# Can’t See the Words for the Letters: Whole Word Processing in Adolescents and a Novel “Visual” Intervention for Dyslexia.

By

Emma Samantha Louise Ashcroft

A thesis submitted to Victoria University of Wellington in fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University of Wellington

2020

# ABSTRACT

Individuals with developmental dyslexia, considered as a group, perform poorly on tasks that involve phonological analysis, such as applying sight-sound rules to read new words, or analysing words into their component sounds (De Groot, Huettig, & Olivers, 2016; Temple & Marshall, 1983). However, dyslexia is also associated with other types of difficulties. For example, in some individuals, reading latencies increase disproportionately with the length of the word (De Luca, Barca, Burani, & Zoccolotti, 2008; Spinelli et al., 2005) suggesting they may have difficulties recognising familiar words as whole units ("whole word" processing).

This thesis examined the relationship between the word length effect and overall reading proficiency in a diverse sample of 49 adolescents. We found that the length effect was a unique predictor of reading proficiency, even after factoring out variance in phonological skills (measured using a nonword reading task). We also tested the recent hypothesis that *visual attention span* - the number of letters a reader can capture in a single glance - is important for efficient whole word reading (Bosse, Tainturier, & Valdois, 2007). Contrary to this hypothesis, we found no association between the word length effect and scores on a standard measure of visual attention span (a partial report task).

We also explored whether reading-delayed adolescents could benefit from an intervention targeting their specific cognitive profile. Five cases demonstrating a selective difficulty with either "phonological" or "whole word" skills completed two interventions. One targeted phonological skills: participants were trained to recognise and apply common sight-sound correspondences. The other targeted whole word skills: we reasoned that training participants to recognise commonly-occurring letter redundancies (e.g. ogue: rogue, synagogue, dialogue) could reduce the load on parallel letter processing. Only one of the five cases showed greater improvement in (untrained) word reading accuracy following their "target" intervention. However, four of the five showed intervention-specific improvements in reading latency. These results suggest that it could be valuable to consider heterogeneity when treating reading delay.

Acknowledgments

First and foremost, this research would have been impossible without the participants' and their willingness to give both me and this research a chance, particularly those who took part in the intervention (HE, MI, FR, OK, PI, and SH). Teachers, forever the unsung heroes, have also been instrumental to the completion of this research and I am in awe of the hard work, compassion and patience they put in. Thanks go to Gemma, but also Hannah, Peter, Sarah, Maureen and Carly.

To the inmates of Easterfield- you've been a lifeline. Particular thanks go to; Kris, Roisin, Sophie, and Michaela. Above all this experience has given me wonderful, supportive and empowering friendships. Thank you to the members of the "neuro-plastic-ity lab"; for never taking yourselves, me or this wacky world of academia to seriously. To Morgan, Niamh, Catherine and Bethan, thank you for all your assistance in this project. A special thank you should also go to Emily, Helen, and Matilda. You're a weird glorious bunch and have helped to constantly ground me (namely through insults), remind me of what is important, and encourage a belief that I may actually be capable of this. To Ruth, I miss you – thank you for everything you've done for me and everything you are.

Thank you to my supervisor Carolyn Wilshire. While you do possess a ridiculously enormous brain, what I have appreciated most is your understanding and compassionate guidance, never judging and always encouraging me. A special thank you must also go to Tirta Susilo for all his support.

What would an acknowledgment section be without a reference to the people that truly ruin your life – parents. Thank you everything. From the excessive number of encyclopaedias and TY beanie babies, to every school trip you were able to send me on and all the support whenever my flat floods or I needed bacon, I know I have been incredibly privileged, and I thank you. I love you.

# Contents

# List of Appendicies

## List of Figures

# List of Tables

"The more you read, the more you'll know.

The more that you learn, the more places you'll go."

Dr. Seuss.

# Chapter 1 : Literature review

Experienced readers need to be robust to variations in the appearance of words. They need to be to able identify a word irrespective of the font, (sometimes questionable) handwriting, and subtle spelling errors. Indeed, experienced readers can also deocde typo-laden texts, sometimes without even being aware of such errors (Gordon, Plummer, & Choi, 2013; Pynte, Kennedy, & Ducrot, 2004). For example, did you notice the typo in "decode" in the previous sentence? Unless you were explicitly looking for errors, it would have been easy to miss. This example illustrates how rapid and robust the word recognition process is in skilled adult readers: they can recognise a word at a single glance, without breaking it down into the individual letters. This ability is essential for the efficient reading of large passages or texts (such a thesis).

But in individuals with developmental dyslexia, reading is not quite so robust. Dyslexia can be defined as a significant reading delay or difficulty in the absence of a general intelligence deficit, which cannot be explained by socio-economic factors (Shaywitz & Shaywitz, 2005; Siegel, 1988; World Health Organisation, 1992). In New Zealand, it is estimated that approximately 70,000 children and adolescents are reading at least two years below what would be expected for their chronological age (DFNZ, 2017). Higher than normal truancy rates, poor academic performance, increased likelihood of depression and anxiety, and even criminal offending are all wider effects associated with dyslexia (Kirk & Reid, 2001; Shaywitz & Shaywitz, 2005). Given that many dyslexic readers experience enduring symptoms throughout their lifetime, it is paramount that we work to understand not just how reading happens, but also why some people show pervasive difficulties. We need to ensure that effective intervention methods are devised and continuously updated as scientific research and theoretical claims evolve. The current research aims to further this effort. Specifically, we wish to advance our understanding of the cognitive processes that contribute to reading – and reading difficulties – in adolescent readers, and to develop new reading interventions based on this understanding. This review will open by describing the key features of dyslexia. Following this, we summarise current theoretical models of reading development, finishing with an overview of theories specifically relating to developmental reading delay.

**Features of Dyslexia**

Research distinguishes between two major forms of dyslexia: acquired and development dyslexia. Acquired dyslexia refers to a reading impairment that becomes apparent after a neurological event, such as a stroke, traumatic brain injury, or tumour-surgery. Developmental dyslexia on the other hand refers to a reading impairment with no discernible neurological cause and is thus considered to have been present from the early stages of development, even though it may not be detected until adolescence, or even adulthood. In developmental dyslexia, the focus of the current study, individuals who meet the criteria generally struggle to read and understand words and/or larger chunks of text and also commonly exhibit spelling difficulties (for a review see Baddeley, Logie, & Ellis, 1988). Further, large group studies have found that these individuals demonstrate difficulties in a number of related cognitive domains. For example, they have been found to perform poorly on rapid automatized naming tasks, which involve speeded naming of pictures and words (Denckla & Rudel, 1976; Zoccolotti et al., 2013). In addition, they may exhibit problems with low-level visual processing tasks, including tasks that require fine-grained discrimination between visual patterns based on their spatial frequency or orientation (Georgiou, Papadopoulos, Zarouna, & Parrila, 2012; Lovegrove, 1993), and tasks that involve identifying the direction of motion in complex moving arrays (Conlon, Lovegrove, Barker, & Chekaluk, 2001; Cornelissen, Richardson, Mason, Fowler, & Stein, 1995; Wilmer, Richardson, Chen, & Stein, 2004). Finally, they have been found to perform poorly on phonemic awareness tasks, which involve separating auditory words into their component phonemes (e.g. *top* -> /t/, /ɒ/, /p/), or creating new words by blending phonemes together (e.g. *s + top* -> “stop”; Lallier, Donnadieu, & Valdois, 2013; Muter, Hulme, Snowling, & Taylor, 1998; Tallal, 1980). Since most of these studies have treated dyslexic readers as a homogenous group, they cannot determine whether every case experiences all these documented difficulties, or whether some cases display a more selective profile.

Acquired dyslexia, while not the focus of the current study, has been instrumental in modelling the cognitive processes of reading, so it is worth reviewing very briefly here. Cases of acquired dyslexia can be broadly divided into “peripheral” dyslexias, in which low-level visual identification and discrimination processes are impaired, and “central” dyslexias, in which the key impairment affects the processes involved in mapping between the visual word form, and its sound and/or meaning (Hinton and Shallice, 1991). Considering the central dyslexias, two dissociating profiles have been observed, and have been fundamental for the development of reading models: they are *phonological dyslexia* and *surface dyslexia*. Acquired phonological dyslexia is marked by difficulty reading novel words, often tested using a list of nonwords which readers are guaranteed to have no prior knowledge of (that is made-up words, e.g. *vun*). Individuals with this profile commonly "lexicalise" these nonwords, meaning they read out a completely unfamiliar made-up word as a familiar real word (e.g. *vun* > "van"; Funnell, 1983; Marshall & Newcombe, 1973). Acquired surface dyslexia on the other hand, is marked by a difficulty with irregular word reading (that is words that do not obey conventional sight-sound correspondence rules that dictate orthographic-phoneme mapping, e.g. *blood*). Individuals with this profile commonly "regularise" these words, meaning they pronounce the word according to the rules for that spelling (e.g. *blood* > "blude"; Patterson & Marcel, 1977; Tree & Kay, 2006). These individuals are also extremely slow readers (Shallice, Warrington, & McCarthy, 1983). Finally, those with surface dyslexia often have difficulty comprehending the meaning of written words; for example, when presented with a written word pair and a probe word, they may fail to correctly identify the target word in the pair that matches the probe (Funnell, 1983). The contrasting patterns of impairment exhibited in acquired phonological and surface dyslexia suggest that the process of reading can be affected in different ways in different individuals. A plausible model of reading must explain how novel or nonword reading may be impaired, while irregular word reading remains intact. Conversely, it must also explain how irregular word reading, reading speed and comprehension impairments can arise while novel word reading remains intact.

