistical interaction was found between spacing and frequency, suggesting that spacing effects do not vary as a function of word frequency.

The study findings suggest that as the spatial proximity between parafoveal and foveal words increases, processing costs in the recognition of the fixated word also increase. Parafoveal processing comes with an intrinsic cost and interword spaces further modulate such interference effects. Exploring the optimal spacing in novice and skilled readers can provide valuable insights into the underlying processes of visual word recognition.

Preface

First and foremost, I would like to express my deepest gratitude to my co-supervisors, Professor Athanasios Protopapas and Dr. Laoura Ziaka for their ongoing support, encouragement, and supervision. Thank you for giving me the resources to develop my ideas and being always available to answer any questions. Words can hardly express how grateful I am for the guidance I received.

I would also like to thank all the members of the BetterReading project for our weekly sessions where we had the chance to talk about our projects, present articles, give and receive feedback.

Many thanks to my fellow student, Ivana for the great communication and teamwork in the data collection. I am also thankful to everyone who took time out of their busy schedule and participated in the study. Your willingness to take part in my experiment is so much appreciated.

The support I received from the Achillopouleio Foundation was invaluable to pursue my master's degree abroad. I will be forever grateful for making my dream come true.

Getting through my master's thesis required more than academic and financial support. A big thanks to my family and close friends, especially my mother for always being there and encouraging me. They were cheering me up and made me laugh along this journey; even during the hardest times of these past two years. Without their support, it would not have been achievable to complete my studies abroad.

Christina, Pantelis thank you for everything.

This thesis is dedicated to the memory of my biggest supporter who couldn't see this work finished. I hope I am making you proud up there, dad.

June 2023,
Stefania A. Kyriakidou

Table of Contents

List of Tables and Figures

Tables

Figures

1. Introduction

1.1 Background and Rationale

Recent studies show that fluent reading requires the efficient processing of successive words (Protopapas et al., 2018). Processing word sequences accurately and quickly involves reader's ability to extract information from both the currently fixated word and the adjacent words during a fixation (Schotter et al., 2012). What remains equivocal in reading research is whether foveal and parafoveal words are processed serially (Reichle et al., 1998) or in parallel (Engbert et al., 2005).

Studies using the flanker task, an experimental design that bridges the gap between word-reading and sentence-reading research, have demonstrated that parafoveal items affect foveal processing efficiency (e.g., Grainger et al., 2014; Snell & Grainger, 2018). Such effects, known as *parafoveal-on-foveal effects,* challenge the assumption of serial distribution of attention (Drieghe et al., 2005) and provide further evidence for the parallel processing of multiple words. If words were processed serially, one word at a time —during a fixation— then the upcoming word should neither interfere with nor facilitate the processing of the currently attended word.

Different lexical properties of parafoveal items have been investigated as to whether they facilitate or inhibit word processing (Dare & Shillcock, 2013). However, little attention has been paid to text properties and whether they further modulate parafoveal-on-foveal effects. Research has shown that interword spaces provide essential information for word boundaries and their removal inevitably interferes with word recognition (Perea & Acha, 2009; Rayner et al., 2013; Sheridan et al., 2016) and saccade programming (Paterson & Jordan, 2010), yet they have not been studied in relation to the interaction between foveal and parafoveal processing.

The present study aims to address this research gap and investigate whether interword-spacing manipulations further modulate the impact of parafoveal words on foveal word recognition. To accomplish this, a novel version of the Visual World Paradigm is used combined with a flanking design and backward masking. This thesis is a follow-up on Zelihić (2020) who documented that the presence of parafoveal words, one space away from the fixated word, interfere with word recognition.

1.2 Research Questions

The study aims to answer the following two questions:

"*To what extent do interword spacing manipulations affect the level of nearby-items interference in visual word recognition?*",

"*Is there an interaction between spacing and word frequency? Is the difference across spacing conditions the same for higher-frequency and lower-frequency words?*"

1.3 Structure of the Study

The paper consists of six chapters. *Chapter 1* gives an introduction to the theoretical background of the study, the research aims and the research questions. *Chapter 2* starts with an introduction to the use of eye tracking in reading, followed by the prerequisites of reading fluency highlighting the role of automaticity and processing of multiple words. Then an overview of the literature on parafoveal processing is presented before ending with the role of interword spacing in reading research. *Chapter 3* outlines the methodological approach and *Chapter 4* presents the main study results. *Chapter 5* comes with a discussion of the findings in light of the relevant theoretical and empirical background the study validity and reliability, potential limitations, implications and directions for future research. Finally, in *Chapter 6* a summary of the key findings, along with some final reflections, is given.

2. Literature Review

2.1 Eye Tracking in Reading

Reading is a multifaceted process requiring the coordination of visual, attentional, perceptual, cognitive, language-related and oculomotor systems (Aghababian & Nazir, 2000; Frey & Bosse, 2018; Kliegl et al., 2006; Pennington, 2006). Linguistic skills, including phonology, morphology, syntax, semantic, pragmatics (Gleason, 2005; Kirby et al., 2008; Norton & Wolf, 2012), along with cognitive aptitudes (i.e., memory, automatization, attention, visual-spatial skills) are necessary to achieve the end goal of reading which is "to recover the intended meaning of each word, phrase, and sentence" (Rayner et al., 2016, p. 5).

Numerous studies aim to disentangle and shed light into the cognitive processes involved in reading using eye tracking (Rayner, 2009). By recoding readers' eye movements, researchers can draw inferences about how brain processes language during reading and identify processes underlying reading (Clifton et al., 2007; Holmqvist et al., 2015; Jarodzka & Brand-Gruwel, 2017).

The eye-tracking measures usually reported are *fixations* and *saccades* (Conklin et al., 2018). Fixations refer to the stationary part of vision while saccades refer to the moving part (Rayner & Reingold, 2015). Both measures are typically reported in relation to specific *Interest Areas* (IA), also known as Regions of Interest (ROI). For instance, an IA in a reading task could be a word of a sentence.

During fixations the skilled readers' eyes stay relatively stable for about 200–250 ms in silent reading. Eye-tracking technology gives information regarding where the participants are looking, how many times they have looked at a specific IA, and for how long. The most commonly reported eye-tracking measures are fixation counts, first fixation duration, gaze duration, total fixation time (see Conklin et al., 2018 p. 68)

During saccades the eyes move from one fixation point to another. Their execution typically takes 20–40 ms and their average length is 7–9 characters (Starr & Rayner, 2001). During saccades readers' vision is suppressed (Matin, 1975). Readers can extract useful information only during a fixation. What triggers these rapid ballistic eye movements are the constraints imposed by visual acuity.

Visual acuity is "a measure of the ability of the eye to distinguish shapes and the details of objects at a given distance" (Marsden et al., 2014, p. 16). However, it varies across the visual field. The visual field is divided into three regions: the *fovea* extending 1° of visual angle on each side of the fixation point covering 3–4 character spaces in alphabetic orthographies; the *parafovea* extending up to 5° of visual angle -horizontally- in each direction; the *periphery* extending beyond 5° of the visual field (McConkie & Rayner, 1975; Rayner et al., 1986, 2016). Visual acuity is maximal in the *fovea,* moderate in the *parafoveal,* low in the *peripheral* region. It is high closer to fovea and gradually decreases as the distance from the fixation point increases. Therefore, readers move their eyes to bring new information into the fovea where the visual acuity is high (Conklin et al., 2018).

Although visual acuity is lower in parafovea, readers can still partially process the upcoming (or parafoveal) word, the "*word n+1*". The upcoming word is located to the right of the fixated (or foveal) word, the "*word n*", in left-to-right alphabetic orthographies. Reading rate of skilled readers decrease when preview of the upcoming word is not available or masked, with overall reading behavior being disrupted (McConkie & Rayner, 1975; Rayner et al., 2006). This indicates that parafoveal processing (i.e., the partial processing of the upcoming words while fixating *word n*), is essential for fluent silent reading (Sperlich et al., 2015). Before delving into parafoveal processing, it is important to define and describe the construct of reading fluency.

2.2 Reading Fluency

Over the last 20 years, much attention has been paid to *Reading Fluency* (RF). The report of the National Reading Panel [NRP] (2000) established RF as an important aspect of reading comprehension and primary educational goal in elementary curricula.

Particularly, RF is defined as "the ability to read a text quickly, accurately, and with proper expression" (NRP, 2000, pp. 3–5). Fuchs et al. (2001) defined oral RF as "the oral translation of text with speed and accuracy" (p. 239). Other definitions of oral RF are more expanded including an additional component namely, proper expression, also known as *prosody* (Rasinski et al., 2012). Therefore, the three core dimensions of RF are: *accuracy*, *speed* (or automaticity), and *prosody* (Hudson et al., 2009; Kuhn et al., 2010; Kuhn & Stahl, 2003; Rasinski, 2006; Wolf & Katzir, 2001).

Accuracy refers to the "correct identification of a word", indicated by correct articulation of the word. Speed concerns the "immediate identification of a word" (Wise et al., 2010, p. 341). Prosody (i.e., phrasing, stress, variations in pitch, intonation, appropriate rhythm, pauses) refers to the expressiveness in oral reading (Frankel et al., 2016; Hudson et al., 2005). Given the focus of the present thesis on the prerequisites of word-level RF, prosody is not included in the working definition of RF (e.g., van Viersen et al., 2022).

Reading skills develop in primary grades through carefully structured instruction (Rayner et al., 2016). Becoming a fluent reader involves developing automaticity in word recognition and requires practice and repeated exposure to reading materials (Giovagnoli et al., 2016). However, it is important to note that automaticity is not limited to speed (Protopapas et al., 2007).

2.2.1 Word Reading Automaticity

Automaticity is an integral component of RF, necessary for the efficient execution of high-level skills (e.g., syntactic parsing) involved in reading comprehension. Automatic low-level lexical skills (e.g., word recognition) are exhibited quickly, effortlessly, autonomously without awareness and attentional engagement (Jenkins et al., 2003). If sub-lexical skills have not reached automaticity, all the attentional resources are allocated there, thereby imposing a bottleneck for comprehension (Hudson et al., 2009, p. 8). According to reading automaticity theories, attentional resources are limited and their allocation affects the overall reading performance (LaBerge & Samuels, 1974; Logan, 1997).

Reaching automaticity takes time. Novice readers read slowly, exerting much effort in decoding (i.e., "the process of accurately and fluently translating print into spoken words or units"; Melby-Lervåg & Lervåg, 2014, p. 409). As students gain more experience (Rasinski et al., 2012), they become faster and start recognizing words by sight (Ehri, 2014), thereby releasing attentional resources for the more advanced aspects of reading (Kirby et al., 2010; Samuels, 1997).

Although automaticity is conceptually well-defined, its operationalisation and assessment remain challenging. Roembke et al. (2018) endeavoured to measure children's word reading automaticity beyond speed with automaticity concerning the rapid mappings between orthographic and phonological or semantic input. To account for potential confounding factors in assessing automaticity (e.g., knowledge of grapheme-phoneme correspondence, speed processing differences among children as a function of age, susceptibility of response times to decision making, response planning), Roembke et al. used backward masking. Two versions of each experimental task were used to assess automaticity: the masked and the unmasked. In the masked version, single words were briefly displayed on the screen and replaced with a visual mask (i.e., ####) after 90 ms. The mask forced the rapid lexical activation for the target stimulus (Hendrickson et al., 2021).

Although Roembke et al. developed an innovative "accuracy-based" tool for assessing word reading automaticity, the design overlooks the complexity of multi-element displays encountered in sentence reading. Recent studies have shown that dealing with word sequences is an indispensable component of fluent reading.

2.2.2 Efficient Processing of Word Sequences

Based on current Reading Fluency (RF) theories (Fuchs et al., 2001; Hudson et al., 2009; Kuhn et al., 2010; Rasinski, 2012) the main difference between novice and skilled readers is their ability to process individual words accurately and quickly (i.e., *individual word recognition efficiency*). To assess RF researchers use word lists ("word list reading fluency" or "*word reading efficiency*"; see Test of Word Reading Efficiency, TOWRE, including words and nonwords; Torgesen et al., 1998) and entire passages ("text reading fluency" or "*oral reading fluency*") (Riedel, 2007). Kim and Wagner (2015) define word reading fluency as "accurate and fast reading of (list of) words in isolation out of context" and text reading fluency as "fast and accurate reading of connected text" (p. 225). Both word-level and text-level fluency are typically assessed by measuring correct words per minute (Jenkins et al., 2003). The common characteristic between these two metrics —and the only one addressed so far— is the accurate and fast recognition of individual words presented either in a list or in a text. The only difference is that text reading fluency also involves supra-lexical skills (e.g., semantic integration of words into a sentence).

In other words, RF theories support that the core factor accounting for the developmental changes in word-list RF is the developmental changes in how readers deal with individual words. If word reading fluency was merely based on fast and accurate recognition of individual words, without benefiting from the surrounding words, then children's differences in processing speed of individually presented words should predict word reading fluency (Altani et al., 2020). However, recent studies contradict this notion.

Unique insight into the role of sequential processing efficiency comes from studies examining the relationship between naming speed and reading outcomes. *Naming speed* refers to "the ability to name quickly a number of highly familiar visual stimuli" (Kirby et al., 2010, p. 342). It is commonly measured with Rapid Automatised Naming (RAN) tasks in which participants name familiar items (e.g., colours, objects, digits, letters, symbols) as fast as possible. Items are presented either in a discrete or a serial format. In serial formats, all items are simultaneously presented on a single sheet or screen. In discrete formats, items are presented one-by-one (isolated naming).

Better performance in RAN tasks is strongly related to greater RF (Georgiou et al., 2012; Lervåg & Hulme, 2009). To interpret this concurrent and longitudinal relationship, researchers explore common underlying processes between naming speed and reading (Norton & Wolf, 2012; Protopapas et al., 2013, 2018) and, particularly between serial naming and reading as the former has been found to be more strongly correlated with reading than discrete naming (de Jong, 2011; Georgiou et al., 2013).

Protopapas et al. (2013) found that readers' (Grade 2 and 6) differences in processing speed of individual words contributed only moderately to word-list RF and only at the beginning stages of reading; the correlation weakened as reading skills increased. They also found that serial naming was a stronger predictor of word reading fluency in Grade 6. Based on these findings, they concluded that efficient processing of word sequences is a dominant factor in word-level RF beyond individual word recognition speed (see also Altani et al., 2020; Georgiou et al., 2023; Protopapas et al., 2018; van Viersen et al., 2022).

Similarly, Logan et al. (2011) explored the differences in performance between serial and discrete naming and their relation to reading and found that serial naming speed was more strongly correlated with reading outcomes than discrete naming after controlling for isolated naming. They interpreted these findings in light of factors uniquely involved in serial tasks and suggested that eye movements and the ability to retrieve information beyond the fovea are two factors required only in serial naming. Likewise, Kuperman et al. (2016) suggested that one of the factors that partly account for the relation between performance in RAN and reading is the requirement for similar eye-movement patterns. They referred to this shared oculomotor control as the "Visual Scanning Hypothesis" (Henry et al., 2018).

Briefly, RF is not exclusively reliant on individual word reading efficiency, but also on rapid sequential processing of successive items (Protopapas et al., 2018). Attending to more than one word is also supported by eye-tracking studies demonstrating effects from parafoveal words on foveal words and vice versa. Building upon the prerequisites of RF, the next section focuses on parafoveal processing and the interaction between foveal and parafoveal processing.

2.3 Processing Beyond the Fovea

During reading individuals retrieve information beyond the fovea. They can partially process the upcoming words that are not fixated yet (Rayner, 1998; Schotter et al., 2012 for a review). This readers' ability is known as "*parafoveal processing*". Readers extract information from a specific region, termed as "*Perceptual Span*" or "span of effective vision for reading" (Grainger et al., 2016, p. 172) during a single fixation.

To measure the amount of information that can be processed during a fixation, or to estimate the size of the perceptual span, researchers have implemented a gaze-contingent technique called *moving window* (McConkie & Rayner, 1975). The amount of available information is controlled each time by reader's fixation such that the computer-generated text on the screen displays only a part of the visual field around the fixation point. A string of Xs or random letters masks the letters outside this region, thereby creating a *virtual window* (Starr & Rayner, 2001). The size of the perceptual span is estimated by comparing the reading rate in the baseline condition (visible line) to that in the experimental condition(s) (window size). When the reading rate in the experimental condition is equivalent to that in the baseline condition, the window size represents the size of the perceptual span (Rayner et al., 2016).

In left-to-right alphabetic orthographies, the perceptual span of skilled readers is asymmetric, extending 3–4 letter spaces to the left and about 13–14 letter spaces to the right of the fixation (McConkie & Rayner, 1976; Rayner, 1986). As readers' low-level lexical skills develop, their ability to efficiently process word sequences increases (Altani et al., 2020) and their perceptual span expands (e.g., Häikiö et al., 2009; Sperlich et al., 2015).

Parafoveal processing is a "key mechanism which enables fluent reading" (Hutzler et al., 2013, p. 7) as studies report decreased reading rates when parafoveal preview is absent, masked, or deliberately manipulated. However, the beneficial effects of parafoveal previewing on reading efficiency represent only a portion of the existing literature on parafoveal processing. To gain a better understanding of the interaction between parafoveal and foveal processing, a rigorous examination of both benefits and costs is needed. The two primary findings revolve around *preview effects* (i.e., the effects of parafoveal previews on processing the word subsequent to the fixated word) and *parafoveal-on-foveal effects* (i.e., the effects of parafoveal words on processing the fixated word).

2.3.1 Parafoveal Preview Effects

Parafoveal preview effects ("*N+1 preview effects*" in Vasilev & Angele, 2017, p. 669) refer to the effects of parafoveally previewing a target word (n+1), while fixating the current word (n), on the recognition of the target when ultimately fixating it. It is documented that having access to valid or invalid previews of parafoveal words affects word processing (Rayner, 2009; Schotter, 2018). Most researchers refer to preview effects as "preview benefits" because valid previews are found to enhance reading efficiency (Schotter et al., 2012). However, recent studies demonstrated processing costs induced by parafoveal previews suggesting an overestimation of preview benefits in the literature (Hutzler et al., 2013, 2019; Kliegl et al., 2013; Marx et al., 2015). Therefore, a more neutral term, "parafoveal preview effects", encompassing both benefits and costs, is preferable (*Note* 1 in Vasilev et al., 2021).

Preview benefits are well established in the field of parafoveal processing. Findings indicating shorter fixation durations on target words that readers have a valid preview of during the preceding fixation are quite robust (e.g., McDonald, 2006). Most evidence derives from studies using the *boundary paradigm* (Rayner, 1975). A word, called *preview*, in a sentence is experimentally manipulated to a given dimension of interest. This preview is either identical to the target word with which the sentence makes sense (*valid preview*) or a different word/nonword/random letters (*invalid preview*). When readers' eyes cross an invisible boundary location the invalid preview is replaced by the target word and remains visible. Most studies using this gaze-contingent technique report that having access to a valid preview of word *n+1*, while fixating the current word (n) leads to shorter fixation durations on that word (*n+1*) when readers fixate it compared to masked, absent or invalid previews (Rayner, 1998, 2009).

The magnitude of preview benefits is estimated by subtracting fixation durations after preprocessing accurate previews from fixation durations after preprocessing invalid previews (Marx et al., 2015; Vasilev et al., 2018). To accurately estimate preview benefits, it is assumed that the baseline condition, which blocks parafoveal processing, should not interfere with the subsequent foveal word recognition (Hutzler et al., 2013).

There is an ongoing debate as to whether parafoveal masks serve as a neutral, baseline condition for estimating the true size of preview benefits (Hutzler et al., 2019; Marx et al., 2015; Vasilev & Angele, 2017) and it has been suggested that preview effects are a complex mixture of benefits and costs (Kliegl et al., 2013). Hutzler et al. (2013) conducted an boundary experiment using X-mask previews and they found that parafoveal mask delayed target word processing compared compared to no parafoveal preview (isolated words presented serially). They concluded that the string of Xs interfered with the target processing leading to an overestimation of preview benefits.

Considering the processing costs induced by parafoveal masks as invalid previews, Marx et al. (2015) applied an incremental boundary technique in young readers (Grade 4 and 6) where the parafoveal preview was gradually decreased. They suggested a salience manipulation of valid preview as a more accurate way to estimate its magnitude. In particular, they applied increasing levels of visual degradation (0%, 10%, 20%) for both valid and invalid previews. They observed that as the degradation of invalid previews increased, the fixation durations on target words decreased presumably because the invalid preview was less clear (see also Hutzler et al., 2019). However, Vasilev et al. (2018) failed to replicate these results concluding that increasingly degraded invalid previews did not yield reduced preview costs for adult readers.

Vasilev and Angele (2017) in a systematic review of 93 boundary paradigm experiments found that the type of parafoveal masks further modulated *n+1* preview effects. Particularly, they divided the different types of masks into two main categories, the masks containing no information about the target word, and masks containing information about the target. The interference induced by less "word-like" masks was larger presumably because "they look more unnatural, and they are more likely to be noticed by participants" (Vasilev & Angele, 2017, p. 681).

Overall, recent studies highlight that preview effects should be interpreted with caution as to whether they reflect facilitation or inhibition. If masked conditions interfere with word recognition, observed benefits may stem from preview costs induced by invalid previews (Kliegl et al., 2013).

Besides invalid previews, high foveal load may also reduce parafoveal processing efficiency. Foveal load refers to the processing demands for foveated word recognition. High foveal load words are harder to be processed and consume attentional resources. In turn, fewer resources are available for parafoveal words and readers subsequently fixate them for a longer time (Henderson & Ferreira, 1990; Kliegl et al., 2006). Perceptual span diminishes with increased foveal processing difficulty (Meixner et al., 2022).

To investigate how foveal load affects parafoveal processing efficiency researchers commonly manipulate the frequency of the target. Word frequency concerns how frequently a word appears in a language, as determined usually from corpus data (Kuperman & Van Dyke, 2013). High-frequency words are recognised more quickly and easily than low-frequency words after controlling for word length (Rayner, 1998). Therefore, the foveal load is lower for high-frequency words and higher for low-frequency words.

The interplay between foveal and parafoveal processing extends beyond the effects of foveal words on parafoveal processing efficiency. Parafoveal words may also affect foveated word recognition.

2.3.2 Parafoveal-on-foveal Effects

Parafoveal-on-foveal effects refer to the influence of word *n+1* on the processing of word *n* (Schotter et al., 2012). Some studies demonstrate immediate effects of parafoveal word properties on foveal fixation times (Dare & Shillcock, 2013, experiment 2, 3, and 4; Kennedy, 1998; Kennedy et al., 2002; Pynte et al., 2004), other studies fail to capture significant effects (e.g., Perea & Acha, 2009; White & Liversedge, 2004), while some studies have raised concerns regarding the methodology of studies reporting parafoveal-on-foveal effects (Rayner & Juhasz, 2004; see also Kennedy, 1998).

Further investigation of the rather controversial parafoveal-on-foveal effects is needed as they have implications for the contentious debate on serial vs. parallel processing. This debate concerns the question whether attention is allocated to words serially or in parallel when reading a sentence/text; whether one word is attended at a time or multiple words simultaneously.

Growing evidence for parafoveal-on-foveal effects and the parallel exploitation of foveal and parafoveal information comes also from flanker tasks probing the identification of individual words parafoveally flanked by task-irrelevant stimuli (e.g., Dare & Shillcock, 2013, Experiment 1; Grainger et al., 2014). Both facilitatory and inhibitory effects have been reported untangling the type of information that readers extract beyond the fovea.

2.3.3 Examining Parafoveal-on-foveal Effects in Flanker Tasks

The flanker task, introduced by Eriksen and Eriksen, initially focused on selective attention (i.e., "individuals' ability to select and focus on particular input for further processing while simultaneously suppressing irrelevant or distractive information"; Stevens & Bavelier, 2012, p. 30). It has since been used in reading research to examine word identification in multiple-element displays, word-parallel processing, parafoveal-on-foveal effects, and visual-spatial attention (Snell & Grainger, 2018).

In the original flanker task, participants were presented, in each trial, with a central letter (target) displayed either alone (no flankers) or flanked bilaterally by other letters ("*noise letters*") (Eriksen & Eriksen, 1974). Subjects were asked to press the key corresponding to the target letter. For instance, letters H and K were assigned to a right response and letters S and C were assigned to a left response. They used various types of flanker letters: *identical letters* (e.g., H flanked by three repetitions of the letter H), *response-congruent flankers* (e.g., H flanked by three repetitions of the letter K; compatible condition), *response-incongruent* flankers (e.g., H flanked by three repetitions of the letter S;

incompatible condition), *a set of mixed flanker letters with features similar to the target letter* (e.g., N W Z **H** N W Z), and *a set of mixed flanker letters with features dissimilar to the target letter* (e.g., G J Q **H** G J Q). The aim was to investigate whether these manipulations affected participants' response times (how fast they pressed the key) and accuracy level (error rate). By comparing the congruent and incongruent conditions, they found that incongruent flanker letters interfered with target identification. Response times were greater and accuracy levels lower for incongruent flankers than for congruent. This robust finding is the so-called *flanker effect*.

The flanker effect suggests that two or more items can be attended and processed simultaneously. If items were processed serially, then it would be expected that nearby stimuli did not affect the target identification. Attention would have shifted towards the next item only after the target item was completely processed. To examine parallel processing and how nearby items affect foveal word recognition, reading studies have implemented the flanker design using reading-like materials (i.e., string of letters).

Researchers have incorporated the lexical decision task within the flanker design aiming to bridges the gap between word-level and text-level reading research. In such experiments, participants are presented with a target string of letters (usually) placed at the centre of the computer screen and flanked on either side by task-irrelevant stimuli (e.g., letters, bigrams, visual symbols, words, nonwords). They are asked to decide as quickly and as accurately as possible whether the string of letters forms a real word or a nonword (Grainger et al., 2014).

Dare and Shillcock (2013) used a lexical decision task combined with the flanking design to investigate whether orthographic processing may occur across multiple words in parallel. Participants were presented with a four-letter target string flanked bilaterally by a pair of letters under three conditions: the first and the last bigram of the target stimulus (e.g., RO ROCK CK), the first and the last bigram of the target stimulus reversed (e.g., CK ROCK RO), unrelated bigrams (e.g., LE ROCK SH) (p. 488). They found that response times were decreased for flanking bigrams included in the target as compared to unrelated flanking bigrams. The order of the bigrams present in the target did not affect response times. There was 0 ms difference between the condition in which the related bigrams were ordered left-to-right (RO ROCK CK) and the condition in which the related bigrams were reversed (CK ROCK RO).

To replicate these findings and test whether the position of letters within the flanking bigram affected participants' response times and accuracy, Grainger et al. (2014) used the same flanking-letter lexical decision. They replicated the previous findings demonstrating facilitatory effects in all conditions where bigrams shared letters with the target stimulus compared to unrelated bigrams. Regarding the letter order within the bigram, reversed letter position (e.g., OR ROCK KC) inhibited target processing compared to the conditions where letter order was the same as appeared in the target (e.g., RO ROCK CK). However, the reversed-letter bigrams still facilitated target processing as compared to unrelated flanking bigrams (e.g., LE ROCK SH). Hence, the interpretation of the results heavily relies on the conditions being compared.

Facilitatory parafoveal-on-foveal effects were also observed by Snell, Vitu, et al. (2017; Experiment 2) using a similar flanking-letter lexical decision paradigm. Orthographically related flankers were produced by using the first and the last bigram of a four-letter long orthographic neighbour (of each target). Orthographic neighbours are "words that have all but one letter in common with each other" (Snell, Vitu, et al., 2017, p. 1986). They also manipulated the flanker lexicality by using bigrams of words and nonwords. They crossed the variables of *flanker relatedness* (related vs. unrelated) and *flanker lexicality* (words vs. nonwords) and the experimental conditions were: related word condition (e.g., WA BARN RN), unrelated word condition (e.g., PI BARN LL), related nonword condition (e.g., KA BARN RN), and unrelated nonword condition (e.g., LI BARN RT). They found that target processing was facilitated by flanker relatedness, but there was no significant relationship between flanker lexicality and flanker relatedness. They concluded that facilitatory effects occurred at a sub-lexical orthographic level. Information was extracted from multiple words presented in fovea and parafovea simultaneously, but researchers could not support that lexical activation of foveal and parafoveal words occurred simultaneously.

However, a study conducted by Snell, Meeter, et al. (2017) probing the effects of syntactically related flankers on foveal word recognition demonstrated that multiple words are processed in parallel not only in a sub-lexical level but also in a higher-order level like syntactic parsing. Researchers used a flanker paradigm similar to the above-mentioned. Participants had to indicate whether the target was a noun (left response) or a verb (right response). To test whether syntactic integration may occur across multiple words in parallel, they manipulated the part-of-speech congruency (congruent vs. incongruent) between target and flankers. In the *congruent condition*, nouns were flanked by nouns (e.g., cops **rack** cops) and verbs were flanked by verbs (e.g., hear **went** hear). In the *incongruent condition*, nouns were flanked by verbs (been **rack** been) and verbs were flanked by nouns (e.g., cops **went** cops). To test whether the integration of syntactic information is a result of a sentence-like context, they used two more conditions where flankers and target formed either a *grammatically correct* or an *incorrect sentence*. For instance, for a target noun the correct sentence was "this **rack** fell" and the incorrect "fell **rack** this". For a target verb, the correct sentence was "they **went** here" and the incorrect "here **went** they". They found significant differences only between the first two conditions for the part-of-speech congruency. However, the facilitatory effects of syntactically congruent flankers on target processing were significant only in Experiment 2 with the flanking design. In Experiment 1 where the task involved sentences, the effects were non-significant. According to the researchers, different cognitive mechanisms required in each task might account for this discrepancy in the results.

Considering the aforementioned findings and interpretations, compatible flankers appear to facilitate foveal word processing (Cauchi et al., 2020). Nevertheless, none of these studies include a no-flanker condition. If participants' responses were faster and more accurate in the no-flanker condition compared to the congruent conditions, these results would indicate that parafoveal stimuli require attentional resources despite their relatedness thereby inevitably affecting foveal word processing —given the parallel word processing—.

Snell and Grainger (2018) acknowledged the importance of a no-flanker condition and reported inhibitory parafoveal-on-foveal effects. Particularly, they tested whether the flanking paradigm resembles normal reading by investigating the distribution of attention to the words. Considering the rightward attentional bias in sentence-reading in a language read left-to-right, they hypothesised that rightward flankers would have a stronger impact on target processing than leftward and used a flanking design in a lexical decision task with 7 experimental conditions: (a) **rock** (no flankers), (b) rock **rock** rock, (c) step **rock** step, (d) rock **rock** – (no rightward flanker), (e) – **rock** rock (no leftward flanker), (f) rock **rock** step, (g) step **rock** rock. They replicated the results that related flankers facilitate target processing as compared to unrelated ones. They also found that participants' performance was facilitated more by the repetition of rightward flankers (shorter response times) than the repetition of leftward flankers when comparing the "mixed flankers" condition (f and g). In additional, they observed that leftward flankers —even when they were identical to the target— resulted in longer response times compared to the no-flanking condition.

The last finding of Snell and Grainger (2018) demonstrates processing costs for the currently fixated word elicited by parafoveal items. As mentioned by Ziaka (2023), the lack of a no-flanker condition challenges the interpretation of research findings as to whether related flankers actually facilitate or just interfere less than unrelated flankers with the foveal processing.

2.3.4 Nearby Interference in Word Recognition

The most robust finding of flanker tasks is undoubtedly the influence of parafoveal items on foveal word recognition indicating the parallel processing of multiple words. However, interpretation of these results is not one-sided. Indeed, parafoveal previews sharing orthographic information with the fixated word facilitate foveal word recognition compared to previews with no orthographic overlap. However, the absence of a no-flanker condition serving as the baseline challenges such interpretations (Hutzler et al., 2019).

Evidence for interference induced by parafoveal words comes from Zelihić (2020). Three conditions were implemented: *no-flanker condition* serving as the neutral baseline, *visual-flanker condition* (i.e., %s), and *word-flanker condition*.

Zelihić (2020) hypothesised that nearby stimuli interfere with foveal word recognition because parafoveal items require attentional resources given the parallel processing of words during reading. As expected, he demonstrated that the word flankers delayed single word recognition. He concluded that recognizing a word in a multi-word context was more demanding than recognizing an isolated word or a word flanked by visual symbols for adult skilled readers. These findings are consistent with the notion of *nearby-items interference* defined as "the impairment in performance due to the simultaneous presentation of items in spatial proximity to the target, with target and nearby items requiring the concurrent execution of multiple processes" (Ziaka, 2023, pp. 30–31).

The target words were divided into two frequency groups, the high-frequency, and low-frequency words after controlling for other lexical variables. *Word frequency* is considered as an index of lexical retrieval (i.e., "the process of accessing information from the word's representation in the mental lexicon"; Dobó et al., 2022, p. 320), indicating "how easily the word is recognized and retrieved from the lexicon" (Conklin et al., 2018, p. 66). By doing so, they tested whether high-frequency words are less susceptible to interference by nearby items given that they are recognized faster than low-frequency words. He found that the frequency effect was exaggerated in the visual-flanker condition. Low-frequency words flanked by visual symbols were recognized more slowly than high-frequency words.

To assess nearby interference, they examined the lexical activation rate for the target; how efficiently participants access the lexical representation of a fixated word across the three conditions. Zelihić (2020) used a novel version of the Visual World Paradigm (VWP) incorporating a backward masking flanking design and eye-tracking technology.

2.3.4.1 The Visual World Paradigm. The VWP (Cooper, 1974; Tanenhaus et al., 1995) is a technique used to investigate language processing in real time (Huettig et al., 2011). The assumption underlying the task is that participants' eye movements convey what individuals think over time as the stimulus unfolds (Conklin et al., 2018).

The VWP for assessing single word recognition —typically— involves presenting four **pictures** and a concurrent **linguistic stimulus** (Altmann, 2004). One picture corresponds to the target; one or more pictures correspond to words that partially match the target due to overlapping letter/phonemes; the remaining pictures represent unrelated items. The linguistic stimuli could be either auditory (spoken words; e.g., Apfelbaum et al., 2021; McMurray et al., 2018; Rigler et al., 2015) or visual (written words; e.g., Hendrickson et al., 2021). Participants are asked to select the picture corresponding to the linguistic input.

With the use of eye-tracking technology, researchers can examine the likelihood of looking at particular objects relative to the stimulus onset. Countless VWP studies have shown the co-activation of words that orthographically or phonologically overlap resulting in what is called "lexical competition" until one word eventually prevails (e.g., Simmons & Magnuson, 2018). In their introduction, Apfelbaum et al. (2021) give an example of spoken word recognition using items from a study conducted by Allopenna et al. (1998; Experiment 1). The target is *sandal*, its phonological (onset) competitor ("cohort") is *sandwich*, its rhyme is *candle,* and the unrelated word/item is *necklace*. Once participants hear the *sa-* in *sandal*, the proportions of their fixations are distributed between *sandal* and *sandwich* whereas the rhyme and the distractor receive no fixations. When they hear more letters, *sanda-*, then fixations on the cohort *sandwich* decrease while the fixations on the target *sandal* increase. As the word unfolds, the rhyme also receives some activation. Finally, participants access the word and both competitors are eventually suppressed. Hence, the eye-movement patterns convey "*which* specific words are active, *when* they are active, and *by how much*" (Hendrickson et al., 2021, p. 1654).

For assigning fixations to the pictures, researchers define *Interest Areas* (IA). Each picture is treated as a separate IA to analyse the proportions of fixations on each object at certain time points during a trial. Ito and Knoeferle (2022, Glossary) define the variable *fixation proportion*, as "the proportion of time spent fixating an IA within a time window (which is then averaged across trials)".

The VWP has mainly been used for spoken word recognition. However, Hendrickson et al. (2021) recently developed a version of the VWP assuming for written word recognition.

Specifically, they investigated lexical activation and competition in written word recognition by using three different competitors (i.e., cohorts, rhymes, and anadromes). The experiment started with a preview of four pictures (e.g., *mug* [target], *gum* [anadrome competitor], *nap* and *pan* [two unrelated items]; p. 1660). Then, a target word appeared at the centre of the screen followed by a visual mask. To block any bottom-up process and ensure that they measure "response-driven" fixations and not "input-driven" (p. 1657), the word was presented only for *100ms*. They also incorporated backward masking in the written word task to make sure that the visual mask (i.e., #######) erased any visual traces of the word on the blank screen (in accordance with Roembke et al., 2018). However, what do researchers miss when they probe the recognition of individual words presented in isolation?

To fill this gap, Zelihić (2020) used adjacent words on either side of the target word, thereby investigating single word recognition in a multi-element context.

2.3.4.2 A Backward-masking Flanking Design. By integrating the flanking design in the VWP, Zelihić (2020) addressed four crucial aspects of reading: the multi-element context encountered in sentence reading forcing participants to efficiently deal with three words; the parallel processing since target and flankers appear simultaneously; the parafoveal processing by using short target words that occupied only a part of the perceptual span; the automatic word recognition by using the backward masking, a method intended to force rapid lexical activation and mapping between orthographic and phonological or semantic representations.

In an effort to address additional aspects of reading and word recognition within the aforementioned study design, the present study explores interword spaces. Eriksen and Eriksen (1974) found that the spatial proximity between target and flankers affects target recognition. With regards to reading research, spacing manipulations have been found to impact reading performance with removal of spaces inhibiting word recognition in spaced alphabetic orthographies. However, most studies explore spacing at a sentence-level, rather than at a word-level in relation to parafoveal effects. This creates a gap in the existing literature that the present thesis aims to address. Before presenting in greater detail the study objectives and methodological approach, a literature review on spacing precedes.

2.3.5 Spacing Effects in a Flanker Task

Evidence for the spacing effects on target recognition while considering nearby interference comes from Eriksen and Eriksen (1974). Apart from the different types of flanking letters (see 2.3.3), they also investigated the effects of spatial proximity by placing the flankers .06°, .5°, and 1° of visual angle away from the target.

In all conditions, response times decreased as between-letter spacing increased. Thus, the interference effect was further modulated by spacing manipulations. They claimed that flanking stimuli at a distance less than 1° from the target are inevitably attended thereby inhibiting target recognition. The observed *spacing effect* was attributed to the ease of spatial discrimination. However, this is not a reading task.

To my knowledge, although numerous studies have used the flanker design to probe parafoveal-on-foveal effects in reading (e.g., Cauchi et al., 2020; Vandendaele & Grainger, 2022), none of them have manipulated text properties. Little do we know about the role of interword spacing in parafoveal-on-foveal effects. What is known is that the removal of interword spaces affects word recognition in sentence reading.

2.4 The Role of Interword Spaces in Reading

Visual manipulations in the typeface of a text (Zorzi et al., 2012) have been found to influence reading outcomes. For instance, numerous studies have investigated the function of spacing information in normal silent reading in spaced and unspaced writing systems. Researchers have explored both *interword spaces*, the spaces between two words, and *interletter spaces*, "the spaces between two adjacent graphemes" of the same word (Van den Boer & Hakvoort, 2015, p. 697).

The effects of interletter spaces on reading processes remain inconclusive. Some authors report increased interletter spaces yielded facilitatory effects leading to better performance in lexical decision tasks (Perea et al., 2011; Perea & Gomez, 2012a) and sentence-reading tasks (Perea & Gomez, 2012b; Zorzi et al., 2012). Other researchers report impaired reading performance due to increased interletter spaces (Galliussi et al., 2020; Paterson & Jordan, 2010) presumably because of the disruption of the physical integrity of words (Van Overschelde & Healy, 2005). Moreover, most researchers have confounded interletter spaces with interword spaces. Consequently, it remains ambiguous which manipulation accounts for the observed effects. Considering these equivocal results and the more consistent effects of interword spaces on reading, the present study focuses on the spatial proximity between words. Similarly, Slattery and Rayner (2013) pinpointed that interword spacing may be more important for readability of texts than interletter spacing.

Various spacing conditions have been examined such as *removing spaces* (Perea & Acha, 2009; Rayner et al., 1998), *filling spaces* with nonlinguistic symbols (i.e., letters, digits, open/closed squares, Xs) (Epelboim et al., 1997; Pollatsek & Rayner, 1982; Sheridan et al., 2016), *highlighting manipulations* (e.g., the alternating**bold** unspaced condition; Bai et al., 2008; Perea & Acha, 2009), and/or *adding spaces* —usually— in naturally unspaced languages (Bai et al., 2008 for Chinese; Kasisopa et al., 2013 for Thai; Sainio et al., 2007 for Japanese). One consistent finding is that word recognition and saccade programming are disrupted when reading unsegmented sentences/texts in an alphabetic orthographies (McGowan et al., 2014; Pollatsek & Rayner, 1982; Spragins et al., 1976).

Specifically, removing interword spaces has led to decreased reading rates, longer fixation durations, reduced skipping rates and shorter saccades (e.g., Rayner et al., 1998). Regarding eye guidance in spaced text, the initial fixation location, known as *preferred viewing location*, is at the word centre and slightly to the left (Rayner, 1979). In unsegmented texts/sentences initial landing positions are shifted closer to the beginning of the words (Paterson & Jordan, 2010 but also McGowan et al., 2015).

Sheridan et al. (2013) used two experimental conditions in which sentences were presented either with normal spacing (e.g., “John decided to sell the table in the garage sale”), or with random numbers replacing spaces (e.g., “John4decided8to5sell9the7table2in3the9garage6sale”). Researchers reported decreased reading rates when spaces were replaced by random digits. The word frequency effect was larger for the gaze duration and the total time in the unsegmented condition than in the normal spacing condition. Due to the increased difficulty in recognizing the low-frequency words when spaces were replaced, researchers concluded that replacing spacing affects word recognition.

Longer sentence-reading times and exaggerated fixations on low-frequency words in unspaced texts were also observed for younger and older adult readers by Rayner et al. (2013). Likewise, McGowan et al. (2014) recruited young and older readers to investigate the effects of removing or filling with nonlinguistic symbols the interword spaces in a sentence. They found that the absence of spacing information slowed down reading times and resulted in larger word frequency effects for both age groups.

In a similar vein, McGowan et al. (2015) implemented more realistic changes in interword spaces including young and older readers. Sentences were presented in one of the following conditions: normal spacing, condensed spacing, and expanded spacing. Expanded spaces increased sentence reading times and resulted in more yet shorter fixations. The closer word proximity resulted in fewer yet longer fixations. No interaction between spacing and frequency was observed.