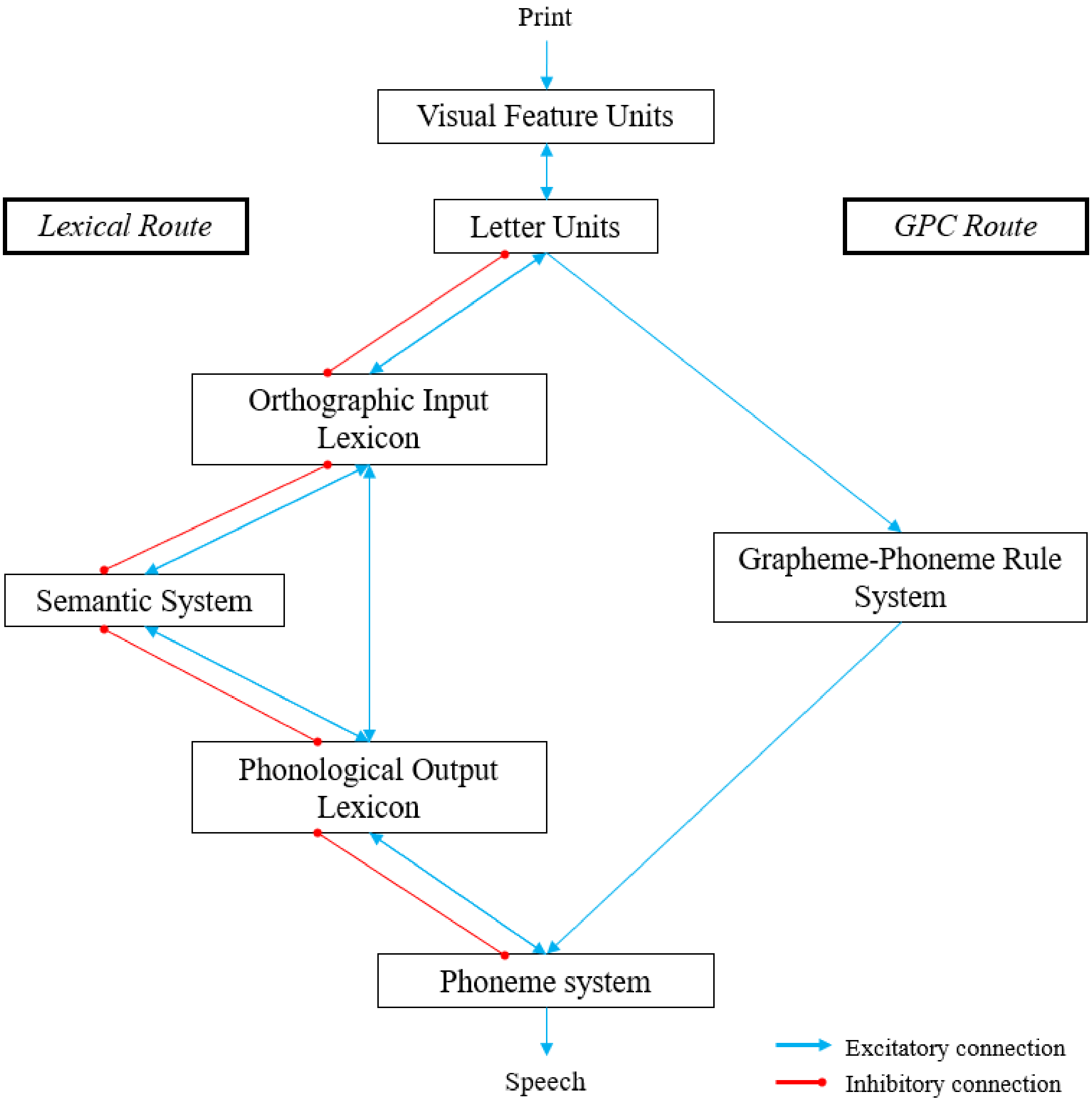
**Theories of Reading**

Ask someone "what is reading?" and they might provide a general answer describing the subjective experience of reading: how it feels to recognise a word, or to comprehend a piece of text, some might even wax poetic and talk about the potential for escapism. These definitions are not wrong, they're just not enough. Changing the question to "*how* do we read?" adds depth. Reading involves mapping a visual stimulus onto an internal representation of the meaning and/or sound form of the word. There are a number of models describing how this mapping takes place. These models vary widely in their features, their scope and the phenomena they aim to explain. With respect to their features, models can be roughly divided into two classes: those that propose reading is accomplished via a unitary process that applies to all words ("single route" models, see Seidenberg & McClelland, 1989), and those that propose that different processes are engaged to differing extents depending upon the word, the reader, and/or the context ("multiple route" or "dual route" models). Below we outline some influential multiple route models.

The most influential contemporary model of reading is the Dual Route Cascaded (DRC) model by (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). The model, illustrated in Figure 1.1, is organised into specialised layers (represented in the figure by rectangle boxes), and these layers are made up of units that represent the smallest possible features within any given layer. Units within a layer are activated when the external stimulus (i.e. written word) matches the internal representations these units encode. Within the *visual feature units* layer, units encode the most basic components of a letter: lines and vertices. As such, one (set of) unit(s) will encode for and be activated by the presence of long vertical lines (e.g. "|"), and another (set) will encode for short horizontal lines (e.g. "–"). Therefore, the letter 'H' would activate the units for the "|" and "–" lines. Unit activation within one-layer spreads to the corresponding units in the next layer in a cascading manner. Thus, the units "|" and "–" at the visual feature level activate units in the *letter units* layer that contain these elements (e.g. "H", "T" and "L"), and the greater the match, the higher the activation of the letter unit. This process continues to propagate down until ultimately the summation of activation across layers gives way to the selection of a phonemic representation.

Once the letters have been identified, there are two processes – or routes – by which information about the word can be recovered. The *phonological route* applies rules about how key letters or letter groups ("graphemes") are commonly pronounced ("phonemes"). These *grapheme-phoneme conversion (GPC) rules* are stored within the *grapheme-phoneme rule system*. To convert a letter string into phonemes, this route operates over individual letter pairs, beginning with the leftmost pair of letters in the word (e.g. "sh" in *sheep*, or "st" in *steep*). There are three distinct processes involved in the conversion. First, the individual graphemes within the word are identified. Second, the GPC rule system will attempt to convert the letter-pair into a phoneme. If the pair has a corresponding phoneme then this is converted (e.g. *sh* > /š/), and the process proceeds through the remaining letter. However, if there is no phoneme for the letter pair, then the process reverts to identifying the phonemic correspondence for the first letter, then the second, and so on (e.g. *s* > /s/, then *t* > /t/). The model 'cycles' through each consecutive grapheme in the word, until all graphemes have been processed and their corresponding phonemes have been identified. Third and finally, the phonemes are blended in the correct order to create a phonological representation of the entire word. Because this route can operate on words that the reader has never encountered before, it is fundamental in the early stages of reading development. However, because this route relies on processing each letter one-by-one, it is slow and laboured, and the more graphemes a word has, the more time the process takes.

Figure 1.1 The Dual-Route Cascaded model of reading




Conversely, the *lexical route* is designed for the rapid recognition of familiar words. In this route, highly familiar letter sequences activate the corresponding orthographic representations via rapid processes that map all the letters in a word against the internal representations simultaneously. Once the appropriate unit within the *orthographic input lexicon* has been identified, the word's phonological representation within the *phonological output lexicon* is activated. At the same time, the word's semantic representation (meaning) in the *semantic system* also becomes activated, which in turn strengthens the activation of the corresponding phonological representation. Once a single phonological representation has been identified, this unit then activates its component phonemes within the *phoneme system*. Within this lexical route, the connections between the layers are bidirectional. That is, activation can spread in a feedforward manner, from the letter unit layer to the orthographic lexicon and so on, but it can also feedback in the reverse direction, for example, from the orthographic lexicon to the letter unit layer. These feedback connections reinforce patterns at each level that correspond to a unit at the next level, thus facilitating top-down processing during reading. Since the lexical route processes the word's letters simultaneously, words can be rapidly identified, as long as they are highly familiar. When a word is new to the reader, and there is no corresponding orthographic representation, this route cannot operate.

In a computational simulation, the DRC model was found to be highly accurate at reading aloud words of up to eight letters and was also capable of mimicking several patterns that experienced readers show, such as frequency effects on reading latencies (Coltheart et al., 2001). "Longer" naming times (measured in cycles) were found for low-frequency words, nonwords, and words with very few neighbours (that is, words that differ from the target word by only one letter e.g. *cart* > *dart, card, part* etc.). A simulation of the model's performance on a lexical decision task – a task that requires the identification, but not production, of a word –also reproduced patterns of accuracy and response times across different frequency words. However, one limitation of the simulated model was that it did not include the lexical semantic route (that is, the pathways between semantic representations and their corresponding orthographic and phonological representations). One small-scale simulation, which modelled the facilitatory effect of colour information on the reading of colour words (e.g. *red, blue, green*), found evidence supporting the model's claims (Coltheart, Woollams, Kinoshita, & Perry, 1999), however this simulation was very limited in its scope.

Simulations of the DRC model are also able to replicate a number of key features observed in acquired surface and phonological dyslexia. To mimic the patterns affiliated with phonological profiles of dyslexia, the researchers increased the amount of fixed time (i.e. number of cycles) between the mapping of each letter and weakened activation in the phoneme system. As such, the GPC rules would be more difficult to apply because the creation of letter pairs is delayed. Meanwhile, the lexical route is working optimally, mapping the input against lexical representations in the orthographic lexicon. By "lesioning" the model's phonological route, Coltheart and colleagues (2001) were able to reproduce the key features of phonological dyslexia: the model performed disproportionately poorly on nonwords and produced a significant number of lexicalisation errors on these stimuli (Coltheart et al., 2001). Conversely, to mimic surface dyslexia, researchers weakened the connections between letter units and their corresponding representations in the *orthographic input lexicon*. As such, it takes longer for activation to reach threshold in the lexical route, while the phonological route is fully functioning, processing words in sequential components. By "lesioning" the lexical route, researchers were able to simulate the key features of acquired surface dyslexia: the model performed disproportionately poorly on irregular words and produced significant numbers of regularisation errors on these words.

Despite replicating some key patterns of acquired dyslexia, the DRC model has difficulty explaining other observations. For example, individuals with acquired phonological dyslexia are disproportionately poor at reading grammatical function words (i.e. *if* or *the*; Patterson, 1978; Shallice & Warrington, 1975). These words are short and lack any concrete meaning, however they are highly frequent and therefore, in theory, they should be processed via the lexical non-semantic route. If the individual has a selective impairment to the phonological route, then they should be able to read these words accurately via the lexical route. Another pattern of impairment that is difficult to describe within the DRC model is deep dyslexia, marked by substantial orthographic-semantic processing impairments, as well as phonological impairments. Shallice and Coughlan (1980) offer an account of one such case that performed poorly on picture-word matching, synonym matching, and semantic categorisation tasks as well as nonword reading, rhyme matching, and phoneme matching tasks. Other studies have found complex patterns of substitution with lexicalisation errors, visual errors (e.g. *skate* > "scale"), derivational errors (e.g. *scientist* > "science") and semantic errors (e.g. *broken* > "snapped"; Marshall & Newcombe, 1973; Patterson & Marcel, 1977; Ruiz, Ansaldo, & Lecours, 1994; Shallice & Coughlan, 1980). According to the DRC model, there is no single defining deficit, instead the impairment would arise from damage to both the phonological route, *and* the lexical route.

Second, the model does not offer an explanation as to how an individual *learns* to read. The model was designed to account for the behaviour of proficient adult readers. As such, its "functional architecture" is pre-determined: the researcher specifies which units are included in each layer (e.g. words, phonemes), and explicitly creates the connections between adjacent layers. The model does not offer an explanation as to how reading actually *develops* in children. This is important, particularly for the current research, given that dyslexia is often conceived of as a pattern of developmental delay.

A recent adaptation of the DRC model, the Self-Teaching Dual-Route Cascaded model (Pritchard, Coltheart, Marinus, & Castles, 2018), has attempted to address this critique by modelling how reading changes across developmental stages. This model uses concepts from the self-teaching hypothesis of reading (Jorm & Share, 1983; Share, 1995), which outlines the stages involved in the acquisition of a new word. The updated model proposes that, upon seeing a word for the very first time, it is assessed for familiarity. If the word does not reach the minimum threshold of familiarity, the GPC rules are applied, while the development of a new representational node is also initiated. Crucially, the model specifies that a threshold must be reached before a word is processed via the lexical route. In computational simulations of the model this threshold is implemented as a minimum level of activation an orthographic node must reach for lexically-based reading to continue. However, if no nodes reach this threshold, then the model determines that the word is novel, and must be "learned". Learning a word commences with the creation of a new representational node in the orthographic lexicon. Bidirectional connections are then established between this representation and the adjacent layers specified in the original DRC model – connections that are gradually strengthened with repeated exposure to the word. The simulated model also implements a highly simplified semantic system.

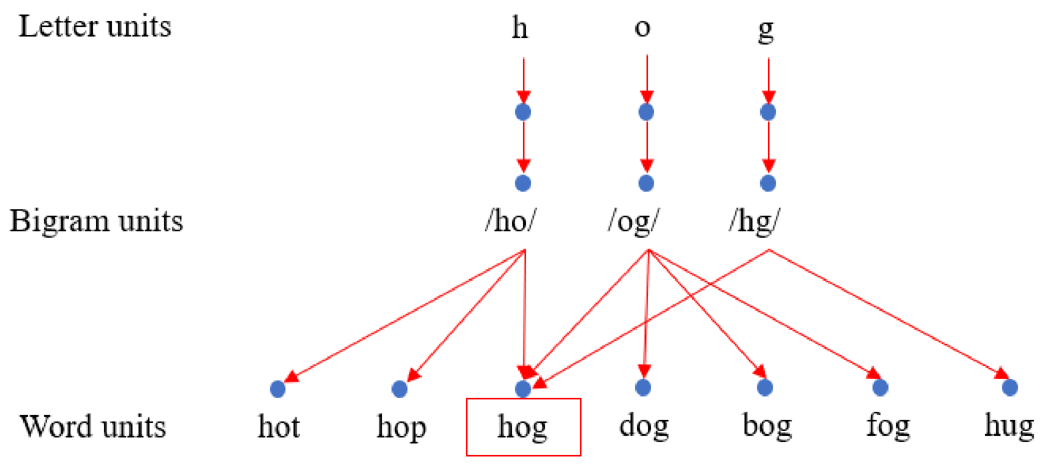
One problem with the simulated self-teaching DRC model was that it could not reliably discriminate between a familiar word and a completely novel word. By manually adjusting the various unit thresholds in the model, the researchers were able to improve the model's lexical decision performance. However, these adjustments had consequences for other aspects of reading performance. Furthermore, the limited representation of semantic processing in the simulation meant that the theoretical role of such mechanisms in reading in general, and the lexical route more specifically, could not be rigorously tested.

**Modelling the Visual Aspects of Reading.**

***The SERIOL model.*** One dual-route model that speaks more directly to the processes involved in whole-word level reading is the SERIOL model (Whitney, 2001, 2008; Whitney & Cornelissen, 2008; Whitney & Marton, 2013). In SERIOL, words are recognised via a feedforward process that first attends to and identifies the individual letters within the stimulus word (e.g. *hog* > "h", "o", "g"). Once activation has accumulated in the *letter units*,

units in the *bigram layer* then become activated. These consist of adjacent and non-adjacent letter pairs (the word *hog* could be broken down to “ho” + “og” + “hg”). This activation then begins to spread to the *lexical (word) layer* via weighted connections (see Figure 1.2). Multiple connections and word units are stimulated but ultimately the word unit with the highest activation is selected. A longer word like *Hogwarts* will undergo the same procedure (“ho” + “og” … “hs”/).

Figure 1.2 A basic schematic of the SERIOL model.

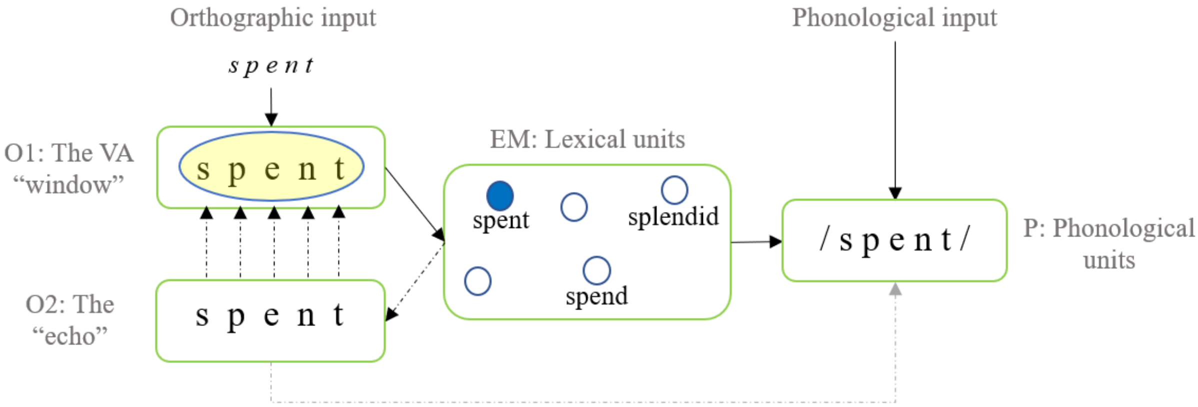



A fundamental part of this model is that, the individual letters of the word are attended to in serial order. As each is processed, the corresponding representation starts to become activated. The serial order of letters is coded using a form of time stamping, in which the activation of each successive letter unit is slightly staggered in time from left to right. So, activation of the leftmost letter unit in the word commences slightly in advance of that for the next letter unit in the series, and so forth until the end of the word is reached. This process effectively preserves the serial order information provided in the written word in an abstract form. In SERIOL, a shorter word will take less time to process at the letter level than a longer one. However, this advantage does not directly translate into shorter overall reading latencies, because the shorter word is likely to have more neighbours (*hog* has 13 for example), and therefore takes longer to become sufficiently activated above its competitors at the lexical level.

***The Multi-trace Memory model.*** Another visual model of reading – the Multi-trace Memory model of polysyllabic reading (henceforth known as the Multi-Trace Model; Ans, Carbonnel, & Valdois, 1998) – emphasises the importance of visual attention during word recognition. This model proposes that there is a Visual Attentional (VA) window designed specifically to facilitate parallel letter processing. This window is capable of operating in two modes: an *analytic* mode where the window is narrow and focussed on one or two letters; and a *global* mode where the window is broad, and focus is spread across the whole word. In other words, the VA window is akin to a flashlight where the diameter of range can be adjusted according to demand. If all the lights in your house go out you might want to adjust the flashlight so that it has a wide breadth, however in doing so the light rays will be spread diffusely, so your eyes will not pick up fine-grained details. Conversely, if you're trying to fix the kitchen sink in daylight, you might want to adjust the flashlight so that it is very narrow, and light is focused onto a specific area, making the details much easier to see. The flashlight itself dictates the way information is processed. An adjustable VA window allows readers to switch between these two modes according to their experience and knowledge regarding each word. When the window is set to analytic mode, part-based processing will ensue, but when this window is set to global mode, the word can be processed as a whole unit. In this model, experienced readers will commonly utilise the global mode, which allows parallel letter processing, whereas inexperienced readers frequently need to fall back on the analytic mode.

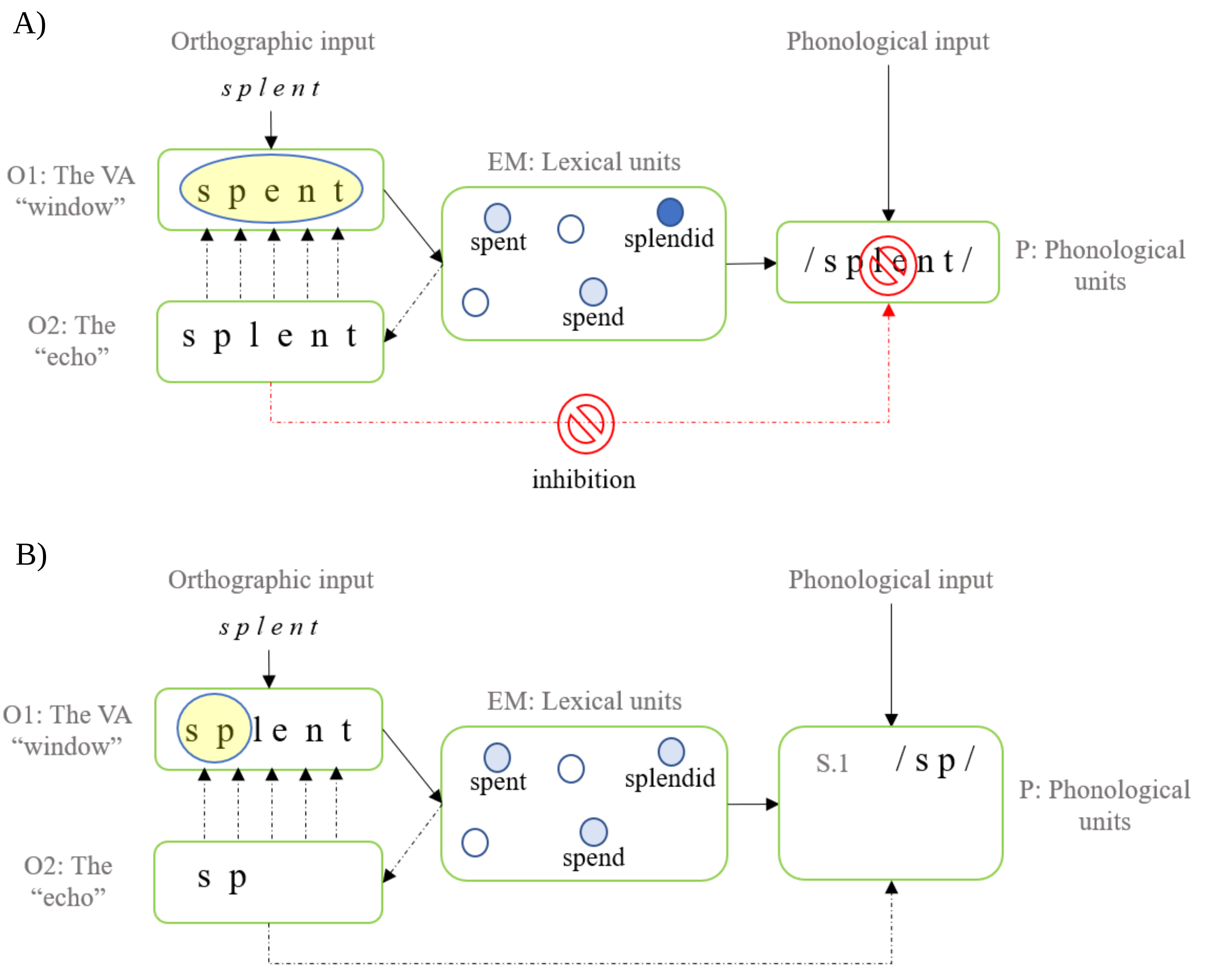
As illustrated in Figure 1.3, the model contains four primary layers of units: there are two layers of orthographic units, a layer of lexical units (referred to as the *episodic memory layer*) and a *phonological layer*. The first of orthographic layers, the *O1*, is effectively the VA window. In this layer, the window will "capture" the orthographic input, activating the corresponding letter units within the layer. This layer is set to global mode by default, allowing all letters from the orthographic input to be processed simultaneously. The pattern of activation at this layer is then mapped onto lexical representations in the episodic memory (EM) layer that partially or fully match the input. The lexical unit that is most highly activated is then selected and that unit in turn activates the corresponding units in the phonological layer (P). Crucially, the model also performs a "check" to ensure the lexical unit selected accurately matches the orthographic input: the lexical unit activates its corresponding orthographic representation in the second orthographic layer, the *O2*, thus creating an "echo" that is checked against the representation in O1. If they match, processing continues, and the phonological form of the word is then fully activated. For example, when an experienced reader encounters a familiar word like *spent*, they will be able to efficiently activate the correct lexical unit, and the echo will confirm that it is a match to the input.

Figure 1.3 Familiar word reading in the Multi-trace Memory model.




However, if the reader encounters a novel word (e.g. *splent*), things go differently (Panel A in Figure 1.4). This novel word has no corresponding lexical representation, but it will partially activate the representations of orthographically similar words within the lexical layer EM (e.g. spent, spend, splendid). In this instance, the most activated unit will be selected (e.g. splendid) and an echo will be created in the O2 layer. This echo is then matched against the original orthographic input representation in O1. The mismatch between *splendid* and *splent* will be detected, and the phonological output will then be inhibited. The model then re-processes the word, this time with the O1 layer set to analytic mode, capturing only one or two letters at a time, building the phonological output one syllable at a time (Panel B in Figure 1.4).

Figure 1.4 Novel (non)word reading in the Multi-trace Memory model; the initial global processing (A), and subsequent analytic processing (B).