Overall, most studies investigating interword spacing have either applied extreme changes (spaced vs. unspaced sentences) which are not likely to be encountered in alphabetic languages or they have combined changes in interword spacing with changes in interletter spacing in the same spacing condition (McGowan et al., 2015). Furthermore, none of these studies have investigated interword spacing in relation to parafoveal-on-foveal effects.

2.5 The Present Study

Zelihić (2020) observed that adjacent words delayed single word recognition indicating that parafoveal processing comes with an intrinsic cost. As a follow-up, the present thesis aims to investigate whether interword spaces further modulate these processing costs.

To explore the effects of spatial proximity, flankers are placed one, half and double space away from the target. We expect that decreased spaces cause a stronger interference effect, induced by adjacent words, leading to slower word recognition. On the other hand, by increasing interword spaces adjacent words interfere less with single word recognition. In the wider spacing condition, participants are expected to suppress competitors and shift their eye gaze towards the correct picture faster than in normal and half spacing.

Furthermore, target words are equally divided into two frequency groups. Our hypothesis for the frequency effect is that high-frequency words exhibit faster lexical activation rate than low-frequency words. We predict that participants recognize high-frequency words more quickly and easily than low-frequency words. For the interaction between frequency and spacing, we anticipate that the frequency effect is exaggerated in the half-spacing condition where the word recognition becomes more challenging.

3. Method

The study employs a quantitative method to explore interword spacing effects on single word recognition in a multi-element display. This chapter describes the experimental design, the sample, the eye-tracking apparatus, the procedure, the data analysis before ending with a description of the validity, reliability and ethical considerations of the study.

3.1 Design

The present thesis is a follow-up on a previous study that reported processing costs in single word recognition induced by adjacent words. Zelihić (2020) compared visual and word flankers to a no-flanking condition and the study is available at http://urn.nb.no/URN:NBN:no-83372. Only word flankers were found to interfere with word recognition. Thus, the present study includes only this type of flankers.

To explore the effects of spacing on the correct and rapid lexical activation for non-isolated fixated words, we use a backward-masking flanker design in a VWP, implementing three experimental conditions: *normal (default) spacing* (Figure 1), *half spacing* (Figure 2), and *double spacing* (Figure 3). To examine whether spacing effects vary as a function of word frequency, target words are divided into two corpus-based frequency groups, the high-frequency and the low-frequency group after controlling for lexical variables known to influence word recognition.

Word flankers are horizontally aligned with the target word to spatially approach a "sentence-like" context (Meade et al., 2021, p. 539). Moreover, backward-masking ensures rapid mappings between orthographic and phonological (or semantic) input as the target word is briefly visible before being masked. Adult readers are known to have reached automaticity in word recognition and, therefore, it is expected that they are able to access word's representation in the mental lexicon within 75 ms. Roembke et al. (2018) assessed automaticity in children by presenting the target word for 90 ms.

Finally, the study design is *within-subjects*, also known as repeated-measures (Lix & Keselman, 2018) or paired design (Cumming & Calin-Jageman, 2017). All participants were exposed to all three conditions and responded to the same stimuli. We focus on the variability *within* each person rather than *between* participants (Shaughnessy et al., 2015). The main advantage of a within-subjects design over a between-subjects design is that individual differences are removed resulting in more efficient inferential statistics (Lix & Keselman, 2018). However, there are numerous threats to validity that researchers should be aware of and implement practices to eliminate them.

Figure 1

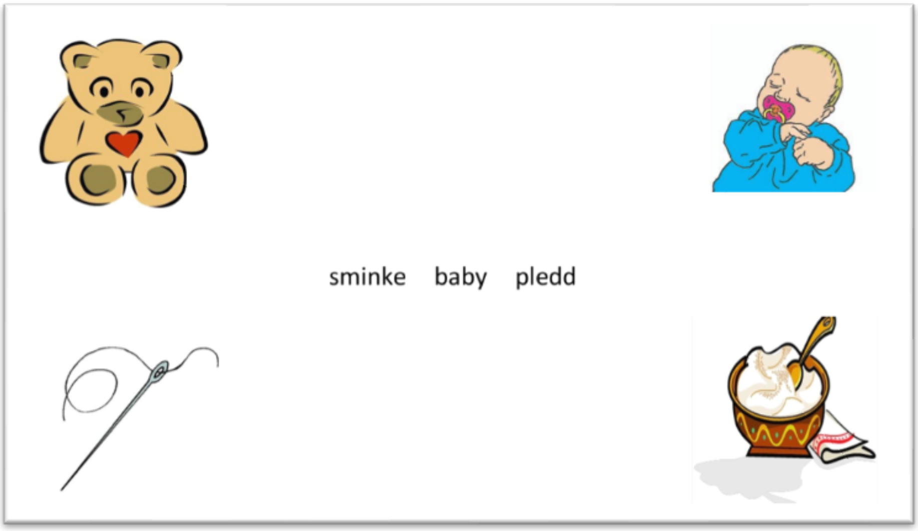
Target Presented in the Normal-spacing Condition

Figure 2

Target Presented in the Half-spacing Condition

Figure 3

Target Presented in the Double-spacing Condition



3.2 Participants

A total of 55 participants ($n_{female} = 41$), mainly university students, took part in the study. Their age ranged from 19 to 39 ($M = 25$ years, $SD = 4.40$). All participants were native speakers of Norwegian. They had normal or corrected-to-normal vision and no reading difficulties (by self-report). Twenty-three were right-eye dominant; thirty-two were left-eye dominant. The dominant eye was the one with the least maximum and average error during the first binocular calibration process. One subject was removed due to low accuracy level. This left data from 54 participants to be analysed.

The sampling method was convenient, a non-probability method which involves recruiting participants easily accessible and available at a certain time and place that researcher proposes. Participants were recruited mainly through personal and professional networks, research flyers on bulletin boards, and social media platforms.

3.3 Apparatus

Participants' eye-movements were recorded with an EyeLink 1000 plus, desktop mount, eye tracker (with 35mm lens) manufactured by SR Research (Toronto, ON, Canada). Recording was monocular with a sampling rate at 1000 Hz. Stimuli were displayed on a Ben-q 24-inch LCD monitor with a display resolution of 1920 × 1080 and a refresh rate of 120 Hz. The computer-based experiment was designed and implemented in the Experiment Builder software (SR Research, version 2.2.38), a drag-and-drop programming environment.

Subjects were seated at a distance of 91 cm away from the screen with one letter subtending 19° of visual angle. Whilst viewing was binocular, only the eye movements from the dominant eye were recorded and analysed. Head movements were minimised by using a chin and forehead rest to maintain a constant viewing distance and improve data precision and accuracy (Carter & Luke, 2020). Participants' eyes were positioned at 75% of the monitor's height. Responses were recorded using a standard computer mouse.

3.4 Materials

The present study uses the same materials as in Zelihić (2020) (Appendix A). The experimental text elements (target words, flankers, and mask) were displayed in black lowercase 22-point Calibri, a proportional width font, on a white full-screen background. Zelihić (2020) used a 20-point Consolas, a fixed width font, yet for practical reasons it was not possible to implement the half spacing with a fixed width font. Thus, a proportional width font resulting in approximately 1 cm width per three characters and resembling the 20-pt Consolas was chosen instead.

3.4.1 Target Words

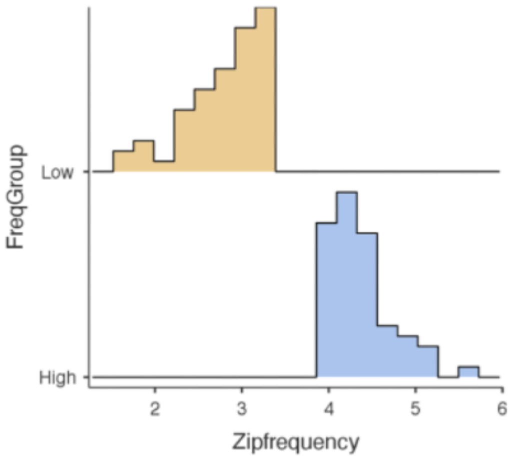
The target words were disyllabic nouns that have multiple orthographic neighbours and occupy only a part of the perceptual span, thereby allowing the processing of adjacent words. All words were Norwegian imageable nouns with a total of 120 words (excluding practice trials). They were equally divided into two frequency groups: (1) low-frequency words and (2) high-frequency words. Orthographic properties of the target words were obtained from the corpus-based Norwegian Orthographic Analyzer available at https://noa.spell.uiocloud.no/.

Many lexical variables (e.g., word frequency, length) are known to influence word recognition (Adelman et al., 2014). For the purposes of the present study, both frequency groups were controlled for the following variables: *word frequency* (Zipfreq), *word length* (Nlet), orthographic *neighbourhood size* (OLD20), and *bigram frequency with end* (Bigram w/end) (Appendix B). Considering Norwegian orthography, nouns with digraphs (e.g., traktor), long vowels (e.g., rake), and short vowels (e.g., tønne) were included.

Starting with the *word frequency* (Clifton et al., 2007), the variable refers to how frequently the target words appear in Norwegian language, as determined from corpus data. The frequency scale ranges from 1 (*lowest*) to 7 (*highest*). Words with a value equal to or lower than 3 are considered as low-frequency, while words with a value equal to or higher than 4 as high-frequency. The average zip frequency is 4.36 (*SD* = 0.35) for the high-frequency words and 2.81 (*SD* = 0.44) for the low-frequency words (Figure 4).

Figure 4

Mean Zipfreq by Frequency

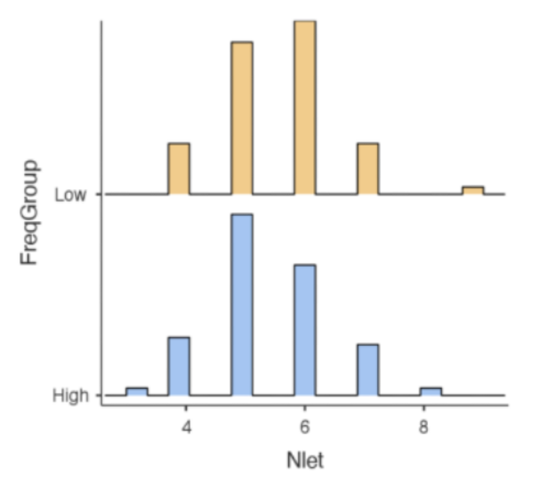


Note. From "Automaticity and the notion of interference. Assessing word reading automaticity as freedom from interference in a Visual World Paradigm," by D. Zelihić, 2020, p. 22. Reprinted with permission.

Word length refers to the number of letters of a single word. Word length affects eye-movement patterns, impacting both fixation durations and saccade programming (Kuperman & Van Dyke, 2011). Longer words tend to elicit longer processing times (Schmidtke et al., 2021). The average word length is 5.58 (*SD* = 0.98) for the high-frequency group and 5.42 (*SD* = 0.96) for the low-frequency group (Figure 5).

Figure 5

Mean Nlet

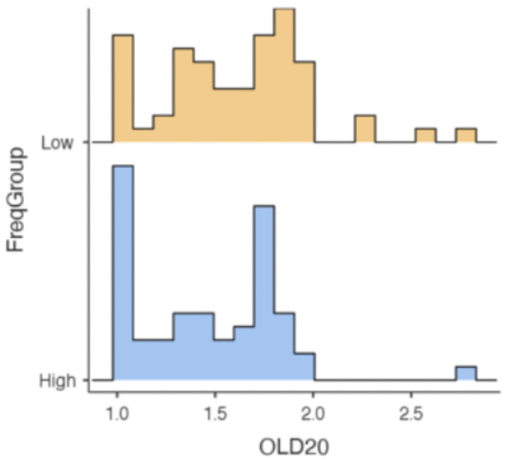


Note. From "Automaticity and the notion of interference. Assessing word reading automaticity as freedom from interference in a Visual World Paradigm," by D. Zelihić, 2020, p. 22. Reprinted with permission.

Regarding *Orthographic Neighborhood Size*, there are two measures of orthographic similarity in the literature: Coltheart's *N* (ON) and orthographic *Levenshtein distance 20* (OLD20). The former refers to the number of words of the same length that differ only by one letter while maintaining letter position. The latter measure is based on the principles of ON but more flexible including orthographic similarity resulting from replacing, inserting or deleting a letter (Yarkoni et al., 2008). The average distance of the nearest 20 words is 1.45 (*SD* = 0.37) for the high-frequency words and 1.61 (*SD* = 0.39) for the low-frequency words (Figure 6).

Figure 6

Mean OLD20

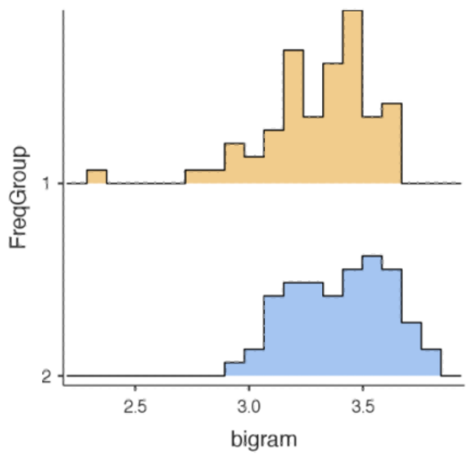


Note. From "Automaticity and the notion of interference. Assessing word reading automaticity as freedom from interference in a Visual World Paradigm," by D. Zelihić, 2020, p. 23. Reprinted with permission.

Lastly, target words were controlled for the variable *Bigram frequency with end*, which refers to the frequency of specific letter combinations occurring at the end of words (Schmalz & Mulatti, 2017). The average bigram frequency is 3.40 (*SD* = 0.22) for the high-frequency words (*SD* = 0.22) and 3.31 (*SD* = 0.24) for the low-frequency words (Figure 7).

Figure 7

Mean Bigram w/end



Note. From "Automaticity and the notion of interference. Assessing word reading automaticity as freedom from interference in a Visual World Paradigm," by D. Zelihić, 2020, p. 23. Reprinted with permission.

Target stimuli were counterbalanced across participants so that each target word was displayed in each condition. After dividing target words into two frequency groups, the 60 words of each group were shuffled, and three lists were created. The first list corresponded to the normal-spacing condition, the second one to the double-spacing and the third one to the half-spacing. This process was repeated as many times as needed to create enough target lists for the desirable sample size, unique for each participant. In doing so, a target word was presented in the normal spacing condition for one participant and in the half (or double) spacing for another. Finally, each list was shuffled so that each target appeared in a different order for each testing session. Each target was presented only once for each participant and in randomised order across sessions.

3.4.2 Flanker Words

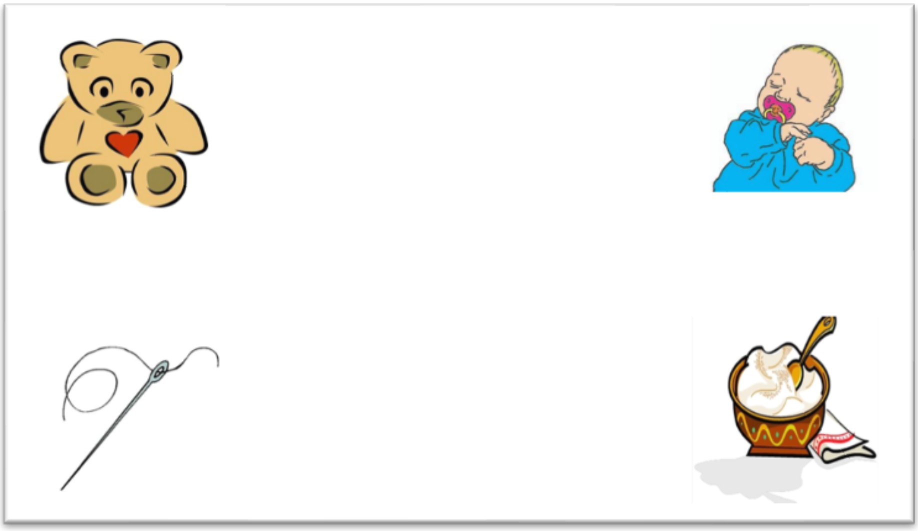
Each target word was assigned a set of two distinct flanker words. Flankers were task-unrelated, spatially distinct, and phonologically dissimilar sharing no similar onset or ending syllables with the corresponding target. Their length ranged from one to three syllables. All flankers were Norwegian nouns.

3.4.3 Pictures

A set of four pictures was selected for each target word (Appendix C). One of the four pictures corresponded to the target, one was the competitor and two were distractors to occupy the screen and challenge the eye gaze. The competitor picture corresponded to a disyllabic word with the same onset as the target word (Figure 8; *baby* [= baby], competitor: *bamse* [= teddy bear], distractor 1: *nål* [= needle], distractor 2: *grøt* [= porridge]). A total of 480 images were used (excluding the 6 practice trials). All pictures were colour drawings with fine lines carefully selected to match the target words. The locations of the target pictures were randomised across trials.

Figure 8

Example of Display Images



3.5 Procedure

All participants were tested individually in a windowless room with artificial light. They were seated in front of the display and instructed to place their chin and forehead in the headrest. Instructions were presented on the screen followed by an oral summary to ensure compliance with the task demands. Time for questions and clarifications was provided.

Before commencing the experiment, pupil and corneal reflection thresholds were adjusted to ensure sufficient tracking of the pupil. The initial calibration was binocular using a 13-point grid of dots that appear individually and consecutively and cover the entire screen. Participants were asked to fixate these points. Calibration was immediately followed by a validation to determine the stability and accuracy of calibration. Validation was successful when the average error was less than .50° and the maximum error less than .99°. Participants' fixations on the dots during the calibration/validation process were accepted manually for increased accuracy. After the first calibration, the eye with the least error was chosen and data were collected only for the dominant eye throughout the whole experimental session.

Following the initial calibration, participants underwent 6 practice trials to familiarise themselves with the task. At that point the experimental trials started. The experiment consisted of 3 blocks of 40 trials each, yielding a total of 120 trials. All conditions were randomly intermixed in each block. Participants were encouraged to take breaks between the blocks. A 9-point calibration/validation was performed after every block and as needed between trials.

Each trial began with a drift correction point in the shape of a white dot on a black background at the centre of the screen, manually accepted as the calibration points. The trial-by-trial drift check was implemented to check the calibration accuracy by "measuring the difference between the computed fixation position and the actual position of the current target" (Zhan, 2018, p. 2). A new calibration was carried out whenever the eye gaze was not detected within 1° of the computed fixation position. Once the drift correction point was accepted, four pictures were displayed, one in each quadrant of the screen.

Participants were provided with a free picture preview for *1500 ms*. This preview was followed by a red dot at the centre of the screen for *520 ms* (Figure 9) immediately replaced by a blue dot (Figure 10) which shifted participants' eye gaze towards the centre of the screen. The blue dot was used as a fixation trigger point and participants were instructed to click on the blue dot to see the target word. The word did not appear unless participants clicked on the blue dot and fixated on it for a minimum *100 ms* thereby ensuring that an inaccurate response was not a result of not seeing the word.

Figure 9

Illustration of the Red Dot

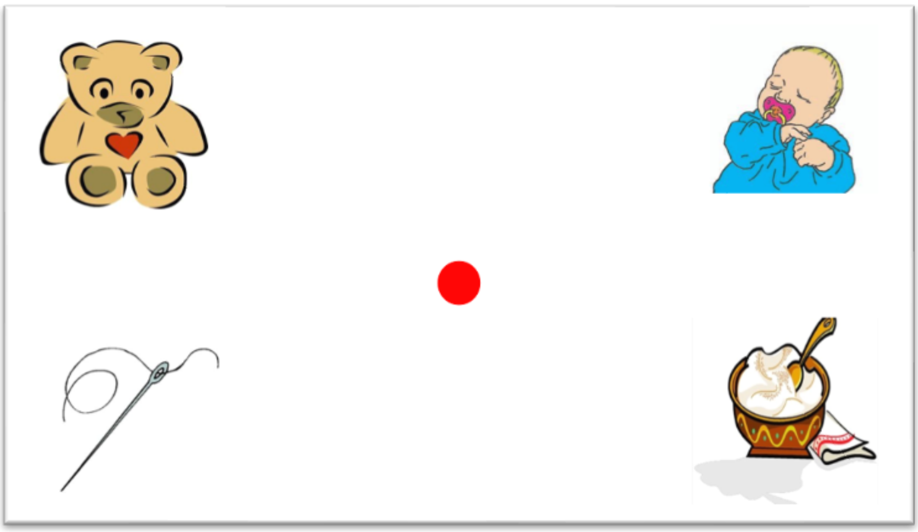
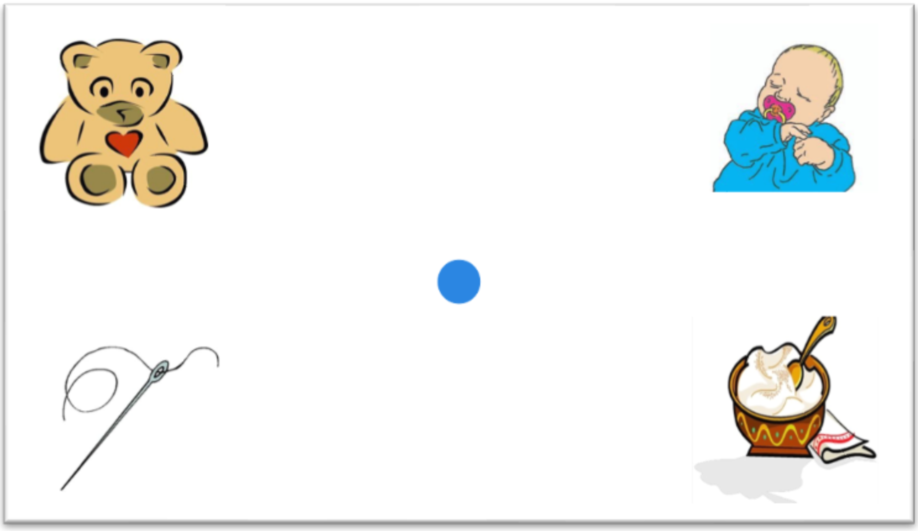


Figure 10

Illustration of the Fixation Trigger



Once participants clicked the fixation trigger, the blue dot disappeared. Immediately the target word was displayed at the centre of the screen for *75 ms*, concurrently with the word flankers (Figure 1). Immediately after, both target and flanker words were covered by a nonlinguistic mask, a string of repeated "#s" which exceeded the number of letters of the longest word in the trial (Figure 11). After *100 ms* the mask disappeared, and participants had to indicate by means of a mouse click the picture that matched the written target word. Participants were not asked to select a picture as fast as possible to avoid accuracy-speed trade off.

To summarise, the sequence of events in each trial was as follows (Figure 12): (1) drift correction point at the centre of the screen, (2) free preview of the four pictures, (3) red dot, (4) blue dot as a fixation trigger, (5) target word flanked by two words, (6) visual mask replacing both target and flanker words, (6) mouse click for response. The whole experimental session lasted approximately 30 minutes, and the testing took place in January to March during the spring 2023 academic semester.

Figure 11

Illustration of the Visual Mask

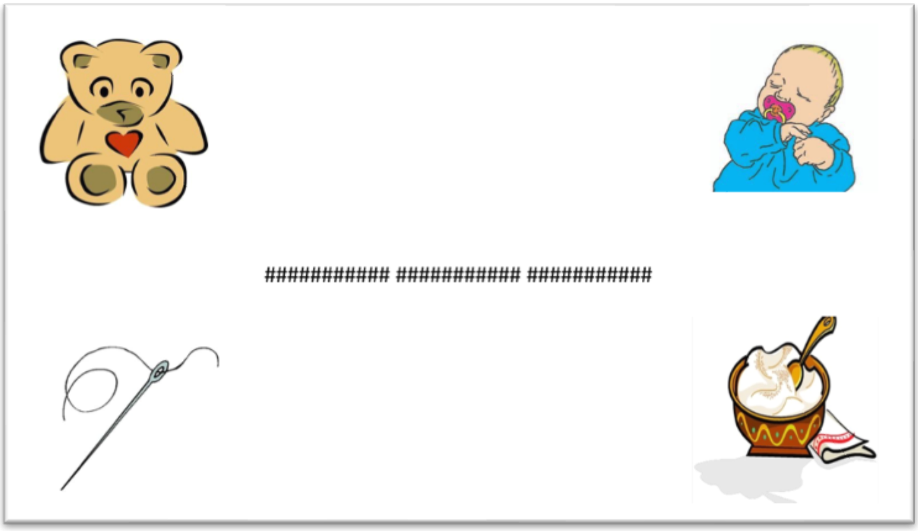
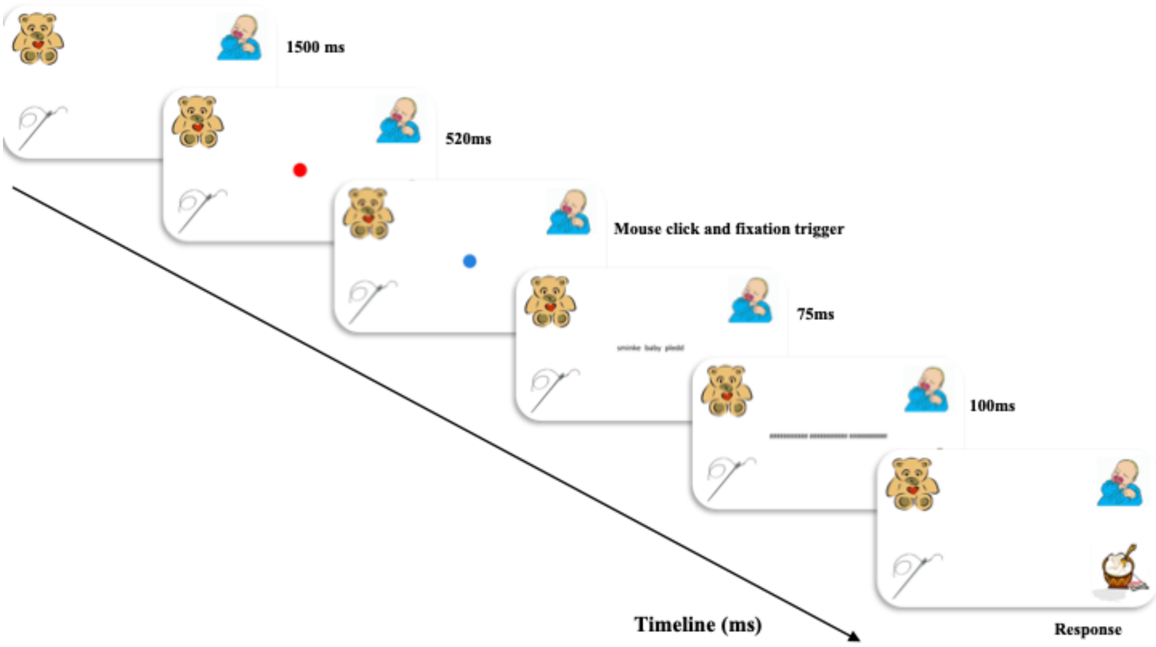


Figure 12

Time Course of the Successive Displays During a Trial



3.6 Data Analysis

As previously stated, the purpose of the study is twofold: (1) to investigate how interword spacing affects nearby interference induced by parafoveal words, (2) to explore whether spacing effects are affected by word frequency. Thus, we are interested in comparing the three spacing conditions and the two frequency groups *within* participants. All statistical analyses are based on the within-subjects design of the study.

Data cleaning and preprocessing were performed before conducting subsequent statistical analyses. For the preparation and plotting of visual-world eye-tracking data the *VWPre* package was used (Porretta et al., 2020).

In most VWP studies, researchers probe the lexical activation rate by analysing the *proportion looks*. This variable refers to “how likely the participants are to look at specific regions of interest at different times during a trial” (Huettig et al., 2011, p. 154). In the present study three dependent variables are analyzed to explore spacing effects: (1) *proportion correct responses*, (2) *proportion target looks*, (3) *time of first sample to target*.

Response times were also calculated but not analysed. Differences in response times may be affected by factors other than spacing manipulations, such as accuracy-speed trade off or individual differences in response speed and decision-making.

The independent variables (also called factors) are two. The first factor is *Spacing* with three levels: half, normal, and double spacing. The second one is the *Word Frequency* with two levels: high-frequency, and low-frequency words. A third factor for the analyses of proportion looks is employed. This factor is *Time* with three levels: 200–400 ms, 400–600 ms, and 600–800 ms relative to stimulus onset.

The time intervals were chosen based on the graphs in section 4.2. The process of assigning looks to the pictures and selecting the desired bin size given the sampling rate of the eye tracker are described in section 4.1.

Data were then analysed using descriptive statistics for the dependent variables and analyses of variance (ANOVA) to investigate the main effects of each factor (i.e., unique effects of each factor on the dependent variable) and statistical interactions between factors that is, whether the effect of one factor is relatively constant or affected by other factors. Checking the assumptions of the parametric test preceded any analysis and corrections were applied when needed.

The dependent variables “proportions of correct responses” and “time of first sample to target” are analysed with a two-way ANOVA (two factors; Spacing and Frequency). The dependent variable “proportion target looks” is analysed with a three-way ANOVA (three factors; Spacing, Frequency, and Time).

For language materials, such as a list of words, it has been argued that any observed effect should be investigated not only in relation to the subject sample, but also in relation to the item sample. In other words, to account for the variability of the outcome if the experiment was conducted with a different sample of participants and a different sample of words, both by-subject (F_1) and by-item (F_2) ANOVAs are conducted (Clark, 1973; Locker et al., 2007; Raaijmakers, 2003). Subject ANOVA accounts for the noise induced by participants as "participants" is a random factor and "words" a fixed factor. Item ANOVA accounts for the noise induced by the target words as "words" is a random factor and "participants" a fixed factor. The F_1 is a *repeated-measures ANOVA*. The F_2 is a *mixed model ANOVA* since the factor *Frequency* is a grouping variable in the item analysis while the spacing remains a repeated-measures factor as in subject analysis. To consider a main effect or an interaction as generalisable, both F_1 and F_2 must be statistically significant.

Statistically significant main effects and interactions were followed by pairwise comparisons to determine which pairs are significantly different. ANOVA denotes a significant difference between —at least— two levels but it does not give information about the particular differences between pairs of means (Strunk & Mwavita, 2021). The post-hoc test, Tukey's Honestly Significant Difference (HSD), was used.

The statistical software JAMOVI (version 2.2.5) was used for all the descriptive and inferential statistics at an alpha level of .05.

3.7 Validity and Reliability

The quality of the methodology and the generalizability of the findings are assessed with respect to the concepts of validity and reliability.

Reliability concerns the "repeatability" of a study (Cumming & Calin-Jageman, 2017, p. 26); how reproducible the study results are when the experiment is carried out over repeated occasions under the same circumstances (De Vaus, 2002). When an experiment elicits consistent results on repeated trials, reliability is considered as high.

Validity refers to the accuracy of a measurement (Cohen et al., 2018). There are four main types of validity: statistical validity, construct validity, internal validity, and external validity (Cook & Campbell, 1979).

Statistical validity refers to the consistency and evidence-based connections among the data, the statistical analyses, and the final conclusions. It concerns the statistical power, potential violations of the assumptions of the statistical analyses, *Type I errors* (i.e. concluding that there is a relationship when the null hypothesis cannot be rejected) and *Type II errors* (i.e., concluding there is no relationship when there is) (Cohen et al., 2018).

Construct validity depends on the operationalization of the theoretical constructs the research study is based on. To fulfil the criterion of construct validity, the study design should comply with the theoretical background and the existing literature.

Internal validity in quantitative research refers to the degree to which researchers' conclusions and explanations can be sufficiently supported by the data (Clark-Carter, 2010). Certain threats to the internal validity of within-subjects designs should be considered. (e.g., fatigue effects, order effects) and minimised with the appropriate modifications.

External validity relates to the transferability and generalizability of the findings to other conditions, participants and experimental items (Clark-Carter, 2010). An unrepresentative and small sample are potential threats to external validity limiting study interpretations.

3.8 Research Ethics

To protect participant's right to privacy, the study adheres to *research ethics* defined as "a wide variety of values, norms, and institutional arrangements" that regulate scientific activities (The National Committee for Research Ethics in the Social Sciences and the Humanities [NESH], 2016).

Every research project in Norway needs approval from the Norwegian Centre for Research Data (NSD). The present study is a part of "BetterReading", a research project carried out by researchers at the department of Special Needs Education of University of Oslo and the department of Psychology and Brain Science of University of Iowa in close collaboration with the Oslo Special Education and Learning Lab. An assessment of the privacy impact was performed and research permission was obtained from the NSD. Studies —within the project— using the VWP with adult participants will be completed 31/12/2023.

The processing of personal data, including collecting, storing, organising, using, and publishing information that can be linked to an individual, follows the principles specified in article 5 of the General Data Protection Regulation (GDPR). These principles are "lawfulness, fairness, and transparency", "purpose limitation", "data minimization", "accuracy", "storage limitation", "integrity and confidentiality", and "accountability".

In compliance with these guidelines, participants were given a consent form before testing (Appendix D). This form included information about the research purposes, the procedure, participants' rights, data storage, data protection, and data anonymization. Participants did not sign the form; consent was given orally. However, silence or inaction was not considered as consent. Participants could withdraw their consent at any point during the testing without any repercussions, but none did so.

To protect the data from unauthorised access, damage and/or data leakage, collected personal data are fully anonymized. Participants were asked only about their age, noted as a number of years, and gender. Personal data like name, or personal identification number was not collected. Every participant was assigned a unique ID number starting with the first two capital letters of their nationality (i.e., Norwegian) followed by four digits (i.e., NOXXXX).

The experiment was non-invasive. A friendly positive environment was ensured for participants. Time for questions before and after the experiment was provided. Participants were encouraged to take breaks and they could adjust the height of the desk where the headrest was placed to feel comfortable during the eye-tracking session.

4. Results

This section presents the main findings of the study based on the preprocessing, plotting, and analysis of the eye-tracking data.

4.1 Data Processing

The VWP data were preprocessed using *VWPre* package (version 1.2.4) available in R (Porretta et al., 2020) as implemented in RStudio (version 2023.03.0+386). This package is suitable for data collected with SR Research EyeLink eye trackers like the one used for the present data collection. Raw eye-tracking data were exported as a Sample Report using the SR Research Data Viewer software.

Four Interest Areas (IAs) were predefined (top right, top left, bottom right, bottom left) corresponding to: *Target*, *Competitor*, *Distractor 1,* and *Distractor 2*. The preprocessing package aligns samples (i.e., looks) to each IA by calculating the probability of looking at each picture in a time window relative to stimulus onset. The time window chosen for the present study is defined from word presentation (0 ms) to 1000 ms post-stimulus onset.

The sampling rate in the data was 1000 Hz; one sample was taken every 1 ms. Samples were grouped into bins using the function "bin_prop". This function calculates the proportion of samples to each IA for a specific time window. Considering that a fixation lasts approximately 200–250 ms, the desired bin size was set at 200 ms. The maximum count of samples in a 200-millisecond interval is 200 as one sample is taken every 1 ms. If 200 samples are in a single quadrant of the screen, the proportion of looks to the remaining IAs adds up to .0 (0%). If a participant looks for 100 ms at one IA and then shifts their gaze to another for 100ms, each quadrant gets equal proportions of looks, .5 (50%) each. Therefore, samples were grouped into the following binning windows (in ms): 0–200 (first bin at 0 ms), 200–400 (second bin at 200 ms), 400–600 (third bin at 400 ms), 600–800 (fourth bin at 600 ms) and 800–1000 (fifth bin at 800 ms).

Looks outside the IAs, blinks and data from the practice trials were excluded from the subsequent data visualisations and statistical analyses.

4.2 Data Plotting

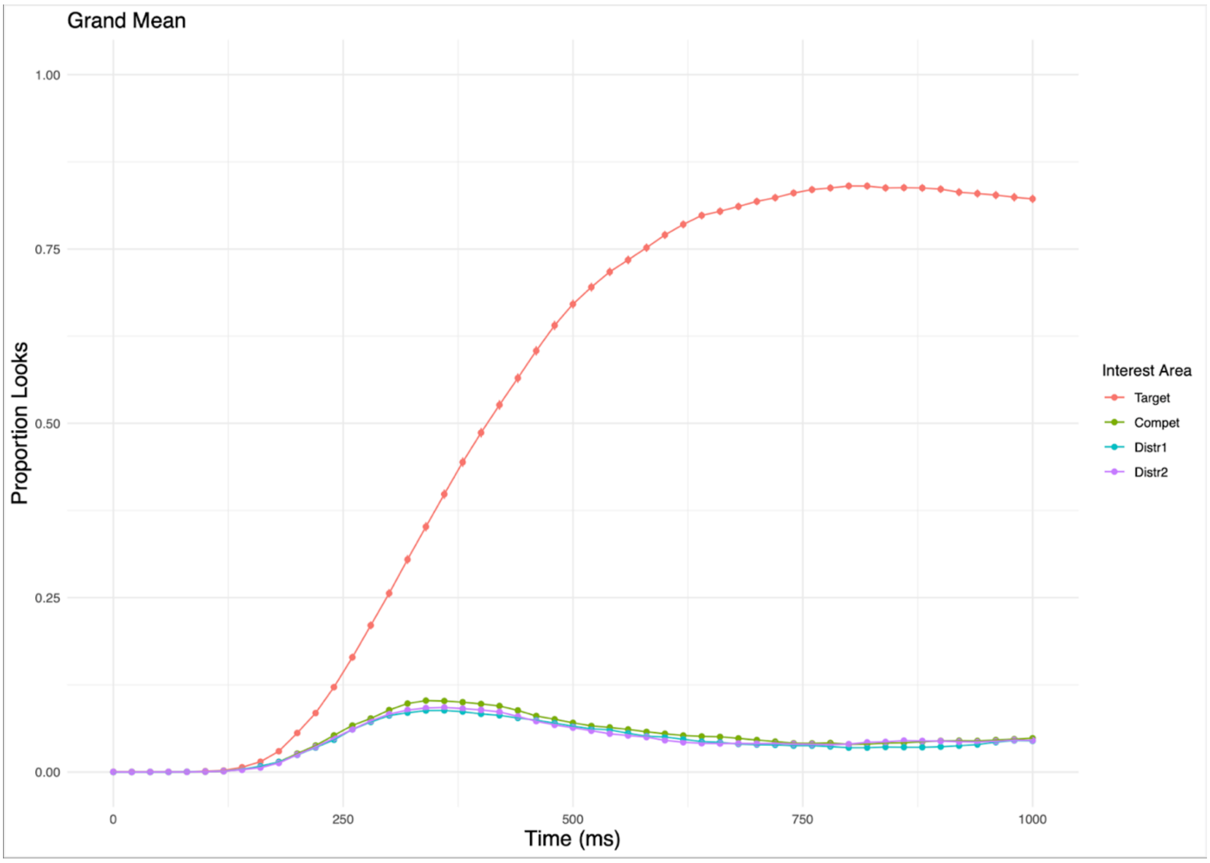
For the data visualisation, the function "plot_avg" was used (Porretta et al., 2020). This function is powered by the plotting package "ggplot2" in R and further customization (e.g., adding colours, modifying labels) was possible. Plotting raw eye-tracking data enables researchers to initially detect patterns and decide the time intervals that will be further analysed.

In all produced graphs, the vertical axis (y) represents "proportion looks" and the horizontal axis (x) represents "time" in ms relative to stimulus onset. Figure 18 represents the grand mean of proportion looks to each IA from the stimulus onset to 1000 ms. Figure 19 demonstrates the mean proportion looks to each IA by spacing condition and frequency group. Figure 20 illustrates the mean proportion looks to the target by spacing condition and frequency group.

Specifically, Figure 18 displays four curves. The red curve shows the grand mean of proportion looks to the target, the green curve the grand mean of proportion looks to the competitor, the blue and the purple curve the grand mean of proportion looks to the distractor 1 and 2 respectively. Overall, the target curve diverges from the other three curves early, approximately 200 ms post-word onset, and gradually increases; within 400 ms participants spend 50% of the time to the target IA and the target curve eventually levels off after reaching a peak around 85% within 800 ms. The more the time unfolds, the more the participants look at the target. Competitor and distractors are largely suppressed by 400 ms.

Figure 13

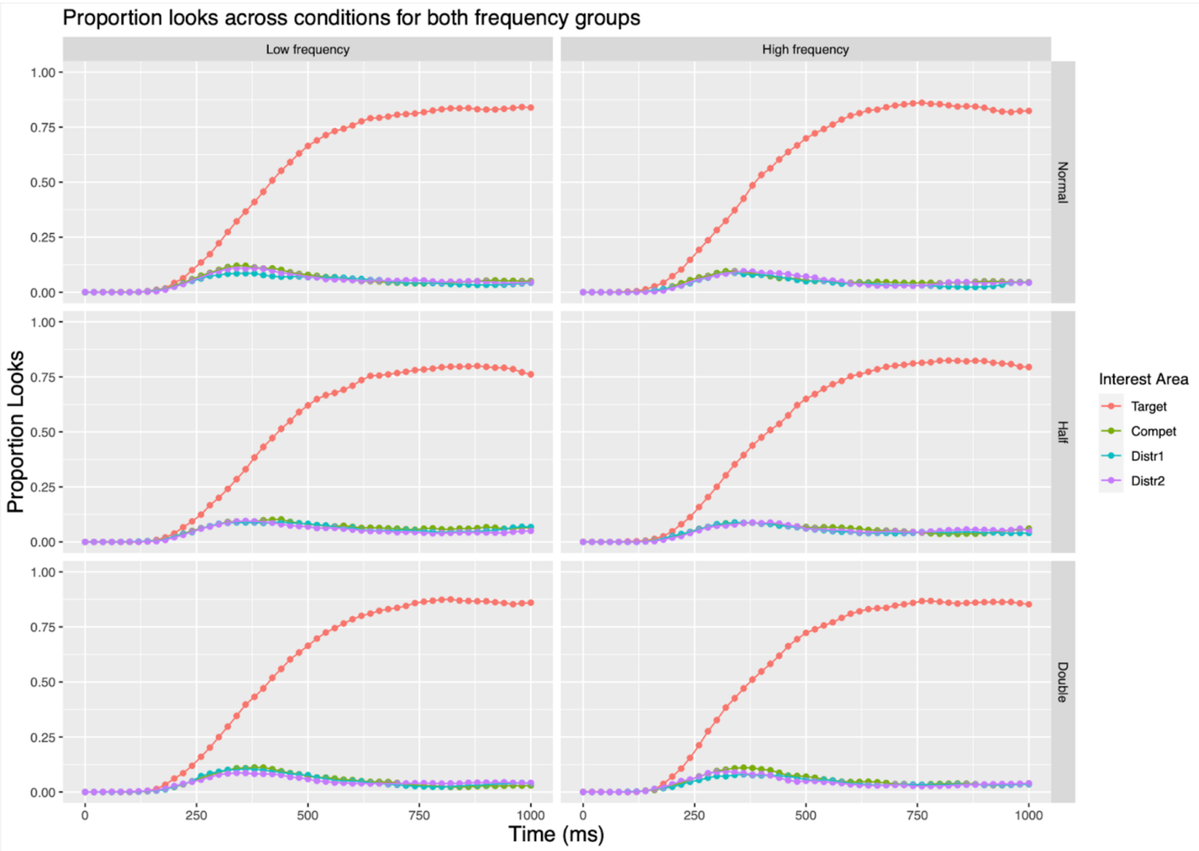
Grand Mean of Proportions Looks to each IA



The peak amount of target looks is nearly 85% within 800 ms yet this percentage appears to vary as a function of interword spacing based on Figure 19. In particular, the plot below summarises the mean proportion looks to each IA across conditions for higher-frequency (HF) and lower-frequency (LF) words.