In a simulation of this model, the mode of processing (global or analytic) crucially affects reading latency: stimuli processed via the global mode are read more rapidly than those processed via the analytic mode. Further, the two different modes generate different reading latency patterns as word length increases. In the global mode, there is little increase in reading latency as word length increases, while in the analytic model, latency increases sharply as word length increases. These results are consistent with the model's assertion that the global mode is the gate-keeper of parallel letter processing. The researchers also attempted to simulate the poor irregular word reading and regularisation errors seen in acquired surface dyslexia. To simulate these effects, the model was "forced" to operate in analytic mode. In this simulation, irregular word reading accuracy decreased, and regularisation errors were produced. Further, and consistent with observation of acquired surface dyslexia, the model demonstrated a frequency by regularity interaction, whereby more errors were produced on infrequent irregular words than on frequent ones.

This emphasis on the visual processes and how they contribute to whole-word reading may be beneficial for considering how developmental dyslexia arises. To address this possibility, we will now turn our discussion to developmental dyslexia. We will look at heterogeneity in developmental cases of dyslexia and then discuss some of the theories that have been proposed to explain developmental dyslexia.

**Developmental Dyslexia: Some Key Observations**

In our earlier discussion of developmental dyslexia, we briefly mentioned that these individuals have been found to exhibit problems in a number of non-reading tasks. However, as we mentioned, these studies treated all dyslexic readers as a homogenous group. There is mounting evidence to suggest considerable individual variability (Castles & Coltheart, 1993; Kohnen et al., 2018; Manis, Seidenberg, Doi, McBride-Chang, & Petersen, 1996; Wybrow & Hanley, 2015; Zoubrinetzky, Bielle, & Valdois, 2014). For example, Castles and Coltheart (1993) examined oral reading of regular words, irregular words and nonwords in a sample of 56 dyslexic readers aged between eight and 15 years-old. They found that a large proportion of participants exhibited difficulties on both irregular words and nonwords. However, a significant number showed more selective difficulties: that is, they scored below the normal range for irregular word reading but showed normal nonword reading, or conversely, below normal nonword reading but normal irregular word reading. Castles and Coltheart concluded that children with dyslexia vary along two dimensions: phonological, or part-based reading ability, as measured by their nonword reading accuracy, and whole-word reading ability, as measured by their irregular word reading accuracy. Most dyslexic readers scored poorly on both dimensions; however, a proportion had a selective difficulty with only one. Interpreting the findings within the context of the DRC model, the researchers suggested that those demonstrating a selective difficulty with nonword reading might have problems utilising the phonological route, while those demonstrating selective difficultly with irregular word reading might have difficulty utilising the lexical route.

Case studies have also reported highly selective patterns of difficulty in individuals with developmental dyslexia that broadly align with the two distinct profiles observed in Castles and Coltheart’s (1993) group study. Specifically, some of these individuals perform normally when reading familiar words but extremely poorly on novel words or nonwords, commonly reading them as real words (lexicalisations, e.g. *boril* > "boil"; Campbell & Butterworth, 1985; Temple & Marshall, 1983). Some of these cases also perform poorly on phonemic awareness tasks, tasks that require them to separate words into individual phonemes or create new words by changing these phonemes. For example, one case who performed disproportionately poorly on nonword reading was also found to be impaired on such tasks, while another case who performed well on nonword reading performed in the normal range on phonemic awareness tasks (Valdois et al., 2003). A number of similar cases have also shown selective difficulty recognising pseudo-homophones, nonwords that sound like real words (e.g. *kote* or "koat"; Coltheart & Leahy, 1996; Nickels, Biedermann, Coltheart, Saunders, & Tree, 2008).

In contrast, some cases perform normally on nonword reading and phonological awareness tasks but poorly on irregular word reading tasks, often sounding them out as they are spelt (regularisations, e.g. *bouquet* > "booket"; Castles & Coltheart, 1996; Coltheart, Masterson, Byng, Prior, & Riddoch, 1983; Wybrow & Hanley, 2015). Such cases often read words extremely slowly, especially long words (Baddeley et al., 1988; De Luca et al., 2008; Judica, De Luca, Spinelli, & Zoccolotti, 2002; Martens & de Jong, 2006; van den Boer, de Jong, & Haentjens-van Meeteren, 2013; Wolf, Bally, & Morris, 1986; Zoccolotti et al., 2005). They may also have difficulty extracting meaning from written words; for example, when asked to define written words, use them in a sentence, or generate a synonym (Castles & Coltheart, 1993; Nation & Snowling, 1998).

In sum, case studies have shown that the variation in reading patterns seen in group studies of developmental dyslexia may be accompanied by quite specific cognitive deficits in other domains. These observations suggest that the two different profiles may be a consequence of different underling cognitive difficulties: a phonological profile marked by poor performance on nonword reading and phonemic awareness tasks; and a whole-word profile marked by poor performance on irregular word reading, reading speed and comprehension tasks.

**Single Deficit Theories.** It is important to remember that the patterns described above characterise only a portion of individuals with developmental dyslexia. Many individuals don't fit into either one of these categories, but rather appear to have a mix of difficulties with *both* phonological and whole word aspects of reading (Manis et al., 1996; Temple & Marshall, 1983). Indeed, some researchers have proposed that the diversity seen within this population may result from a single underlying deficit that manifests as different patterns of reading errors depending on the severity. In one such proposal, Griffiths and Snowling (2002) draw on recent simulations of parallel distributed processing models of reading, which are capable of generating qualitatively different patterns of errors depending on the severity of the "lesion" introduced (Seidenberg & McClelland, 1989). Their hypothesis, the *Phonological Severity Hypothesis*, proposes that all cases of developmental dyslexia are caused by weakly coded phonological representations (Griffiths & Snowling, 2002; Snowling, 1995). When this phonological deficit is moderate, reading difficulties are largely confined to nonwords, however when the deficit is severe, it impacts upon real words.

To explain cases showing selectively poor irregular word reading but intact nonword reading, the authors propose two other factors may be at play. The first is the degree of experience an individual has with the written word; those with less overall exposure to written words will tend to read irregular words more poorly than would be predicted based on their nonword reading. In support of this proposal, they showed that a measure of prior reading exposure – but not phonemic awareness – uniquely predicted irregular word reading accuracy. The second factor is the individual's ability to draw on other cognitive resources to compensate for their problem with nonwords. For example, those with strong verbal short-term memory skills can retain more phonological information online, which can help support nonword reading.

One strength of the phonological severity hypothesis is the recognition that the relationship between error patterns and their underlying causal mechanisms may not always be transparent. This hypothesis also acknowledges the importance of reading experience, not just in the development of overall reading proficiency, but specifically to the development of whole-word processes and reading skills (Griffiths & Snowling, 2002). The hypothesis generates novel, testable hypotheses about this relationship between reading experience and irregular word reading. Indeed, they found reading experience, as estimated using author and title name recognition tasks, to be a unique predictor of irregular word reading, whereas measures of phonemic awareness and verbal short-term memory were not.

Finally, the phonological deficit hypothesis may be important when considering reading intervention. Multiple deficit models, such as the DRC, would suggest that interventions for dyslexic readers with opposing profiles should have different theoretical cores. However, single deficit models and hypotheses, such as this one, would suggest that interventions for these opposing profiles do not need different theoretical cores because they target shared underlying causes. Given this possibility it is important to establish whether interventions should align with single or multiple deficit theories We return to this issue later in Chapters 3 and 4.

Nevertheless, it is not entirely clear as to where the model locates the proposed phonological deficit. Griffiths and Snowling (2002) interchangeably refer to "weak phonological representations" and weak mappings between orthographic and phonological representations. Second, the model's explanation for disproportionately poor irregular word reading is complex: not only do these cases share the "weak phonological representations" of all other dyslexia cases, but they also need to possess at least one further attribute (either unusually limited reading experience and/or unusually strong verbal short-term memory skills). Indeed, it could be argued that this model is not a single deficit hypothesis, given that it seems to propose that individuals may vary in at least two cognitive domains (strength of their phonological representations and their verbal short-term memory skills), as well as one experiential domain (print exposure). As such, the model could be considered a multiple-deficit account. Related to this, one further problem with the model is that the explanation of whole-word cases assumes that it is possible to have unusually poor phonological representations but at the same time unusually strong "verbal short-term memory", a combination that seems paradoxical.

**Multiple Deficit Theories.**

***The DRC Model.*** The DRC model seems to offer a good account for the "phonological" profile in developmental dyslexia (disproportionately poor nonword reading, often accompanied by poor phonemic awareness). It attributes this pattern to a selective difficulty within the phonological route. Given that the "phonological" pattern of reading impairment is often accompanied by problems with auditory phonemic awareness tasks, the difficulty could in principle occur at one of two stages within this route: within the grapheme-phoneme conversion process itself, or within the processes that blend the identified phonemes into a phonological representation of the word. According to the DRC model, nonwords are lexicalised because the activation from the lexical route to the phoneme system is faster than that from the phonological route.

The DRC model has several broad limitations when it comes to developmental dyslexia. First, the interactive nature of the various reading routes means that, in practice, we cannot be certain how much each reading route contributes to any particular reading attempt. Second, the model's lexical route focuses primarily on those processes that map between orthographic representations, and their corresponding meaning and phonological representations. It does not offer a detailed account as to how the reader arrives at the correct orthographic representation based on highly variable visual information, and often with just a single glance at the word. For example, the model assumes that in the lexical route, the letters of a word are processed in parallel, but clearly, accurate word recognition depends not just on which letters a word has, but what order they appear in. The model provides no information about how letter order is coded. Consequently, the model is ill-equipped to deal with the difficulties in rapid whole word-level recognition that can occur in developmental dyslexia.

Furthermore, the whole-word profile in developmental dyslexia is difficult to explain within the DRC model. Although this profile shares some similarities with acquired surface dyslexia – for example, in both disorders irregular word reading is disproportionately poor – there are also some important differences. On the one hand, individuals with acquired surface dyslexia commonly demonstrate accompanying impairments in semantic processing (Woollams, Ralph, Plaut, and Patterson (2007). On the other hand, individuals with the whole word subtype of developmental dyslexia are more variable; some demonstrate normal levels of comprehension (Castles & Coltheart, 1996; Coltheart et al., 1983; Zoccolotti et al., 1999), however some do not (Castles & Coltheart, 1993; Nation & Snowling, 1998). Whole-word cases also commonly perform poorly on nonverbal visual tasks that involve copying complex geometric designs and object patterns, recalling the position of objects in a scene from a previous image, and letter identification (Rowse, 2005; Valdois et al., 2003). Also, while these individuals read nonwords relatively accurately, their response latencies to these stimuli are abnormally and disproportionately long, a finding which would appear inconsistent with a difficulty restricted to the DRC's lexical route (Valdois et al., 2003; Zabell & Everatt, 2002; Zoccolotti et al., 1999).

***The SERIOL model.*** The authors of the SERIOL model suggest two underlying deficits (Whitney & Cornelissen, 2005). First, similar to the DRC, phonological deficits where there are difficulties associating graphemes to the corresponding phonemes, result in weak GPC representations. Second, visual deficits where difficulty focusing visual attention during development results in poor or incorrect grapheme-phoneme associations. The inability to attend to fine-details in written stimuli (i.e. letters) will affect the ability to activate the correct phonological letter representations in serial. The key here is that each letter must, in serial, be the focus of attention. The authors of this model argue that this ability is impaired in some cases of developmental dyslexia. That is, the serial order of letters in a word will not be maintained, and subsequent whole-word mapping at the lexical layer will be affected. Although this proposal would allow for the differential patterns of difficulty we see in dyslexia, the authors do not offer any specific details as to what these “visual” impairments are or how they may be impaired. Instead, their proposal is a summary of evidence that all points towards a visual impairment in dyslexia, but there is little in the way of novel testable predictions.

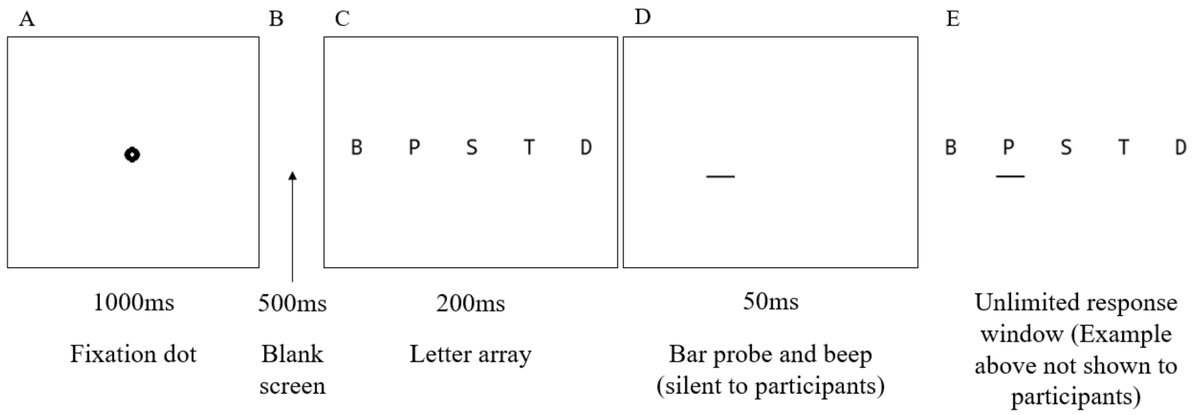
***The Multi-Trace Model and the Visual Attention Span Hypothesis.*** Similar to the DRC model, the Multi-Trace Memory model Figure 1.3maintains the notion that there are two independent processes involved in reading – one which capitalises on stored knowledge of words, and the other which capitalises on common correspondences between orthography and phonology (Ans et al., 1998). Crucially, it also postulates the existence of two different visual attention modes that can be used during reading: a global and an analytic mode. In the global mode, the reader’s VA window extends over the whole-word, enabling the letters of a word to be processed in parallel. Conversely, in the analytic mode, the window narrows down to focus attention on smaller units one-at-a-time (Bosse et al., 2007). The choice of attentional mode impacts upon whether the word is read primarily using whole-word strategies (that is, activation of the word’s lexical representation) or part-based strategies (that is, utilising more general information about letter-sound correspondences).

Within the Multi-trace model, developmental cases of phonological and whole-word dyslexia are seen as reflecting distinctly different cognitive impairments (Ans et al., 1998). However, the explanation for the “whole-word” profile is somewhat different from that offered within the DRC model. This pattern of impairment arises when the VA window cannot operate fully in global mode, and so must revert to analytic mode. Consequently, instead of processing all the letters of a word in parallel, the reader is forced to revert to a part-based processing strategy. This proposal is referred to as the *Visual Attention Span Deficit hypothesis* (Bosse et al., 2007).

The VA span deficit hypothesis (Bosse et al., 2007) has a lot of potential as an explanation of selective reading difficulties. As a theory it goes well beyond the simple box-and-arrow approach, detailing how the components are organised and operate with respect to one another, clearly outlining what mechanism(s) are affected in impaired readers. The hypothesis is also nested within a computational model (the Multi-Trace Memory model; Ans et al., 1998), meaning that the predictions can and have been tested via computer simulation and then compared against actual participant performance.

Support for the VA span hypothesis primarily comes from studies using a variation of the visual partial report task. In this task, an array of letters is presented for a very brief time, followed by a bar probe which appears under the position of one of the previously shown letters (see Figure 1.5). Participants must verbally identify the probed letter. Bosse et al. (2007) found scores in this task were positively associated with a general measure of reading proficiency (reading age). Further, partial report scores reliably predicted irregular word reading accuracy, whereas scores on the phonemic awareness tasks did not. These findings support the proposal that visual attentional processes are critical for efficient reading and contribute to reading proficiency independent of phonological processing skills.

Figure 1.5 The visual partial report tasks used to determine VA span score.




Empirical evidence suggests that impaired VA span might be underpinned by a range of different kinds of impairments. Individuals who score poorly on VA span tasks have also been found to exhibit impairments in: a) a rapid serial visual presentation choice task, (Shih & Sperling, 2002); b) the number of visual elements processed per second (Bogon et al., 2014; Stenneken et al., 2011); and c) target orientation detection (Joo, White, Strodtman, & Yeatman, 2018; Moores, Cassim, & Talcott, 2011; Moores, Tsouknida, & Romani, 2015). Although, it is not yet clear exactly how VA span impairments arise, and while more research addressing this issue is needed, the current research will focus on whether the broader concept of VA span is applicable to reading.

**Summary.** There is mounting evidence that development dyslexia is a heterogeneous condition. Many individuals have particular difficulty on tasks that require phonological or part-based processing of words, suggesting an underlying difficulty with phonological processing. However, not all of those with developmental dyslexia perform poorly on phonological tasks; some show other types of difficulties, such as abnormally slow reading times for familiar words – particularly long words – or difficulties with tasks involving visual analysis or visual memory. Theories differ as to how they explain this heterogeneity. Within conventional dual route models of reading, the "phonological" profile can be explained as a deficit to the phonological reading route, which utilises sight-sound correspondences (Coltheart et al., 2001; Pritchard et al., 2018). Conversely, the "whole-word" profile can be seen as a deficit to the lexical route, which operates by activating the word's representation within the mental lexicon. While this type of model provides a good account of the features in the phonological profile, it fails to capture some of the key features in the whole-word profile, such as difficulties with visual tasks. A more recent proposal is that the latter profile may reflect a restricted window of visual attention, which limits the number of letters the individual can simultaneously capture and map onto internal lexical representations (Bosse et al., 2007).

## Current research

The current research has two primary goals. The first is to investigate the heterogeneity of reading impairments and evaluate the VA span deficit hypothesis (Bosse et al., 2007) as a framework for understanding whole-word impairments in developmental dyslexia. The second is to develop a novel intervention targeting these impairments and evaluate its effectiveness.

In this study, we aim to explore individual variation in reading and reading-related skills in a relatively unselected adolescent sample. We will focus on investigating the role of whole-word processes in reading proficiency. Specifically, we will test whether commonly used measures of whole-word processes uniquely contribute to variance in overall reading proficiency, independent from measures of phonological processing. We will also explore the role of the VA window in whole-word processing, examining the relationship between scores on a VA span task and general measures of reading proficiency as well as more specific measures of reading-related processes.

In the second part of the current research we will develop a novel visual intervention that targeted whole-word reading. This intervention will use a series of techniques aimed at increasing readers' awareness of the common letter patterns within words. The effectiveness of this intervention will be directly compared to that of a more conventional phonological programme focusing on implicit GPC training. For this study, we will recruit adolescents demonstrating relatively pure cases of phonological and whole-word impairment, with both cases completing the novel visual intervention, as well as the more conventional GPC intervention.

# Chapter 2 : Investigating Heterogeneity in Reading Processes

As discussed in Chapter 1, there is considerable ongoing disagreement about the origins of developmental reading difficulties. According to one perspective, the primary underlying cause of developmental reading delay is a deficit in phonological processing – that is, a problem segmenting words into phonemes, and/or mapping between graphemes and phonemes (Griffiths & Snowling, 2002; Snowling, 1995). However, it has also been suggested that some reading-impaired individuals may have a distinctly different set of cognitive problems, ones that affect whole-word level processing, rather than phonological processing (Ans et al., 1998; Coltheart et al., 2001; Pritchard et al., 2018). If this is the case, then reading proficiency should be predicted, not only by traditional measures of phonological processing, but also by measures that index reading at a whole-word level. In this study, we examined several commonly used indices of phonological and whole-word processing skills and their relationship to overall reading proficiency in a relatively unselected adolescent sample. We also examined whether the patterns align with the VA span deficit hypothesis (Bosse et al., 2007).

## The word length effect

Proficient adult readers can process a text extremely rapidly, glancing at each word only momentarily before moving on to the next (for a review see Rayner, 2009). Achieving this level of proficiency involves transitioning from a phonological, part-based reading approach to a whole-word approach. One skill that is generally considered critical for successfully achieving this transition is the ability to process multiple letters simultaneously (Björnström, Hills, Hanif, & Barton, 2014; Weekes, 1997). One common index of this ability is the *Word Length Effect*. The word length effect is a measure of how reading latencies (or lexical decision times) change as a function of word length. If a reader is processing the letters of words in a serial manner, their reading latencies will increase stepwise with word length. Conversely, if they are processing multiple letters at once, then increasing the length of a word will have much less impact on their reading latencies. Indeed, it has been found that proficient adult and adolescent readers show little decrement in their word recognition times or reading latencies as word length increases, not at least for words up to seven letters long (Bijeljac-Babic, Millogo, Farioli, & Grainger, 2004; Juphard, Carbonnel, & Valdois, 2004; Kwok, Cuetos, Avdyli, & Ellis, 2017; Spinelli et al., 2005). These findings suggest that proficient readers are capable of processing multiple letters simultaneously.

Interestingly, this minimal effect of word length is observed only for familiar words presented in canonical fashion. Weekes (1997) measured reading latencies on low and high frequency words ranging three to six letters long, in an undifferentiated sample of adult readers. They found an interaction between word length and familiarity; word length affected reading latency for low frequency words, but not high frequency words. When proficient readers are presented with nonwords, stimuli we can be certain they are unfamiliar with, they exhibit a clearly discernible length effect (Kwok et al., 2017; Martens & de Jong, 2006; Weekes, 1997). A similar effect occurs when they are presented with words in which the letter order has been reversed, so that the letters are processed from right to left (e.g. *t u n o c o c*; Björnström et al., 2014). These results suggest that the parallel letter processes utilised for whole-word reading are distinct from those utilised for sub-lexical or part-based reading.

In contrast, in impaired readers, reading times often increase substantially with word length (reading latencies: De Luca et al., 2008; Marinus & de Jong, 2010; Spinelli et al., 2005; van den Boer et al., 2013; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003; Zoccolotti et al., 2005; Lexical decision times: Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Juphard et al., 2004; Martens & de Jong, 2006). Furthermore, the difference between length effect differences for real and nonwords are substantially greater in impaired readers, compared to unimpaired readers (Martens & de Jong, 2006).

However, much of this research treats impaired readers as a homogenous group and does not look at how word length effects may relate to or predict, phonological or other processes related to reading. There are a couple of notable exceptions. De Luca et al. (2008) and Spinelli et al. (2005) found that impaired readers (10 to 14 years-old) fell into one of two categories: those whose reading latencies increased markedly as a function of length (i.e. they display a word length effect), and those whose reading latencies increased only on extremely long words, similar to unimpaired readers. Interestingly, Spinelli et al. (2005) found that those with a marked word length effect showed a more severe pattern of reading difficulty overall, with poorer reading latencies across both real and nonwords. As the effect did not predict selective reading difficulty, length effects may be a general marker of poor reading and may not index any specific process associated with whole-word recognition.

In our study, we used this simple measure – the word length effect – as an index of the degree to which readers process words as a whole. We examined the relationship between this measure and commonly used measures of reading proficiency, such as reading age. We also examined whether this measure accounted for unique variance in reading proficiency, over and above what could be explained by commonly used measures of phonological processing skills, such as a measure of nonword reading accuracy. It is important to verify whether parallel letter processing is a unique predictor of whole-word, but not phonological, processing; without this investigation, arguments as to whether reading is achieved via a single or multi-route complex of cognitions cannot be settled.