Figure 14

Mean Proportion Looks to each IA by Spacing and Frequency



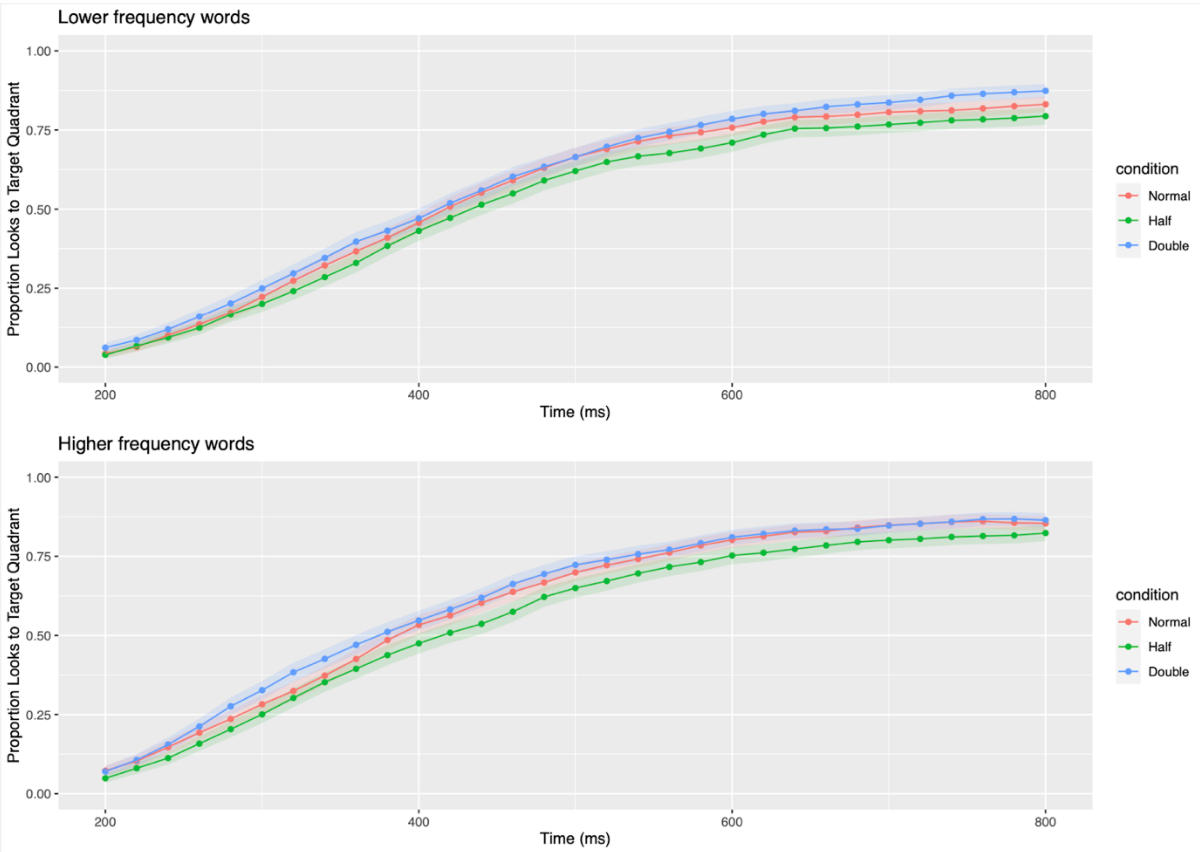
Regarding the trials with a target word of LF, in the normal-spacing condition average proportions of target looks peak at around 85% within 800 ms. In the half-spacing condition, they peak at 80%, a bit lower than in the normal-spacing, within 800 ms. In the double-spacing condition, the target curve reaches its highest point at approximately 90% within 800 ms. Thus, it looks like participants consider the target object faster in the double-spacing condition than in the normal- and the half-spacing. By contrast, it takes longer to look at the target object when the fixated word is presented in the half-spacing condition.

Regarding the trials with a target word of HF, in the normal-spacing condition average proportions of target looks peak at around 90% within 750 ms. In the half-spacing condition, they peak at 85%, a bit lower than in the normal spacing, within 800 ms. In the double-spacing condition, the target curve reaches its highest point at approximately 90% within 750 ms. The lexical activation for the target in the half-spacing condition seems to be slower while the target curves in the normal spacing and double spacing for the HF words seem to be almost the same. Participants spend 50% of the time looking at the target IA corresponding to a HF word within 400 ms in both normal- and double-spacing condition. Additionally, both curves reach a maximum level of target looks at approximately 90% within 750 ms.

Based on the above graph, it seems that participants' eye gaze starts shifting from the centre to the target quadrant after 200 ms and it settles there by the end as the proportion of target looks levels off after 800 ms. Judging from the graphs in Figures 18 and 19, the time bins selected to be analysed are 200–400 ms, 400–600 ms, and 600–800 ms. Figure 20 displays the mean proportion of target looks by spacing condition and frequency group as a function of time including the three aforementioned time bins.

Figure 15

Mean Proportions of Target Looks by Spacing and Frequency



It seems that participants take longer to look at the target IA when flankers are half space away as compared to trials that flankers are one or double space away; a pattern observed for both frequency groups. By contrast, it takes participants less time to look at the target object when flankers are double space away; a pattern observed for both frequency groups.

Regarding the differences between frequency groups, HF and LF words differ in the *maximum level* of target looks in the half-spacing and the normal-spacing condition. Frequency groups also differ in *when* they reach a peak in the normal condition (i.e., at approximately 750 ms for the HF words and at 800 ms for the LF words).

In the double-spacing condition both frequency groups reach a maximum level of nearly 90% of target looks within 750 ms. However, participants start considering the target object corresponding to a HF word earlier than target words corresponding to a LF word. They spend 50% of the time looking at the target picture corresponding to a HF word within almost 380 ms whereas it takes them longer to look at the target picture corresponding to a LF word spending 50% of the time fixating the target within 420 ms. This indicates a proximate difference of 40 ms within the first time-interval between the two frequency groups for targets presented in the double-spacing condition.

Moreover, what was noted in Figure 19 regarding the peaks of HF words in the normal- and double-spacing condition is also observed in Figure 20; target looks reach a peak of 90% within 750 ms in both spacing conditions. In contrast with the LF words panel (top) where the curves of double spacing and normal spacing maintain their differences throughout the time window, in the HF panel (bottom) it looks that the two curves overlap in the last time interval (600–800 ms).

4.3 Accuracy

4.3.1 Descriptive Statistics

Before using any inferential statistics to explore the effects and interactions, it is necessary to calculate basic descriptives as the distribution of a variable determines which statistical analyses should be performed. Tables of descriptive statistics summarise the measures of central tendency (e.g., mean), variability (e.g., standard deviation), skewness (asymmetry) and kurtosis (tail weight) giving information for the distribution of numeric variables (Navarro & Foxcroft, 2018). To run parametric tests such as ANOVA, the distribution should approximate the normal distribution (*assumption of normality*).

In normal distribution data points are evenly distributed around the mean; they are more tightly packed close to the mean and taper off as the distance from the mean increases (Navarro & Foxcroft, 2018). There are three approaches to evaluate normality: graphical evaluation by looking at the histograms and the corresponding Quantile-Quantile (Q-Q) plots, calculating two numerical indices namely, skewness and kurtosis, running the Shapiro-Wilk test of normality.

As a preliminary screening, the total accuracy of participants' responses was calculated by averaging the proportion of correct responses across conditions and frequency groups (M = 0.91, SD = 0.08). Table 1 shows the descriptive statistics of participants' accuracy by spacing condition and frequency group including Shapiro-Wilk p values.

Table 1

Descriptives for Accuracy Across Participants

	Low-Frequency			**High-Frequency**		
	Normal	**Half**	**Double**	**Normal**	**Half**	**Double**
M	0.92	0.88	0.94	0.92	0.88	0.95
SD	0.09	0.13	0.10	0.08	0.12	0.06
Skewness	–1.27	–1.38	–1.56	–0.90	–1.72	–1.49
Kurtosis	1.35	1.69	1.33	0.42	3.15	2.38
Shapiro-Wilk	0.83	0.84	0.70	0.87	0.81	0.78
p	< .001	< .001	< .001	< .001	< .001	< .001

Low-frequency Words

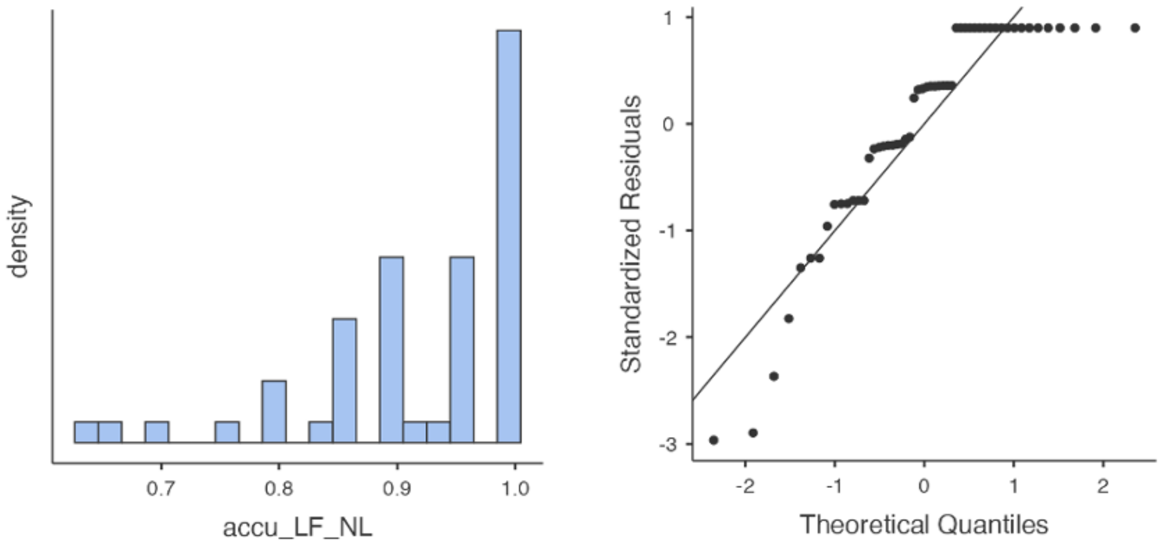
Accuracy in normal spacing:

The variable is distributed over a range from 0.64 to 1.00 with a mean of 0.92 (SD = 0.09). The histogram indicates a deviation from normality with a left-skewed peak. Skewness is –

1.27. Kurtosis is 1.35 indicating too many data points far from the mean compared to the expected normal distribution. The variable fails the Shapiro-Wilk test ($W = 0.83$, $p < .001$) indicating that it is unlikely that this variable is sampled from a normal distribution. Additionally, the standardised residuals on the Q-Q plot deviate from the reference line (i.e., the line indicating the theoretically expected normal distribution).

Figure 16

Accuracy in Normal Spacing for LF Words

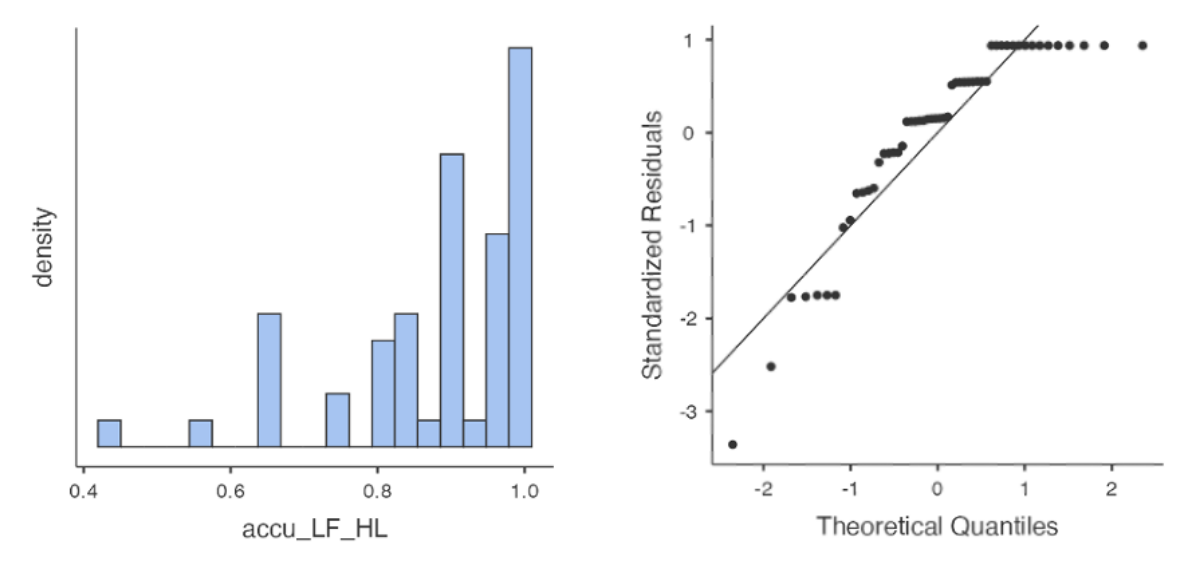


Accuracy in half spacing:

The variable ranges from 0.44 to 1.00 with a mean of 0.88 ($SD = 0.13$). Skewness is –1.38 indicating left-skewed data. Kurtosis is 1.69 indicating an excess of data points far from the mean. The Shapiro-Wilk test produces a statistically significant result ($W = 0.84$, $p < .001$), and therefore, the variable is not considered as normally distributed.

Figure 17

Accuracy in Half Spacing for LF Words

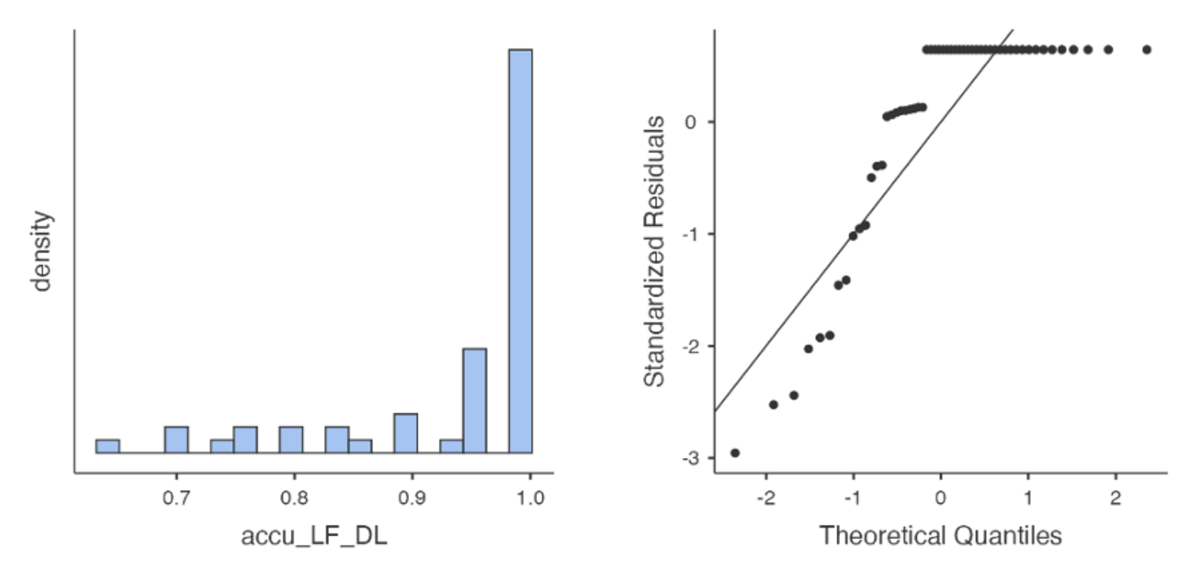


Accuracy in double spacing:

The variable ranges from 0.65 to 1.00 with a mean of 0.94 (*SD* = 0.10). Skewness is –1.56 indicating a pile-up on the right side of the curve. Kurtosis is 1.33 indicating an excess of data far from the mean. The variable fails the Shapiro-Wilk test ($W = 0.70$, $p < .001$) indicating a violation of normality.

Figure 18

Accuracy in Double Spacing for LF Words



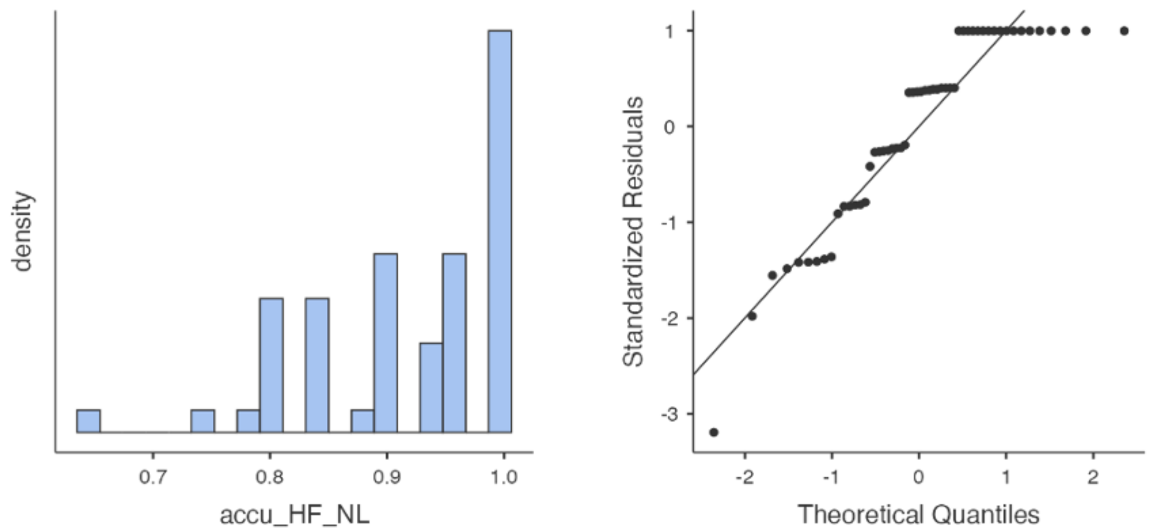
High-frequency Words

Accuracy in normal spacing:

The variable ranges from 0.65 to 1.00 with a mean of 0.92 (*SD* = 0.08). The histogram indicates a deviation from normality with a left-skewed peak. Both skewness (–0.90) and kurtosis (0.42) are between –1 and 1 indicating a mild deviation from normality as far as these two indices are concerned. The Shapiro-Wilk test produces a statistically significant result ($W = 0.87$, $p < .001$) indicating a violation of normality.

Figure 19

Accuracy in Normal Spacing for HF Words

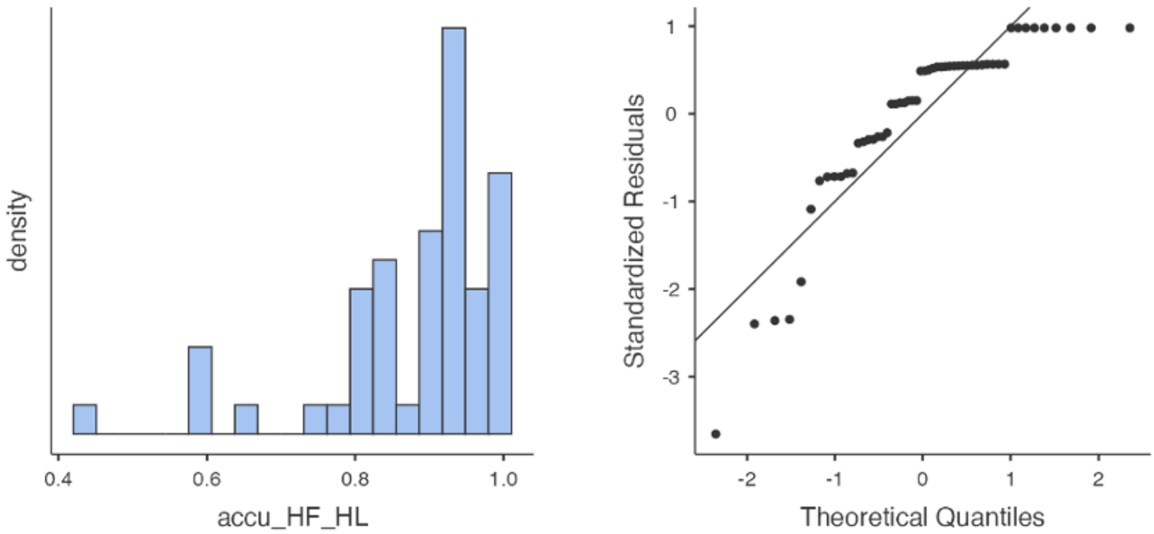


Accuracy in half spacing:

The variable is distributed over a range from 0.44 to 1.00 with a mean of 0.88 (*SD* = 0.12). Skewness is negative –1.72 indicating left-skewed data. Kurtosis is 3.15 indicating an abundance of data far from the mean. The variable fails the Shapiro-Wilk test ($W = 0.81$, $p < .001$).

Figure 20

Accuracy in Half Spacing for HF Words

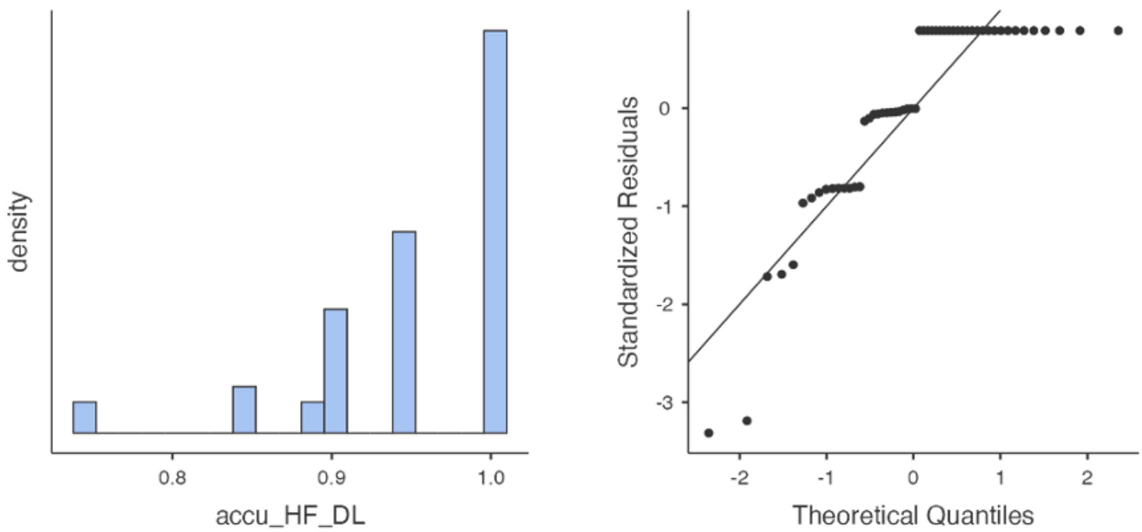


Accuracy in double spacing:

The variable ranges from 0.74 to 1.00 with a mean of 0.95 (*SD* = 0.06). Skewness is –1.49 indicating a left-skewed distribution. Kurtosis is 2.38 indicating an excess of data far from the mean. The Shapiro-Wilk test produces a statistically significant result ($W = 0.78$, $p < .001$) indicating a violation of normality.

Figure 21

Accuracy in Double Spacing for HF Words



Overall, it looks like participants' proportions of correct responses are lower in the half-spacing condition than in the normal- and double-spacing. The highest accuracy percentages are observed in the double-spacing condition for both frequency groups. However, just by looking at the descriptives, it is not possible to tell whether these differences are significant. As the assumption of normality is violated, we proceed to ANOVA with caution considering that ANOVA is quite robust to deviations from normality (Cribbie & Klockars, 2018). We expected that the variable would be skewed, as the values of accuracy range from 0 to 1 and only participants with a high accuracy level are included in the analyses.

It should be noted that in cases of skewed data, the median and the interquartile range should be reported as measures of central tendency and variability respectively as they are less affected by outliers and skewed data. Nevertheless, for the purposes of the subsequent analyses we report the *M* and the *SD*.

Descriptive statistics of the proportion of correct responses in the item sample were also performed. The table of descriptives, along with the corresponding histograms and Q-Q plots, can be found in Appendix E. The proportion of correct responses across items were similar to the percentages of correct responses across participants. Therefore, both F_1 and F_2 should be interpreted with caution. It is also necessary to check the remaining assumptions.

4.3.2 Assumptions of ANOVA

Apart from the normality check presented in the above section, the assumption of *sphericity* for the within-subjects factors and the assumption of *homogeneity* of variance across the levels of the between-subjects factor (Navarro & Foxcroft, 2018), before running a repeated-measures and a mixed-design ANOVA respectively, should be checked.

Figures 27 and 28 display the Q-Q plots to evaluate normality in the subject and item sample respectively. As expected, and mentioned earlier, data are left-skewed violating the assumption of normality. The distribution of the residuals in both data sets deviate from the reference line.

Figure 22

Q-Q Plot: Proportion Correct Responses Across Subjects

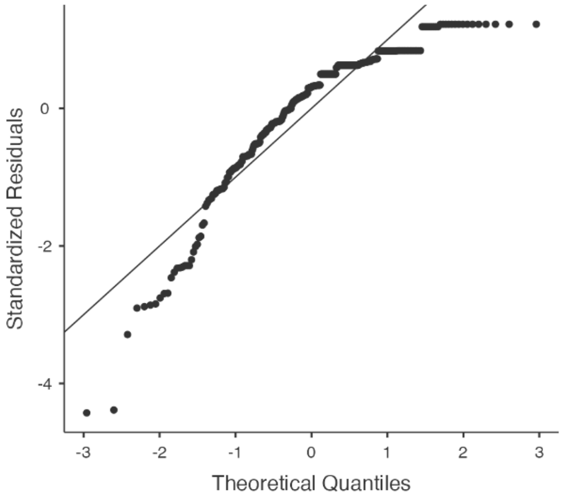
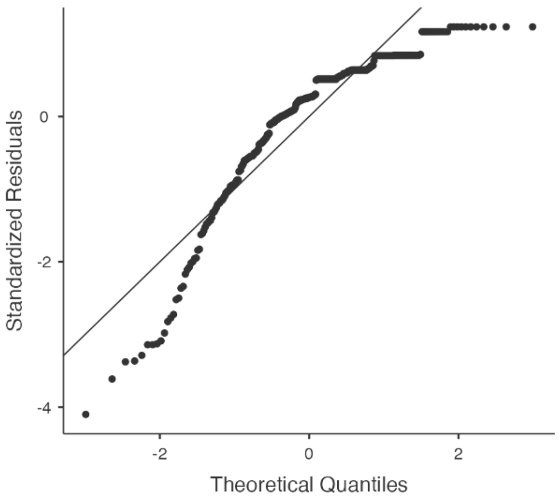


Figure 23

Q-Q Plot: Proportions Correct Responses Across Items



The assumption of sphericity for the repeated-measures variables concerns the variance of differences between levels and applies for factors with three or more levels. Sphericity is checked with the Mauchly's test (Lix & Keselman, 2018). If it is violated, a correction is applied using the Huynh-Feldt ε (Huynh & Feldt, 1976). As *Spacing* is a repeated-measures variable in both F_1 and F_2, we checked whether the variances of differences across the three levels (of all possible pairs) are equal. Mauchly's W test indicated that the null hypothesis (H_0) of equal variances of differences between levels is rejected (*p* = .002 for the F_1; *p* = .009 for the F_2). Hence, Huynh-Feldt correction is applied (Tables 2 and 3).

Frequency is a within-subjects variable in the subject sample, but it has two levels. The assumption of sphericity applies for factors with three or more levels.

The assumption of sphericity is fulfilled for the interaction between *Spacing* and *Frequency* (*p* = .130).

Table 2

Tests of Sphericity for the Repeated-measures Factor of F_1 (accuracy)

	Mauchly's W	***p***	**Huynh-Feldt ε**
Spacing	.79	.002	.85
Frequency*Spacing	.92	.130	.96

Table 3

Test of Sphericity for the Repeated-measures Factor of F_2 (accuracy)

	Mauchly's W	***p***	**Huynh-Feldt ε**
Spacing	.79	.002	.85

Frequency in the item sample is a grouping, between-subjects variable and, thus, the assumption for homogeneity of variance is tested with Levene's test. As indicated in Table 4, we cannot reject the H_0 for equal variances. The assumption of homogeneity is not violated.

Table 4

Levene's Test for the Between-subjects Factor in F_2 (accuracy)

	F	***df*1**	***df*2**	***p***
Half	0.22	1	118	.638

Normal	0.01	1	118	.903
Double	0.51	1	118	.478

4.3.3 Two-way ANOVA

To investigate the effects of *Spacing* on participants' accuracy, statistical interaction between *Spacing* and *Frequency* while accounting for the noise induced not only by the participants but also by the items two ANOVA are carried out, F_1 and F_2. Both *F* tables present the effects of factors, the effects of interactions and the residuals from the within-subjects sum of squares (and the residuals from the total sum of squares in the mixed-design ANOVA).

The effect size estimate for the repeated-measures ANOVA, partial eta squared (η^2_p), is also reported. The η^2_p denotes the proportion of the variability not accounted for by any other factors (or combination of factors).

Subject Analysis-F_1

Starting with the main effects (i.e., the average effects of each factor), the within-subjects factor *Frequency* has a non-significant main effect, $F(1,53) = 0.35$, $p = .555$. We cannot reject the H_0 that there is no difference in the means of the two frequency groups for this subject sample.

The within-subjects factor *Spacing* has a statistically significant main effect after correcting for sphericity, $F(1.71,90.39) = 21.04$, $p < .001$, $\eta^2_p = .28$. Spacing accounts for 28% of the variability that is not accounted for by other factors. Calculating the eta square for *Spacing* reveals that 6% of the total variability is explained by *Spacing* ($\eta^2 = .06$).

There is no statistically significant interaction between *Frequency* and *Spacing*, $F(1.92,102.00) = 0.45$, $p = .633$. This suggests that whatever the effect of spacing is, it is not different for HF and LF words. Although the interaction does not violate the assumption of sphericity, we report the corrected values because *Spacing* violates it.

Table 5

F_1 for the Within-subjects Effects on Proportion Correct Responses

	SS	*df*	MS	*F*	*p*	η^2_p

Frequency	0.00	1.00	0.00	0.35	.555	.01
Residual	0.35	53.00	0.01			
Spacing	0.22	1.71	0.13	21.04	**< .001**	.28
Residual	0.56	90.39	0.01			
Frequency*Spacing	0.00	1.92	0.00	0.45	.633	.01
Residual	0.30	102.00	0.00			

Note. Values are adjusted using the Huynh-Feldt correction.

Item Analysis-F_2

To find whether the main effect of *Spacing* are generalizable beyond the experimental word set, we run an item analysis (Table 6).

Spacing has a statistically significant main effect, after correcting for sphericity, $F(1.88,222.26) = 25.56$, $p < .001$, $\eta^2_p = .18$. Spacing accounts for 18% of the variability that is not accounted for by other factors. Calculating the eta square for *Spacing* reveals that 6% of the total variability is explained by *Spacing* ($\eta^2 = .06$). The H_0 that there is no difference in the means of the different spacing conditions for this set of words is rejected.

The main effect of *Frequency* is not statistically significant, $F(1,118) = 0.18$, $p = .669$. We cannot reject the H_0 that there is no difference in the means of the two frequency groups for this set of words (Table 7).

Turning to the interaction between *Spacing* and *Frequency*, it seems that the differences across spacing conditions do not differ for HF and LF words after correcting for sphericity, $F(1.88,222.26) = 0.24$, $p = .775$.

Table 6

F_2 for the Within-subjects Effects on Proportion Correct Responses

	SS	*df*	MS	*F*	*p*	η^2_p
Spacing	0.24	1.88	0.13	25.56	**< .001**	.18
Frequency*Spacing	0.00	1.88	0.00	0.24	.775	.00
Residual	1.09	222.26	0.00			

Note. Values are adjusted using the Huynh-Feldt correction.

Table 7

F_2 for the Between-subjects Effect on Proportion Correct Responses

	SS	*df*	MS	*F*	*p*	η^2_p
Frequency	0.00	1	0.00	0.18	.669	.00
Residual	2.50	118	0.02			

4.3.4 Post-hoc Comparisons

Statistically significant effects of factors with three (or more) levels are followed by pairwise comparisons using the Tukey's HSD. Regarding *Spacing*, the post-hoc analysis of F_1 revealed significant differences in proportions of correct responses between half and normal spacing ($p = .004$), half and double ($p < .001$), between normal and double ($p = .001$) (Table 8). Post-hoc analysis of F_2 also revealed significant differences in the proportions of correct responses between half and normal spacing ($p < .001$), half and double ($p < .001$), between normal and double ($p = .003$) (Table 9).

Figure 24 shows that participants achieved higher accuracy percentages in the double-spacing condition than in the half- and normal-spacing. By contrast, they achieved significantly lower accuracy percentages in the half-spacing condition than in the normal and double. The same differences emerged for the item sample (Figure 25).

Table 8

Tukey's HSD Comparisons of Spacing (F_1)

Comparison							
Spacing		**Spacing**	**Mean Difference**	**SE**	***df***	***t***	p_{tukey}
Half	-	Normal	–0.04	0.01	53.00	–3.35	**.004**
	-	Double	–0.06	0.01	53.00	–5.83	**< .001**
Normal	-	Double	–0.03	0.01	53.00	–3.74	**.001**

Figure 24

Estimated Marginal Means for Accuracy by Spacing Across Participants

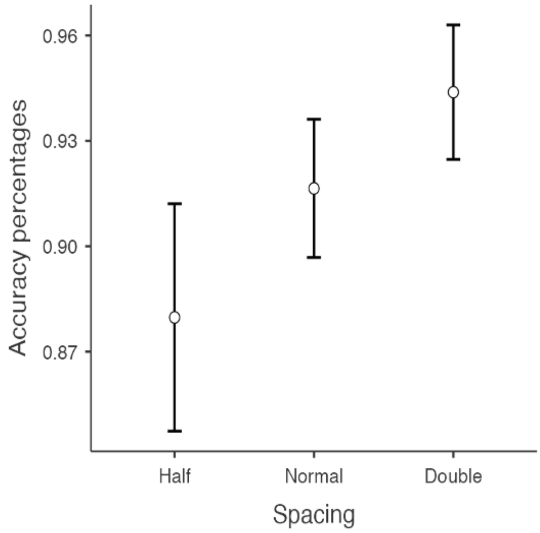


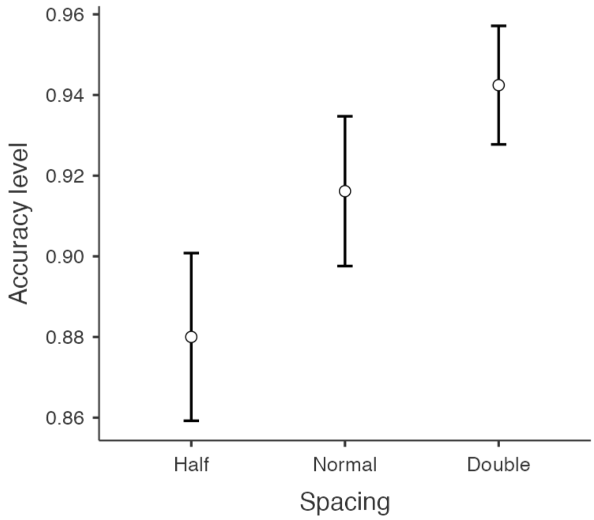
Table 9

Tukey's HSD Comparisons of Spacing (F_2)

Comparison							
Spacing		**Spacing**	**Mean Difference**	**SE**	***df***	***t***	p_{tukey}
Half	-	Normal	–0.04	0.01	118.00	–4.19	**< .001**
	-	Double	–0.06	0.01	118.00	–6.35	**< .001**
Normal	-	Double	–0.03	0.01	118.00	–3.40	**.003**

Figure 25

Estimated Marginal Means for Accuracy by Spacing Across Items



4.4 Proportions Looks to Target

4.4.1 Descriptive Statistics

The effects of spacing on lexical activation rate for the target words were investigated by analyzing participants' *proportions of target looks* over three time intervals: 200–400 ms, 400–600 ms, 600–800 ms. Tables 10, 11 and 12 present the descriptives of the dependent variable by spacing condition and frequency group across participants starting with the time interval at 200 ms (200–400), then at 400 ms (400–600) and finally at 600 ms (600–800).

Table 10

Descriptives for Proportion Target Looks at 200 ms

	Low-Frequency			**High-Frequency**		
	Normal	**Half**	**Double**	**Normal**	**Half**	**Double**
M	0.20	0.19	0.23	0.26	0.23	0.29
SD	0.12	0.11	0.14	0.14	0.12	0.13
Skewness	0.35	0.68	1.15	0.80	0.70	0.38

Kurtosis	–0.32	–0.14	1.91	0.32	0.39	–0.43
Shapiro-Wilk	0.97	0.95	0.93	0.94	0.95	0.98
p	.272	.026	.003	.009	.040	.340

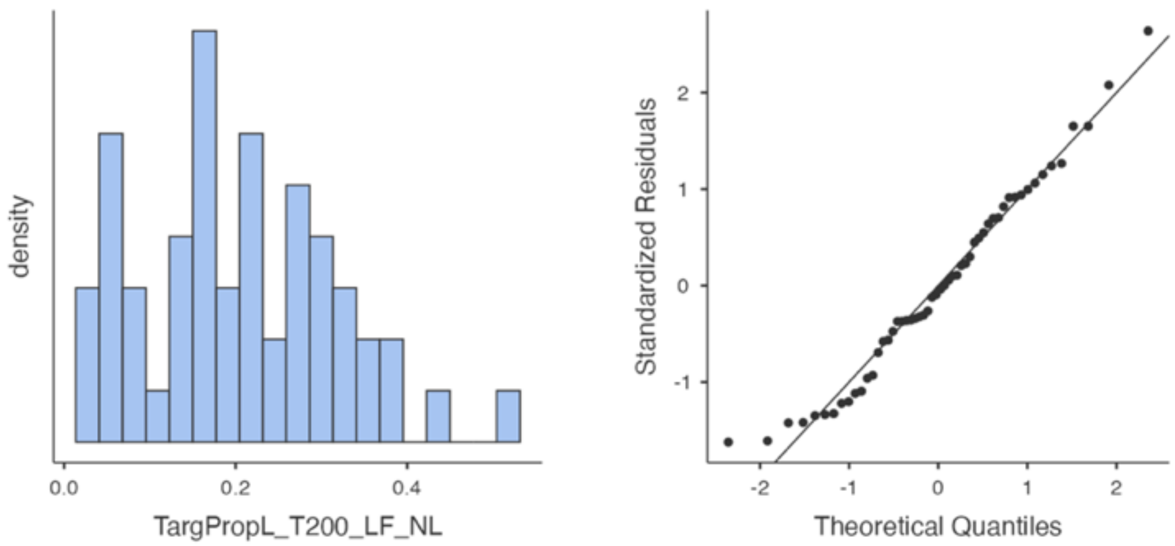
Low-frequency Words (200ms)

Proportion target looks in normal spacing:

The distribution seems to be slightly right-skewed and trails off to the right. The variable is distributed from 0.02 to 0.51 ($M = 0.20$, $SD = 0.12$). Both skewness (0.35) and kurtosis (–0.32) are between –1 and 1 indicating a mild deviation from normality. Shapiro-Wilk test produces a non-significant result ($W = 0.97$, $p = .272$). Overall, the distribution looks to fit quite well with the normal distribution.

Figure 26

Proportion Looks to LF Target in Normal Spacing (200 ms)



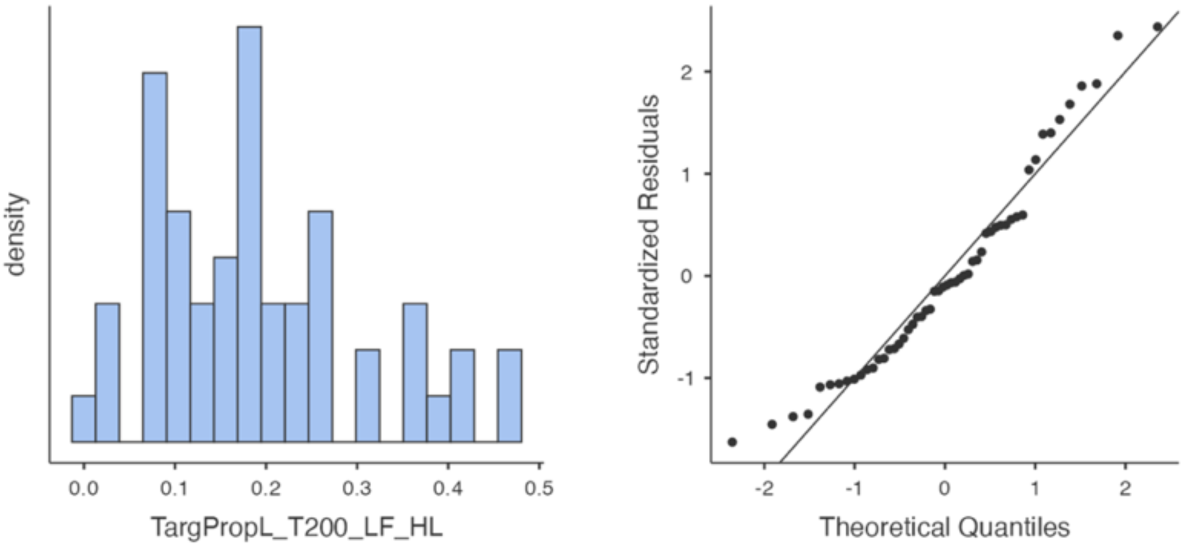
Proportion target looks in half spacing:

The distribution seems to be somewhat asymmetric. The variable ranges from 0.00 to 0.47 ($M = 0.19$, $SD = 0.11$). Skewness is 0.68 indicating right-skewed data. Kurtosis is –0.14

indicating fewer values in the tails than the expected normal distribution. Shapiro-Wilk test produces a significant result ($W = 0.95$, $p = .026$) implying a violation of normality.

Figure 27

Proportion Looks to LF Target in Half Spacing (200 ms)

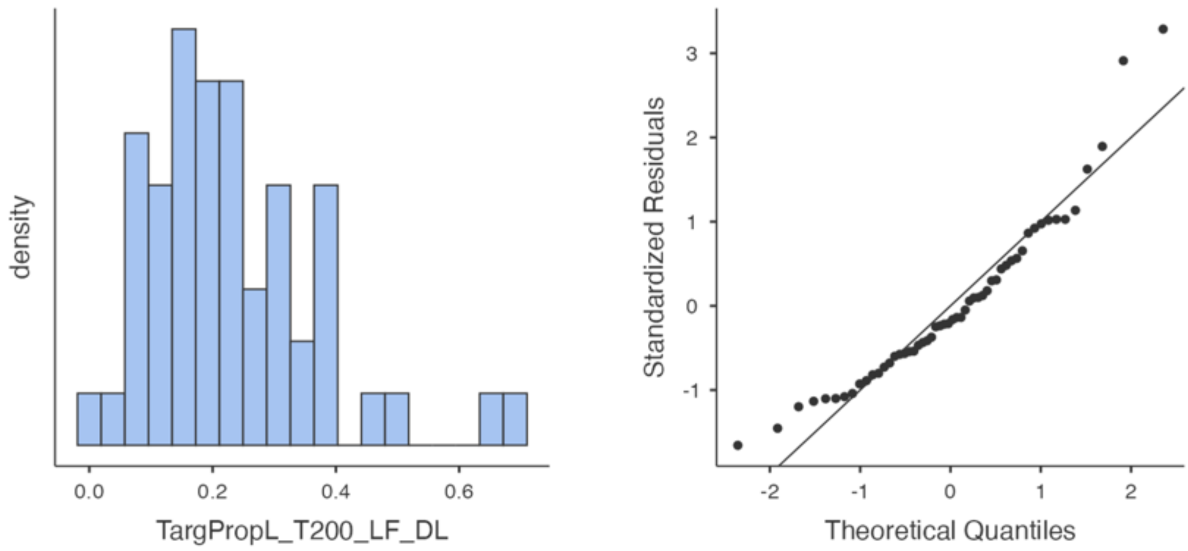


Proportion target looks in double spacing:

The variable ranges from 0.00 to 0.69 ($M = 0.23$, $SD = 0.14$) and few outliers on the right side. Skewness is 1.15 indicating a right-skewed distribution. Kurtosis is 1.91 indicating an excess of values in the tails. The variable fails the Shapiro-Wilk test ($W = 0.93$, $p = .003$).

Figure 28

Proportion Looks to LF Target in Double Spacing (200 ms)



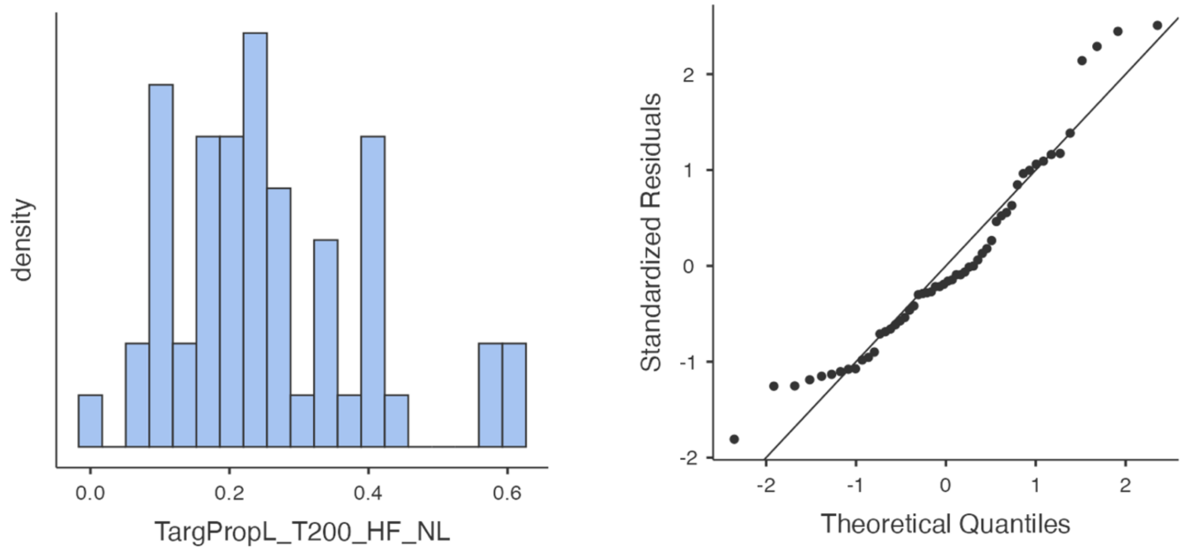
High-frequency Words (200 ms)

Proportion target looks in normal spacing:

The variable is distributed over a range from 0.00 to 0.61 (M = 0.26, SD = 0.14) with a humped peak at around 0.25 (25%) and few gaps between high points. Skewness is 0.80 indicating right-skewed data. Kurtosis is 0.32 indicating a slight excess of distant data. Shapiro-Wilk test produces a statistically significant result (W = 0.94, p = .009). Based on the output of the normality test, the variable is not considered as normally distributed.

Figure 29

Proportion Looks to HF Target in Normal Spacing (200 ms)

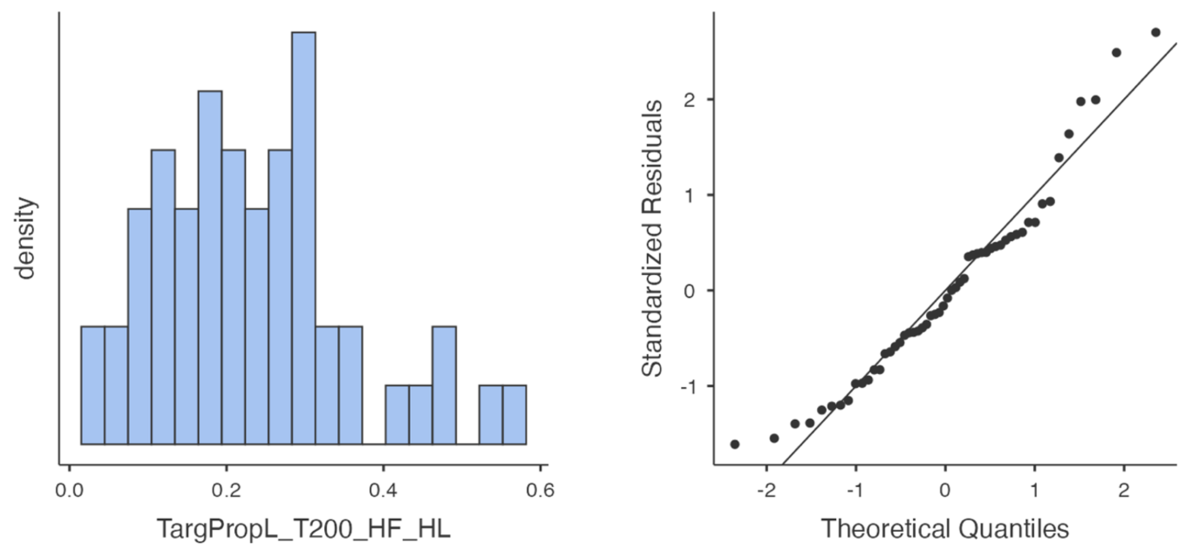


Proportion target looks in half spacing:

The distribution looks asymmetric as only a minor portion of values is located to the right of the peak. The variable ranges from 0.03 to 0.57 (M = 0.23, SD = 0.12). Skewness is 0.70 and kurtosis is 0.39 indicating a mild deviation from normality. The Shapiro-Wilk produces a significant result (W = 0.95, p = .040) implying a violation of normality.

Figure 30

Proportion Looks to HF Target in Half Spacing (200 ms)



Proportion target looks in double spacing:

The variable yielded a quite symmetric distribution and ranges from 0.07 to 0.59 ($M = 0.29$, $SD = 0.13$). Skewness is 0.38 indicating a slight skew to the right. Kurtosis is –0.43 indicating fewer values in the tails. Both numeric indices are within the range from –1 to 1 denoting a mild deviation from normality. Shapiro-Wilk test produces a non-significant result ($W = 0.98$, $p = .340$).

Figure 31

Proportion Looks to HF Target in Double Spacing (200 ms)

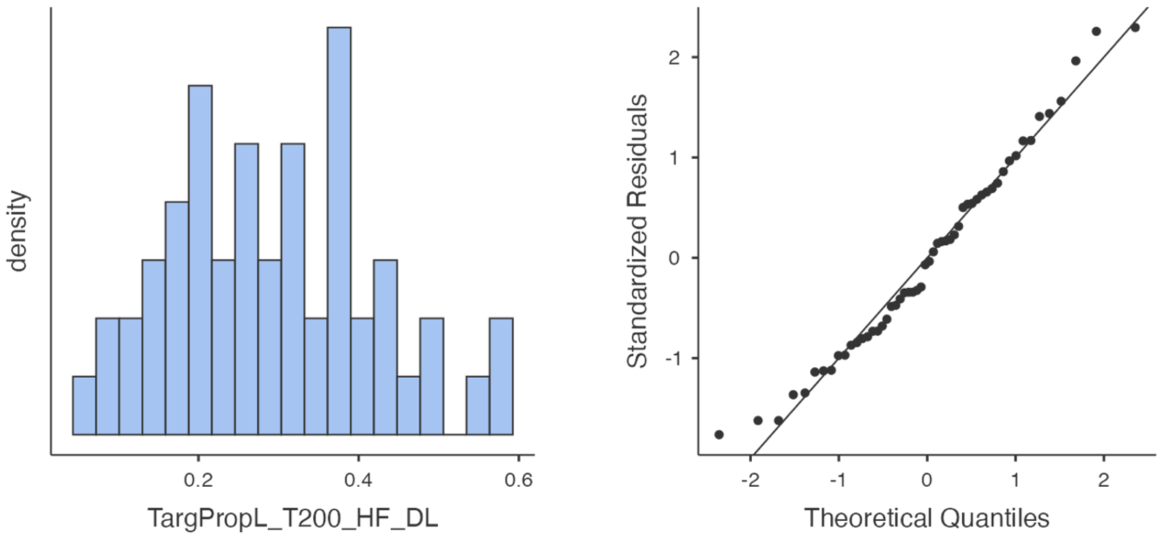


Table 11

Descriptives for Proportion Target Looks at 400 ms

	Low-Frequency			High-Frequency		
	Normal	**Half**	**Double**	**Normal**	**Half**	**Double**
M	0.62	0.58	0.63	0.67	0.61	0.69
SD	0.14	0.14	0.15	0.12	0.14	0.13
Skewness	–0.09	0.37	–0.31	–0.91	–0.08	0.01
Kurtosis	–0.84	–0.31	–0.41	2.41	–0.47	–0.08

Shapiro-Wilk	0.97	0.98	0.98	0.95	0.99	0.99
p	.290	.593	.703	.022	.833	.963

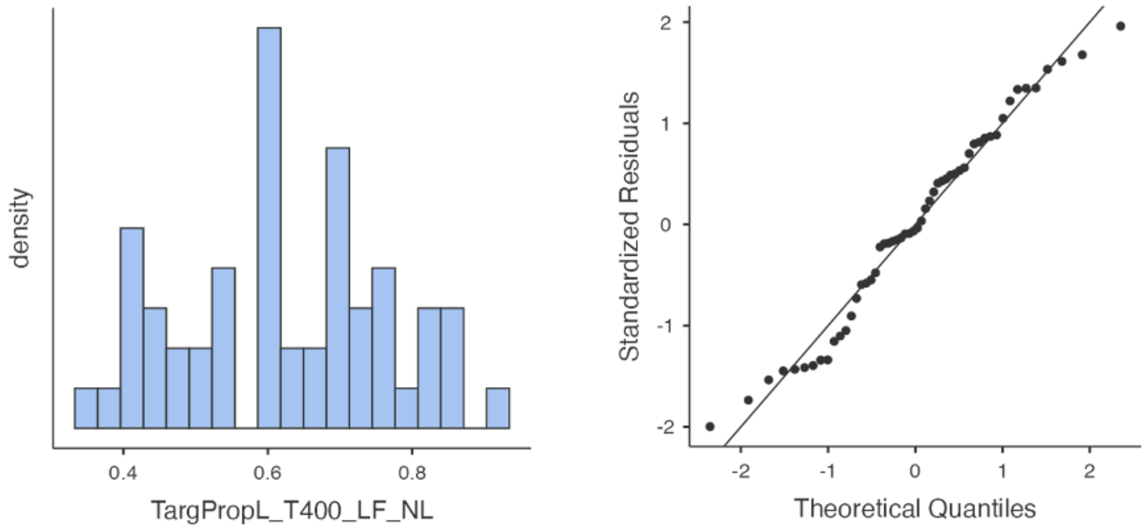
Low-frequency Words (400ms)

Proportion target looks in normal spacing:

The variable yielded a somewhat symmetric distribution with a few prominent spires and a distinct peak at 0.6 (60%). It ranges from 0.34 to 0.91 with a mean of 0.62 (*SD* = 0.14). Skewness is –0.09 and kurtosis is –0.84 indicating a mild deviation from normality. Shapiro-Wilk test produces a non-significant result (*W* = 0.97, *p* = .290).

Figure 32

Proportion Looks to LF Target in Normal Spacing (400 ms)

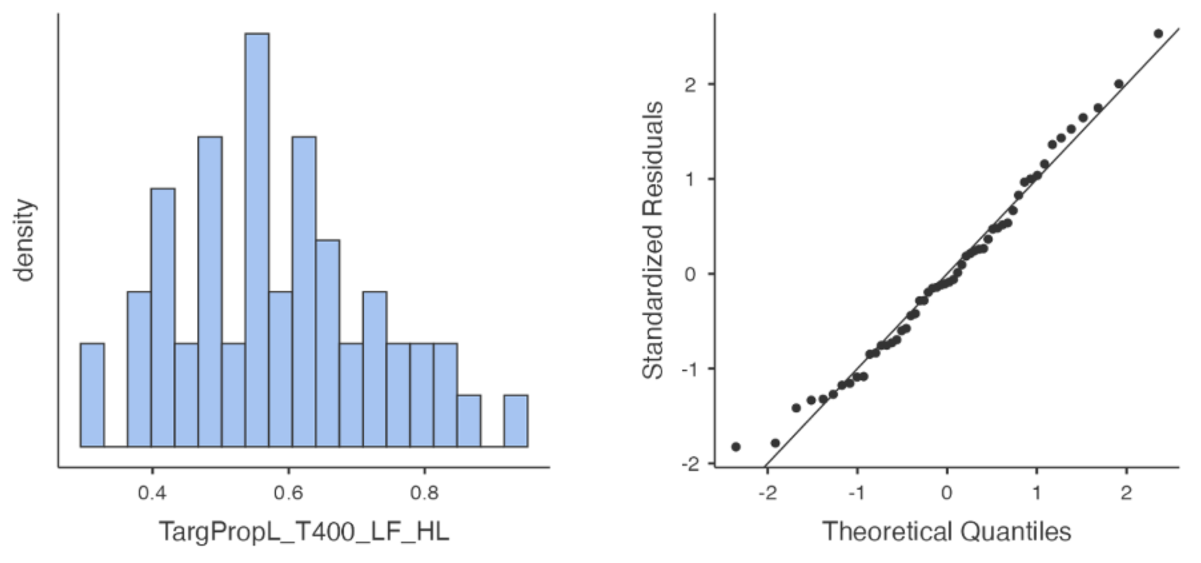


Proportion target looks in half spacing:

The variable yielded a symmetric distribution with few gaps between high points and a prominent peak at 0.55 (55%). It ranges from 0.32 to 0.94 with a mean of 0.58 (*SD* = 0.14). Skewness is 0.37 indicating a slight skew to the right. Kurtosis is –0.31 indicating fewer values in the tails than the expected normal distribution. Shapiro-Wilk test produces a non-significant result (*W* = 0.98, *p* = .593).

Figure 33

Proportion Looks to LF Target in Half Spacing (400 ms)

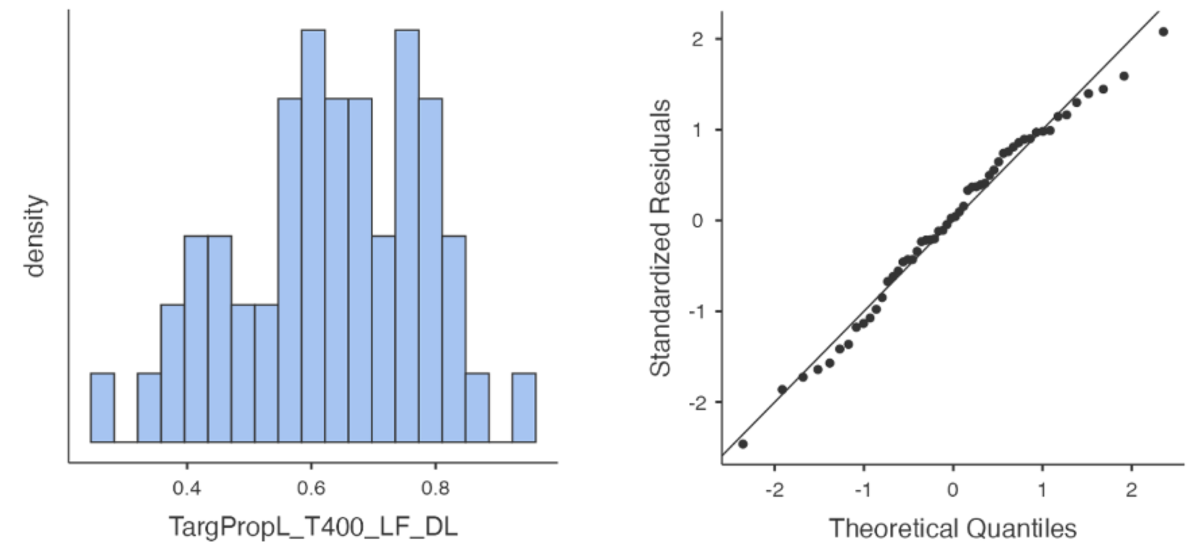


Proportion target looks in double spacing:

The distribution of the variable approximates normal distribution with a slight left skew (skewness = –0.31), fewer values in the tails (kurtosis = –0.41) and two clear peaks around 0.6 (60%) and 0.75 (75%). The variable ranges from 0.26 to 0.94 with a mean of 0.63 (SD = 0.15). The Shapiro-Wilk produces a non-significant result (W = 0.98, p = .703).

Figure 34

Proportion Looks to LF Target in Double Spacing (400 ms)



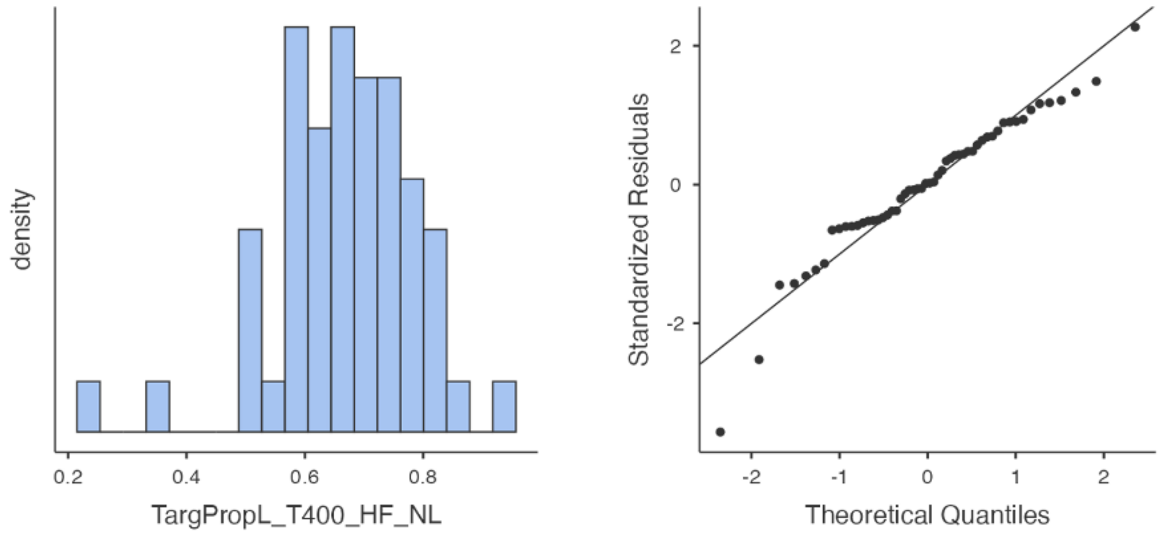
High-frequency Words (400ms)

Proportion target looks in normal spacing:

The variable yielded an asymmetric distribution with left-skewed data (negative skewness) and an excess of values far from the mean (positive kurtosis). The values range from 0.24 to 0.94 with a mean of 0.67 (*SD* = 0.12). The Shapiro-Wilk indicates a significant deviation from normality (*W* = 0.95, *p* = .022).

Figure 35

Proportion Looks to HF Target in Normal Spacing (400 ms)

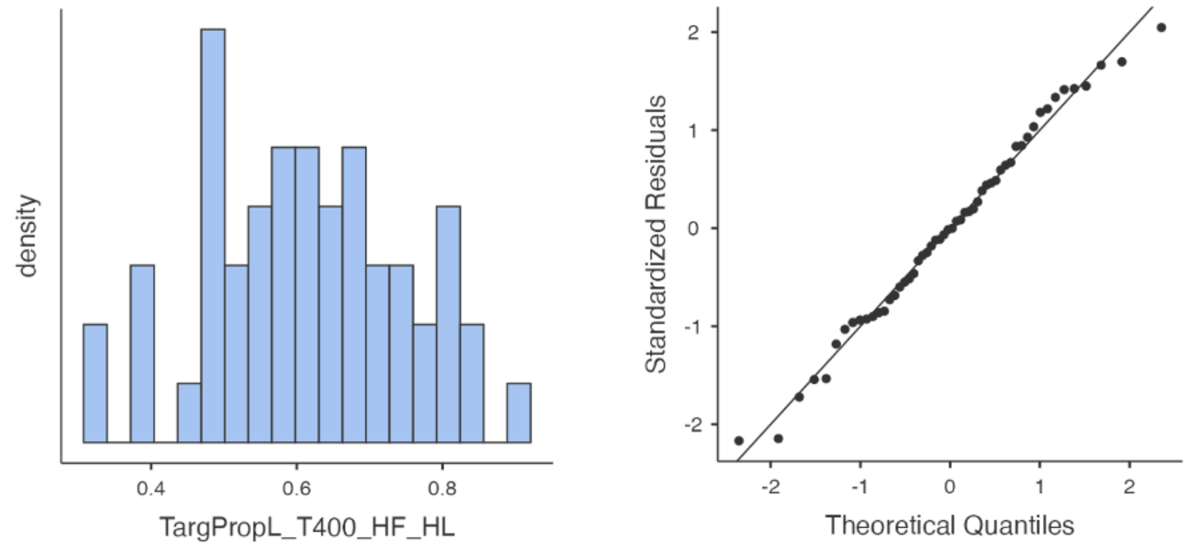


Proportion target looks in half spacing:

The histogram shows a relatively symmetric distribution with a distinct peak at around 0.5 (50%) on the left side of the distribution and a minor proportion of values located to the left side of this peak. The variable ranges from 0.31 to 0.89 with a mean of 0.61 (*SD* = 0.14). Skewness is –0.08 and kurtosis is –0.47. Shapiro-Wilk shows that the variable does not deviate from normality (*W* = 0.99, *p* = .833).

Figure 36

Proportion Looks to HF Target in Half Spacing (400 ms)



Proportion target looks in double spacing:

The histogram shows a symmetric distribution with values ranging from 0.39 to 1.00 and a mean of 0.69 (*SD* = 0.13). Skewness is 0.01 indicating symmetric data and kurtosis is –0.08. Shapiro-Wilk comes up non-significant denoting that the variable fits well with the normal distribution (*W* = 0.99, *p* = .963).

Figure 37

Proportion Looks to HF Target in Double Spacing (400 ms)

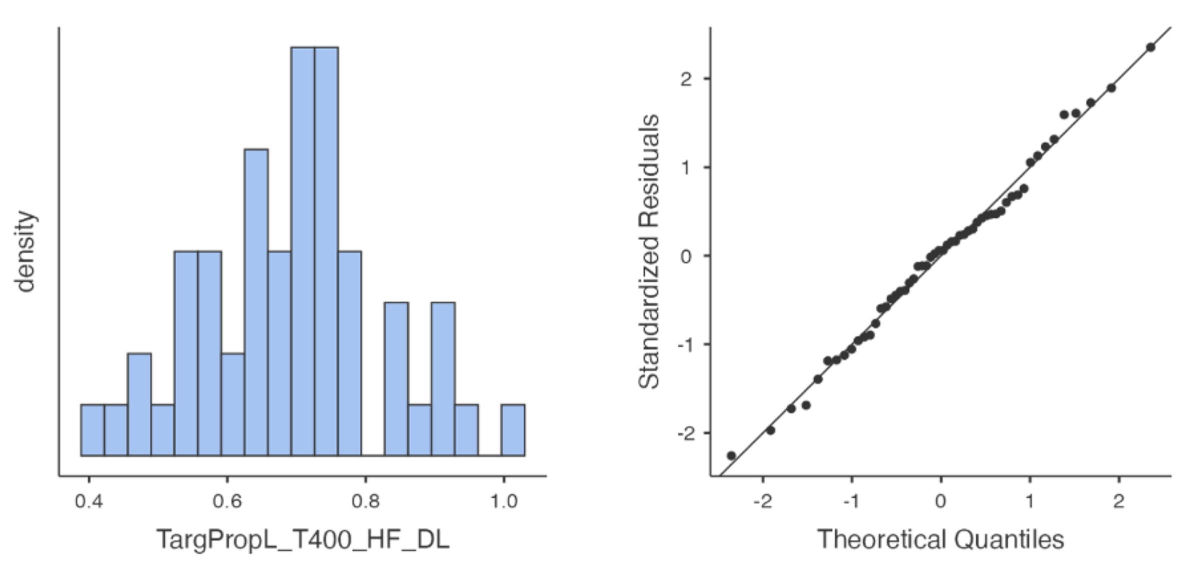


Table 12

Descriptives for Proportion Target Looks at 600 ms

	Low-Frequency			High-Frequency		
	Normal	**Half**	**Double**	**Normal**	**Half**	**Double**
M	0.79	0.75	0.82	0.84	0.78	0.84
SD	0.11	0.13	0.11	0.10	0.10	0.10
Skewness	–0.90	–0.31	–0.85	–0.90	–0.49	–0.56
Kurtosis	1.33	–0.22	1.48	2.14	–0.61	–0.08
Shapiro-Wilk	0.95	0.96	0.95	0.94	0.95	0.96
p	.022	.053	.026	.012	.034	.071

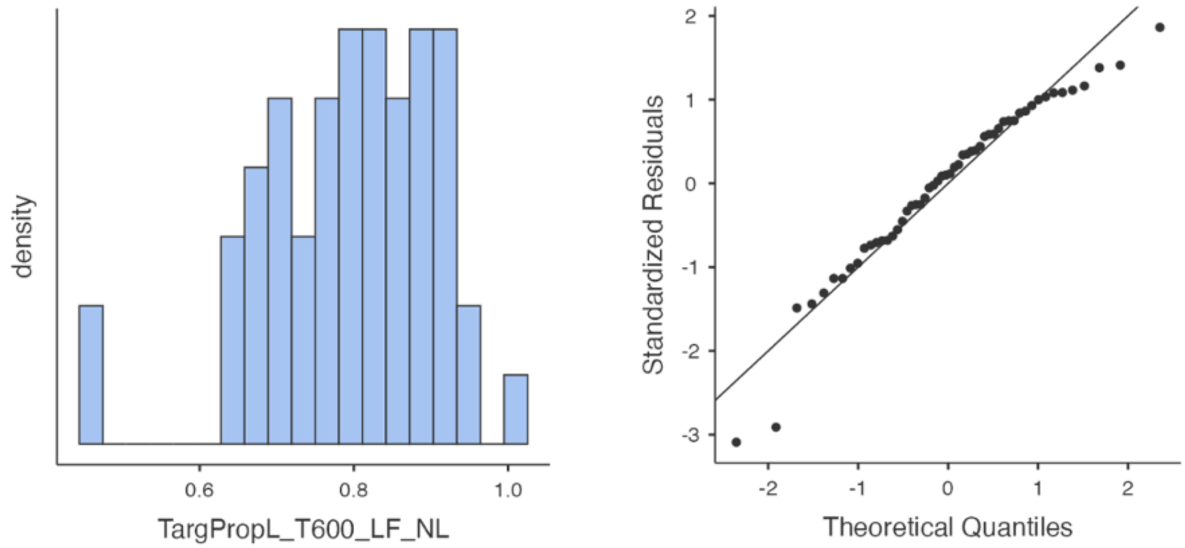
Low-frequency Words(600ms)

Proportion target looks in normal spacing:

The distribution seems to deviate from normality with values clustered towards the right side of the distribution and somewhat thin tails. The variable ranges from 0.45 to 1.00 (M = 0.79, SD = 0.11). Skewness is –0.90 and kurtosis is 1.33. Shapiro-Wilk test (W = 0.95, p = .022) produced a statistically significant result implying a violation of normality.

Figure 38

Proportion Looks to LF Target in Normal Spacing (600 ms)

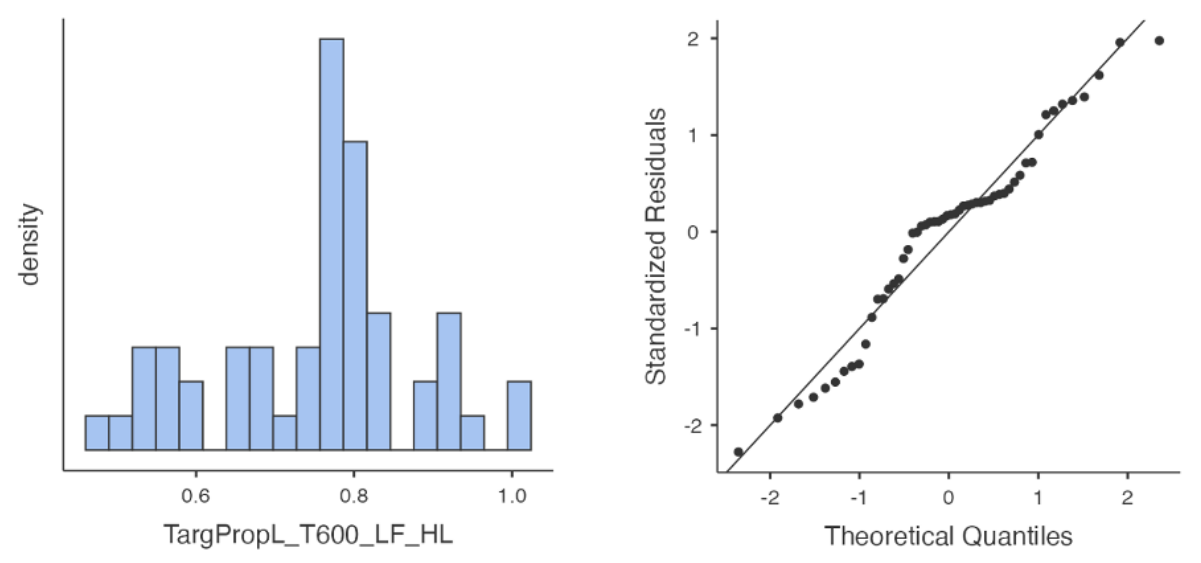


Proportion target looks in half spacing:

Based on the displayed graphs, the variable yielded a rather flat distribution with two prominent peaks at around 0.8 (80%). The variable ranges from 0.47 to 1.00 (M = 0.75, SD = 0.13). Skewness is –0.31 and kurtosis is –0.22. Shapiro-Wilk comes up non-significant (W = 0.96, p = .053).

Figure 39

Proportion Looks to LF Target in Half Spacing (600ms)

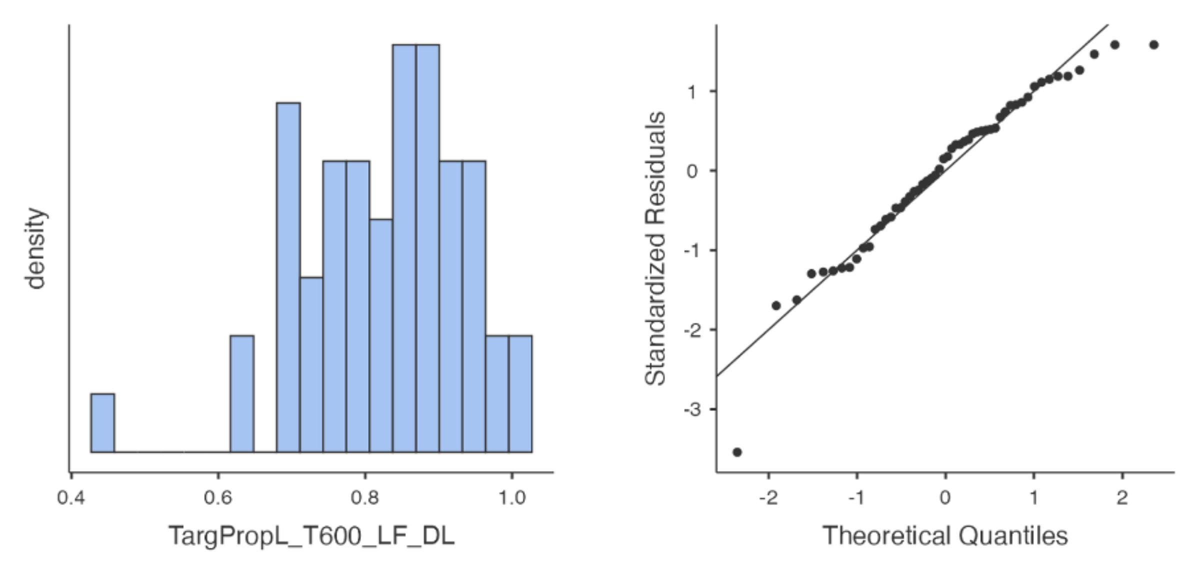


Proportion target looks in double spacing:

The histogram shows a somewhat asymmetric distribution with values ranging from 0.43 to 1.00 (M = 0.82, SD = 0.11). Values are clustered towards the end of the right tails with a negative skewness of –0.85 and a kurtosis of 1.48. The variable fails the normality test (W = 0.95, p = .026).

Figure 40

Proportion Looks to LF Target in Double Spacing (600ms)



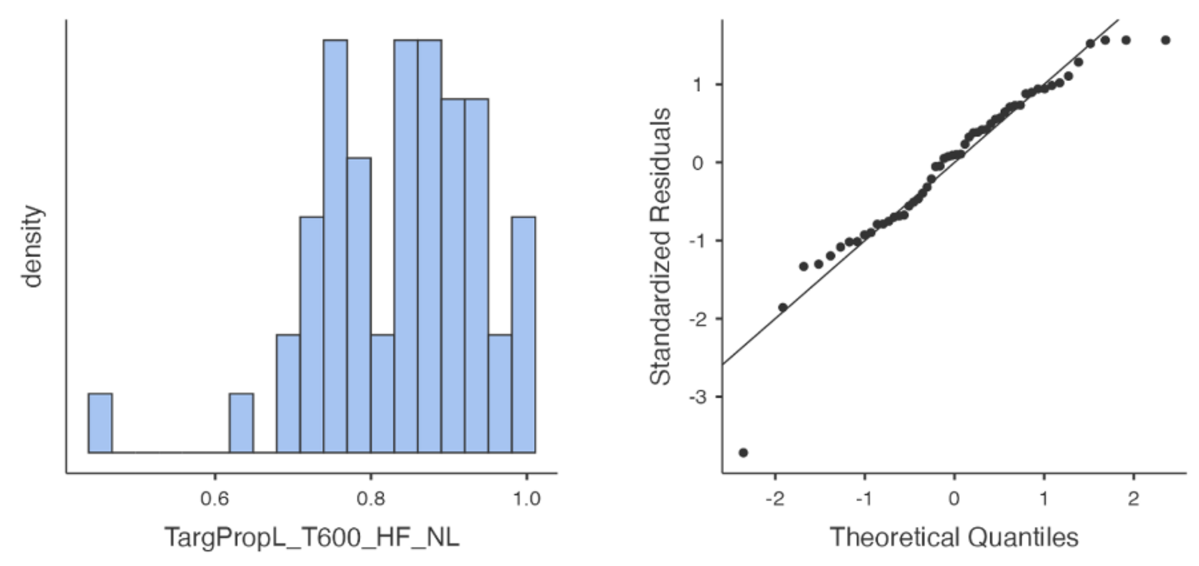
High-frequency Words (600ms)

Proportion target looks in normal spacing:

Based on the graphs, the distribution seems to deviate from normality with values clustered towards the end of the right tail. The variable ranges from 0.46 to 1.00 ($M = 0.84$, $SD = 0.10$). Skewness is –0.90 and kurtosis is 2.14. Shapiro-Wilk test ($W = 0.94$, $p = .012$) produced a statistically significant result implying a violation of normality.

Figure 41

Proportion Looks to HF Target in Normal Spacing (600 ms)

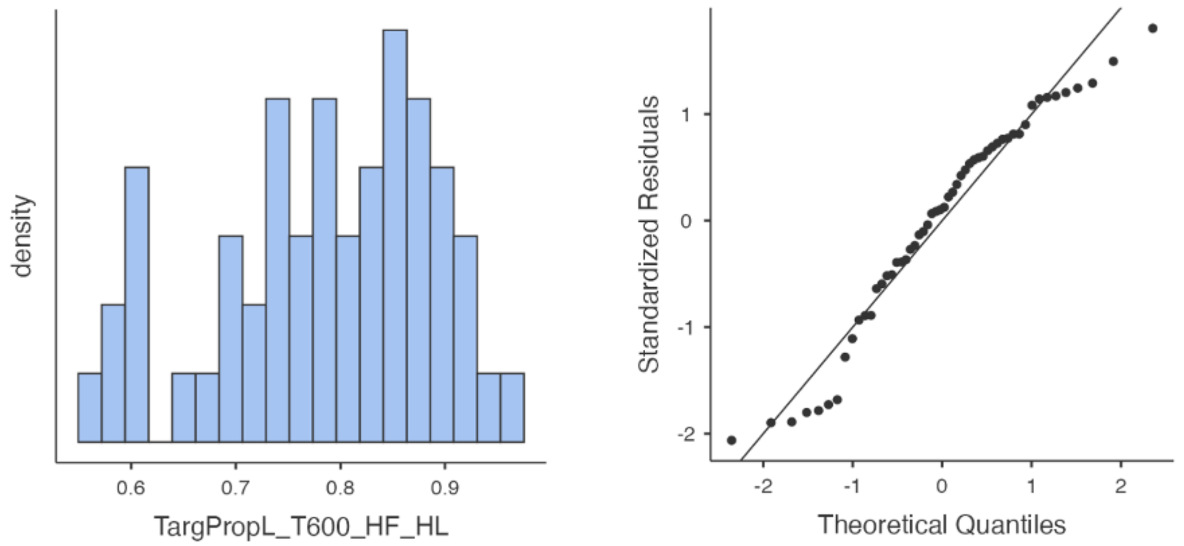


Proportion target looks in half spacing:

The variable ranges from 0.57 to 0.97 ($M = 0.78$, $SD = 0.10$) and a clear peak at 0.85 (85%). A smaller proportion of values is located to the right of the peak than to the left and the distribution trails off a bit to the left (negative skewness). Positive kurtosis indicates too few data away from the mean. The variable fails the Shapiro-Wilk test ($W = 0.95$, $p = .034$).