**Visual attention and parallel letter processing**

Given the research presented in Chapter 1, it is incredibly likely that whole-word processes, such as the ability to process letters in parallel, would contribute significantly to reading proficiency. However, little is actually known about how these processes work, or how they may fail. One suggestion is that whole-word reading relies heavily on visual attention, the ability to distribute visual processing resources across multiple stimuli simultaneously. As we discussed in Chpater 1, one of the most developed theories building on this suggestion is the VA span deficit hypothesis (Bosse et al., 2007). This hypothesis proposes that there is a window of attention (VA span) that determines how many letters can be processed in parallel, and that if this window is restricted, then the reader must fall back on more sequential processes. Theoretically, VA is posited as a precursor to parallel processing; if the VA window can operate in global mode and capture multiple elements simultaneously, then parallel letter processing can proceed. However, if the VA window must operate in analytic mode, because the word is novel or the VA window has been restricted, then parallel letter processing will be affected.

**Empirical research pertaining to VA span.** Much of this research regarding the VA span utilises the visual partial report task. In this task, participants must identify a target letter from a visual array. The target is indicated by a bar probe located where one of the letters had been. Bosse et al. (2007) found that readers (eight to 16 years-old) VA span scores from this partial report task were positively associated with a measure of reading proficiency, reading age, but not scores on a phonemic awareness assessment. Importantly, VA span scores reliably predicted irregular word reading accuracy, whereas phonemic awareness scores did not. These findings suggest that VA span and phonemic awareness measures index different cognitive processes that both contribute to reading proficiency, and that whole-word processing skills may be dependent on visual attention. Bosse and Valdois (2009) found a similar pattern when they examined VA span and phonemic awareness skills in an undifferentiated sample of readers from a range of age groups (seven to 11 years-old). For the youngest age group (seven years-old), factorial analysis identified two major factors contributing to reading proficiency: a phonological factor comprised of phonemic awareness scores, and a VA span factor comprised of visual report scores. These findings suggest that variation across these skills can be detected at a very early age, even when there has been little opportunity for reading experience to influence scores. Further, these same factors were also found to be independent predictors of reading accuracy in the older age groups, and the relationships between measures became more selective as age increased. For example, for the oldest group (11 years-old), only the VA span factor was specifically associated with irregular word accuracy, and not with other reading measures such as nonword accuracy. These findings suggest that measures of VA span are relatively robust to individual differences in experience and learning.

Further, Bosse, Chaves, Largy, and Valdois (2015) investigated whether VA span had any implications for orthographic processing. In their study, novice readers (eight to 11 years-old) were trained to read bi-syllabic pseudo-words in either a large window condition where the first syllable appeared on-screen followed shortly by the second syllable, or a small window condition whereby the two syllables appeared on-screen separately. Participants in the large window condition demonstrated improvements on an orthographic choice task, while those in the small window condition did not. The authors concluded that the opportunity to process multiple stimuli simultaneously was beneficial for developing processes that map whole-words. These results suggest that restricting the VA window will directly impact orthographic processing.

Case studies with impaired readers also show how phonological and visual attentional processes may differentiate. Valdois et al., (2003) looked at performance on phonemic awareness and VA span tasks in two dissociating cases of developmental dyslexia – one phonological case defined according to their poor nonword reading and one whole-word case defined according to their poor irregular word reading. The phonological participant (14 years-old) had selective difficulties on phonemic awareness tasks, while the whole-word participant (13 years-old) had selective difficulties on the VA span tasks. Overall, these findings demonstrate that VA span may be important for reading development, specifically for whole-word level reading skills.

To the best of our knowledge, only one study has explicitly addressed the theoretical relationship between VA span and parallel letter processing (van den Boer et al., 2013). This study investigated the relationship between the word length effect and several measures of phonological and visual attentional skills in a group of unimpaired, novice readers of Dutch (seven to eight years-old). They found significant, negative correlations between the word length effect and measures of both phonemic awareness, and VA span. These results suggest that good phonemic processing and visual attention are both associated with greater parallel letter processing ability. Moreover, the association between word length effect and VA span was independent of phonemic awareness, suggesting that VA span may make a unique contribution to parallel letter processing abilities. Of course, novice readers do not have the reading experience to have developed expertise in whole word reading. Indeed, by the time readers have developed two to three years reading proficiency, length effects are considerably diminished (Bijeljac-Babic et al., 2004; Gagl, Hawelka, & Wimmer, 2015). Strong word length effects in older children and adults are likely to be more indicative of genuine cognitive difficulties at the whole word-level. Consequently, it may be possible to see more specific associations between these various measures in older readers.

In our study, we examined the relationship between VA span scores and standard measures of reading proficiency, such as reading age. We also examined whether VA span accounted for unique variance in reading proficiency, over and above what could be explained by commonly used measures of phonological processing skills, such as a measure of nonword reading accuracy. We then further explored the relationship between the VA span measure and the word length effect measure, assessing whether the concept of a visual attention window could help explain how whole-word mechanisms work or, as maybe the case in certain cases of developmental dyslexia, don’t work.

**Study Outline and Hypotheses**

As summarised above, the primary aims of this research were to investigate whole-word reading processes: a) to see whether they provided unique contributions to overall reading proficiency that could be distinguished from phonological processes; and b) to test the VA span deficit hypothesis’ explanation of these whole-word processes. To address these aims, we recruited secondary school students to complete the word identification task, the

Linguistic Properties Effects, the Coltheart and Leahy task, and the Partial Report task. We used measures obtained from these tasks to create markers of overall reading proficiency as well as indices of phonological processing, whole-word processing and visual attention. The rationale of each of these tasks is described in more detail below.

**Predictions.** We predicted that both indices of phonological and whole-word processing would be unique predictors of reading proficiency. That is, individuals' scores on the phonological index, nonword accuracy, would be negatively associated with the primary reading proficiency measure, Reading Age, and positively associated with the secondary reading proficiency measure, real word reading latency. A unique proportion of the variance in these reading proficiency measures should be accounted for by the nonword accuracy measure. Also, it was predicted that individuals' scores on the whole-word index, word length effect, would be negative associated with Reading Age, and positively associated with real word reading latency. A unique proportion of the variance in these reading proficiency measures should be accounted for by the word length effect measure, even after taking into consideration the effect of phonological processing skills.

Based on the VA span deficit hypothesis (Bosse et al., 2007), we predicted that visual attention would be relevant to reading proficiency, and more specifically, whole-word processes; i.e. parallel letter processing. We predicted that the VA span measure would be positively associated with Reading Age and negatively associated with real word reading latencies. Further, a unique proportion of the variance in these reading proficiency measures should be accounted for by the VA span measure, even after considering the effect of phonological processing skills. We also predicted VA span would be negatively associated with the whole-word index, but not the phonological index. Such a pattern of results would speak to both the heterogeneity in reading processes, *and* the mechanisms of such processes more specifically.

## Method

**Participants.** Recruitment of participants took place in two phases. In the first phase, the lead researcher (EA) approached five secondary schools within Wellington, New Zealand, and asked if they would allow her to describe the study to Year 9 students during school time and distribute letters calling for volunteers (see Appendix A for a copy of the information sheet sent to principals). During this phase, recruitment was aimed at, but not restricted to, students who were enrolled in reading remediation classes, or otherwise suspected of having reading difficulties by their teachers. Three of the schools consented. Schools A and B were co-education state schools (decile 8 and 9 respectively), and School C was a boys-only state school (decile 6). The researcher (EA) verbally described the study to each eligible class, then distributed the parents' information letters and consent forms, and students' information/assent forms (see Appendix B through D respectively; note that the information sheets sent out during this recruitment phase also mentioned that some students may be invited to participate in a subsequent intervention study – this will be discussed in later chapters). The only formal criteria for inclusion in the study was that the individual was currently completing Year 9 and had normal or corrected vision. There were no additional exclusion criteria. Twenty-six participants consented to take part during this recruitment phase.

In the second recruitment phase, the lead researcher (EA) returned to School A to recruit further participants. During this phase, Year 9 students at all levels of reading were targeted equally (i.e. non-reading remediation students were recruited; see Appendix E through H for the letters, information sheets and consent forms used). Thirty-one participants consented to take part during this phase.

Out of the 57 total participants recruited in both phases of this study, five were excluded from the current analysis because they did not complete all tasks, and a further three were excluded due to technical problems (the audio recorder failed during the session). Of the 49 remaining participants who completed all tasks, 20 were female and 29 were male, with a mean age of 13 years and 6 months (*SD*= 5 months; Appendix I contains additional demographic information). Small tokens of appreciation (USB sticks, water bottles, pens) were offered as a thank you for taking part and were not dependent on the number of tasks completed, or performance during the session.

**Materials.** Each participant completed the following tasks: a) the Woodcock Reading Mastery Test; b) the Coltheart and Leahy task; c) the Linguistic Properties Effect task; and d) the Partial Report task. Each of these tasks is described in more detail below. No word was duplicated across any of the three reading tasks. Test reliability and validity information is for all tasks is included in Appendix J.

***Woodcock Reading Mastery Test.*** Reading age was determined using the word identification sub-test from the Woodcock Reading Memory Test-III (Woodcock, 2011). The test was administered according to the standard procedure, except participants started at the very beginning of the test, rather than the recommended starting point for the age group. Reading Age was calculated according to the standardised norms.

***Coltheart and Leahy task.*** This task consisted of 90 words taken from Coltheart and Leahy (1996)'s standardised item list (see Appendix K). The list contains 30 regular words (e.g. *wedding*), 30 irregular words (e.g. *yacht*), and 30 nonwords (e.g. *baft*). Each trial commenced with a fixation dot presented in centre screen, followed 1000ms later by a single stimulus word presented in size 32 mono black font on a white background. The stimulus was accompanied by an beep audible only on the recording and not during the session. The word remained on-screen until participants responded verbally. The task was completed in one block, pseudo-randomised so that the order for all participants was identical. Five practice trials were completed with three real and two nonwords for demonstration. Feedback during these trials was given to encourage reading attempts, but no correction was provided. No feedback was provided during the main task.

***Linguistic Properties Effect task.*** This task comprised of 140 words of three to nine letters long (20 words at each letter length), specially selected for this study. Words in the list were fully counterbalanced for letter length, frequency (the subtitle frequency of each word according to Brysbaert and New, 2009), and imageability (scores from the MRC Psycholinguistic database by Coltheart, 1981 where available, or otherwise, from the Bristol norms Stadthagen-Gonzalez & Davis, 2006). Imageability is a measure of how concrete a word's meaning is. Specifically, highly imageable words are those whose meaning is consistently associated with a clear mental image (e.g. *computer, tree, wand*), whereas low or non-imageable words represent concepts that are more difficult to generate a mental image for (e.g. *despair, judgement, at, the*).  Specifically, words of each length were organised into four Frequency x Imageability groups, based on the cut-off scores presented in Table 2.1: 1) High Imageability / High Frequency; High imageability / Low Frequency; 3) Low Imageability / High Frequency, and 4) Low imageability / Low Frequency. Appendix L provides a full word list, with a full summary of word properties according to length.

Table 2.1 Cut-off boundaries for categorising frequency and imageability.

| | High | Low |
|---|---|---|
| Frequency | 17.06-377.49 | 0.61-5.39 |
| Imageability | 557-659 | 100-403 |

For each word, we also obtained age of acquisition scores from Gilhooly and Logie (1980), and average bigram frequency scores and orthographic neighbour scores (*Ortho_N*) from the English Lexicon Project (Balota et al., 2007; note Age of Acquisition scores were not available for two words: *destroyed* and *impressed*). Orthographic neighbourhood size refers to the number of words that differ from the "target" word by only one letter (e.g. *hat, has, ham, cat, hot & hat*). Table 2.2 shows the mean values for each of these scores.

Table 2.2 Mean scores according to various word properties: length, frequency group, imageability group, and frequency X imageability groups.

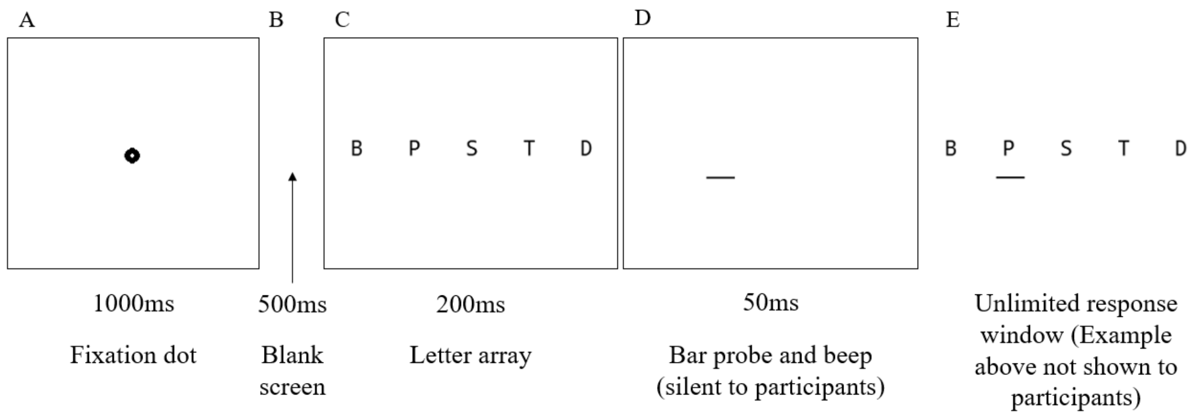
| | Frequency | Imageability | AoA[a] | Bigram | Ortho_N |
|---|---|---|---|---|---|
| Length | | | | | |
| 3 | 85.69 | 456.83 | 6.01 | 2784.85 | 12.55 |
| 4 | 22.72 | 465.58 | 7.66 | 3442.87 | 7.05 |
| 5 | 20.57 | 433.41 | 7.90 | 3341.16 | 6.3 |
| 6 | 29.39 | 432.13 | 7.51 | 3986.5 | 4.35 |
| 7 | 31.71 | 418.75 | 7.87 | 3967.65 | 0.95 |
| 8 | 17.23 | 464.83 | 7.48 | 4001.79 | 1.15 |
| 9 | 14.90 | 450.69 | 8.97 | 4042.99 | 0.78 |
| Frequency | | | | | |
| High | 60.78 | 447.53 | 6.49 | 3873.24 | 5.37 |
| Low | 2.39 | 450.36 | 8.63 | 3427.00 | 4.23 |
| Imageability | | | | | |
| High | 22.57 | 593.30 | 6.55 | 3602.81 | 4.75 |
| Low | 40.99 | 314.50 | 8.63 | 3689.70 | 4.83 |
| Group | | | | | |
| HFHI | 79.72 | 297.93 | 7.20 | 3940.81 | 5.51 |
| HFLI | 41.17 | 602.46 | 5.77 | 3801.58 | 5.21 |
| LFHI | 2.54 | 583.42 | 7.28 | 3415.40 | 4.31 |
| LFLI | 2.25 | 331.07 | 9.97 | 3438.59 | 4.14 |
| Total | 32.11 | 448.92 | 7.59 | 3646.89 | 4.79 |

Note: a) AoA= Age of Acquisition

Each trial commenced with a fixation dot presented in centre screen for 1000ms, followed by a single word presented in size 32 in mono black font, accompanied by a beep, which remained on-screen until participants responded verbally. The task was split into two blocks, each comprising of 70 words which had been pseudo-randomised so that the order for all participants was identical. No practice trials were completed given the similarity between this and the earlier single word reading task. No feedback on performance was given.

***Partial Report task.*** This task, taken from Bosse et al. (2007), required participants to verbally recall a probed target letter from a previously presented horizontal array of five letters. As shown in Figure 2.1, the letter stimuli were presented across the central y-axis, displayed in size 24 font with 1cm between each letter. All strings were made up from a 10-letter pool of consonants (B, D, F, H, L, M, P, R, S, T), with each letter being used in 25 trials, and no repetitions within each string. Each letter was used as the target in five trials, once in each of the five positions of the string. See Appendix M for the stimuli list.

Figure 2.1 The Partial Report task.




Each trial began with a fixation dot presented onscreen for 1000ms (A in Figure 2.1), followed by a blank screen for 500ms (B), and then the five-letter array, which remained visible for 200ms (C). This array was then replaced by a 1cm horizontal – the probe – which remained visible for 50ms (D). The location of the probe along the x-axis aligned with that of one of the five previously presented letters (E). The participant's task was to verbally name the letter that had just appeared in that position. When the probe was removed, a blank screen appeared and remained present until the participant responded. In addition to the 50 experimental trials, five practice trials containing novel letter strings using vowels were administered. There was no feedback in any of the trials as the purpose was merely to demonstrate the procedure and speed, offering an opportunity for questions before completing the main task.

**Procedure.** All tasks, other than the word identification task from the Woodcock Reading Mastery Test, were created via OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). These tasks were administered on a Samsung galaxy tab 10.1. All sessions were recorded using a digital audio recorder.

Each testing session lasted up to one hour, taking place in normal school hours. Sessions were conducted on a one-to-one basis; however, an observer (a research assistant) was also present throughout. All sessions in the first recruitment phase were run by the lead researcher (EA). In the second recruitment phase some of the sessions were run by a secondary researcher (MR) with the lead researcher present as the observer. Each session ran as follows. First, the researcher introduced themselves and the observer, explaining that the latter would not actively participate in the session. They then summarised the research project verbally, and explained how participants' confidential data would be treated, drawing on the printed information disseminated during recruitment.[1] During this initial conversation, the lead researcher also spent a couple of minutes talking to the participant about subjects they enjoyed at school and other general topics to build rapport. Participants were encouraged to ask questions, both during the briefing and throughout the remainder of the session. They were told that they could take a break or stop participating at any point, and that there would be no penalisation for doing so. All participants also verbally assented at the start of the session, after the study had been outlined to them.

Tasks were administered in the same order for each participant: 1) the Coltheart and Leahy task; 2) the Partial Report task; 3) the Linguistic Properties Effect task; and 4) the word identification task from the Woodcock Reading Mastery Test. Some participants also completed additional reading or cognitive tasks designed to assess their suitability for the reading intervention study. These are described in Chapter 3 and will not be discussed here.

On all tasks, participants were told to try their best, but not to panic if they did not know or were unsure at any point. Between each task, participants were given a short break.

[1] Participants recruited in the first phase of the study were given the option to de-identified group reports made available to them, their parents and their teachers after the study was complete. Participants recruited in the second phase of the study were given the option to have individualised reports made available.

During these breaks, the lead researcher made general conversation with the participant. Throughout all tasks, no feedback on performance was given, however occasionally if a student showed concern or appeared nervous, re-assurance was given (e.g.., "you're trying really hard", or "read how you think it might sound"), without affirming performance. At the end of each session participants were de-briefed and offered the opportunity to ask any questions they might have. They were told that if they thought of any questions after they had left, their parents could email the lead researcher (EA) or the supervisor (CW). They were thanked for their participation and given a small token of gratitude (i.e. water bottle, USB stick).

**Data Collection and Analysis.** For the Coltheart and Leahy and Linguistic Properties Effects tasks, reading accuracy was recorded during the session by the lead researcher, and where appropriate, cross checked against the audio-recording of the session. For these tasks, reading latencies were calculated manually from the digital recordings of the sessions using Audacity 2.1.2 (Audacity-Team, 2014), measuring time from the onset of the beep that accompanied the stimulus to the onset of the first correct response. For the Partial Report task, accuracy was recorded manually by the experimenter during the session, and, where appropriate cross checked against the audio-recording of the session. Full details as to how these measures were pre-processed and analysed are provided in the Results section below.

## Results

**Group measures.** The current set of analyses were primarily concerned with how the various measures of reading skills and outcomes related to, and subsequently predicted, one another. In order to do this, individual scores on such measures were calculated and will be discussed shortly. First however, group analyses on selected measures were completed in order to check the general patterns. Assumption checks were completed using SPSS, however, unless otherwise stated, all other analyses were completed in R[2]. Model fit for accuracy (binomial logistic regression) and response time data (linear mixed-effects modelling) were performed using the "*glmer"* and "*lmer"*" functions respectively (via the *lme4*, *lmerTest*, and *lmTest* packages). For the logistic regression, the z-values obtained were the Wald-test statistics calculated by dividing the regression coefficient by the standard error.

---

[2]Packages used: Rstudio: Team (2015), lme4: Bates, Maechler, Bolker, and Walker (2014), lmerTest: Kuznetsova, Brockhoff, and Christensen (2017), lmTest: Hothorn et al. (2018), aod: Lesnoff and Lancelot (2012), Hmisc: Harrell Jr and Harrell Jr (2019), Tidyverse: Wickham (2017), corrplot: Wei and Simko (2017), irr: (Gamer, Lemon, Gamer, Robinson, & Kendall's, 2012).

For the linear mixed-effects modelling, the F-values obtained (via the "*ANOVA*" function of the lmerTest package) were Type III hypotheses tests for each fixed effect, while degrees of freedom were estimated via Satterthwaite's approximation (rounded to the nearest integer). For these model fits, as well as all other analyses, the alpha level was $p < .05$

***VA span task results.*** Assumption checks were completed for Partial Report task accuracy. The data was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), with no apparent skew, kurtosis, or outliers. Figure 2.2 shows the serial position curve, created from the average accuracy of the group at each position in the five-letter array. As can be seen, the curve shows a 'W' function. A one-way analysis of variance (function "*aov*") revealed that overall, there was a significant effect of target position on accuracy ($F(240,4) = 21.34$, $p < .001$). Post-Hoc Tukey tests (function "*TukeyHSD*") revealed significant pairwise differences between all positions ($p < .05$); except positions 2 and 4 ($p = .206$); and 3 and 5 ($p = 1.00$).

Figure 2.2 Serial position curve for the VA span partial report task with group mean (red line) as well as individual mean scores (green-purple dots).


***The effect of word properties on reading accuracy.*** Logistic regression analysis was completed to assess the effect of word properties, such as length, frequency and imageability, on reading accuracy in the Linguistic Properties Effect task. Prior to analysis, 0.16% of trials were excluded due to technical problems that prevented either accuracy, RT or both from being encoded. Figure 2.3 presents a summary of the results for length, while Figure 2.4 presents the results for frequency and imageability. The predictor variables within the regression model were length, frequency category, and imageability category (length was coded as a continuous variable, the latter two were coded as ordinal variables). Participant and word (i.e. trial) were inputted as random effects. No interactions were included in the model because we did not wish to reduce power for this purpose. The Wald z-statistics associated with frequency ($z = 5.79$, $p < .001$) and imageability ($z = 4.19$, $p < .001$) were significant, however length was not ($z = 1.90$, $p = .057$). Overall, high imageability words were read more accurately than low imageability words, and high frequency words were also read more accurately that low frequency words. There were no accuracy changes as a function of length.

Figure 2.3 Linguistic Properties Effect Task: Mean proportion correct as a function of the number of letters. Individual black dots represent individual participants' mean scores, with group mean represented by the yellow dot.