Figure 42

Proportion Looks to HF Target in Half Spacing (600 ms)

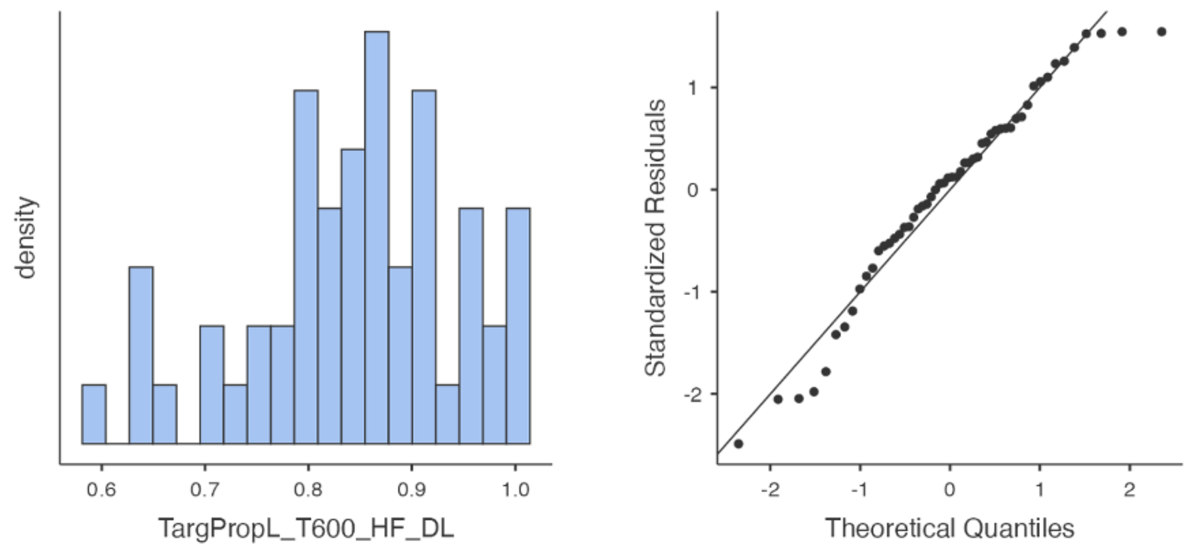


Proportion target looks in double spacing:

The histogram shows a relatively symmetric distribution with values ranging from 0.59 to 1.00 (M = 0.84, SD = 0.10). Skewness is –0.56 and kurtosis is –0.08. Shapiro-Wilk comes up non-significant implying that the variable fits well with the normal distribution (W = 0.96, p = .071).

Figure 43

Proportion Looks to HF Target in Double Spacing (600 ms)



Descriptive statistics of the proportion target looks were also performed for the item sample (Appendix E). As in the subject sample, there were violations of normality. We proceed to ANOVA considering its robustness to violations of normality while keeping in mind that the results of the following analyses must be interpreted with caution.

4.4.2 Assumptions of ANOVA

Figures 44 and 45 display the Q-Q plots to evaluate whether the dependent variable, in the subject and item sample respectively, approximates normal distribution. The distribution of the residuals in both samples somewhat line up with the reference line. The assumption of normality seems to be tenable for the subject sample and there might be a mild deviation from normality for the item sample as the residuals seem to be slightly left-skewed. Be that as it may, ANOVA is said to be quite robust to modest deviations from normality.

Figure 44

Q-Q Plot: Proportion Target Looks Across Subjects

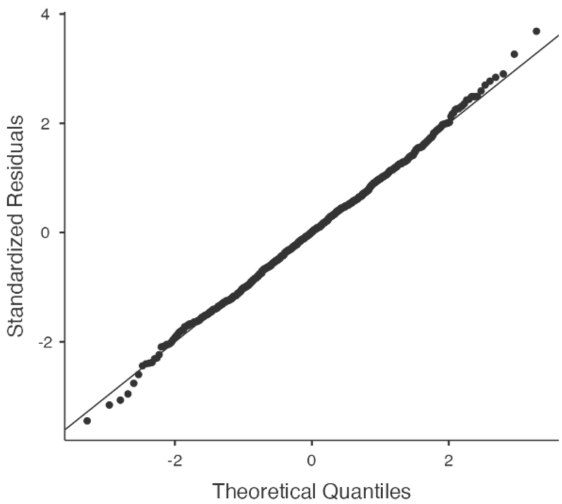
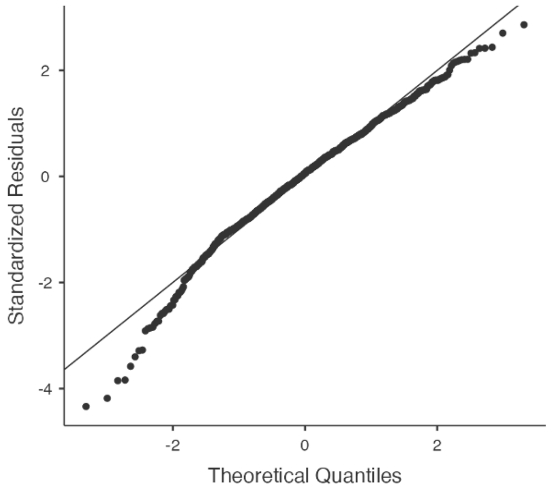


Figure 45

Q-Q Plot: Proportion Target Looks Across Items



Turning to the assumption of sphericity, the Mauchly's W tests for each repeated-measures factor and the corresponding interactions were performed. When *p* values were lower than the α level (.05), the Huynh-Feldt ε correction was applied.

Regarding the within-subjects factors of the subject sample (Table 13), Mauchly's W test indicated that the H_0 of equal variances of differences between levels cannot be rejected for *Spacing* ($p = .925$). However, the assumption of sphericity is violated for the factor *Time* ($p < .001$) and, thus, Huynh-Feldt correction is applied. The interaction between *Frequency* and *Spacing* does not violate the assumption of sphericity ($p = .313$). The assumption of sphericity is violated for the interaction between *Frequency* and *Time* ($p = .008$), the interaction between *Spacing* and *Time* ($p < .001$) and the three-way interaction among *Frequency*, *Spacing* and *Time* ($p = .011$). Therefore, Huynh-Feldt ε correction is applied.

Table 13

Tests of Sphericity for the Repeated-measures Factors of F_1 (proportion target looks)

	Mauchly's W	***p***	**Huynh-Feldt ε**
Spacing	1.00	.925	1.00
Time	.74	< .001	.81

Frequency*Spacing	.96	.313	.99
Frequency*Time	.83	.008	.88
Spacing*time	.58	< .001	.84
Frequency*Spacing*Time	.66	.011	.90

Regarding the within-subjects factors of the item sample (Table 14), Mauchly's W test indicated that *Spacing* does not violate the assumption of equal variances of differences across its three levels ($p = .542$). However, the assumption of sphericity is violated for the factor *Time* ($p < .001$) and for the interaction between *Spacing* and *Time* ($p = .001$). Huynh-Feldt correction is applied.

Table 14

Tests of Sphericity for the Repeated-measures Factors of F_2 (proportion target looks)

	Mauchly's W	*p*	Huynh-Feldt ε
Spacing	.99	.542	1.00
Time	.87	< .001	.90
Spacing*Time	.79	.001	.93

The homogeneity of variance is tested with Levene's test for the between-subjects factor, *Frequency*, which is a grouping variable in the item sample. As it is presented in Table 15, we cannot reject the H_0 for equal variances as none of the results is statistically significant. The assumption of homogeneity is not violated.

Table 15

Levene's Test for the Between-subjects Factor in F_2 (proportion target looks)

		F	***df*1**	***df*2**	***p***
	200 ms	0.61	1	118	.438
Half	400 ms	0.47	1	118	.495
	600 ms	1.20	1	118	.276
	200 ms	0.91	1	118	.341
Normal	400 ms	0.01	1	118	.911
	600 ms	0.01	1	118	.925
	200 ms	1.44	1	118	.233
Double	400 ms	0.10	1	118	.750
	600 ms	0.08	1	118	.778

4.4.3 Three-way ANOVA

By-subject and by-item ANOVAs were carried out to explore the effects of spacing on the mean proportion of target looks and whether these effects are affected by word frequency and time. This section begins with presenting the main effects and then the interactions between and among factors. First the results from the F_1 are reported and then the results from the F_2.

Subject Analysis-F_1

Frequency has a statistically significant main effect, $F(1,53) = 26.77$, $p < .001$, $\eta^2_p = .34$, accounting for 34% of the variability that is not accounted for by other factors. Calculating the eta square reveals that 1% of the total variability is explained by *Frequency* ($\eta^2 = .01$).

Spacing also has a statistically significant main effect, $F(2,106) = 32.84$, $p < .001$, $\eta^2_p = .38$. Spacing accounts for 38% of the variability that is not accounted for by other factors. Calculating the eta square reveals that 1% of the total variability is explained by *Spacing* ($\eta^2 = .01$).

Time has also a statistically significant main effect after correcting for sphericity, $F(1.63,86.19) = 1828.74$, $p < .001$, $\eta^2_p = .97$. Time accounts for 97% of the variability that is not accounted for by other factors. Calculating the eta square reveals that 77% of the total variability is explained by *Time* ($\eta^2 = .77$).

There is no statistically significant interaction between *Frequency* and *Spacing*, $F(2,106) = 0.29$, $p = .745$. This suggests that whatever the effect of spacing is, it is not different for HF and LF words.

There is no statistically significant interaction between *Frequency* and *Time* after correcting for sphericity, $F(1.76,93.41) = 1.39$, $p = .254$. This suggests that the differences between frequency groups are the same in every time interval.

The interaction between *Spacing* and *Time* after correcting for sphericity is not significant either, $F(3.38,179.06) = 2.01$, $p = .107$. This suggests that the differences across spacing conditions are indistinguishable in every time interval.

Finally, the three-way interaction among *Frequency*, *Spacing* and *Time* after correcting for sphericity is not significant $F(3.59,190.18) = 1.23$, $p = .301$. The differences across spacing conditions are the same in every time interval and frequency levels.

Table 16

F_1 for the Within-subjects Effects on Proportion Target Looks

	SS	***df***	**MS**	***F***	***p***	**η^2_p**
Frequency	0.44	1.00	0.44	26.77	**< .001**	.34
Residual	0.87	53.00	0.02			
Spacing	0.60	2.00	0.30	32.84	**< .001**	.38

Residual	0.97	106.00	0.01			
Time	55.66	1.63	34.23	1828.74	**< .001**	.97
Residual	1.61	86.19	0.02			
Frequency* Spacing	0.01	1.99	0.00	0.29	.744	.01
Residual	1.51	105.29	0.01			
Frequency* Time	0.01	1.76	0.01	1.39	.254	.03
Residual	0.54	93.41	0.01			
Spacing* Time	0.03	3.38	0.01	2.01	.107	.04
Residual	0.82	179.06	0.00			
Frequency* Spacing* Time	0.02	3.59	0.01	1.23	.301	.02
Residual	0.81	190.18	0.00			

Note. Values are adjusted using the Huynh-Feldt correction.

Item Analysis-F_2

F_2 is a mixed-design ANOVA. *Frequency* is a between-subjects factor while *Spacing* and *Time* are repeated-measures factors. Table 17 and 18 presents the results with respect to main effects, interactions, and corresponding residuals.

The main effect of *Frequency* is non-significant, $F(1,118) = 3.76$, $p = .055$. There are no statistically significant differences by frequency group for this item list.

Spacing has a statistically significant main effect, $F(2,236) = 22.16$, $p < .001$, $\eta^2_p = .16$. Spacing accounts for 16% of the variability that is not accounted for by other factors. Calculating the eta square reveals that 1% of the total variability is explained by *Spacing* ($\eta^2 = .01$). The H_0 for equal means of proportion target looks across spacing conditions is rejected concluding that there are differences between spacing conditions on average for this item list.

Time has also a statistically significant main effect after correcting for sphericity, $F(1.80,212.41) = 3112.53$, $p < .001$, $\eta^2_p = .96$. Time accounts for 96% of the variability that is not accounted for by other factors. Calculating the eta square reveals that 74% of the total variability is explained by *Time* ($\eta^2 = 0.74$).

Turning to the interactions, there is no significant interaction between *Spacing* and *Frequency*, $F(2,236) = 0.22$, $p = .799$. This suggests that the differences between the spacing conditions are the same for HF and LF words.

The interaction between *Time* and *Frequency* is not significant either, $F(1.80,212.41) = 1.24$, $p = .288$, after correcting for sphericity. This suggests that the frequency effects do not differ across time intervals for this item list.

There is no statistically significant interaction between *Spacing* and *Time* after correcting for sphericity, $F(371,437.36) = 1.27$, $p = .284$. Whatever the effect of spacing is, it is not different across time intervals for this item list.

Finally, the three-way interaction among *Frequency*, *Spacing* and *Time* is not significant, $F(371,437.36) = 0.91$, $p = .450$, after correcting for sphericity indicating that the differences across spacing conditions are the same in every time interval and frequency group.

Table 17

F_2 for the Within-subjects Effects on Proportion Target Looks

	SS	*df*	MS	*F*	*p*	η^2_p
Spacing	0.72	2.00	0.36	22.16	**< .001**	.16
Spacing* Frequency	0.01	2.00	0.00	0.22	.799	.00
Residual	3.85	236.00	0.02			
Time	61.16	1.80	33.98	3112.53	**< .001**	.96
Time* Frequency	0.02	1.80	0.01	1.24	.288	.01
Residual	2.32	212.41	0.01			
Spacing* Time	0.02	3.71	0.01	1.27	.284	.01
Spacing *Time* Frequency	0.02	3.71	0.00	0.91	.450	.01
Residual	2.27	437.36	0.01			

Note. Values are adjusted using the Huynh-Feldt correction.

Table 18

F_2 for the Between-subjects Effect on Proportion Target Looks

	SS	*df*	MS	*F*	*p*	η^2_p
Frequency	0.37	1	0.37	3.76	.055	.03
Residual	11.63	118	0.10			

4.3.4 Post-hoc Comparisons

Statistically significant main effects are followed by pairwise comparisons. As there is only one pair of levels to compare for a two-level factor, post-hoc tests were performed only for the main effects of *Spacing* and *Time*.

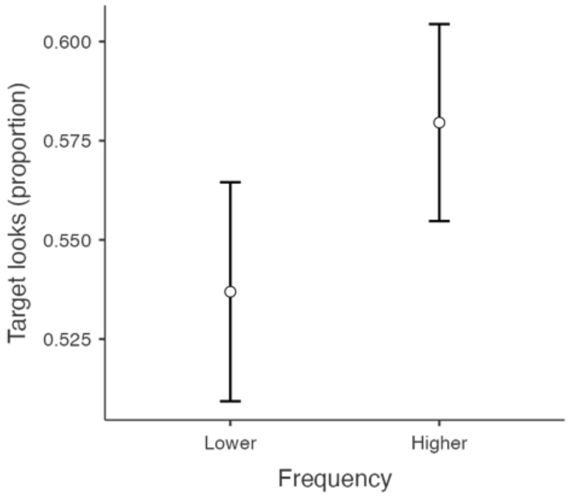
Regarding *Time*, F_1 and F_2 revealed a statistically significant main effect. This was expected as the initial plots, presented in section 4.2, showed that proportions of looks to the target quadrant increased over time. However, the results from the post-hoc comparisons are not included in this section due to the focus of the present thesis on the effects of spacing.

Subject analysis

The two-level factor *Frequency* had a significant main effect in F_1 yet post-hoc analyses is not carried out as there is only one difference to look at: the difference between HF and LF words. Based on the line plot (Figure 46), the average proportions of looks to HF target words were higher than to LF target words. However, this finding cannot be generalised beyond the experimental word set, as F_2 revealed a non-significant main effect (p = .055).

Figure 46

Estimated Marginal Means for Target Looks by Frequency Across Participants



Pairwise comparisons using Tukey's HSD identified significant differences in the proportion target looks from 200 ms to 800 ms between half and normal spacing ($p < .001$), half and double ($p < .001$), as well as between normal and double ($p = .019$) (Table 19). Based on the line plot (Figure 47), the mean proportion of looks is higher in the double-spacing condition than in the normal- and half-spacing. Moreover, the mean proportion of looks is lower in the half-spacing than in the normal- and double-spacing.

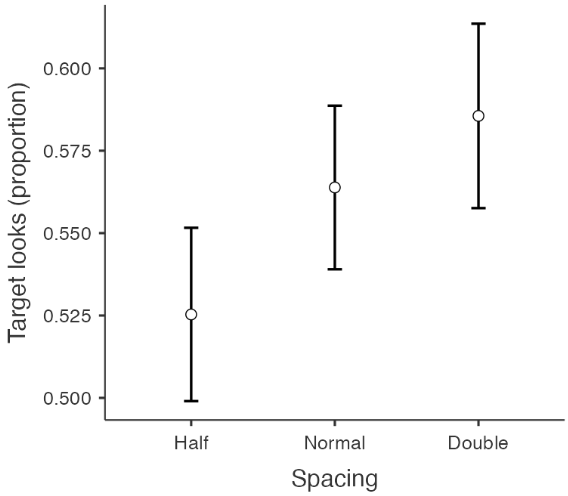
Table 19

Tukey's HSD Comparisons of Spacing (F_1)

Comparison							
Spacing		**Spacing**	**Mean Difference**	**SE**	***df***	***t***	p_{tukey}
Half	-	Normal	–0.04	0.01	53.00	–5.23	**< .001**
	-	Double	–0.06	0.01	53.00	–8.04	**< .001**
Normal	-	Double	–0.02	0.01	53.00	–2.81	**.019**

Figure 47

Estimated Marginal Means for Target Looks by Spacing Across Participants



Item analysis

Post-hoc comparisons of the results obtained from F_2 revealed similar differences for the effects of *Spacing* and the effects of *Time*. Tukey's HSD identified significant differences in the proportion of target looks between half and normal spacing ($p < .001$), half and double ($p < .001$), normal and double ($p = .032$) (Table 20). Based on the line plot (Figure 48), the mean proportion of target looks is higher in the double-spacing condition than in the normal- and half-spacing and lower in the half-spacing condition than in the normal- and double-spacing.

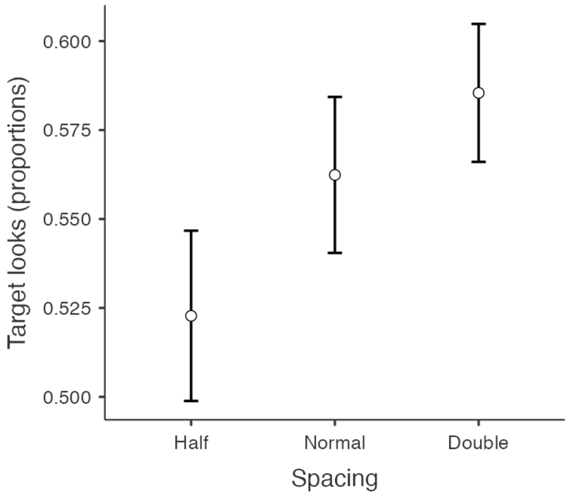
Table 20

Tukey's HSD Comparisons of Spacing (F_2)

Comparison							
Spacing		**Spacing**	**Mean Difference**	**SE**	***df***	***t***	***p*tukey**
Half	-	Normal	–0.04	0.01	118.00	–4.05	**< .001**
	-	Double	–0.06	0.01	118.00	–6.43	**< .001**
Normal	-	Double	–0.02	0.01	118.00	–2.56	**.032**

Figure 48

Estimated Marginal Means for Target Looks by Spacing Across Items



4.5 Time of the First Sample to Target

4.5.1 Descriptive Statistics

The last variable which was analysed to explore the effects of spacing on word recognition is the time of the first sample (i.e., look) to the target; how fast participants start considering the correct object. Table 21 presents the descriptives for the variable, in the subject sample, by spacing condition and frequency group.

Table 21

Descriptives for Time to Target Across Participants

	Low-Frequency			**High-Frequency**		
	Normal	**Half**	**Double**	**Normal**	**Half**	**Double**
M	444.01	447.72	429.49	411.56	439.00	406.49
SD	59.38	63.70	70.67	58.29	69.97	64.27
Skewness	–0.06	–0.75	0.02	–0.37	–0.25	–0.06
Kurtosis	–0.78	1.04	0.35	–0.54	–0.45	–0.88

Shapiro-Wilk	0.97	0.97	0.98	0.96	0.98	0.97
p	.313	.134	.573	.088	.513	.155

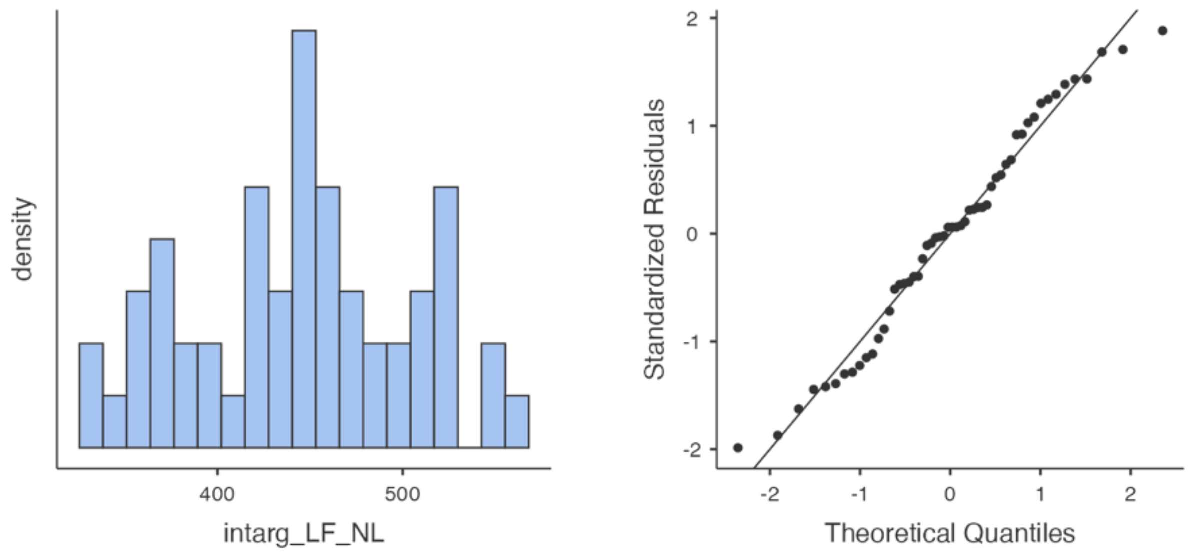
Low-frequency Words

Time to target in normal spacing:

The variable yielded a rather symmetric distribution with values ranging from 326.05 to 555.82 (*M* = 444.01, *SD* = 59.38) and a clear peak around 450 (ms). Skewness is –0.06, close to zero. Kurtosis is –0.78 indicating a lack of enough data away from the mean. Both numeric indices denote a mild deviation from normality. Shapiro-Wilk test produces a non-significant result (*W* = 0.97, *p* = .313).

Figure 49

Time to Target in Normal Spacing for LF Words

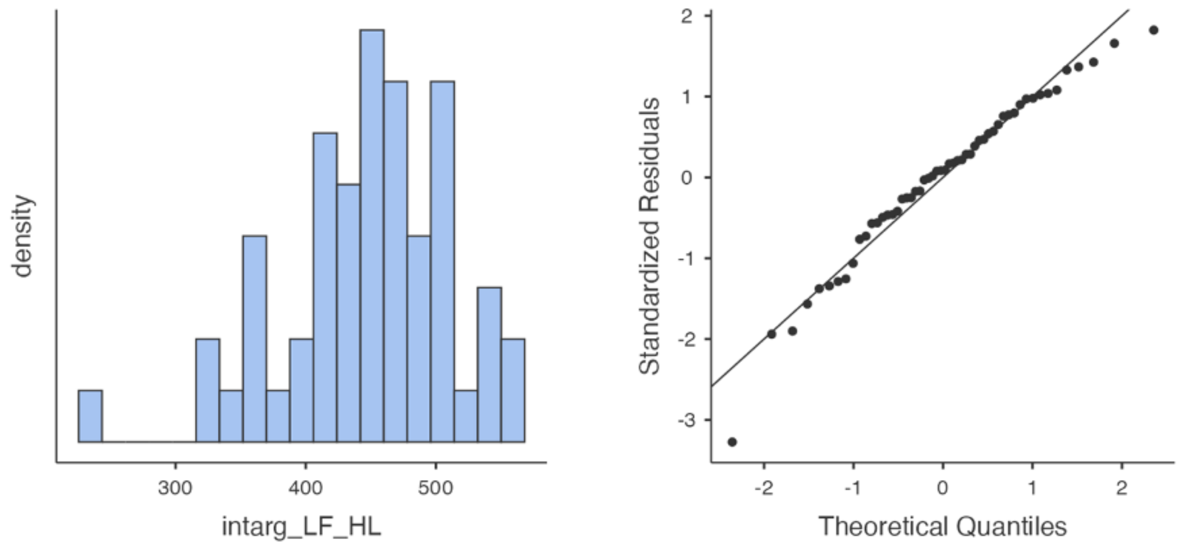


Time to target in half spacing:

The distribution of the variable seems to be slightly skewed to the left. The variable ranges from 239.11 to 563.82 (*M* = 447.72, *SD* = 63.70) and a humped peak around 450 (ms). Skewness is –0.75 and kurtosis is 1.04. The Shapiro-Wilk test produces a non-significant result (*W* = 0.97, *p* = .134).

Figure 50

Time to Target in Half Spacing for LF Words

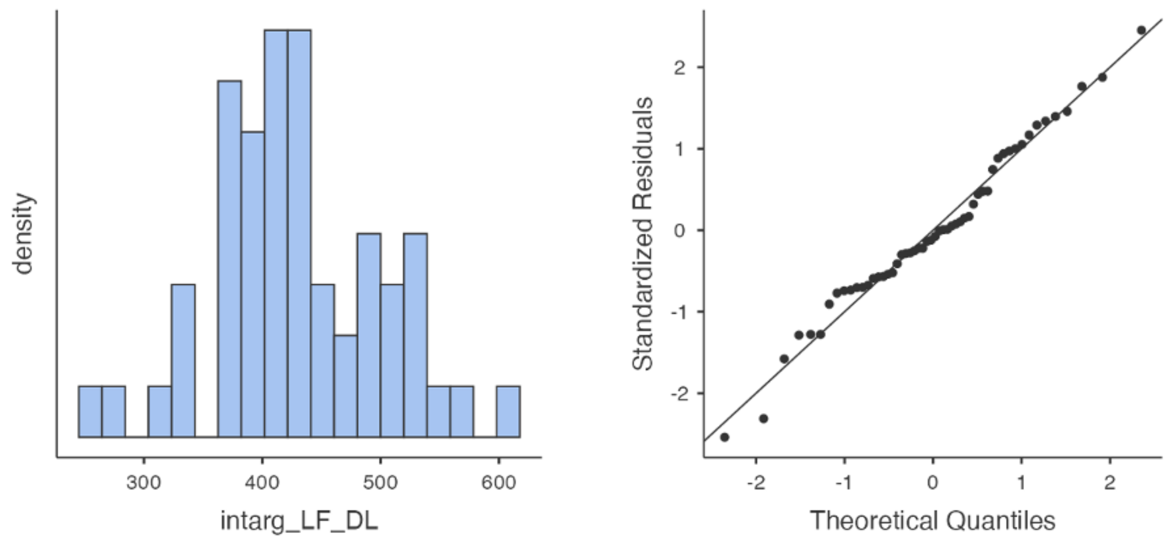


Time to target in double spacing:

The variable ranges from 250.00 to 602.89 (M = 429.49, SD = 70.67). Skewness is 0.02 close to zero. Kurtosis is 0.35 indicating a slight excess of data far from the mean. Shapiro-Wilk test produces a non-significant result (W = 0.98, p = .573).

Figure 51

Time to Target in Double Spacing for LF Words



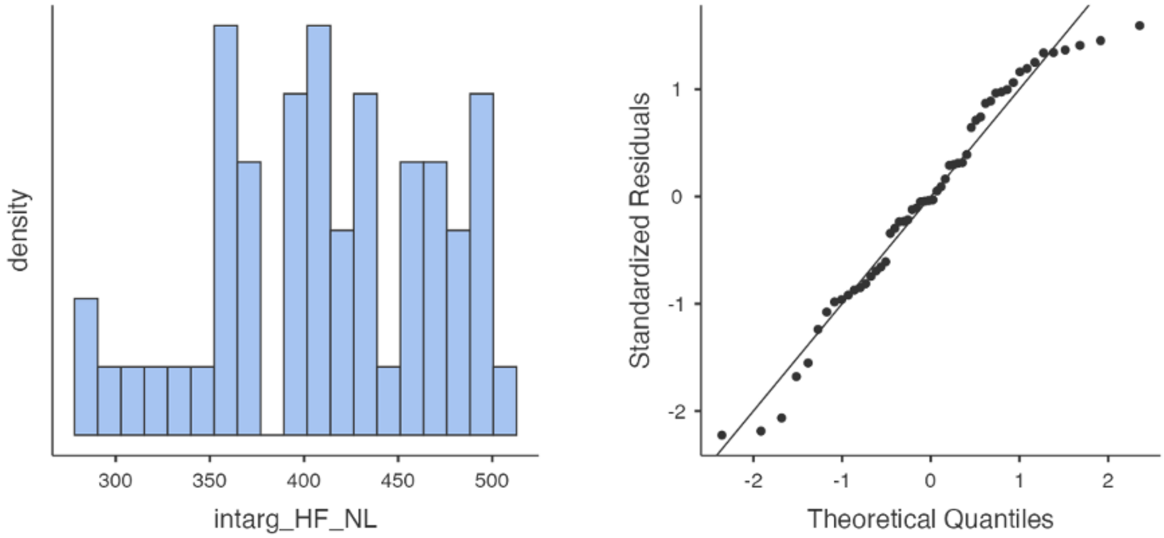
High-frequency Words

Time to target in normal spacing:

By looking at the graphs the distribution seems to be relatively asymmetric. The variable is distributed over a range from 281.89 to 504.36 (*M* = 411.56, *SD* = 58.29). Skewness is –0.37 and kurtosis is –0.54 indicating a mild deviation from normality. Shapiro-Wilk test produces a non-significant result (*W* = 0.96, *p* = .088).

Figure 52

Time to Target in Normal Spacing for HF Words

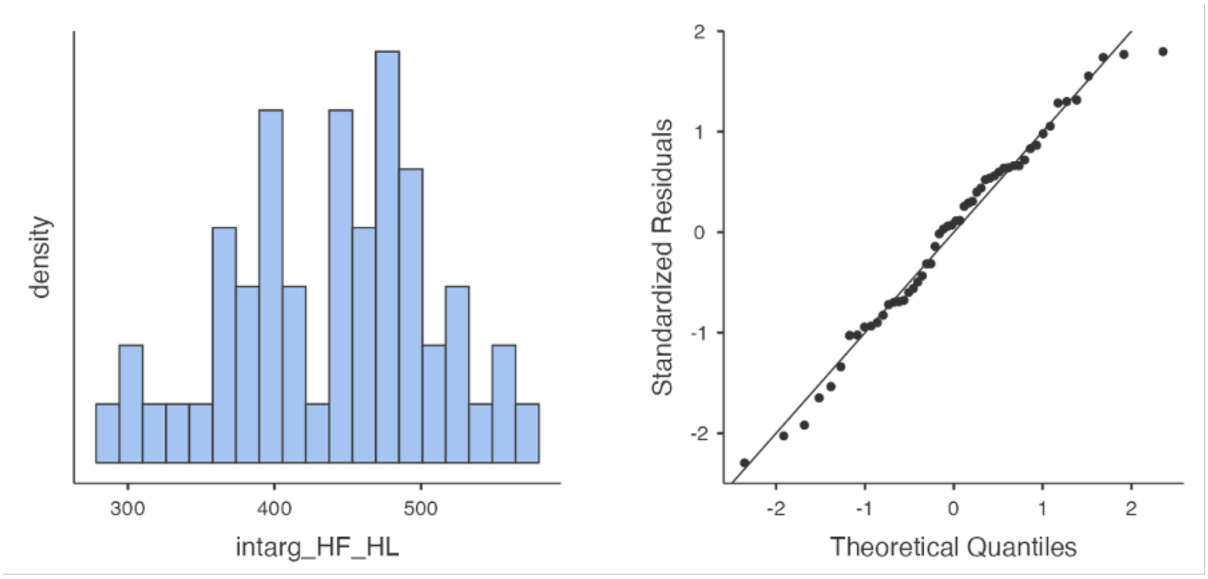


Time to target in half spacing:

The variable yielded a quite symmetric distribution with values ranging from 278.44 to 564.70 (*M* = 439.00, *SD* = 69.97), with a peak around 480 (ms) and few gaps between high points. Skewness is –0.25 indicating slightly left-skewed data. Kurtosis is –0.45 indicating fewer values in the tails than the expected normal distribution. Shapiro-Wilk test produces a non-significant result (*W* = 0.98, *p* = .513).

Figure 53

Time to Target in Half Spacing for HF Words

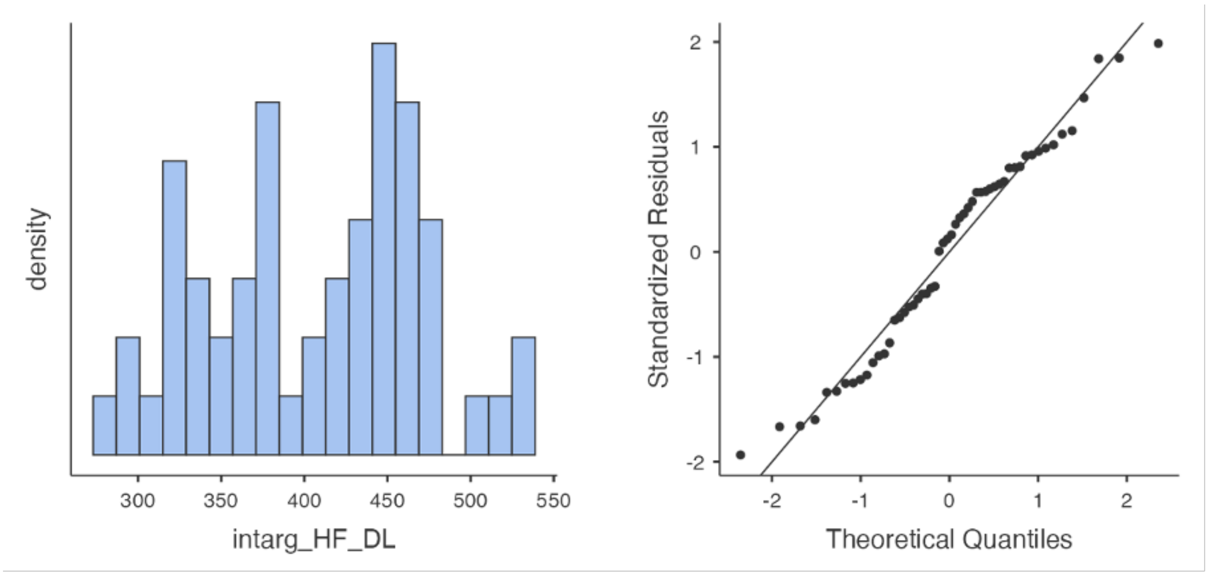


Time to target in double spacing:

The variable ranges from 282.17 to 534.12 (M = 406.49, SD = 64.27), a negative, close to zero and a negative kurtosis. Both indices indicate a mild deviation from normality. Shapiro-Wilk test produces a non-significant result (W = 0.97, p = .155).

Figure 54

Time to Target in Double Spacing for HF Words



Descriptive statistics of the dependent variable across items were also performed (Appendix E). Time (of first sample) to target in normal spacing for HF words ($p = .005$) and in double spacing for HF words ($p = .011$) were the only variables that failed the Shapiro-Wilk test. Therefore, results from the item analysis must be interpreted with caution.

4.5.2 Assumptions of ANOVA

Figures 55 and 56 display the Q-Q plot to evaluate whether the dependent variable, in the subject and item sample respectively, approximates normal distribution. The distribution of the residuals in both data sets somewhat line up with the reference line.

Figure 55

Q-Q Plot: Time (of first sample) to Target Across Participants

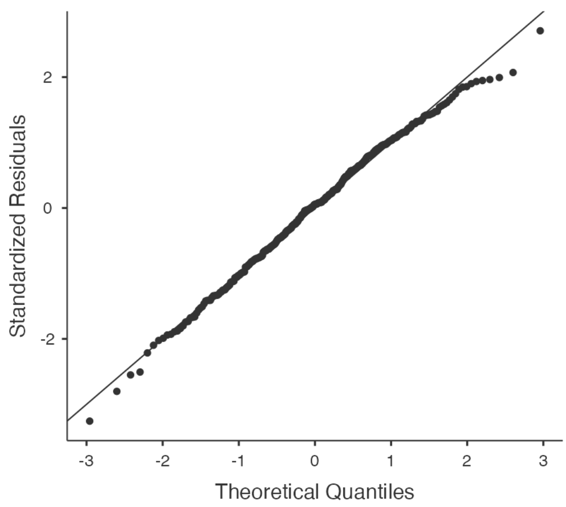
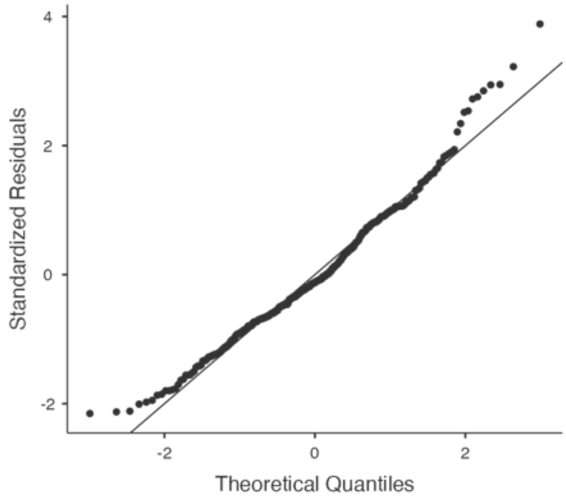


Figure 56

Q-Q Plot: Time (of first sample) to Target Across Items



Turning to the assumption of sphericity for the within-subjects factors with more than two levels, Table 22 displays the results obtained from the Mauchly's W test for the factors and interactions of F_1. Table 23 displays the sphericity test for the repeated-measures factor of F_2.

Mauchly's W test indicated that the H_0 of equal variances of differences between levels cannot be rejected for *Spacing* (p = .145 in the subject sample; p = .210 in the item sample). However, the assumption of sphericity is violated for the interaction between *Spacing* and *Frequency* (p = .021). Huynh-Feldt correction is applied.

Table 22

Tests of Sphericity for the Repeated-measures Factors of F_1 (time to target)

	Mauchly's W	***p***	**Huynh-Feldt ε**
Spacing	.93	.145	.97
Frequency*Spacing	.86	.021	.91

Table 23

Test of Sphericity for the Repeated-measures Factor of F_2 (time to target)

	Mauchly's W	***p***	**Huynh-Feldt ε**
Spacing	.97	.210	.99

The homogeneity of variance is tested with Levene's test for the between-subject factor, *Frequency*, in the item sample. As presented in Table 24, we cannot reject the H_0 for equal variances. The assumption of homogeneity is not violated.

Table 24

Levene's Test for the Between-subject Factor in F_2 (time to target)

	F	***df*1**	***df*2**	***p***
Half	0.00	1	118	0.981
Normal	0.00	1	118	0.998
Double	0.00	1	118	0.992

In brief, when the assumption of sphericity for the repeated-measures variables is not tenable appropriate, correction is applied. The assumption of homogeneity for the between-subjects factor in the item sample is not violated. Regarding the assumption of normality, it seems that the distributions do not deviate significantly from normality. All considered, we proceed with caution to two two-way ANOVAs, by-subjects and by-items.

4.5.3 Two-way ANOVA

Subject Analysis-F_1

Starting with the main effects yielded from the subject analysis (F_1), *Frequency* has a statistically significant main effect, $F(1,53) = 22.77$, $p < .001$, $\eta^2_p = .30$, accounting for 30% of the variability that is not explained by other factors. Calculating the eta square reveals that 3% of the total variability is explained by *Frequency* ($\eta^2 = .03$).

Spacing has also a statistically significant main effect, $F(2,106) = 11.03$, $p < .001$, $\eta^2_p = .17$. Spacing accounts for 17% of the variability that is not accounted for by other factors. Calculating the eta square reveals that 3% of the total variability is explained by *Spacing* ($\eta^2 = .03$).

There is no statistically significant interaction between *Frequency* and *Spacing* after correcting for sphericity, $F(1.81,96.10) = 2.06$, $p = .138$, indicating that whatever the effect of spacing is, it does not differ across frequency groups.

Table 25

F_1 for the Within-subjects Effects on Time to Target

	SS	*df*	MS	*F*	*p*	η^2_p
Frequency	37065.10	1.00	37065.10	22.77	**< .001**	.30
Residual	86256.96	53.00	1627.49			
Spacing	35357.15	1.93	18303.10	11.03	**< .001**	.17
Residual	169863.58	102.38	1659.10			
Frequency*Spacing	7711.18	1.81	4252.96	2.06	.138	.04
Residual	198391.55	96.10	2064.51			

Note. Values are adjusted using the Huynh-Feldt correction.

Item Analysis-F_2

Turning to F_2 to investigate whether effects can be generalised beyond the item sample, *Spacing* has a statistically significant main effect, $F(2,236) = 12.00$, $p < .001$, $\eta^2_p = .09$. Spacing accounts for 9% of the variability that is not accounted for by other factors. Calculating the eta square reveals that 3% of the total variability is explained by *Spacing* ($\eta^2 = .03$).

The main effect of the between-subjects factor *Frequency* is not statistically significant, $F(1,118) = 3.68$, $p = .057$. There are no statistically significant differences by frequency group.

Turning to the interaction between *Spacing* and *Frequency*, it seems that the differences across spacing conditions are the same for HF and LF words, $F(2,236) = 1.35$, $p = .260$.

Table 26

F_2 for the Within-subjects Effects on Time to Target

	SS	***df***	**MS**	***F***	***p***	**η^2_p**
Spacing	54963.60	1.98	27746.07	12.00	**< .001**	.09
Frequency*Spacing	6207.86	1.98	3133.78	1.35	.260	.01
Residual	540662.50	233.75	2312.97			

Note. Values are adjusted using the Huynh-Feldt correction.

Table 27

F_2 for the Between-subjects Effect on Time to Target

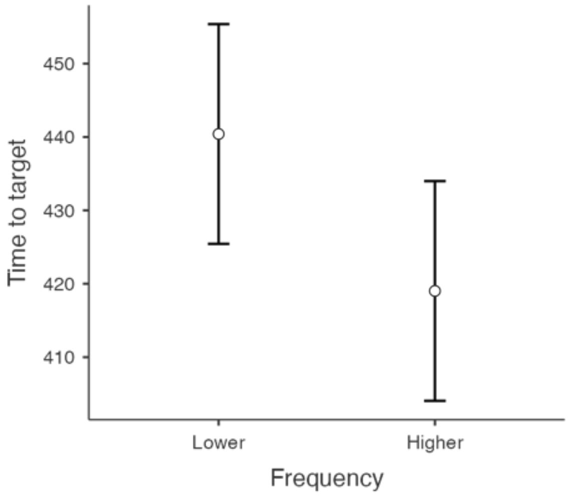
	SS	***df***	**MS**	***F***	***p***	**η^2_p**
Frequency	30876.79	1	30876.79	3.68	.057	.03
Residual	989591.58	118	8386.37			

4.5.4 Post-hoc Comparisons

Frequency in F_1 had a significant main effect yet post-hoc analysis is not carried out as there is only one difference to look at: the difference between HF and LF words. Based on the line plot (Figure 57), participants take more time to first look at a target object corresponding to a LF word compared to the time it takes them to look at a target object corresponding to a HF word. However, we cannot generalise this finding beyond the experimental word set as the F_2 yielded a non-significant main effect (p = .057).

Figure 57

Estimated Marginal Means for Time to Target by Frequency Across Participants



Tukey's HSD tests for the statistically significant main effects of Spacing were performed to determine which pairs of levels differ. The post-hoc analysis of F_1 revealed significant differences in time to target between half and normal (p = .030), half and double (p < .001). The difference between normal and double spacing was non-significant (p = .101) (Table 28). Post-hoc analysis of F_2 also revealed significant differences between half and normal spacing (p = .006), half and double (p < .001). The difference between normal spacing and double spacing was non-significant (p = .198) (Table 29).

Figures 58 and 59 show that eye gaze shifted earlier towards target objects in the double-spacing condition compared to the half-spacing in subject and item sample respectively.

Table 28

Tukey's HSD Comparisons of Spacing (F_1)

Comparison							
Spacing		**Spacing**	**Mean Difference**	**SE**	***df***	***t***	p_{tukey}
Half	-	Normal	15.58	5.93	53.00	2.63	**.030**
	-	Double	25.37	5.65	53.00	4.49	**< .001**
Normal	-	Double	9.79	4.68	53.00	2.09	.101

Figure 58

Estimated Marginal Means for Time to Target by Spacing Across Participants

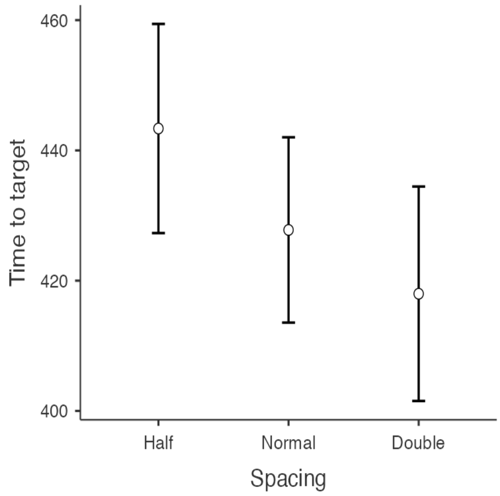


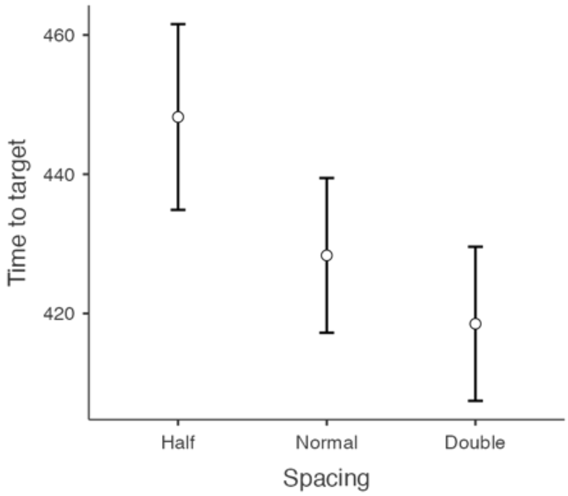
Table 29

Tukey's HSD Comparisons of Spacing (F_2)

Comparison							
Spacing		**Spacing**	**Mean Difference**	**SE**	***df***	***t***	**p_{tukey}**
Half	-	Normal	19.88	6.30	118.00	3.15	**.006**
	-	Double	29.70	6.53	118.00	4.55	**< .001**
Normal	-	Double	9.82	5.68	118.00	1.73	.198

Figure 59

Estimated Marginal Means for Time to Target by Spacing Across Items



5. Discussion

This section presents the main findings in reference to the research questions, hypotheses, and existing literature, and discusses validity, reliability, limitations, and future research directions before ending with theoretical and practical implications of the study.

It is important to mention that previous studies have explored interword spaces in sentence reading rather than in word reading. Additionally, studies probing word recognition by using the flanker design have focused on lexical properties of parafoveal stimuli rather than visual cues. To my knowledge, only Eriksen and Eriksen (1974) have addressed whether the distance of nearby stimuli affects target recognition. However, their experiment explored selective attention processes using non-reading-like materials. Due to the lack of studies investigating the role of interword spaces at a word level, the results are discussed in light of sentence-reading studies and Eriksen and Eriksen's experiment.

5.1 First Research Question

"To what extent do interword spacing manipulations affect the level of nearby-items interference in visual word recognition?"

In text-reading studies, removing or replacing spaces inhibits word recognition and saccade programming (e.g., McGowan et al., 2015; Paterson & Jordan, 2010; Perea & Acha, 2009; Rayner et al., 2013; Sheridan et al., 2013, 2016). In addition, Eriksen and Eriksen (1974) found that reduced spacing interfered with target recognition. Based on these findings, we hypothesised that adjacent words half a space away would elicit greater processing costs in foveal word efficiency compared to default spacing. Conversely, double spacing would result in reduced interference and faster lexical activation rates for the target words as compared to default spacing.

These hypotheses were corroborated by the study results. Although the main effects of spacing on the dependent variables were relatively small, they were all statistically significant at a significance level of .001 in both F_1 and F_2.

Specifically, lower proportions of correct responses were reported in the half-spacing condition compared to normal- and double-spacing while higher accuracy levels were observed in the double-spacing condition compared to normal-spacing. Post-hoc analyses indicated that all pairwise comparisons were statistically significant.

Regarding the lexical activation rate, average proportions of target looks were lower in the half-spacing condition compared to normal- and double-spacing. This finding indicates that flankers half a space away elicit greater processing costs than flankers one or double space away. Conversely, double spacing led to faster lexical activation compared to normal spacing.

Finally, spacing also affected the time it took participants to look at the target for the first time yet only the comparisons between half vs. normal spacing and half vs. double spacing were statistically significant. Participants took more time to look at the target picture in the half-spacing condition than in normal- and double-spacing. These results provide further evidence for a slower lexical activation rate due to the closer proximity between parafoveal and foveal words.

In light of the study results, reduced interword spacing leads to delayed and more fragile word recognition. This finding is consistent with the literature examining the role of interword spacing by removing or replacing spaces. Such studies have shown that decreasing the clarity of word boundaries makes word processing more challenging indicated by longer fixation durations and disrupted saccade programming.

To investigate nearby interference in selective attention, Eriksen and Eriksen (1974) manipulated the distance between targets and flanking stimuli and found that response times were delayed by flankers in closer proximity. The interference observed when flankers were present compared to a no-flanking condition was further modulated by spatial proximity. They concluded that flankers at a distance less than 1° from the target are attended thereby impeding target recognition. In contrast, increased spaces resulted in decreased response times.

In reading research, while many studies have investigated the role of interword spaces by obscuring word boundaries, only a few studies have examined the effects of wider spacing (Drieghe et al., 2005; Rayner et al., 2013). Drieghe et al. (2005) inserted an extra blank space to the right of the target word and found that wider spacing resulted in faster word recognition indicated by shorter fixation durations compared to normal spacing. Our findings align with the improved reading efficiency observed in Drieghe et al. as double spacing enhanced single word recognition indicated by faster and less fragile word recognition compared to default spacing.

One factor which might account for the reduced costs elicited by flankers placed at an increased distance could be visual acuity constraints. Given the lower visual acuity in the parafoveal region, it is plausible that the flankers were not (pre)processed or even perceived by the participants (Holmqvist et al., 2015).

However, Miellet et al. (2009) offers an alternative perspective on what affects parafoveal processing efficiency. Researchers explored whether the size of the perceptual span depends on visual acuity constraints or the distribution of attention by using a novel version of the moving window called “parafoveal-magnification task”. They enlarged the parafoveal letters by gradually increasing the font size to compensate for the lower visual acuity in parafovea. They reported a window size of 14–15 letters, similar to the size observed before parafoveal magnification thereby concluding that perceptual span is determined by visual attention limitations rather than visual acuity. Visual attention limitations refer to the allocation of more attentional resources around the fixation resulting in less available resources for parafoveal processing.

McGowan et al. (2015) investigated how subtle increases and decreases in normal spacing affect young and old adult readers’ eye movements in text-reading. They reported shorter fixation durations in the wider-spacing condition compared to normal-spacing. They suggested that this facilitation may be accounted for by “the reduction in crowding for the exterior letters” of the target words (p. 618).

At this point, it should be highlighted that the present study investigates single word recognition using a flanking-word VWP. Participants were asked to identify only the fixated word and no other concurrent processing was required. When reading a text, where parallel word processing is required, increased interword spaces may have an adverse effect due to the reduced parafoveal preview (e.g., Paterson & Jordan, 2010; Van Overschelde & Healy, 2005; Yu et al., 2007).

Moreover, we cannot conclude that double spacing benefits word recognition due the absence of a no-flanking condition serving as the baseline. By comparing double spacing to normal spacing, parafoveal words at a greater distance appear to facilitate foveal word recognition. However, if a no-flanker condition led to higher proportions of correct responses and target looks compared to double spacing, then wider spacing actually interferes with word recognition. As Hutzler et al. (2013, 2019) pointed out, a baseline condition that yields no effects is necessary for valid assessments of costs and benefits.

Based on the design and the results of the present study, the presence of adjacent words at a greater distance appears to interfere less with foveal word processing compared to words one space away. Apart from flankers’ lexical properties (Dare & Shillcock, 2013; Snell, Meeter, et al., 2017), flankers’ spatial proximity further modulate word recognition.

5.2 Second Research Question

“Is there an interaction between spacing and word frequency? Is the difference across spacing conditions the same for higher-frequency and lower-frequency words?”

The second research question aimed to investigate statistical interactions which refer to "the effect of one independent variable differs across the levels of at least one other independent variable" (Marczyk et al., 2005, p. 133). Neither F_1 nor F_2 revealed a statistical interaction between spacing and frequency in any of the dependent variables suggesting that the interword spacing affects foveal efficiency equally across higher-frequency and lower-frequency words.

Prior to discussing the lack of an interaction between frequency and spacing, the main effects of frequency are briefly discussed. Research has shown that high-frequency words are recognized faster than low-frequency words after controlling for word length indicated by shorter fixations on high-frequency words than low-frequency words (Conklin et al., 2018; Rayner, 1998; Rayner et al., 2004). Similarly, VWP studies have documented *word frequency effects* with high-frequency words being fixated more and earlier than low-frequency words (Dahan et al., 2001; Magnuson et al., 2007). Therefore, we predicted that participants recognise high-frequency words more quickly and easily than low-frequency words.

Our findings diverge from previous studies as frequency had no significant main effect on the dependent variables. In particular, word frequency did not have an effect on the proportions of correct responses neither across participants ($p = .555$) nor across items ($p = .669$). Regarding the lexical activation rate, proportions of looks to high-frequency target words were on average higher than those to low-frequency target words across participants ($p < .001$). However, we cannot generalise this finding beyond the experimental word set as the item analysis revealed a non-significant result ($p = .055$). Similarly, high-frequency words attracted eye gaze earlier than low-frequency words across participants ($p < .001$) yet the result was non-significant across items ($p = .057$) and, thus, non-generalizable.

The discrepancy in the results may be attributed to differences in the methodological approaches regarding the study design and the eye-tracking measures. Most eye-tracking studies documenting the robust word frequency effect involve sentence reading. Researchers typically analyse fixation durations rather than time-series eye-tracking data obtained from the VWP. On the other hand, VWP studies reporting frequency effects did not include phonological competitors (Dahan et al., 2001) and investigated spoken word recognition. The present study included phonological competitors sharing the same onset with the targets and investigated written word recognition.

Turning to the interaction between word frequency and spacing, sentence-reading studies investigate spacing effects on word recognition by comparing the size of the frequency effect across different spacing conditions (McGowan et al., 2014). If the frequency effect is exaggerated after removing spaces (i.e., larger effects for low-frequency words), researchers conclude that the absence of spaces inhibits word recognition (Rayner et al., 1998; Sheridan et al., 2013, 2016). If manipulating spaces does not interfere with word recognition, the size of the word frequency effect is the same across spacing conditions (Perea & Acha, 2009).

However, studies reporting exaggerated frequency effects in unspaced conditions have applied extreme changes not likely to be encountered in normal reading conditions (McGowan et al., 2014; Perea & Acha, 2009; Sheridan et al., 2016). McGowan et al. (2015) applied subtle yet naturalistic changes in interword spaces and found no interaction between spacing and frequency. Spacing effects did not vary as a function of additional difficulties in word recognition. Based on this result, they concluded that interword spacing affects "early stages of feature encoding" (p. 619).

As mentioned earlier, to my knowledge, there are no previous studies using the flanker task and manipulating interword spaces to probe the single word recognition. The absence of an interaction between spacing and frequency may also stem from the different task demands in individual word recognition and sentence reading. For single-word reading, we conclude, albeit tentatively given the limitations in statistical analyses, that frequency does not appear to have a main effect and the spacing effects do not differ for high-frequency and low-frequency words at a word level.

5.3 Discussion of Validity and Reliability

The section discusses study results in light of potential threats to validity and reliability of the study describing techniques we implemented to mitigate them.

Reliability

A pilot study for the first VWP study (Zelihić, 2020) within the BetterReading project was conducted to identify plausible problems with the design, the materials, and the procedure. A pilot study for the present experiment was not carried out due to time constraints. The materials and the target words were the same as those used in the previous study.

By comparing the grand average proportions of target looks observed in the present study and the ones observed in Zelihić (2020, p. 34), it seems that there is a consistency in participants' eye movements. Target curves diverge from the other three within around 200 ms (post-word onset). They steadily increase and reach a peak within approximately 800 ms.

However, this comparison is just a preliminary screening to initially evaluate reliability. The experimental conditions in these two studies are different.

Construct Validity

Construct validity refers to the extent that a study measures “what it purports to measure” (Mueller & Knapp, 2018, p. 397). Factors that lead to a weak connection between theory and practice are potential threats to construct validity (Bielenia-Grajewska, 2018).

The central theoretical concept of the study is *interference on single word recognition* as defined by Ziaka (2023) that is, “the impairment in performance due to the simultaneous presentation of items in spatial proximity to the target, with target and nearby items requiring the concurrent execution of multiple processes” (pp. 30–31). In the present study, performance was assessed by exploring the accuracy and efficiency in word recognition.

As the processing costs induced by adjacent words were already observed by Zelihić (2020), we focused on the effects of “spatial proximity” by manipulating interword spacing using the flanking design to create the interference effect. The presence of nearby items is core in the theoretical construct of interference and a multi-element display was required. The backward masking forced the rapid lexical activation as adult readers are known to have reached automaticity in word recognition.

However, response times were not analysed. Instructing participants to answer as fast as possible would have added a bias in the final data due to accuracy-speed trade off. The study focused on the *latency to look at the correct object* as this reflects language processing over time, rather than on *response latency*.

Statistical Validity

Regarding statistical validity, all assumptions of the repeated-measures and mixed-design ANOVA were checked before running any inferential statistics. In particular, tests for the assumption of *normality*, *sphericity* (for the within-subjects factors with more than two levels) and *homogeneity of variance* (for the between-subjects factors) were carried out (Strunk & Mwavita, 2021).

The assumption of normality was violated for the proportions of correct responses as the data were left-skewed. This was somewhat expected as only participants with high accuracy levels were included in the final analyses. A mild deviation from normality was observed for the variable the other two numeric variables. Considering that ANOVA is quite robust to modest deviations from normality (Cribbie & Klockars, 2018), we proceeded to ANOVA interpreting the results with caution. Regarding the remaining assumptions, whenever the assumption of sphericity was violated, the Huynh-Feldt ε correction was applied. The assumption of homogeneity of variance for the between-subjects factor in the item sample was not violated.

The selection of the statistical analysis needs also addressing. Statistical analyses like ANOVA and *t*-test are no longer frequently used for analysing VWP data. Researchers use (linear) mixed-effects models for time-course analyses of eye-tracking data treating items and participants as crossed random effects (Baayen et al., 2008). However, ANOVA has been used in many articles and it is considered as "a first step into the analysis of VWP data for beginners" (Ito & Knoeferle, 2022). Although ANOVA is not relevant anymore for analysing VWP data, it is the most relevant among the statistical analyses included in the master's curriculum.

Statistical validity can also be discussed with respect to *Type I* and *Type II error*. Type I error rate refers to the probability of rejecting the H_0 and reporting that there is an effect when the research hypothesis H_1 is false. Type II error rate refers to the probability of failing to reject the H_0 based on the data when H_1 stating that there is an effect is actually true (Cumming & Calin-Jageman, 2017). All statistical analyses used to explore the effects of *Spacing* on word recognition were statistically significant at the level .001 ($p < .001$) which is lower than the most commonly used alpha level (α) at .05. Essentially, the α level (i.e., significance level) determines whether the H_0 should be rejected or not (Cumming & Calin-Jageman, 2017). Setting the α level at .001 means that the H_0 will be falsely rejected 0.1% of the time.

Unaccounted item variability might also cause inflated Type I error (Clark, 1973). By carrying out both by-subject and by-item analyses and generalising the findings that come up significant in both analyses, this threat is eliminated. Nevertheless, reducing Type I error rate by setting a more rigorous level of significance can lead to increased Type II error rate. Type II error might also be inflated by a small sample size or a non-representative sample. The relatively small study sample may weaken the statistical power that is, the probability of detecting an effect when it exists in the population (Lix & Keselman, 2018).

Internal Validity

To ensure internal validity in repeated-measures designs, researchers implement various techniques to minimise potential threats. The aim is to guarantee that the observed effects are not influenced by other extraneous factors. Any change during a testing session or any factor that might affect participants' attention and/or well-being can be a potential threat to internal validity.

In within-subjects experiments, participants' performance may change across trials because of the repeated exposure to the experimental conditions (Shaughnessy et al., 2015). To balance such *practice effects* participants were completing a block of 6 practice trials before proceeding to the experimental trials. By doing so, they had the chance to familiarise themselves with the task demands. Practice trials were not included in the preprocessing and later analyses.

Apart from practice effects and increased efficiency, repeated exposure may also lead to *fatigue effects*. In that case, performance is decreased over time because participants are getting tired, bored, or distracted. To eliminate such effects, participants were encouraged to take breaks between the experimental blocks or whenever they needed it (Lix & Keselman, 2018).

Anticipation effects may also threaten internal validity in a sense that participants anticipate the correct picture in a particular quadrant of the screen in every trial, thereby looking at this Interest Area without having actually processed the target word. By randomising the locations of the target pictures across trials, such effects were eliminated. Participants also see each target word only once, thereby eliminating effects of repetition.

Counterbalancing also increases internal validity of experiments. Target stimuli were counterbalanced across participants to ensure that each target word is displayed in every condition.

External Validity

External validity refers to the generalisability of results to other subjects, items, and settings (Bielenia-Grajewska, 2018). Main effects of spacing can be generalised beyond the subject sample and the experimental set of target words because both F_1 and F_2 yielded a statistically significant effect (Locker et al., 2007).

A primary factor that can render external validity low is the small and/or unrepresentative sample. Random sampling is said to be the best method to obtain a representative sample as every member of the relevant population has the same probability of being chosen and its selection is independent of any other member of the sample (Cumming & Calin-Jageman, 2017). As these two requirements were not practically achievable, a convenience sample was used instead (Marczyk et al., 2005).

Regarding the sample size, no power analysis was conducted to estimate the desired sample size due to time constraints. However, within-subjects designs are less affected by small sample sizes as the comparisons are within subjects (Shaughnessy et al., 2015).

Participants still vary within themselves (i.e., fatigue effects), but the variability is less than in between-subjects designs (Lix & Keselman, 2018).

5.4 Limitations

Although the study addresses a gap in word-reading research providing new insights into the role of interword spacing, there are certain limitations that should be acknowledged. The primary limitation to generalising the findings concerns the sampling method and the sampling size. Moreover, limitations arising from violations of assumptions should not be overlooked when interpreting the findings. Future studies running more complex statistics are needed.

The findings are also subject to limitations imposed by the lack of a neutral baseline condition. The absence of a no-flanking condition does not allow us to conclude whether double spacing actually facilitates or just interferes less with word recognition compared to default spacing. Other limitations that should be considered are the sole use of disyllabic words and the lack of direct comparisons with studies implementing similar study designs to investigate how spatial proximity between foveal and parafoveal words affect word reading efficiency.

5.5 Future Directions

This study is written in association with the BetterReading project. Regarding the VWP studies of the project, what has already been researched are two different types of flankers (i.e., *visual* flankers and *word* flankers) compared to a no-flanking condition. What is currently being investigated are the effects of flanker lexicality and visual complexity on visual word recognition in skilled adult readers (see also Vandendaele & Grainger, 2022). Data collection from elementary students is also underway. A comparison between novice and skilled readers is intriguing considering that novice readers have a smaller perceptual span (Häikiö et al., 2009). Differences between skilled and less skilled readers would suggest that assessing individual word recognition in multi-element displays could identify reading difficulties.

As studies have observed beneficial effects of increased interletter spacing on students' with dyslexia reading outcomes (Perea et al., 2012; Zorzi et al., 2012), while other studies report deficits in suppressing irrelevant or distractive information among individuals with dyslexia (Sperling et al., 2005), future studies could examine whether spacing effects are larger for dyslexics. Students with dyslexia may face greater difficulties in processing fixated words when adjacent stimuli are placed in greater proximity compared to the default spacing.

5.6 Implications for Education

The research findings suggest that text properties affect foveal efficiency in a sense that interword spaces modulate the processing costs induced by nearby words. Although parafoveal processing has been mainly studied through the lens of preview benefits, recent studies including the present thesis provide evidence that nearby stimuli might actually interfere with the processing of the fixated word.

These findings have considerable implications for education. Delving into the function of interword spaces can contribute to design age-appropriate physical and digital educational materials adapted to students' special needs thereby optimizing their reading experiences. For instance, exploring the optimal spacing for students with visual impairment will improve reading accessibility and support learning processes.

Moreover, the spacing effects have implications for the reading research and the experimental designs employed. Spacing manipulations can be used in word-reading experiments as well as in reading interventions programs aiming to enhance reading fluency.

6. Conclusion

The present study investigated the effects of interword spacing on single word recognition in a multi-element context resembling normal reading conditions. Adjacent words in closer proximity to the fixated word inhibit foveal word recognition to a greater extent compared to the default spacing. Conversely, adjacent words at a greater distance from the fixated word interfere with single word recognition to a lesser extent than default spacing. These results suggest that interword spacing impacts single word recognition by modulating the processing costs induced by parafoveal words.

Interword spaces uniquely accounted for the variability in participants' lexical activation rate and accuracy in word recognition even though adult readers are known to have reached automaticity in reading. Future studies delving into the effects of spacing on young readers' word recognition and exploring the optimal spacing between words could provide evidence-based guidelines to educators and designers for developing age-appropriate educational materials, thereby enhancing reading experiences and outcomes.

Refence List

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects, and some models. *Cognitive Psychology*, *68*, 113–160. https://doi.org/10.1016/j.cogpsych.2013.11.001

Aghababian, V., & Nazir, T. A. (2000). Developing Normal Reading Skills: Aspects of the Visual Processes Underlying Word Recognition. *Journal of Experimental Child Psychology*, *76*(2), 123–150. https://doi.org/10.1006/jecp.1999.2540

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology*, *112*(1), 22–39. https://doi.org/10.1037/edu0000359

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The ‘blank screen paradigm’. *Cognition*, *93*(2), B79–B87. https://doi.org/10.1016/j.cognition.2004.02.005

Apfelbaum, K. S., Klein-Packard, J., & McMurray, B. (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm. *Journal of Memory and Language*, *121*, 104279. https://doi.org/10.1016/j.jml.2021.104279

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1277–1287. https://doi.org/10.1037/0096-1523.34.5.1277

Bielenia-Grajewska, M. (2018). Threats to Research Validity. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc. https://doi.org/10.4135/9781506326139

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

Cauchi, C., Lété, B., & Grainger, J. (2020). Orthographic and phonological contributions to flanker effects. *Attention, Perception, & Psychophysics*, *82*(7), 3571–3580. https://doi.org/10.3758/s13414-020-02023-0

Clark, H. H. (1973). The Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Clark-Carter, D. (2010). *Quantitative psychological research: The complete student's companion* (3rd ed.). Psychology Press.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements* (pp. 341–371). Elsevier. https://doi.org/10.1016/B978-008044980-7/50017-3

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Houghton Mifflin.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Cribbie, R. A., & Klockars, A. J. (2018). Analysis of Variance: Between-Groups Designs. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd edition.). Routledge.

Cumming, G., & Calin-Jageman, R. (2017). *Introduction to the new statistics: Estimation, open science, and beyond*. Routledge, Taylor & Francis Group.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time Course of Frequency Effects in Spoken-Word Recognition: Evidence from Eye Movements. *Cognitive Psychology*, *42*(4), 317–367. https://doi.org/10.1006/cogp.2001.0750

Dare, N., & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *Quarterly Journal of Experimental Psychology*, *66*(3), 487–504. https://doi.org/10.1080/17470218.2012.703212

de Jong, P. F. (2011). What discrete and serial rapid automatized naming can reveal about reading. *Scientific Studies of Reading*, *15*(4), 314–337. https://doi.org/10.1080/10888438.2010.485624

De Vaus, D. A. (2002). *Surveys in Social Research* (5th ed.). Routledge.

Dobó, D., Ladányi, E., Szőllősi, Á., Lukics, K. S., Németh, K., & Lukács, Á. (2022). The relationship between cognitive control and lexical conflict resolution in developmental dyslexia. *Clinical Linguistics & Phonetics*, *36*(4–5), 319–340. https://doi.org/10.1080/02699206.2021.1998632

Drieghe, D., Brysbaert, M., & Desmet, T. (2005). Parafoveal-on-foveal effects on eye movements in text reading: Does an extra space make a difference? *Vision Research*, *45*(13), 1693–1706. https://doi.org/10.1016/j.visres.2005.01.010

Ehri, L. C. (2014). Orthographic Mapping in the Acquisition of Sight Word Reading, Spelling Memory, and Vocabulary Learning. *Scientific Studies of Reading*, *18*(1), 5–21. https://doi.org/10.1080/10888438.2013.819356

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, *112*(4), 777–813. https://doi.org/10.1037/0033-295X.112.4.777

Epelboim, J., Booth, J. R., Ashkenazy, R., Taleghani, A., & Steinman, R. M. (1997). Fillers and spaces in text: The importance of word recognition during reading. *Vision Research*, *37*(20), 2899–2914. https://doi.org/10.1016/S0042-6989(97)00095-3

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Frankel, K., Becker, B., Rowe, M., & Pearson, P. (2016). From ‘What is Reading?’ to What is Literacy? *Journal of Education*, *196*, 7–17. https://doi.org/10.1177/002205741619600303

Frey, A., & Bosse, M.-L. (2018). Perceptual span, visual span, and visual attention span: Three potential ways to quantify limits on visual processing during reading. *Visual Cognition*, *26*(6), 412–429. https://doi.org/10.1080/13506285.2018.1472163

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading*, *5*(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

Galliussi, J., Perondi, L., Chia, G., Gerbino, W., & Bernardis, P. (2020). Inter-letter spacing, inter-word spacing, and font with dyslexia-friendly features: Testing text readability in people with and without dyslexia. *Annals of Dyslexia*, *70*(1), 141–152. https://doi.org/10.1007/s11881-020-00194-x

Georgiou, G. K., Papadopoulos, T. C., Fella, A., & Parrila, R. (2012). Rapid naming speed components and reading development in a consistent orthography. *Journal of Experimental Child Psychology*, *112*(1), 1–17. https://doi.org/10.1016/j.jecp.2011.11.006

Georgiou, G. K., Parrila, R., Cui, Y., & Papadopoulos, T. C. (2013). Why is rapid automatized naming related to reading? *Journal of Experimental Child Psychology*, *115*(1), 218–225. https://doi.org/10.1016/j.jecp.2012.10.015

Georgiou, G. K., Tao, S., Romero, S., Ma, L., Chen, R., Li, Y., Liu, N., Wang, L., & Protopapas, A. (2023). Serial and discrete naming and reading in Chinese first graders: Testing predictions from the cascaded processing hypothesis. *Journal of Experimental Child Psychology*, *231*, 105650. https://doi.org/10.1016/j.jecp.2023.105650

Giovagnoli, G., Vicari, S., Tomassetti, S., & Menghini, D. (2016). The Role of Visual-Spatial Abilities in Dyslexia: Age Differences in Children's Reading? *Frontiers in Psychology*, *7*(1997), 1–9. https://doi.org/10.3389/fpsyg.2016.01997

Gleason, J. B. (2005). *The development of language* (6th edition). Pearson Education, Inc.

Grainger, J., Dufau, S., & Ziegler, J. C. (2016). A Vision of Reading. *Trends in Cognitive Sciences*, *20*(3), 171–179. https://doi.org/10.1016/j.tics.2015.12.008

Grainger, J., Mathôt, S., & Vitu, F. (2014). Tests of a model of multi-word reading: Effects of parafoveal flanking letters on foveal word recognition. *Acta Psychologica*, *146*, 35–40. https://doi.org/10.1016/j.actpsy.2013.11.014

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*(2), 167–181. https://doi.org/10.1016/j.jecp.2008.04.002

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*(3), 417–429. https://doi.org/10.1037/0278-7393.16.3.417

Hendrickson, K., Apfelbaum, K., Goodwin, C., Blomquist, C., Klein, K., & McMurray, B. (2021). The profile of real-time competition in spoken and written word recognition: More similar than different. *Quarterly Journal of Experimental Psychology (2006)*, *75*(9), 1653–1673. https://doi.org/10.1177/17470218211056842

Henry, R., Van Dyke, J., & Kuperman, V. (2018). Oculomotor planning in RAN and reading: A strong test of the visual scanning hypothesis. *Reading and Writing*, *31*(1). https://doi.org/10.1007/s11145-018-9856-3

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2015). *Eye tracking: A comprehensive guide to methods and measures*. University Press.

Hudson, R. F., Lane, H., & Pullen, P. (2005). Reading Fluency Assessment and Instruction: What, Why, and How? *Reading Teacher - READ TEACH*, *58*(8), 702–714. https://doi.org/10.1598/RT.58.8.1

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *25*(1), 4–32. https://doi.org/10.1080/10573560802491208

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Hutzler, F., Fuchs, I., Gagl, B., Schuster, S., Richlan, F., Braun, M., & Hawelka, S. (2013). Parafoveal X-masks interfere with foveal word recognition: Evidence from fixation-related brain potentials. *Frontiers in Systems Neuroscience*, *7*(33). https://www.frontiersin.org/articles/10.3389/fnsys.2013.00033

Hutzler, F., Schuster, S., Marx, C., & Hawelka, S. (2019). An investigation of parafoveal masks with the incremental boundary paradigm. *PloS One*, *14*(2), e0203013. https://doi.org/10.1371/journal.pone.0203013

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box Correction for Degrees of Freedom from Sample Data in Randomized Block and Split-Plot Designs. *Journal of Educational Statistics*, *1*(1), 69–82. https://doi.org/10.3102/10769986001001069

Ito, A., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01969-3

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: Towards a model of real-world reading. *Journal of Computer Assisted Learning*, *33*(3), 193–201. https://doi.org/10.1111/jcal.12189

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of Individual Differences in Reading Comprehension and Reading Fluency. *Journal of Educational Psychology*, *95*(4), 719. https://doi.org/10.1037/0022-0663.95.4.719

Kasisopa, B., G. Reilly, R., Luksaneeyanawin, S., & Burnham, D. (2013). Eye movements while reading an unspaced writing system: The case of Thai. *Vision Research*, *86*, 71–80. https://doi.org/10.1016/j.visres.2013.04.007

Kennedy, A. (1998). The Influence of Parafoveal Words on Foveal Inspection Time: Evidence for a Processing Trade-Off. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 149–179). Elsevier Science Ltd. https://doi.org/10.1016/B978-008043361-5/50008-0

Kennedy, A., Pynte, J., & Ducrot, S. (2002). Parafoveal-on-foveal interactions in word recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *55*(4), 1307–1337. Scopus. https://doi.org/10.1080/02724980244000071

Kim, Y.-S. G., & Wagner, R. K. (2015). Text (Oral) Reading Fluency as a Construct in Reading Development: An Investigation of its Mediating Role for Children from Grades 1 to 4. *Scientific Studies of Reading : The Official Journal of the Society for the Scientific Study of Reading*, *19*(3), 224–242. https://doi.org/10.1080/10888438.2015.1007375

Kirby, J. R., Desrochers, A., Roth, L., & Lai, S. S. V. (2008). Longitudinal predictors of word reading development. *Canadian Psychology / Psychologie Canadienne*, *49*(2), 103–110. https://doi.org/10.1037/0708-5591.49.2.103

Kirby, J. R., Georgiou, G., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. Reading Research Quarterly, 45, 341–362. *Reading Research Quarterly*, *45*(3), 341–362.

Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *Quarterly*

Journal of Experimental Psychology, *66*(3), 581–600. https://doi.org/10.1080/17470218.2012.658073

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning Theory and Assessment of Reading Fluency: Automaticity, Prosody, and Definitions of Fluency. *Reading Research Quarterly*, *45*(2), 230–251. https://doi.org/10.1598/RRQ.45.2.4

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*(1), 3–21. https://doi.org/10.1037/0022-0663.95.1.3

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

Kuperman, V., Van Dyke, J. A., & Henry, R. (2016). Eye-Movement Control in RAN and Reading. *Scientific Studies of Reading: The Official Journal of the Society for the Scientific Study of Reading*, *20*(2), 173–188. https://doi.org/10.1080/10888438.2015.1128435

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

Lane, D. (2016). The assumption of sphericity in repeated-measures designs: What it means and what to do when it is violated. *The Quantitative Methods for Psychology*, *12*(2), 114–122. https://doi.org/10.20982/tqmp.12.2.p114

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, *20*(8), 1040–1048. https://doi.org/10.1111/j.1467-9280.2009.02405.x

Lix, L. M., & Keselman, H. J. (2018). Analysis of Variance: Repeated Measures Designs. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd edition., p. 514). Routledge.

Locker, L., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*, *39*(4), 723–730. https://doi.org/10.3758/bf03192962

Logan, G. D. (1997). Automaticity and Reading: Perspectives from the Instance Theory of Automatization. *Reading & Writing Quarterly*, *13*(2), 123–146. https://doi.org/10.1080/1057356970130203

Logan, J. A. R., Schatschneider, C., & Wagner, R. K. (2011). Rapid serial naming and reading ability: The role of lexical access. *Reading and Writing*, *24*(1), 1–25. https://doi.org/10.1007/s11145-009-9199-1

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The Dynamics of Lexical Competition During Spoken Word Recognition. *Cognitive Science*, *31*(1), 133–156. https://doi.org/10.1080/03640210709336987

Marczyk, G. R., DeMatteo, D., & Festinger, D. (2005). *Essentials of research design and methodology* (1st ed., Vol. 2). John Wiley & Sons.

Marsden, J., Stevens, S., & Ebri, A. (2014). How to measure distance visual acuity. *Community Eye Health*, *27*(85), 16.

Marx, C., Hawelka, S., Schuster, S., & Hutzler, F. (2015). An incremental boundary study on parafoveal preprocessing in children reading aloud: Parafoveal masks overestimate the preview benefit. *Journal of Cognitive Psychology (Hove, England)*, *27*(5), 549–561. https://doi.org/10.1080/20445911.2015.1008494

Matin, E. (1975). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899. https://doi.org/10.1037/h0037368

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586. https://doi.org/10.3758/BF03203972

McConkie, G. W., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, *8*(5), 365–368. https://doi.org/10.3758/BF03335168

McDonald, S. A. (2006). Parafoveal preview benefit in reading is only obtained from the saccade goal. *Vision Research*, *46*(26), 4416–4424. https://doi.org/10.1016/j.visres.2006.08.027

McGowan, V. A., White, S. J., Jordan, T. R., & Paterson, K. B. (2014). Aging and the use of interword spaces during reading: Evidence from eye movements. *Psychonomic Bulletin & Review*, *21*(3), 740–747. https://doi.org/10.3758/s13423-013-0527-8

McGowan, V. A., White, S. J., & Paterson, K. B. (2015). The effects of interword spacing on the eye movements of young and older readers. *Journal of Cognitive Psychology*, *27*(5), 609–621. https://doi.org/10.1080/20445911.2014.988157

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, *54*, 1472–1491. https://doi.org/10.1037/dev0000542

Meade, G., Grainger, J., & Declerck, M. (2021). Friend or foe? Flankers reverse the direction of orthographic neighbourhood effects. *Language, Cognition and Neuroscience*, *36*(5), 535–542. https://doi.org/10.1080/23273798.2020.1849750

Meixner, J. M., Nixon, J. S., & Laubrock, J. (2022). The perceptual span is dynamically adjusted in response to foveal load by beginning readers. *Journal of Experimental Psychology. General*, *151*(6), 1219–1232. https://doi.org/10.1037/xge0001140

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, *140*(2), 409–433. https://doi.org/10.1037/a0033890

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science*, *20*, 721–728. https://doi.org/10.1111/j.1467-9280.2009.02364.x

Mueller, R. O., & Knapp, T. R. (2018). Reliability and Validity of Instruments. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd edition., pp. 397–401). Routledge.

National Reading Panel. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Reports of the subgroups.* (NIH Publication No. 00-4754). U.S. Department of Health and Human Services.

Navarro, D. J., & Foxcroft, D. R. (2018). *Learning statistics with jamovi: A tutorial for psychology students and other beginners* (Version 0.70). https://doi.org/10.24384/HGC3-7P15

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*(1), 427–452. https://doi.org/10.1146/annurev-psych-120710-100431

Paterson, K. B., & Jordan, T. R. (2010). Effects of increased letter spacing on word identification and eye guidance during reading. *Memory & Cognition*, *38*(4), 502–512. https://doi.org/10.3758/MC.38.4.502

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385–413. https://doi.org/10.1016/j.cognition.2006.04.008

Perea, M., & Acha, J. (2009). Space information is important for reading. *Vision Research*, *49*(15), 1994–2000. https://doi.org/10.1016/j.visres.2009.05.009

Perea, M., & Gomez, P. (2012a). Increasing interletter spacing facilitates encoding of words. *Psychonomic Bulletin & Review*, *19*(2), 332–338. https://doi.org/10.3758/s13423-011-0214-6

Perea, M., & Gomez, P. (2012b). Subtle increases in interletter spacing facilitate the encoding of words during normal reading. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0047568

Perea, M., Moret-Tatay, C., & Gómez, P. (2011). The effects of interletter spacing in visual-word recognition. *Acta Psychologica*, *137*(3), 345–351. https://doi.org/10.1016/j.actpsy.2011.04.003

Perea, M., Panadero, V., Moret-Tatay, C., & Gómez, P. (2012). The effects of inter-letter spacing in visual-word recognition: Evidence with young normal readers and developmental dyslexics. *Learning and Instruction*, *22*(6), 420–430. https://doi.org/10.1016/j.learninstruc.2012.04.001

Pollatsek, A., & Rayner, K. (1982). Eye movement control in reading: The role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(6), 817–833. Scopus. https://doi.org/10.1037/0096-1523.8.6.817

Porretta, V., Kyröläinen, A.-J., Rij, J. van, & Järvikivi, J. (2020). *VWPre: Tools for Preprocessing Visual World Data* (1.2.4). https://CRAN.R-project.org/package=VWPre

Protopapas, A., Altani, A., & Georgiou, G. K. (2013). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology*, *116*(4), 914–929. https://doi.org/10.1016/j.jecp.2013.08.004

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, *54*(3), 251–282. https://doi.org/10.1016/j.cogpsych.2006.07.003

Protopapas, A., Katopodi, K., Altani, A., & Georgiou, G. K. (2018). Word Reading Fluency as a Serial Naming Task. *Scientific Studies of Reading*, *22*(3), 248–263. https://doi.org/10.1080/10888438.2018.1430804

Pynte, J., Kennedy, A., & Ducrot, S. (2004). The influence of parafoveal typographical errors on eye movements in reading. *European Journal of Cognitive Psychology - EUR J COGN PSYCHOL*, *16*(1), 178–202. https://doi.org/10.1080/09541440340000169

Raaijmakers, J. G. W. (2003). A further look at the 'language-as-fixed-effect fallacy'. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *57*(3), 141–151. https://doi.org/10.1037/h0087421

Rasinski, T. V. (2006). Reading Fluency Instruction: Moving Beyond Accuracy, Automaticity, and Prosody. *The Reading Teacher*, *59*(7), 704–706. https://doi.org/10.1598/RT.59.7.10

Rasinski, T. V. (2012). Why Reading Fluency Should Be Hot! *The Reading Teacher*, *65*(8), 516–522. https://doi.org/10.1002/TRTR.01077

Rasinski, T. V., Reutzel, C. R., Chard, D., & Linan-Thompson, S. (2012). Reading fluency. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. Afflerbach (Eds.), *Handbook of Reading Research* (Vol. IV, pp. 286–319). Routledge.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*(1), 65–81. https://doi.org/10.1016/0010-0285(75)90005-5

Rayner, K. (1979). Eye guidance in reading: Fixation locations within words. *Perception*, *8*(1), 21–30. https://doi.org/10.1068/p080021

Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, *41*(2), 211–236. https://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), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology (2006)*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

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

Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *40*(4), 473–483. https://doi.org/10.1037/h0080111

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, *38*(8), 1129–1144. https://doi.org/10.1016/s0042-6989(97)00274-5

Rayner, K., & Juhasz, B. J. (2004). Eye movements in reading: Old questions and new directions. *European Journal of Cognitive Psychology*, *16*(1–2), 340–352. Scopus. https://doi.org/10.1080/09541440340000385

Rayner, K., Liversedge, S. P., & White, S. J. (2006). Eye movements when reading disappearing text: The importance of the word to the right of fixation. *Vision Research*, *46*(3), 310–323. https://doi.org/10.1016/j.visres.2005.06.018

Rayner, K., & Reingold, E. (2015). Evidence for direct cognitive control of fixation durations during reading. *Current Opinion in Behavioral Sciences*, *1*, 107–112. https://doi.org/10.1016/j.cobeha.2014.10.008

Rayner, K., Schotter, E. R., Masson, M. E. J., 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., Yang, J., Schuett, S., & Slattery, T. J. (2013). Eye movements of older and younger readers when reading unspaced text. *Experimental Psychology*, *60*(5), 354–361. https://doi.org/10.1027/1618-3169/a000207

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. https://doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445–476. https://doi.org/10.1017/S0140525X03000104

Riedel, B. (2007). The Relation Between DIBELS, Reading Comprehension, and Vocabulary in Urban First-Grade Students. *Reading Research Quarterly International Reading Association*, *42*(4), 546–567. https://doi.org/10.1598/RRQ.42.4.5

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, *51*(12), 1690–1703. https://doi.org/10.1037/dev0000044

Risse, S., Hohenstein, S., Kliegl, R., & Engbert, R. (2014). A Theoretical Analysis of the Perceptual Span based on SWIFT Simulations of the n + 2 Boundary Paradigm. *Visual Cognition*, *22*(3), 283–308. https://doi.org/10.1080/13506285.2014.881444

Roembke, T. C., Hazeltine, E., Reed, D. K., & McMurray, B. (2018). Automaticity of word recognition is a unique predictor of reading fluency in middle-school students. *Journal of Educational Psychology*, *111*(2), 314–330. https://doi.org/10.1037/edu0000279

Sainio, M., Hyönä, J., Bingushi, K., & Bertram, R. (2007). The role of interword spacing in reading Japanese: An eye movement study. *Vision Research*, *47*(20), 2575–2584. https://doi.org/10.1016/j.visres.2007.05.017

Samuels, S. J. (1997). The Method of Repeated Readings. *Reading Teacher*, *50*(5), 376–381.

Schmalz, X., & Mulatti, C. (2017). Busting a myth with the Bayes Factor: Effects of letter bigram frequency in visual lexical decision do not reflect reading processes. *The Mental Lexicon*, *12*(2), 263–282. https://doi.org/10.1075/ml.17009.sch

Schmidtke, D., Van Dyke, J. A., & Kuperman, V. (2021). CompLex: An eye-movement database of compound word reading in English. *Behavior Research Methods*, *53*(1), 59–77. https://doi.org/10.3758/s13428-020-01397-1

Schotter, E. R. (2018). Reading ahead by hedging our bets on seeing the future: Eye tracking and electrophysiology evidence for parafoveal lexical processing and saccadic control by partial word recognition. In K. D. Federmeier & D. G. Watson (Eds.), *The psychology of learning and motivation: Current topics in language*. Elsevier Academic Press.

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

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2015). *Research methods in psychology* (Tenth edition). McGraw-Hill Education.

Sheridan, H., Rayner, K., & Reingold, E. M. (2013). Unsegmented text delays word identification: Evidence from a survival analysis of fixation durations. *Visual Cognition*, *21*(1), 38–60. https://doi.org/10.1080/13506285.2013.767296

Sheridan, H., Reichle, E. D., & Reingold, E. M. (2016). Why does removing inter-word spaces produce reading deficits? The role of parafoveal processing. *Psychonomic Bulletin & Review*, *23*(5), 1543–1552. https://doi.org/10.3758/s13423-015-0997-y

Simmons, E. S., & Magnuson, J. S. (2018). Word length, proportion of overlap, and phonological competition in spoken word recognition. *Proceedings of the Cognitive Science Society*, *40*, 1062–1067.

Slattery, T. J., & Rayner, K. (2013). Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text. *Attention, Perception & Psychophysics*, *75*(6), 1275–1292. https://doi.org/10.3758/s13414-013-0463-8

Snell, J., & Grainger, J. (2018). Parallel word processing in the flanker paradigm has a rightward bias. *Attention, Perception, & Psychophysics*, *80*(6), 1512–1519. https://doi.org/10.3758/s13414-018-1547-2

Snell, J., & Grainger, J. (2019). Readers Are Parallel Processors. *Trends in Cognitive Sciences*, *23*(7), 537–546. https://doi.org/10.1016/j.tics.2019.04.006

Snell, J., Meeter, M., & Grainger, J. (2017). Evidence for simultaneous syntactic processing of multiple words during reading. *PLoS ONE*, *12*(3), e0173720. https://doi.org/10.1371/journal.pone.0173720

Snell, J., Vitu, F., & Grainger, J. (2017). Integration of parafoveal orthographic information during foveal word reading: Beyond the sub-lexical level? *Quarterly Journal of Experimental Psychology (2006)*, *70*(10), 1984–1996. https://doi.org/10.1080/17470218.2016.1217247

Sperlich, A., Schad, D. J., & Laubrock, J. (2015). When preview information starts to matter: Development of the perceptual span in German beginning readers. *Journal of Cognitive Psychology*, *27*(5), 511–530. https://doi.org/10.1080/20445911.2014.993990

Sperling, A. J., Lu, Z.-L., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience*, *8*(7), 862–863. https://doi.org/10.1038/nn1474

Spragins, A. B., Lefton, L. A., & Fisher, D. F. (1976). Eye movements while reading and searching spatially transformed text: A developmental examination. *Memory & Cognition*, *4*(1), 36–42. Scopus. https://doi.org/10.3758/BF03213252

Starr, M. S., & Rayner, K. (2001). Eye movements during reading: Some current controversies. *Trends in Cognitive Sciences*, *5*(4), 156–163. https://doi.org/10.1016/s1364-6613(00)01619-3

Stevens, C., & Bavelier, D. (2012). The role of selective attention on academic foundations: A cognitive neuroscience perspective. *Developmental Cognitive Neuroscience*, *2*(Suppl 1), S30–S48. https://doi.org/10.1016/j.dcn.2011.11.001

Strunk, K. K., & Mwavita, M. (2021). *Design and Analysis in Educational Research Using jamovi: ANOVA Designs*. Routledge. https://doi.org/10.4324/9781003154297

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. https://doi.org/10.1126/science.7777863

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1998). *Test of word reading efficiency*. PRO-ED.

Van den Boer, M., & Hakvoort, B. E. (2015). Default spacing is the optimal spacing for word reading. *Quarterly Journal of Experimental Psychology (2006)*, *68*(4), 697–709. https://doi.org/10.1080/17470218.2014.964272

Van Overschelde, J. P., & Healy, A. F. (2005). A Blank Look in Reading: The Effect of Blank Space on the Identification of Letters and Words During Reading. *Experimental Psychology*, *52*(3), 213–223. https://doi.org/10.1027/1618-3169.52.3.213

van Viersen, S., Protopapas, A., & de Jong, P. F. (2022). Word- and Text-Level Processes Contributing to Fluent Reading of Word Lists and Sentences. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.789313

Vandendaele, A., & Grainger, J. (2022). Now you see it, now you don't: Flanker presence induces the word concreteness effect. *Cognition*, *218*, 104945. https://doi.org/10.1016/j.cognition.2021.104945

Vasilev, M. R., & Angele, B. (2017). Parafoveal preview effects from word N + 1 and word N + 2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin & Review*, *24*(3), 666–689. https://doi.org/10.3758/s13423-016-1147-x

Vasilev, M. R., Slattery, T. J., Kirkby, J. A., & Angele, B. (2018). What are the costs of degraded parafoveal previews during silent reading? *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *44*(3), 371. https://doi.org/10.1037/xlm0000433

Vasilev, M. R., Yates, M., Prueitt, E., & Slattery, T. J. (2021). Parafoveal degradation during reading reduces preview costs only when it is not perceptually distinct. *Quarterly Journal of Experimental Psychology (2006)*, *74*(2), 254–276. https://doi.org/10.1177/1747021820959661

White, S., & Liversedge, S. (2004). Orthographic familiarity influences initial eye fixation positions in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 52.

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., Wolf, M., Kuhn, M., Meisinger, B., & Schwanenflugel, P. (2010). The relationship between different measures of oral reading fluency and reading comprehension in second-grade students who evidence different oral reading fluency difficulties. *Language, Speech, and Hearing Services in Schools*, *41*(3), 340–348. https://doi.org/10.1044/0161-1461(2009/08-0093)

Wolf, M., & Katzir, T. (2001). Reading Fluency and Its Intervention. *Scientific Studies of Reading*, *5*(3), 211–239. https://doi.org/10.1207/S1532799XSSR0503_2

Yan, M. (2015). Visually complex foveal words increase the amount of parafoveal information acquired. *Vision Research*, *111*, 91–96. https://doi.org/10.1016/j.visres.2015.03.025

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Yu, D., Cheung, S.-H., Legge, G. E., & Chung, S. T. L. (2007). Effect of letter spacing on visual span and reading speed. *Journal of Vision*, *7*(2), 1–10. https://doi.org/10.1167/7.2.2

Zelihić, D. (2020). *Automaticity and the notion of interference. Assessing word reading automaticity as freedom from interference in a Visual World Paradigm* [Master's Thesis, University of Oslo]. http://hdl.handle.net/10852/80348

Zhan, L. (2018). Using Eye Movements Recorded in the Visual World Paradigm to Explore the Online Processing of Spoken Language. *Journal of Visualized Experiments : JoVE*, *140*, e58086. https://doi.org/10.3791/58086

Ziaka, L. (2023). *Interference and cognitive control dynamics in the course of serial naming tasks* [Doctoral dissertation, University of Oslo]. https://www.duo.uio.no/handle/10852/99258

Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., Bravar, L., George, F., Pech-Georgel, C., & Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), 11455–11459. https://doi.org/10.1073/pnas.1205566109

Appendix A: Experimental Text Elements

Table 30.

Target words with the corresponding Competitors, Distractors and Flankers

Target	Competitor	Distractor 1	Distractor 2	Flanker 1	Flanker 2
hale	hare	biff	pennal	logo	lunsj
nese	neve	heks	magnet	leilighet	optiker
robot	reke	prest	astronaut	orkan	bølge
maske	mage	klovn	bibel	ørken	sykehus
drage	dyne	lokk	sushi	måned	skjørt
sykkel	sitron	perm	bobil	linje	slave
verktøy	vinyl	skål	brille	mynt	smil
pølse	panda	strikk	høne	klem	spion
pilot	pinne	garn	hane	løype	lomme
skygge	skjære	skall	dynamitt	klasse	sommer
finger	pille	kritt	peanøtt	spøkelse	høst
ridder	rifle	geit	edderkopp	maskin	bokstav
teppe	terning	tårn	spindelvev	vinter	snor
tiger	tygges	lykt	manual	skum	syre
rotte	rustning	krans	pinsett	tegn	lemur
sukker	salat	ørn	bikini	premie	kobra
jakke	jolle	egg	trillebår	dress	spytt
mobil	madrass	negl	hyssing	brosjyre	toll
demon	diplom	ost	påfugl	krybbe	åker
pose	pote	kran	ris	innlegg	fjær
klokke	kabel	tog	parfyme	fryser	motor
vampyr	vinge	fly	kleshenger	politi	tornado
slange	skuter	lue	kalender	grein	mansjett
billett	bjelle	penn	tastatur	regning	kode
vindu	veske	and	gorilla	gladiator	sikte
alarm	albue	elg	propell	uniform	spill
søppel	spade	løk	gulrot	gribb	menneske
kirke	kjelke	sko	bringebær	sirkus	plen
olje	okse	bål	hamburger	klima	kontroll

Target	Competitor	Distractor 1	Distractor 2	Flanker 1	Flanker 2
pakke	parykk	øse	sløyfe	samfunn	skog
drosje	dråpe	slott	rekkert	president	avløp
nøkkel	navle	tanks	syltetøy	media	blære
papir	padde	veps	løve	nebb	chili
kino	kikkert	bro	knekkebrød	syndrom	ungdom
kake	kajakk	spray	hyene	piknik	pels
stjerne	stativ	gips	kirsebær	gress	melding
fotball	flette	mus	sebra	saus	planet
kylling	kjetting	saks	drone	system	univers
kjole	kjerre	skilt	mandel	valg	ledning
engel	eple	hjort	marsvin	vann	adresse
bombe	bolle	fyr	flaggermus	mønster	video
felle	flamme	nisse	mikroskop	vinkel	krydder
monster	måke	sau	teleskop	gjest	vitne
kule	kube	kork	yoghurt	fisk	brus
rose	rede	iglo	trekkspill	venn	kosthold
dronning	drue	caps	tuba	verden	student
skole	sklie	smør	ubåt	voks	dikt
soldat	singlet	rev	satellitt	pedagog	navn
pistol	pirat	kort (spill)	presang	natur	trikk
ansikt	ankel	disk	omelett	himmel	maling
konge	koffert	sekk	åre	middag	visker
øye	øre	dør	puma	vorte	meisel
baby	bamse	nål	grøt	sminke	pledd
hjerte	hjerne	peis	skilpadde	ikon	blad
doktor	dommer	ved	pass	bakterie	signal
kaffe	kanne	øks	muffins	dato	stasjon
nummer	nudel	sag	pensel	kanal	pumpe
fengsel	filter	drill	troll	bryllup	spark
penger	perle	tre	serviett	grense	konsoll
finne	fakkel	bok	krakk	seil	oter
snegle	stubbe	pisk	sprøyte	gardin	konfekt

Target	Competitor	Distractor 1	Distractor 2	Flanker 1	Flanker 2
badstu	bunad	knapp	nonne	skulder	drops
føner	fyrstikk	seng	elefant	bensin	eksamen
lefse	leppe	glass	badekar	ansvar	krutt
sandal	svane	stein	krystall	drøm	teori
skøyte	skjelett	vekt	brygge	krutong	anker
rake	ribbe	mopp	batteri	hekk	vogn
vimpel	vifte	teip	viskelær	fest	resultat
silo	suppe	speil	ilder	bygning	frisør
flygel	flue	dusj	diamant	saft	port
deksel	data	gris	paraply	balkong	flåte
kongle	kåpe	bjørn	grevling	flytevest	jord
giraff	skjorte	stol	baguette	furu	vodka
gaupe	gitar	bord	låve	krem	granat
kjegle	kjevle	vott	pizza	brud	stav
termos	tunnel	skjell	hake	kaviar	gløgg
gevir	gebiss	hval	hette	figur	tiara
linjal	lilje	katt	fløyte	busk	grus
pokal	ponni	hund	honning	frimerke	kors
kiwi	kringle	spyd	bacon	rein	tøffel
børste	blyant	hjelm	radio	fjøl	krok
skorstein	skute	skjegg	radar	heis	reim
kjele	kiste	trapp	garasje	undulat	netting
fele	flaske	fjær	toalett	jerv	knagg
tablett	tvilling	hatt	lader	skinke	gjelle
valnøtt	vulkan	hjul	koala	bonde	nakke
komfyr	kompass	buss	kenguru	grotte	melk
lasso	larve	hest	salami	butikk	balsam
rosin	rakett	brød	paprika	tåke	mose
tromme	tavle	kjeks	globus	kartong	sylinder
moped	måne	skjerf	medalje	dirigent	dessert
strømpe	såpe	smokk	flodhest	kalv	korsett

Target	Competitor	Distractor 1	Distractor 2	Flanker 1	Flanker 2
vaffel	vugge	struts	appelsin	katalog	røyskatt
binders	bleie	sopp	ananas	plakat	teater
krage	krykke	grill	kamera	atlas	skap
kvise	krone	kniv	stadion	gang	flokk
planke	pingvin	bart	lampe	verdi	middel
humle	hytte	skje	mikrofon	karamell	autograf
pære	pipe	slips	parasoll	kjeller	alfabet
bever	belte	brev	lysekrone	sigarett	religion
kaktus	kanon	kost	knute	manet	fossil
mugge	motor	benk	leopard	akvarium	reptil
bestikk	ballong	telt	viking	gruve	potte
bøffel	bukse	tang	gaffel	oppskrift	medisin
delfin	dukke	ring	esel	sang	dans
trompet	tommel	vest	hammer	meny	atom
øgle	ørret	shorts	lego	melon	skjerm
støvel	stige	flagg	basseng	kommode	loft
traktor	turban	kurv	genser	antenne	juletre
bluse	blomkål	tann	popkorn	lakris	molekyl
kjede	kjøkken	sverd	spiker	parti	brikke
truse	tunge	munk	gjerde	offer	lypsyl
kano	kanin	skjold	hylle	fjell	mulighet
potet	puddel	mark	håndkle	skorpe	sorg
hanske	høvel	maur	pute	labyrint	klut
kamel	kalkun	frosk	jeger	stue	nyre
muskel	musling	visp	smultring	kantine	ramme
panne	palme	svamp	spagetti	kryss	røyk
krabbe	kvadrat	mygg	gele	bevis	angrep
tønne	tomat	ratt	sennep	fortau	rekkverk

Note. Adapted from "Automaticity and the notion of interference. Assessing word reading automaticity as freedom from interference in a Visual World Paradigm," by D. Zelihić, 2020, p. 57–59. Copyright 2020 by Dzan Zelihić. Adapted with permission.

Appendix B: Orthographic Properties of Target Words

The orthographic properties of the following words are available at https://noa.spell.uiocloud.no.

High-frequency Target Words

Word	Letter no.	Bigram token w/end	OLD20	Zip Frequency	English translation
hale	4	3.628	1.00	3.981	tail
nese	4	3.824	1.00	3.981	nose
robot	5	3.210	1.80	3.981	robot
maske	5	3.625	1.00	3.987	mask
drage	5	3.481	1.35	3.995	dragon
sykkel	6	3.214	1.85	4.011	bicycle
verktøy	7	3.165	2.75	4.025	tool(s)
pølse	5	3.076	1.65	4.029	sausage
pilot	5	3.294	1.65	4.032	pilot
skygge	6	3.183	1.65	4.053	shadow
finger	6	3.705	1.00	4.063	finger
ridder	6	3.541	1.45	4.070	knight
teppe	5	3.498	1.20	4.076	rug
tiger	5	3.687	1.00	4.077	tiger
rotte	5	3.573	1.00	4.084	rat
sukker	6	3.419	1.05	4.094	sugar
jakke	5	3.385	1.00	4.107	jacket
mobil	5	3.077	1.75	4.115	cell phone
demon	5	3.512	1.10	4.117	demon
pose	4	3.467	1.00	4.121	plastic bag
klokke	6	3.341	1.25	4.132	watch
vampyr	6	2.996	1.85	4.136	vampire
slange	6	3.603	1.40	4.148	snake
billett	7	3.540	1.70	4.150	ticket
vindu	5	3.125	1.75	4.164	window
alarm	5	3.283	1.90	4.190	alarm
søppel	6	3.036	1.95	4.191	garbage

Word	Letter no.	Bigram token w/end	OLD20	Zip Frequency	English translation
kirke	5	3.347	1.30	4.232	church
olje	4	3.260	1.35	4.236	oil
pakke	5	3.471	1.10	4.245	package
drosje	6	3.362	1.75	4.261	cab / taxi
nøkkel	6	3.108	1.75	4.264	key
papir	5	3.198	1.70	4.285	paper
kino	4	3.211	1.00	4.331	cinema
kake	4	3.585	1.00	4.351	cake
stjerne	7	3.543	1.75	4.351	star
fotball	7	3.159	1.95	4.358	soccer ball / football
kylling	7	3.313	1.85	4.398	chicken
kjole	5	3.439	1.45	4.414	dress
engel	5	3.676	1.50	4.418	angel
bombe	5	3.248	1.35	4.434	bomb
felle	5	3.611	1.00	4.437	mouse trap
monster	7	3.701	1.55	4.469	monster
kule	4	3.443	1.00	4.503	marble
rose	4	3.562	1.00	4.511	rose
dronning	8	3.473	1.80	4.539	queen
skole	5	3.611	1.15	4.578	school
soldat	6	3.285	1.75	4.653	soldier
pistol	6	3.354	1.70	4.726	gun
ansikt	6	3.470	1.85	4.752	face
konge	5	3.656	1.45	4.779	king
øye	3	3.084	1.00	4.889	eye
baby	4	2.893	1.35	4.895	baby
hjerte	6	3.558	1.60	4.896	heart
doktor	6	3.313	1.80	4.965	doctor
kaffe	5	3.175	1.40	5.043	coffee
nummer	6	3.334	1.20	5.055	number

Word	Letter no.	Bigram token w/end	OLD20	Zip Frequency	English translation
fengsel	7	3.524	1.50	5.129	prison
penger	6	3.828	1.00	5.622	money
finne	5	3.637	1.00	4.420	(fish) fin

Low-frequency Target Words

Word	Letter no.	Bigram token w/end	OLD20	Zip Frequency	English translation
snegle	6	3.371	1.65	1.824	snail
badstu	6	3.051	2.25	1.824	sauna
føner	5	3.382	1.35	1.949	hair dryer
lefse	5	3.276	1.45	2.046	(Norwegian dish)
sandal	6	3.375	1.60	2.301	sandal
skøyte	6	3.120	1.60	2.301	skate
rake	4	3.631	1.00	2.389	rake
vimpel	6	3.215	1.95	2.389	pennant
silo	4	3.212	1.00	2.426	silo
flygel	6	3.095	1.90	2.426	grand piano
deksel	6	3.453	1.90	2.493	phone cover
kongle	6	3.522	1.75	2.523	pinecone
giraff	6	2.857	1.90	2.523	giraffe
gaupe	5	3.173	1.55	2.578	lynx
kjegle	6	3.339	1.75	2.578	cone
termos	6	3.491	2.00	2.602	thermos
gevir	5	3.333	1.85	2.602	antlers
linjal	6	3.162	1.90	2.669	ruler
pokal	5	3.212	1.90	2.709	trophy
kiwi	4	2.360	1.65	2.838	kiwi
børste	6	3.386	1.55	2.852	hairbrush
skorstein	9	3.630	2.55	2.879	chimney

Word	Letter no.	Bigram token w/end	OLD20	Zip Frequency	English translation
kjele	5	3.567	1.00	2.891	pot
fele	4	3.640	1.00	2.891	fiddle
tablett	7	3.435	1.95	2.903	pill
valnøtt	7	2.974	2.75	2.903	walnut
komfyr	6	2.936	2.25	2.915	stove
lasso	5	3.247	1.70	2.915	lasso
rosin	5	3.519	1.40	2.938	raisin
tromme	6	3.451	1.45	2.938	drum
moped	5	3.202	1.95	2.949	moped
strømpe	7	3.234	1.80	2.949	stocking
vaffel	6	3.077	1.90	2.960	waffle
binders	7	3.604	1.75	2.960	paper clips
krage	5	3.486	1.30	2.960	collar
kvise	5	3.422	1.35	2.981	pimple
planke	6	3.457	1.45	2.991	plank
humle	5	3.140	1.20	3.000	bumblebee
pære	4	2.926	1.15	3.019	pear
bever	5	3.571	1.00	3.046	beaver
kaktus	6	3.273	2.00	3.046	cactus
mugge	5	3.155	1.05	3.080	pitcher
bestikk	7	3.422	1.75	3.146	cutlery
bøffel	6	2.795	1.85	3.160	buffalo
delfin	6	3.440	1.75	3.167	dolphin
trompet	7	3.438	1.75	3.186	trumpet
øgle	4	3.018	1.55	3.186	lizard
støvel	6	3.196	1.85	3.216	gumboot
traktor	7	3.457	1.90	3.234	tractor
bluse	5	3.349	1.35	3.261	blouse
kjede	5	3.479	1.55	3.266	necklace
truse	5	3.480	1.35	3.271	underpants

Word	Letter no.	Bigram token w/end	OLD20	Zip Frequency	English translation
kano	4	3.274	1.00	3.277	canoe
potet	5	3.531	1.30	3.297	potato
hanske	6	3.628	1.40	3.325	glove
kamel	5	3.407	1.30	3.325	camel
muskel	6	3.337	1.95	3.334	muscle
panne	5	3.643	1.00	3.343	pan
krabbe	6	3.232	1.45	3.347	crab
tønne	5	3.273	1.25	3.602	barrel

Appendix C: Experimental Trial Displays

Figure 1

Target Presented in the Normal-spacing Condition

Figure 2

Target Presented in the Half-spacing Condition

Figure 3

Target Presented in the Double-spacing Condition

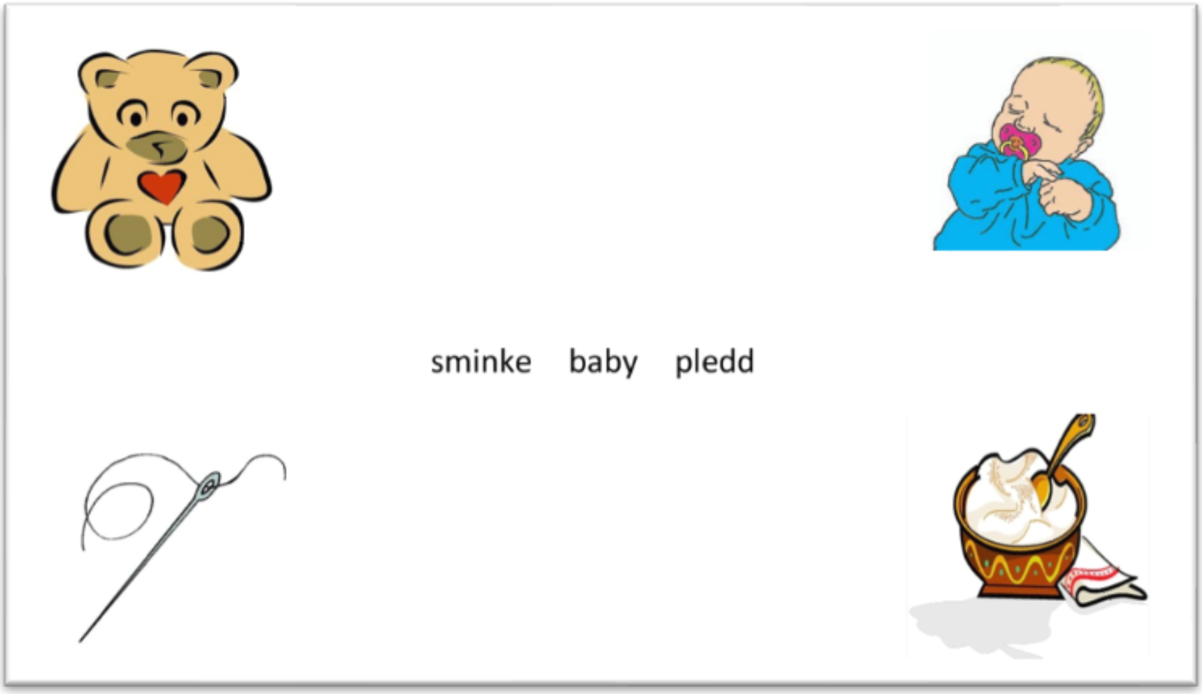


Figure 8

Example of Display Images

Figure 9

Illustration of the Red Dot

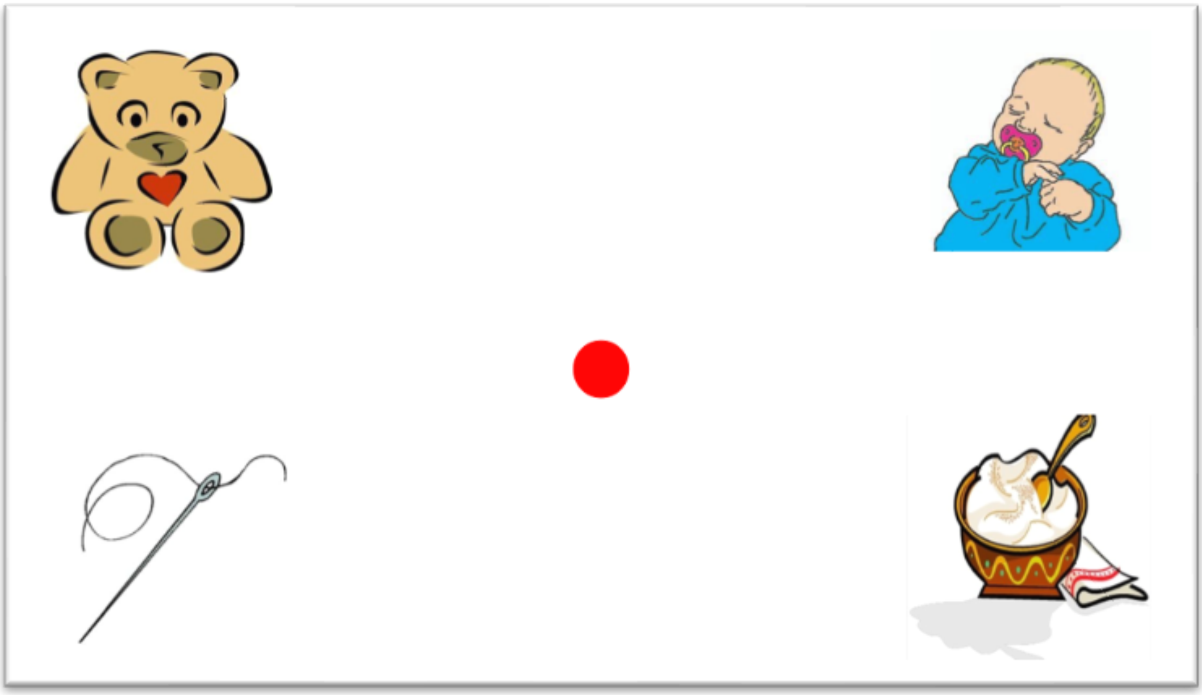


Figure 10

Illustration of the Fixation Trigger

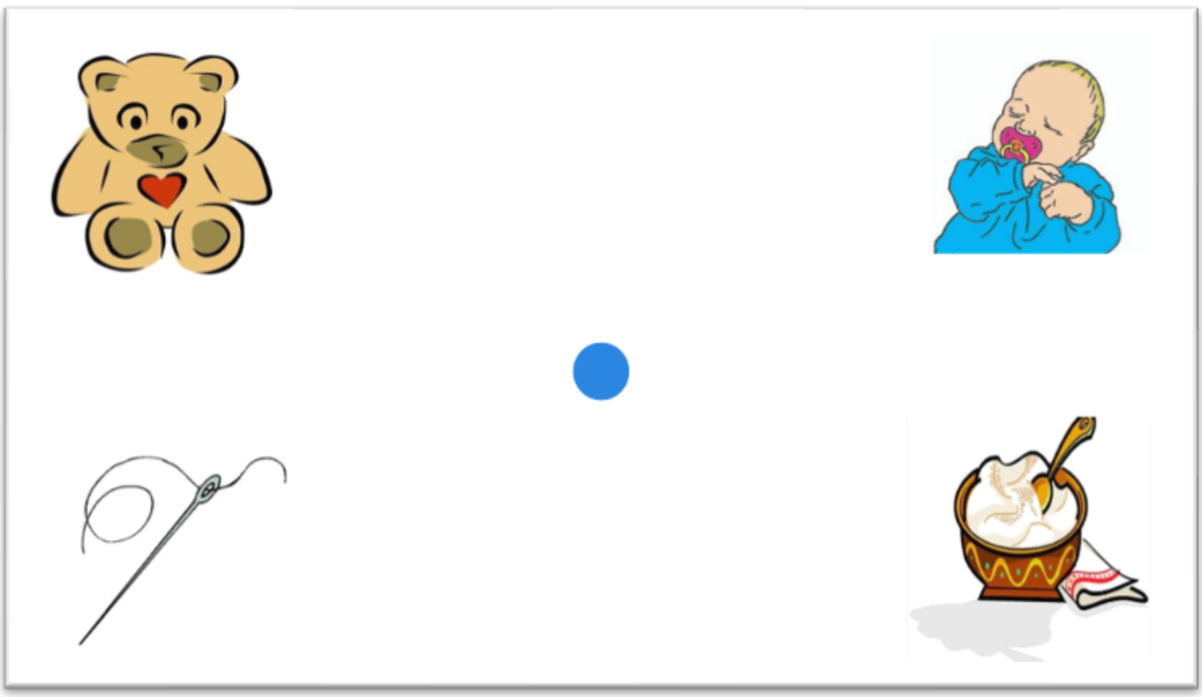


Figure 11

Illustration of the Visual Mask

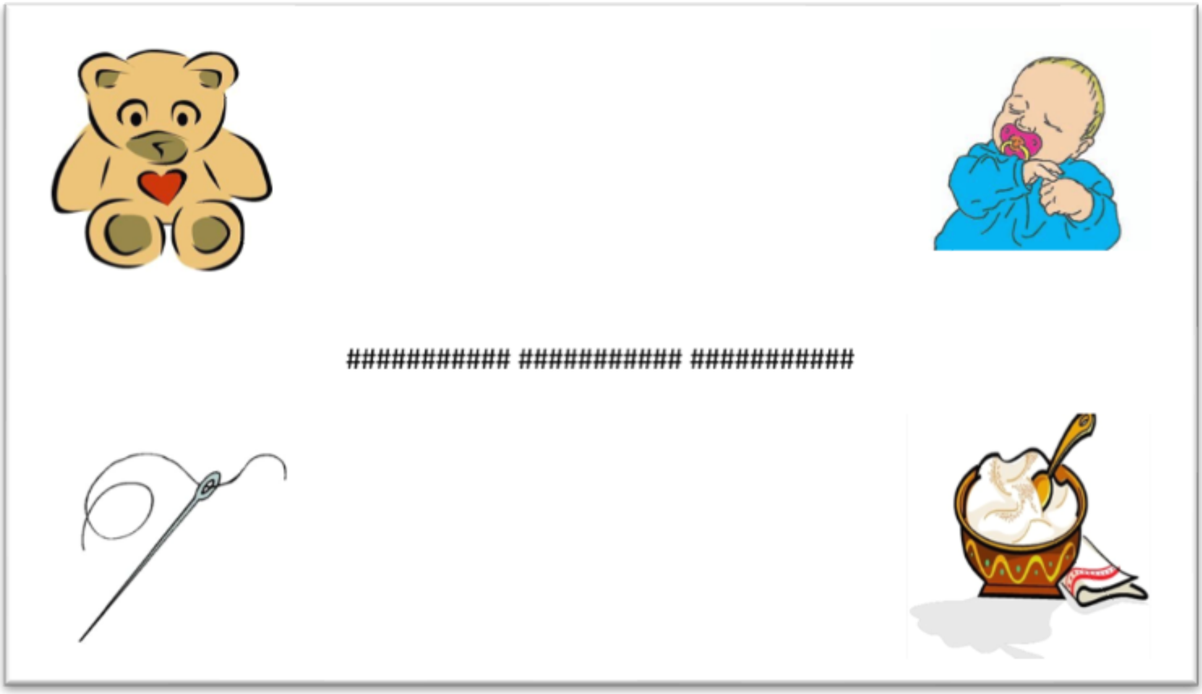
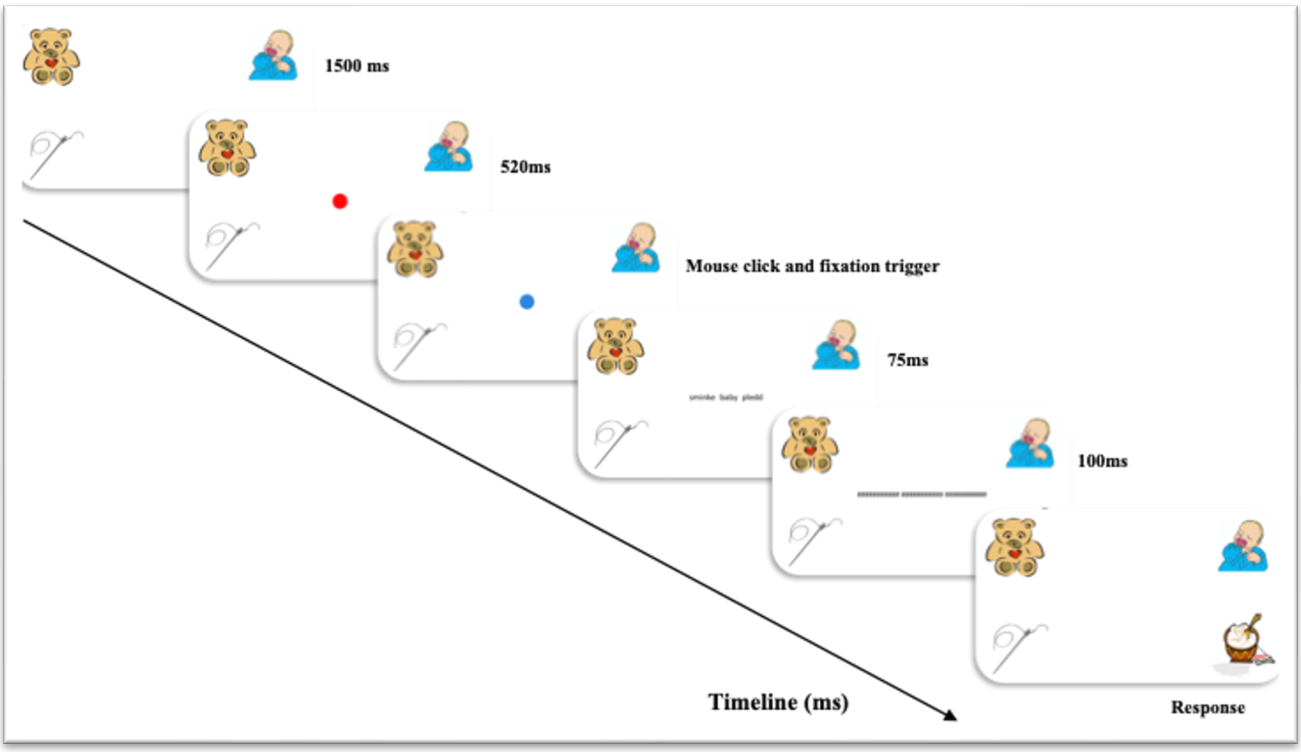


Figure 12

Time Course of the Successive Displays During a Trial



Appendix D: Consent Form

Vil du delta i forskningsprosjektet
«Måling av ordautomatisering ved lesing på norsk»?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å undersøke hvorvidt automatisk ordgjenkjenning av et ord blir påvirket av andre nærliggende ord. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Formålet med forskningsprosjektet er å bidra til å belyse ideen om at nærliggende ord kan forstyrre lesing av enkeltord. Dette er viktig fordi automatisert ordgjenkjenning er nødvendig for å oppnå leseflyt. Tidligere forskning har vist at evne til automatisk ordgjenkjenning vil kunne predikere leseflyt, uavhengig av ord- og bokstavkunnskap. Forskningen som er gjort har tatt utgangspunkt i lesing av isolerte enkeltord, noe som ikke gjenspeiler en reell lesesituasjon der ord vil være omringet av nærliggende ord i setningen. Det er imidlertid antatt at nærliggende ord kan skape forstyrrelser som vil påvirke automatisk ordgjenkjenning, men det er ikke gjort noe direkte forskning på akkurat dette.

Forskningsprosjektet baserer seg på en større doktorgradstudie ved Universitetet i Oslo.

Hvem er ansvarlig for forskningsprosjektet?

Universitetet i Oslo, Institutt for spesialpedagogikk. Studien er et samarbeidsprosjekt mellom universitetet i Oslo og Department of Psychological and Brain Sciences ved University of Iowa.

Hvorfor får du spørsmål om å delta?

Utvalget er trukket ut ifra følgende kriterier:

- Alder (18-35)
- Språk (Har norsk som hovedspråk)
- Antall deltakere som vil få henvendelse er beregnet til å være 60

Rekruttering av deltakere for prosjektet vil hovedsakelig bestå av både kjente og ukjente studenter, samt unge voksne i arbeid.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det først at du gjennomfører to enkel tester: TOWRE, en lesetest som måler effektiv ordlesing av både enkeltord og non-ord, og Korrekturlesing, en test som måler rettskriving. Deretter vil du få presentert ord på en skjerm med tilhørende bilder, hvor oppgaven din blir å klikke på det bildet som tilsvarer ordet. Lesing på skjerm vil bli kartlagt ved bruk av eye-tracking, dette innebærer at vi bruker sensorteknologi for å kartlegge dine øyebevegelser under lesing. Formålet ved bruk av eye-tracking er å kunne måle hastigheten av informasjonsprosessering, med og uten forstyrrelser. Det vil ta deg ca. 45 minutter å gjennomføre hele prosessen.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg er anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- De som vil ha tilgang til dine data er prosjektansvarlig/veileder og to studenter.
- Ingen personlige data vil bli registrert, og alt av data er anonymisert. Vi registrer bare alder og kjønn med brukerkoder.
- Anonymiserte data vil kunne deles med andre forskere til videre forskning.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes 31.12.2023. Datainnsamlingen avsluttes 31.05.2023, og deretter vil resultatene analyseres og bearbeides i forbindelse med å besvare forskningsspørsmål. Anonymiserte forskningsdata vil etter dette bli tatt vare på i forbindelse med videre forskning, og muligens delt med andre forskere både i og utenfor EU.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt muntlig samtykke.

På oppdrag fra Universitetet i Oslo, institutt for spesialpedagogikk har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Institutt for spesialpedagogikk ved Athanasios Protopapas, på e-post: athanasios.protopapas@isp.uio.no
- Vårt personvernombud: Roger Markgraf-Bye, på e-post: personvernombud@uio.no
- NSD – Norsk senter for forskningsdata AS, på epost (personverntjenester@nsd.no) eller telefon: 55 58 21 17.

Med vennlig hilsen

Prosjektansvarlig	Masterstudent
Athanasios Protopapas	Stefania Kyriakidou

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «Måling av ordautomatisering ved lesing på norsk», og har fått anledning til å stille spørsmål. Jeg samtykker til:

- å delta i en ordgjenkjenning studie der mine øyebevegelser blir registrert
- å delta i test (Norsk versjon av TOWRE og Korrekturlesing)
- at anonymiserte dataopplysninger lagres på ubestemt tid etter prosjektslutt, og vil kunne deles med andre forskere til eventuell bruk i videre forskning i Norge eller andre land i eller utenfor EU

Jeg samtykker til at mine opplysninger behandles frem til uavgrenset tid.

(uten signatur; samtykke oppgis muntlig)

Appendix E: Descriptive Statistics for the Item Sample

Dependent Variable: Proportion correct answers

Table 31

Descriptives for Proportion Correct Responses by Frequency Group

	Frequency Group	Normal	Half	Double
M	HF[a]	0.92	0.88	0.95
	LF[b]	0.92	0.88	0.94
SD	HF	0.10	0.12	0.09
	LF	0.10	0.11	0.07
Min	HF	0.58	0.47	0.63
	LF	0.55	0.55	0.70
Max	HF	1.00	1.00	1.00
	LF	1.00	1.00	1.00
Skewness	HF	–1.66	–1.49	–2.29
	LF	–1.67	–0.97	–1.43
Kurtosis	HF	2.42	1.93	5.16
	LF	2.79	0.71	2.21
Shapiro-Wilk	HF	0.78	0.82	0.63
	LF	0.79	0.90	0.81
p	HF	< .001	< .001	< .001
	LF	< .001	< .001	< .001

[a] lower-frequency; [b] higher-frequency

Figure 60

Histogram of accuracy in normal spacing for HF (top) and LF (bottom) words across items

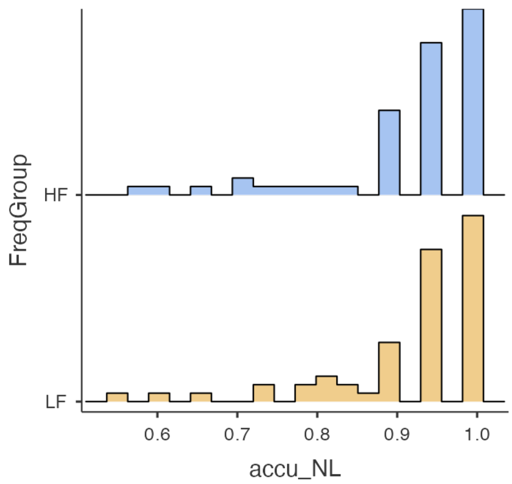


Figure 61

Q-Q plot of accuracy in normal spacing for HF (left plot) and LF (right plot) words across items

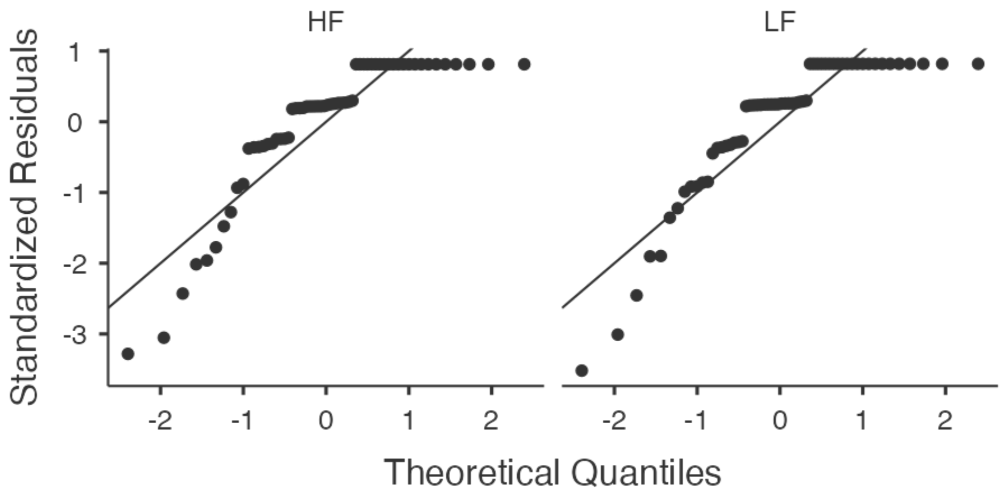


Figure 62

Histogram of accuracy in half spacing for HF (top) and LF (bottom) words across items

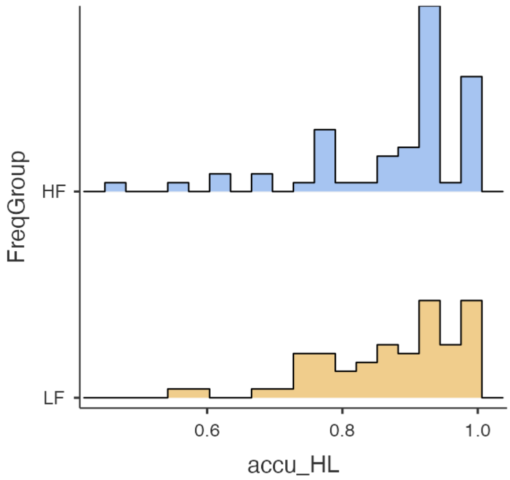


Figure 63

Q-Q plot of accuracy in half spacing for HF (left plot) and LF (right plot) words across items

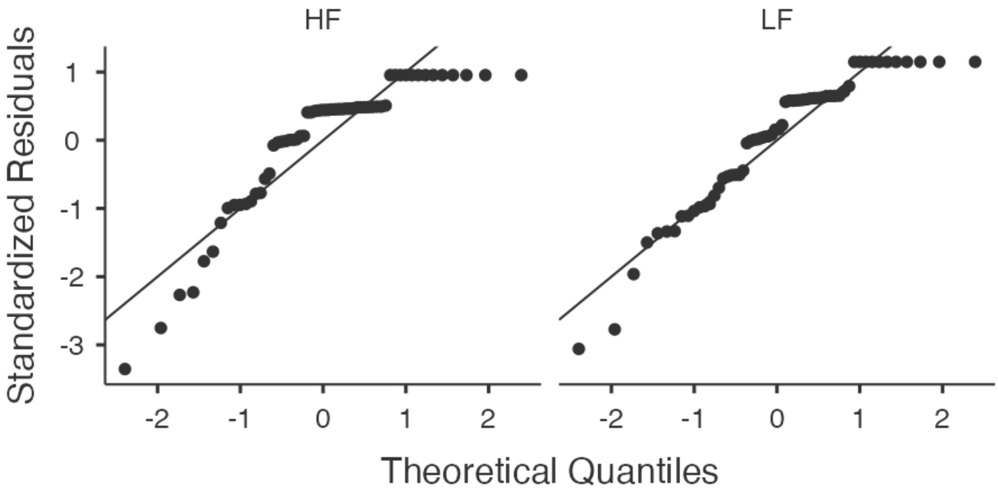


Figure 64

Histogram of accuracy in double spacing for HF (top) and LF (bottom) words across items

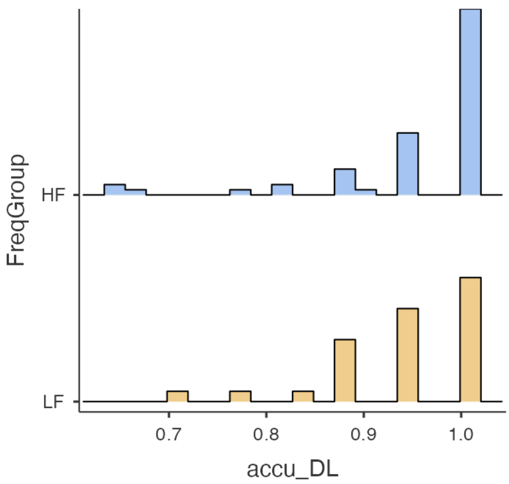
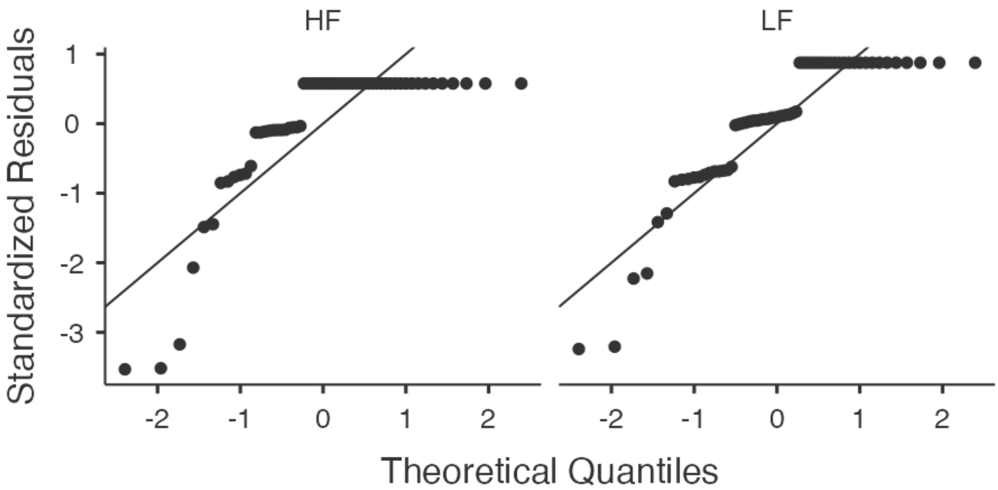


Figure 65

Q-Q plot of accuracy in double spacing for HF (left plot) and LF (right plot) words across items



Dependent Variable: Proportion looks to target

Table 32

Descriptives for Proportion Target Looks at 200ms by Frequency Group

	Frequency Group	Normal	Half	Double
M	HF	0.26	0.23	0.29
	LF	0.21	0.19	0.24
SD	HF	0.11	0.11	0.11
	LF	0.09	0.10	0.10
Min	HF	0.03	0.04	0.06
	LF	0.04	0.00	0.02
Max	HF	0.59	0.53	0.55
	LF	0.44	0.45	0.46
Skewness	HF	0.43	0.71	0.26
	LF	0.37	0.72	0.24
Kurtosis	HF	0.24	0.48	–0.33
	LF	–0.70	0.59	–0.28
Shapiro-Wilk	HF	0.98	0.96	0.98
	LF	0.97	0.95	0.98
p	HF	0.533	0.057	0.599
	LF	0.130	0.021	0.604

[a] lower-frequency; [b] higher-frequency

Figure 66.

Histogram of proportion target looks in normal spacing for HF (top) and LF (bottom) words across items (200ms)

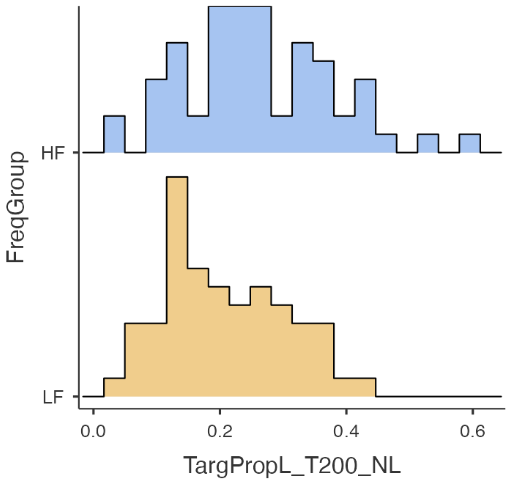


Figure 67

Q-Q plot of proportion target looks in normal spacing for HF (left plot) and LF (right plot) words across items (200ms)

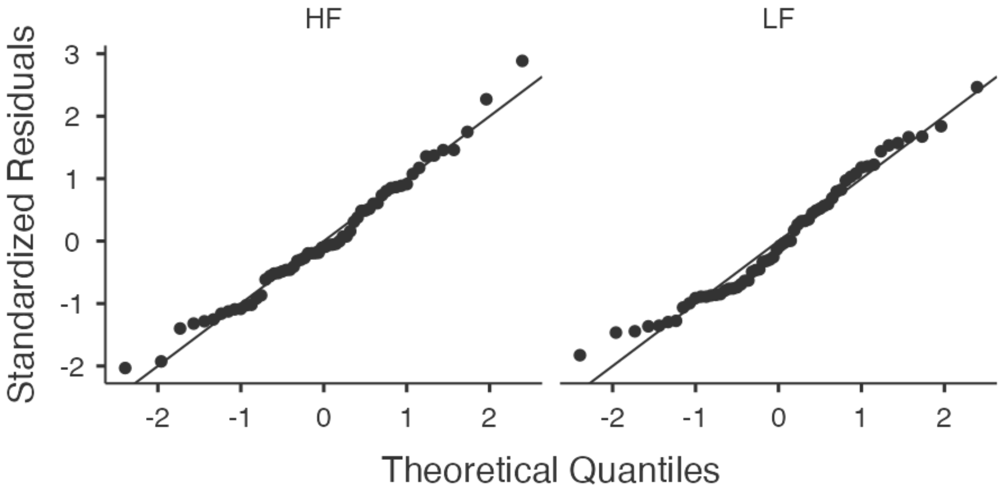


Figure 68

Histogram of proportion target looks in half spacing for HF (top) and LF (bottom) words across items (200ms)

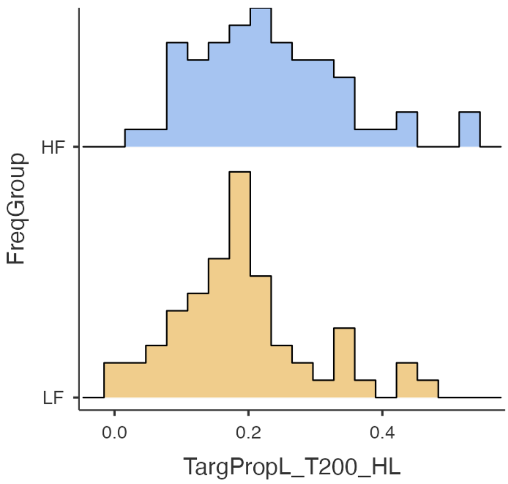


Figure 69

Q-Q plot of proportion target looks in half spacing for HF (left plot) and LF (right plot) words across items (200ms)

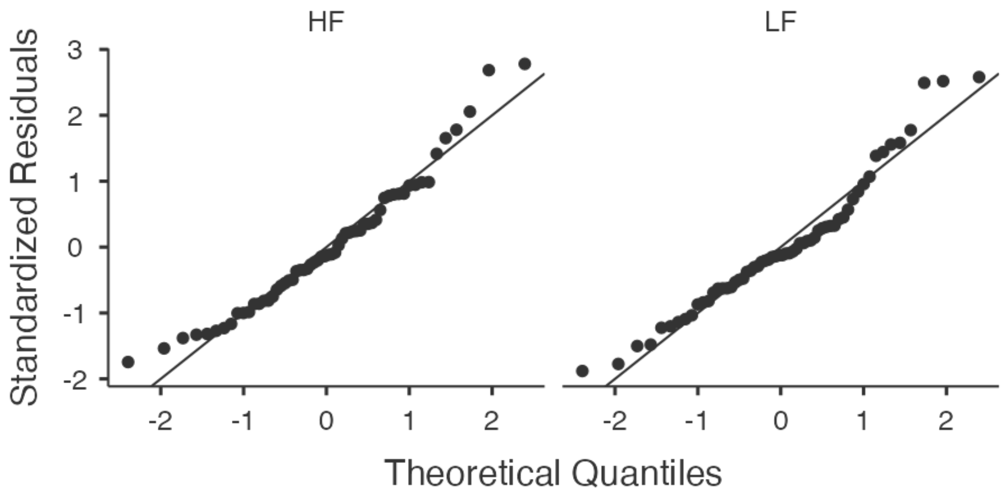


Figure 70

Histogram of proportion target looks in double spacing for HF (top) and LF (bottom) words across items (200ms)

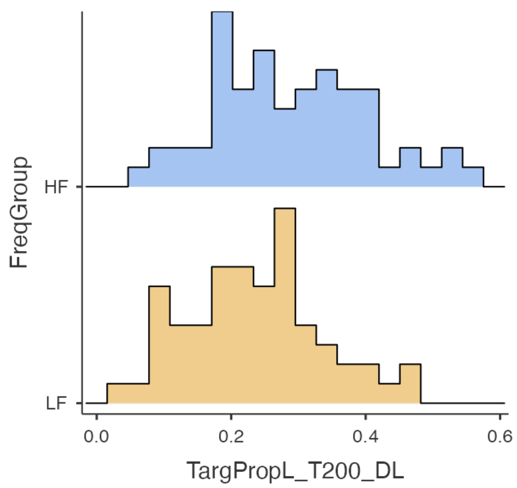


Figure 71

Q-Q plot of proportion target looks in double spacing for HF (left plot) and LF (right plot) words across items (200ms)

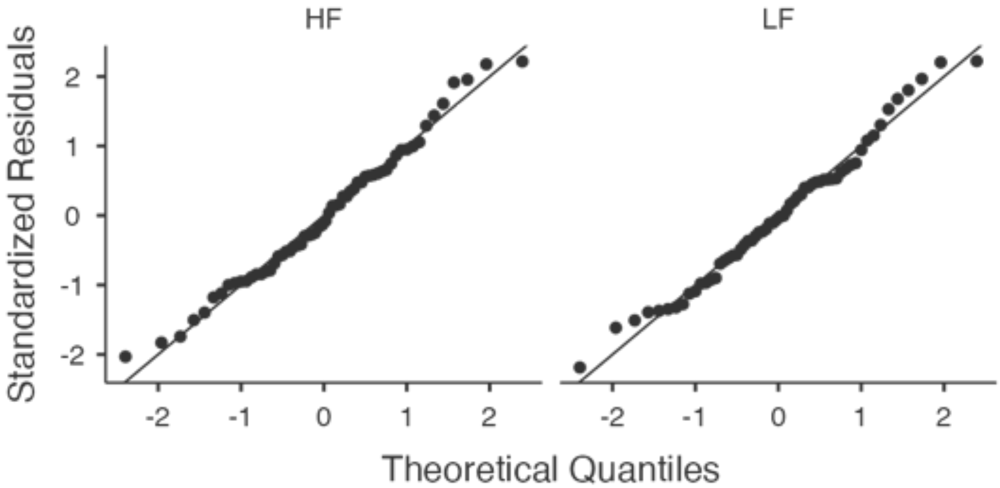


Table 33

Descriptives for Proportion Target Looks at 400 ms by Frequency Group

	Frequency Group	Normal	Half	Double
M	HF	0.66	0.60	0.68
	LF	0.62	0.58	0.64
SD	HF	0.16	0.17	0.14
	LF	0.16	0.19	0.15
Min	HF	0.14	0.21	0.29
	LF	0.16	0.09	0.28
Max	HF	0.96	0.92	0.94
	LF	0.95	0.97	0.97
Skewness	HF	–0.56	–0.37	–0.65
	LF	–0.42	–0.29	–0.26
Kurtosis	HF	0.91	–0.23	0.19
	LF	0.28	0.08	–0.29

	Frequency Group	Normal	Half	Double
Shapiro-Wilk	HF	0.97	0.98	0.97
	LF	0.98	0.99	0.99
p	HF	0.198	0.255	0.098
	LF	0.385	0.802	0.671

[a] lower-frequency; [b] higher-frequency

Figure 72

Histogram of proportion target looks in normal spacing for HF (top) and LF (bottom) words across items (400ms)

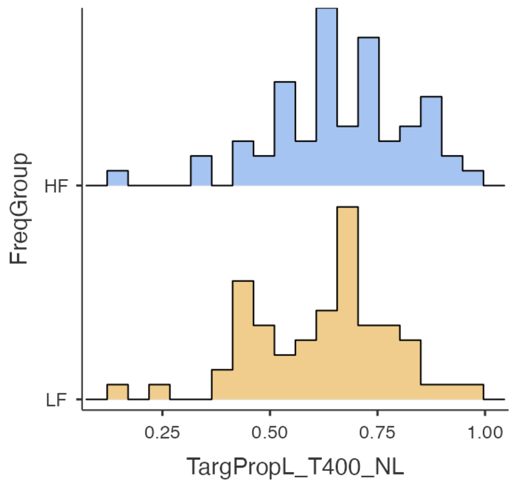


Figure 73

Q-Q plot of proportion target looks in normal spacing for HF (left plot) and LF (right plot) words across items (400ms)

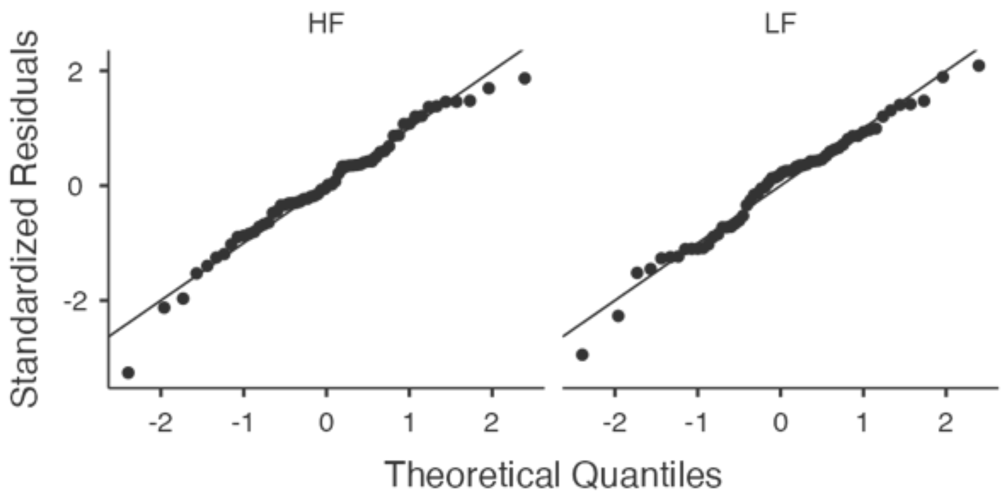


Figure 74

Histogram of proportion target looks in half spacing for HF (top) and LF (bottom) words across items (400ms)

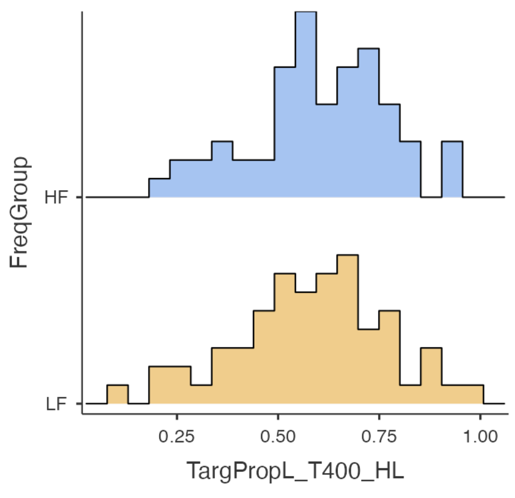


Figure 75

Q-Q plot of proportion target looks in half spacing for HF (left plot) and LF (right plot) words across items (400ms)

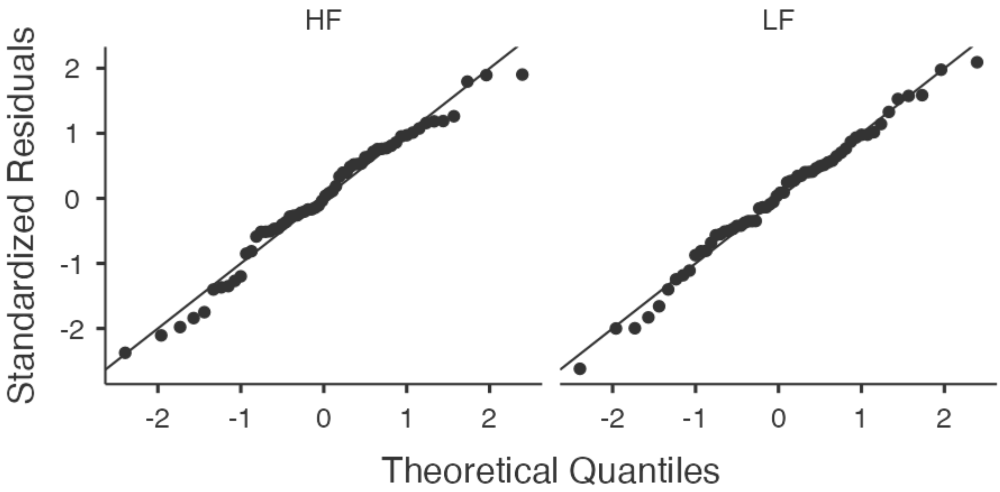


Figure 76

Histogram of proportion target looks in double spacing for HF (top) and LF (bottom) words across items (400ms)

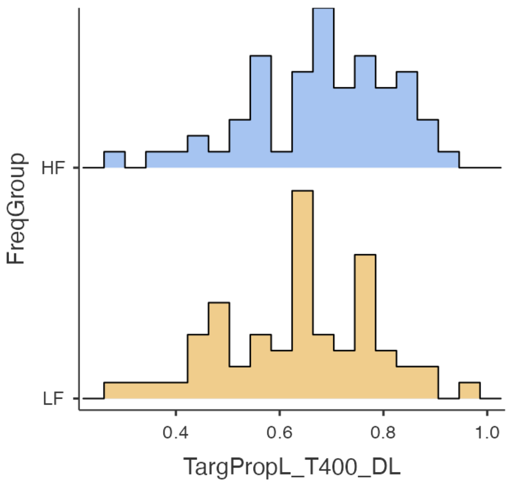


Figure 77

Q-Q plot of proportion target looks in double spacing for HF (left plot) and LF (right plot) words across items (400ms)

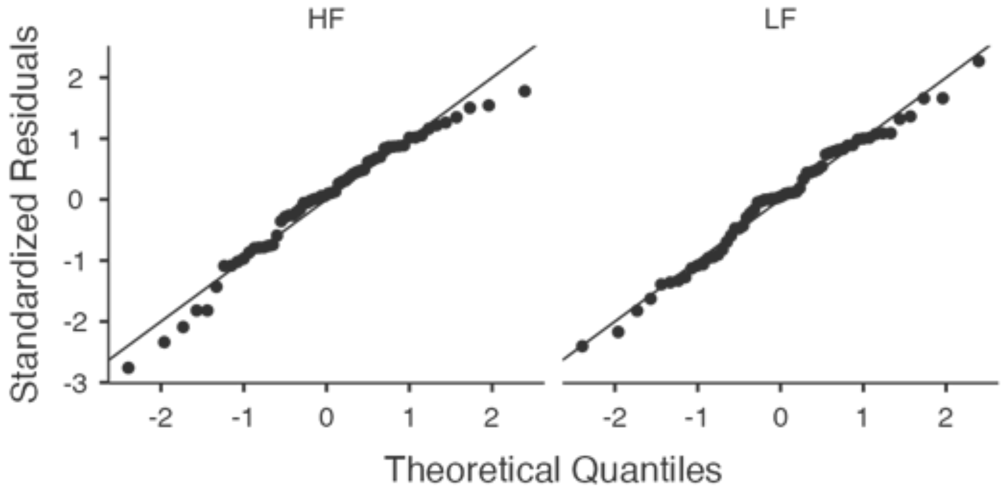


Table 34

Descriptives for Proportion Target Looks at 600ms by Frequency Group

	Frequency Group	Normal	Half	Double
M	HF	0.83	0.78	0.84
	LF	0.79	0.75	0.83
SD	HF	0.15	0.15	0.11
	LF	0.13	0.17	0.12
Min	HF	0.26	0.39	0.59
	LF	0.53	0.16	0.52
Max	HF	1.00	1.00	1.00
	LF	1.00	1.00	1.00
Skewness	HF	–2.03	–0.91	–0.48
	LF	–0.39	–1.15	–0.78
Kurtosis	HF	4.39	0.50	–0.57
	LF	–0.53	2.24	0.30

	Frequency Group	Normal	Half	Double
Shapiro-Wilk	HF	0.78	0.92	0.95
	LF	0.96	0.91	0.94
p	HF	< .001	< .001	0.021
	LF	0.073	< .001	0.006

[a] lower-frequency; [b] higher-frequency

Figure 78

Histogram of proportion target looks in normal spacing for HF (top) and LF (bottom) words across items (600ms)

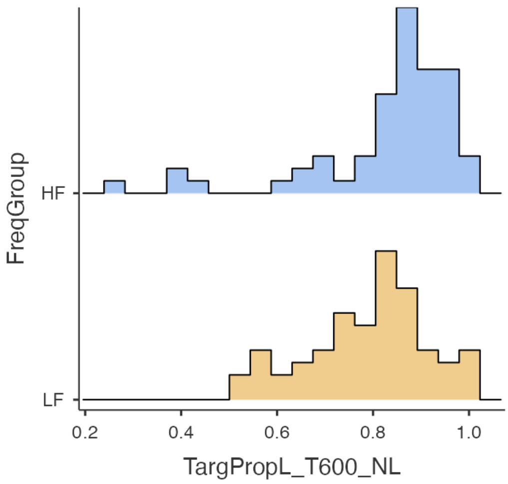


Figure 79

Q-Q plot of proportion target looks in normal spacing for HF (left plot) and LF (right plot) words across items (600ms)

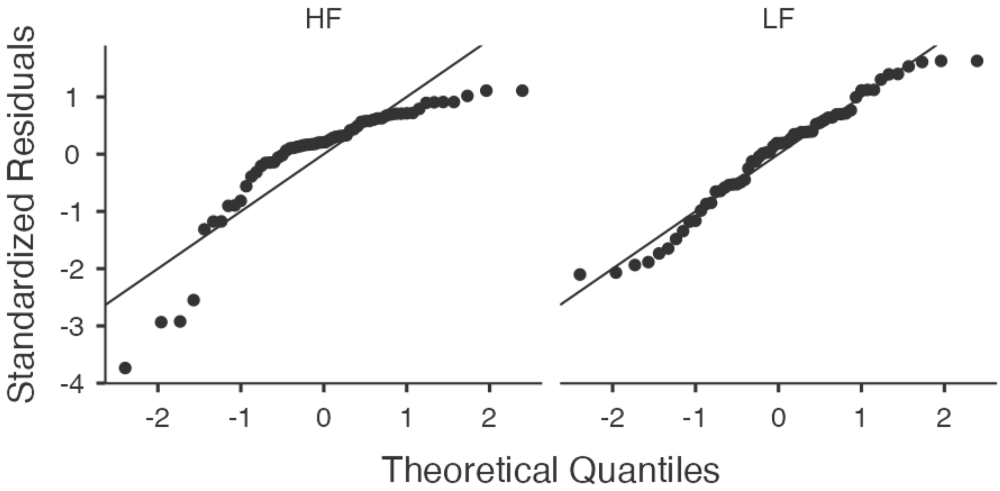


Figure 80

Histogram of proportion target looks in half spacing for HF (top) and LF (bottom) words across items (600ms)

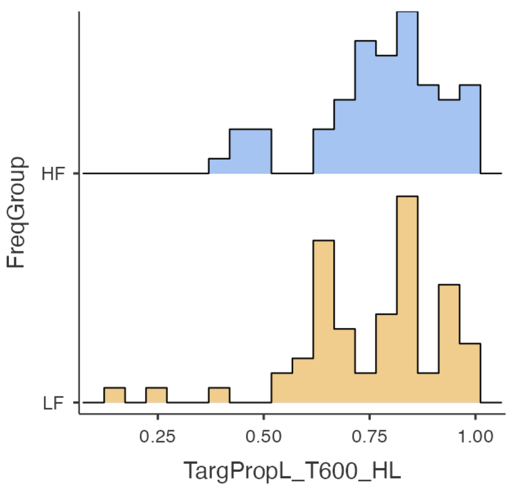


Figure 81

Q-Q plot of proportion target looks in half spacing for HF (left plot) and LF (right plot) words across items (600ms)

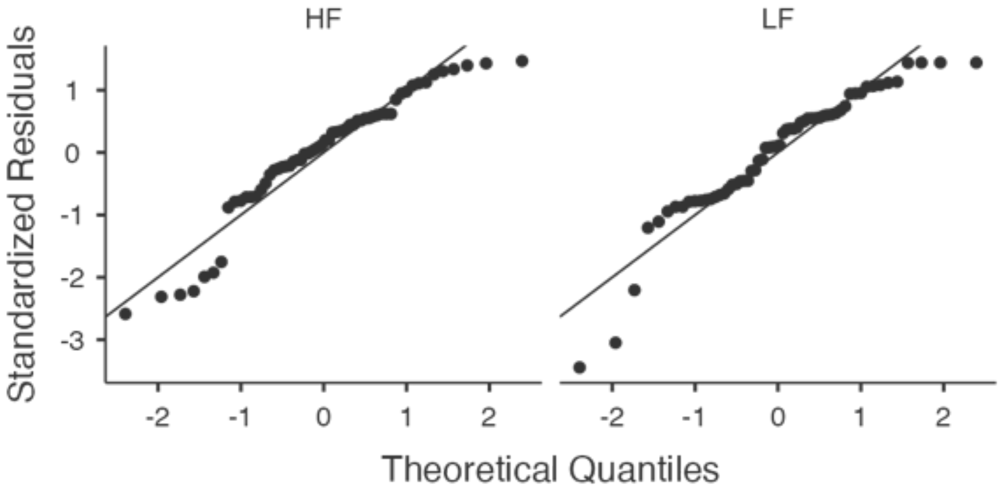


Figure 82

Histogram of proportion target looks in double spacing for HF (top) and LF (bottom) words across items (600ms)

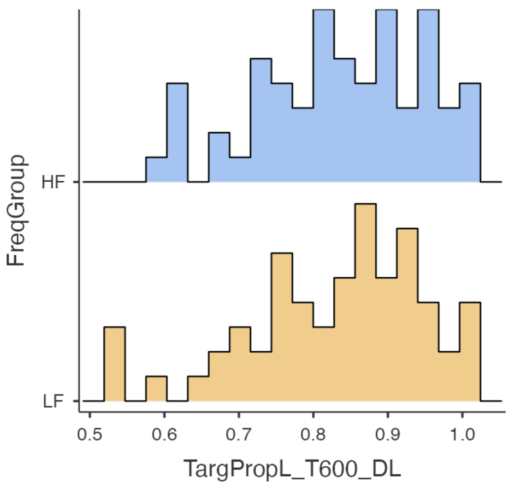
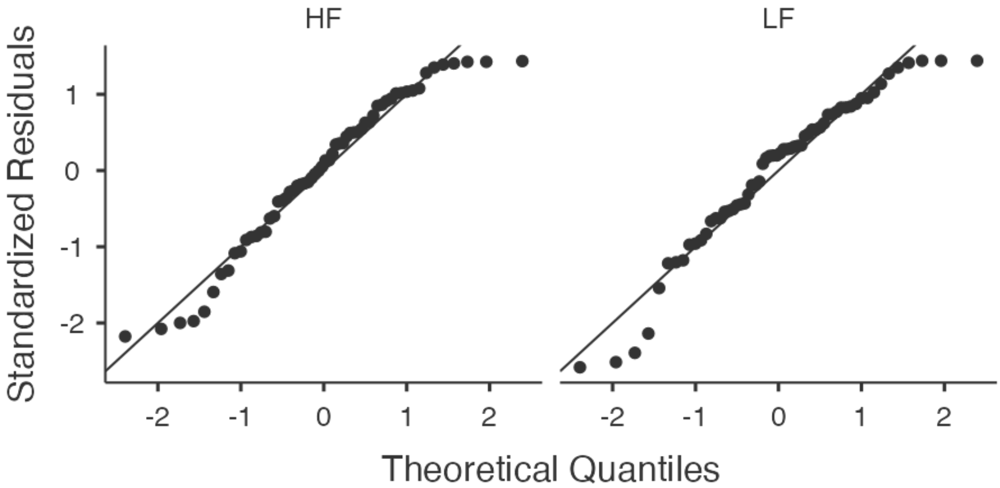


Figure 83

Q-Q plot of proportion target looks in double spacing for HF (left plot) and LF (right plot) words across items (600ms)



Dependent Variable: Time of first sample to target

Table 35

Descriptives for Time to Target Across Items by Frequency Group

	Frequency Group	Normal	Half	Double
M	HF	414.40	444.37	408.53
	LF	442.28	452.07	428.51
SD	HF	64.45	74.08	60.82
	LF	58.32	73.49	61.59
Min	HF	293.33	303.81	306.76
	LF	320.20	313.18	321.69
Max	HF	668.00	630.33	600.50
	LF	568.78	662.50	620.91
Skewness	HF	1.10	0.39	0.89

	Frequency Group	Normal	Half	Double
	LF	0.16	0.39	0.70
Kurtosis	HF	3.18	0.14	0.98
	LF	–0.74	0.35	0.66
Shapiro-Wilk	HF	0.94	0.97	0.95
	LF	0.98	0.98	0.97
p	HF	0.005	0.163	0.011
	LF	0.457	0.453	0.115

[a] lower-frequency; [b] higher-frequency

Figure 84

Histogram of time to target in normal spacing for HF (top) and LF (bottom) words across items

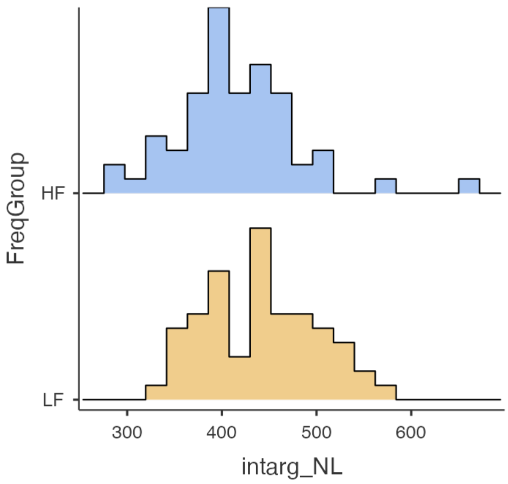


Figure 85

Q-Q plot of time to target in normal spacing for HF (left plot) and LF (right plot) words across items

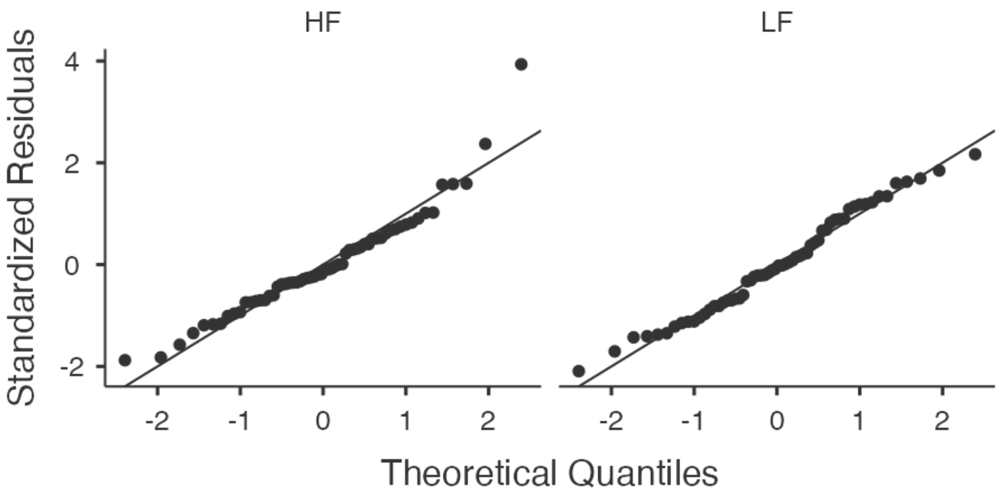


Figure 86

Histogram of time to target in half spacing for HF (top) and LF (bottom) words across items

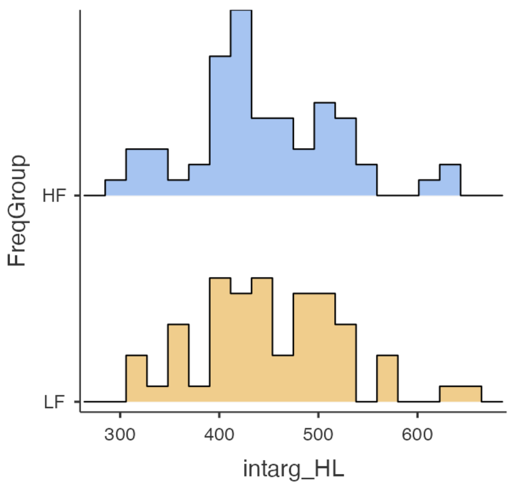


Figure 87

Q-Q plot of time to target in half spacing for HF (left plot) and LF (right plot) words across items

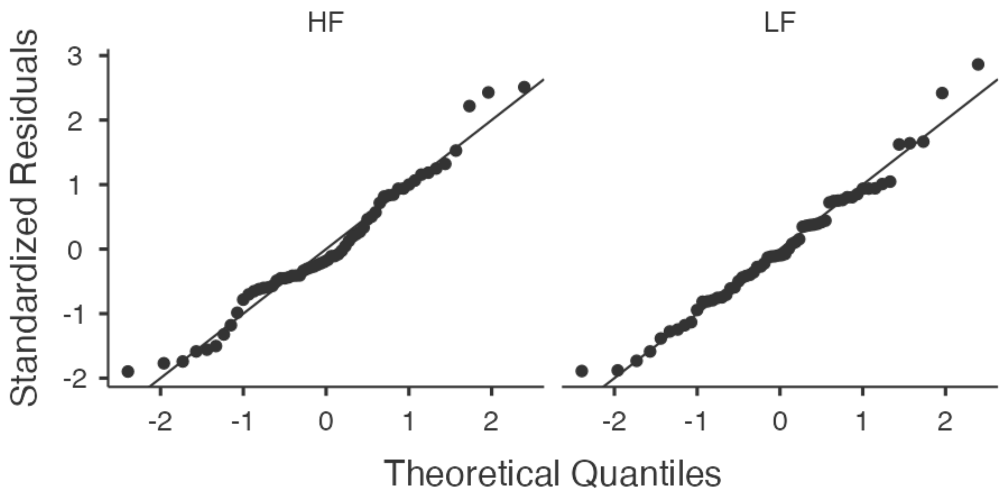


Figure 88

Histogram of time to target in double spacing for HF (top) and LF (bottom) words across items

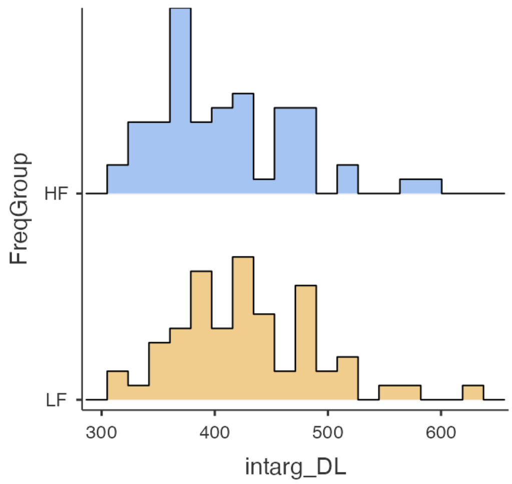


Figure 89

Q-Q plot of time to target in double spacing for HF (left plot) and LF (right plot) words across items