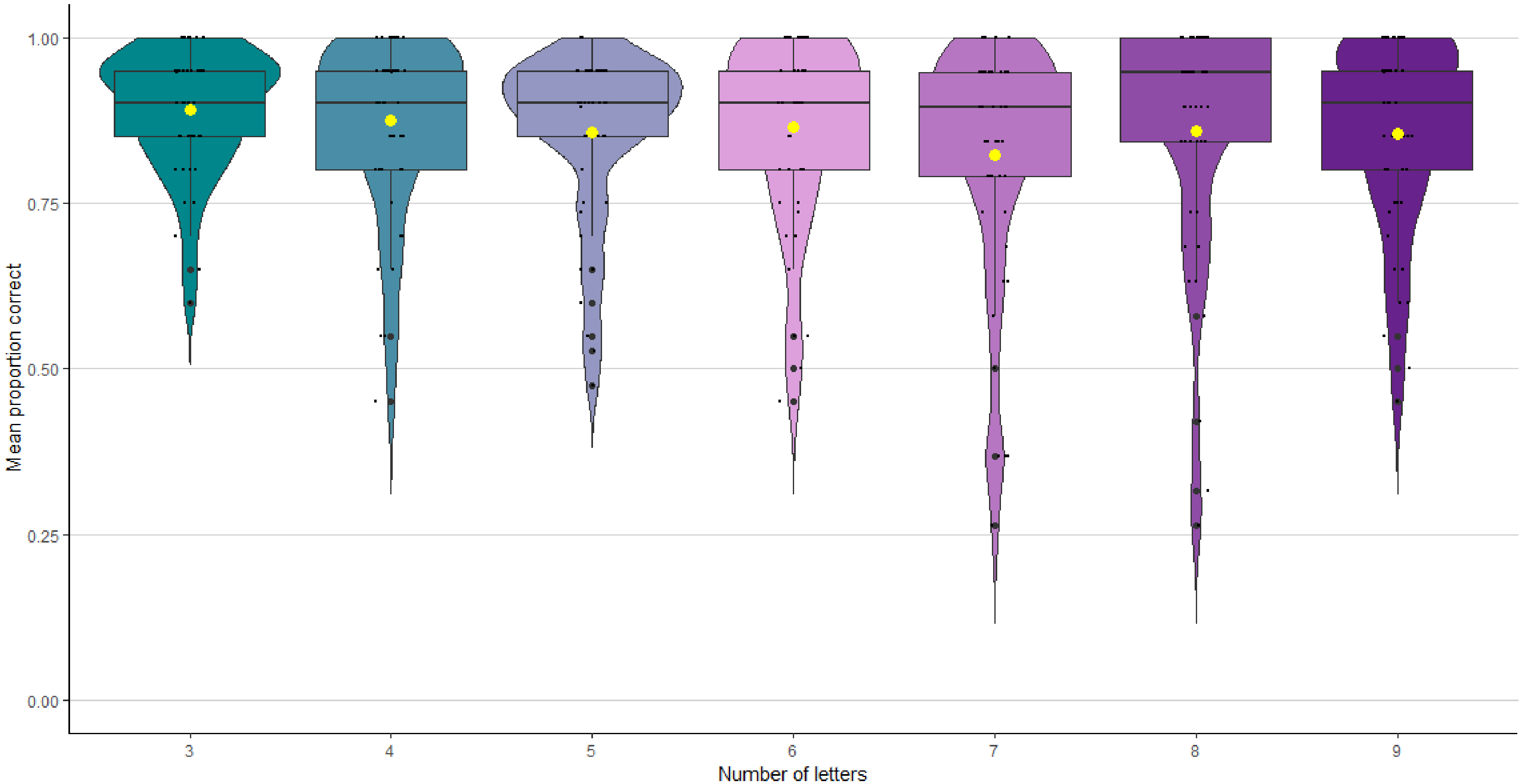
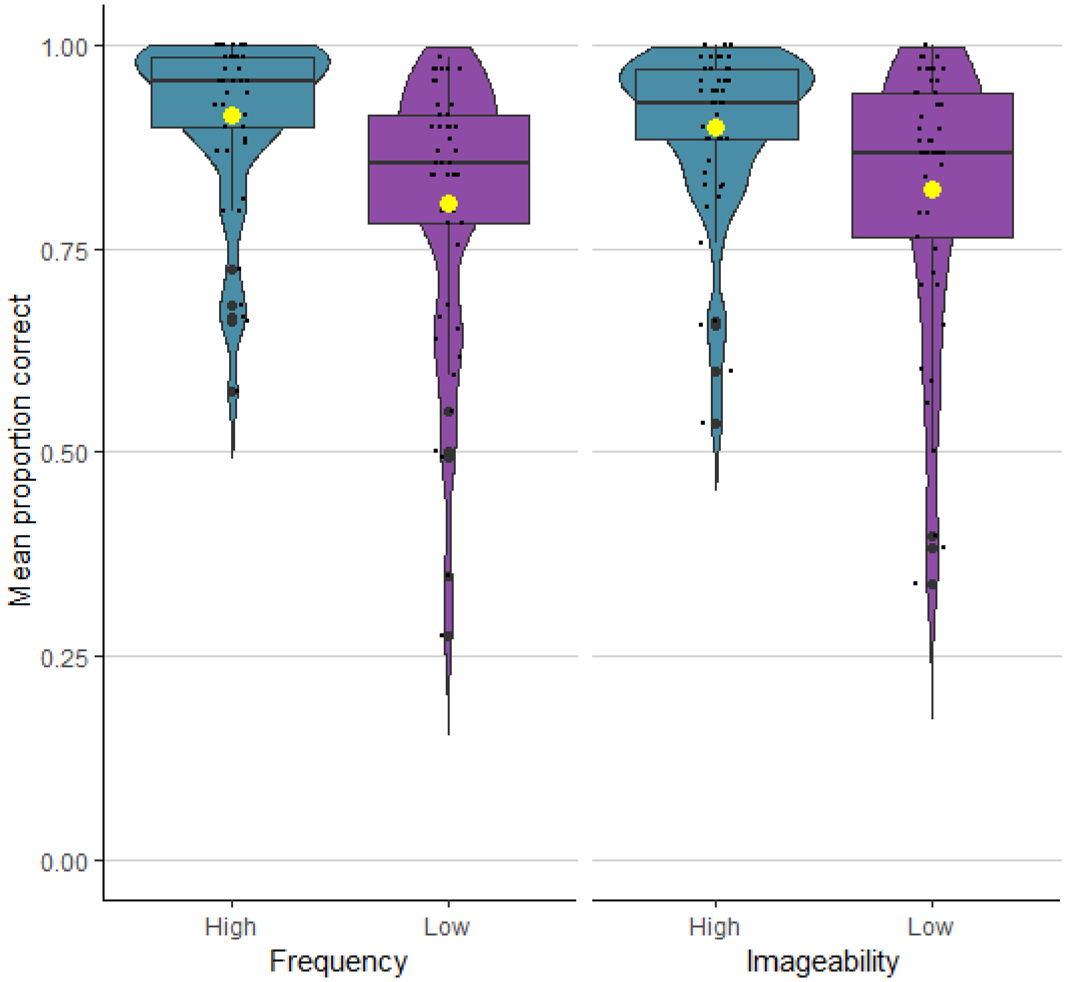


Figure 2.4 Linguistic Properties Effect Task: Mean proportion correct as a function of frequency and imageability. Individual black dots represent individual participant mean scores, with group mean represented by the yellow dot.




*Alternative regression analysis.* To explore the possible effect age of acquisition had on accuracy, we ran two alternative models. The first was identical to the original model described above, except that we replaced frequency category with age of acquisition. In this model, the Wald z-statistic for age of acquisition as a predictor variable was significant (z = 4.90, $p$ < .001), as was length (z = 2.22, $p$ = .027), and imageability (z = 9.59, $p$ < .001). The fit of this model was significantly better than the original frequency-based model ($\chi2$= 1514.7, $p$< .001). In the second alternative model, we included both frequency category and age of acquisition as predictor variables, along with length and imageability. In this combined model, all predictor variables were statistically significant (length: z = 2.32, $p$ = .021, imageability: z = 9.74, $p$ < .001, frequency: z = 12.69, $p$ < .001, age of acquisition: z = 5.09, $p$ < .001). Again, the fit of this model was significantly better than the original model that only included frequency, length and imageability ($\chi2(1) = 1341.4$, $p < .001$), but poorer than the alternative age of acquisition model which replaced frequency with age of acquisition ($\chi2(1) = 173.3$, $p < .001$). In sum, key word properties, such as frequency, age of acquisition, length, and imageability, influence reading accuracy. However, in subsequent analysis the current study primarily relies on latency for measuring the effect of different word properties in the Linguistic Properties Effect task. Therefore, we repeated our analysis with latency and present the findings below.

***The effect of word properties on reading latencies.*** Linear mixed-effects modelling was completed to investigate the effects of length, frequency category, and imageability category on reading latencies in the Linguistic Properties Effect task. Only accurately read trials were included, leading to the removal of 13.86% trials. During analysis, reading latencies were assessed for independence of observations, homogeneity in the error residuals (homoscedasticity), skew, and outliers. The Durbin-Watson statistic (Savin & White, 1977) ($M$= 1.982, range: 1.649 - 2.319) was within the lower (1.383) and upper limits (2.334), indicating an independence of observations. Visual inspection of the scatter plots[3] for each person indicated heterogeneity in the error residuals (heteroscedasticity), skew and multiple outliers. More formal investigation with individual Q-Q plots[4] confirmed positive skew and positive kurtosis in most participants' data. For these reasons, the data was transformed in accordance with guidelines by Baayen and Milin (2010). First, the most extreme outliers were removed. Given the high variability in latencies both within participants (i.e. from word-to word) and between participants, we used a Winsdorising procedure, rather than an absolute cut-off. This procedure involved replacing the longest latency in each length category with the second longest latency in that category, and then replacing the latter with the third longest latency. Second, latencies were log transformed (natural log base). Third, for each participant, we regressed word length, frequency and imageability, onto the log transformed latencies to calculate residuals. Finally, any data points with standardised residuals below -2.5 or above 2.5 were removed (a further 2.63% of the accurate data points). Q-Q plots[5] were recreated for each person and substantial improvement to normality and homoscedasticity was found. All subsequent analyses were completed with these log-transformed latencies (logRT).

[3] Not included in this thesis.
[4] Not included in this thesis.
[5] Not included in this thesis.

Figure 2.5 and Figure 2.6 show the summary results for main effects of length, and frequency and imageability respectively, with Table 2.3 summarises the effects (and interactions) from the linear mixed-effects model(s). The fixed effects included length, frequency category, imageability category (all coded as an ordinal variables), and all possible interactions, and participant and word were entered as random (non-nested) effects. Intercepts were specified for each random effect, but not slopes. We removed non-significant interactions involving fixed effects in a stepwise manner, starting with those yielding the highest *p*-value. The final model included the main effects of length, frequency, imageability, and no interactions. There were significant main effects of length ($F(1, 131) = 47.39$, $p < .001$), frequency ($F(1, 131) = 70.64$, $p < .001$), and imageability ($F(1, 131) = 19.98$, $p < .001$). Frequency and imageability showed identical patterns; it took longer to start reading low frequency or imageability words, compared to high words. Reading latencies also increased as word length increased, so the longer the word, the longer it took to start reading. In sum, key word properties, such as length, frequency and imageability, influence reading latencies, and together with the accuracy findings, these results demonstrate the importance of controlling for word properties, even when not of primary investigative value.

Figure 2.5 Linguistic Properties Effect Task: Mean logRT as a function of the number of letters. Individuals black dots represent single trials, with group mean represented by the yellow dot.

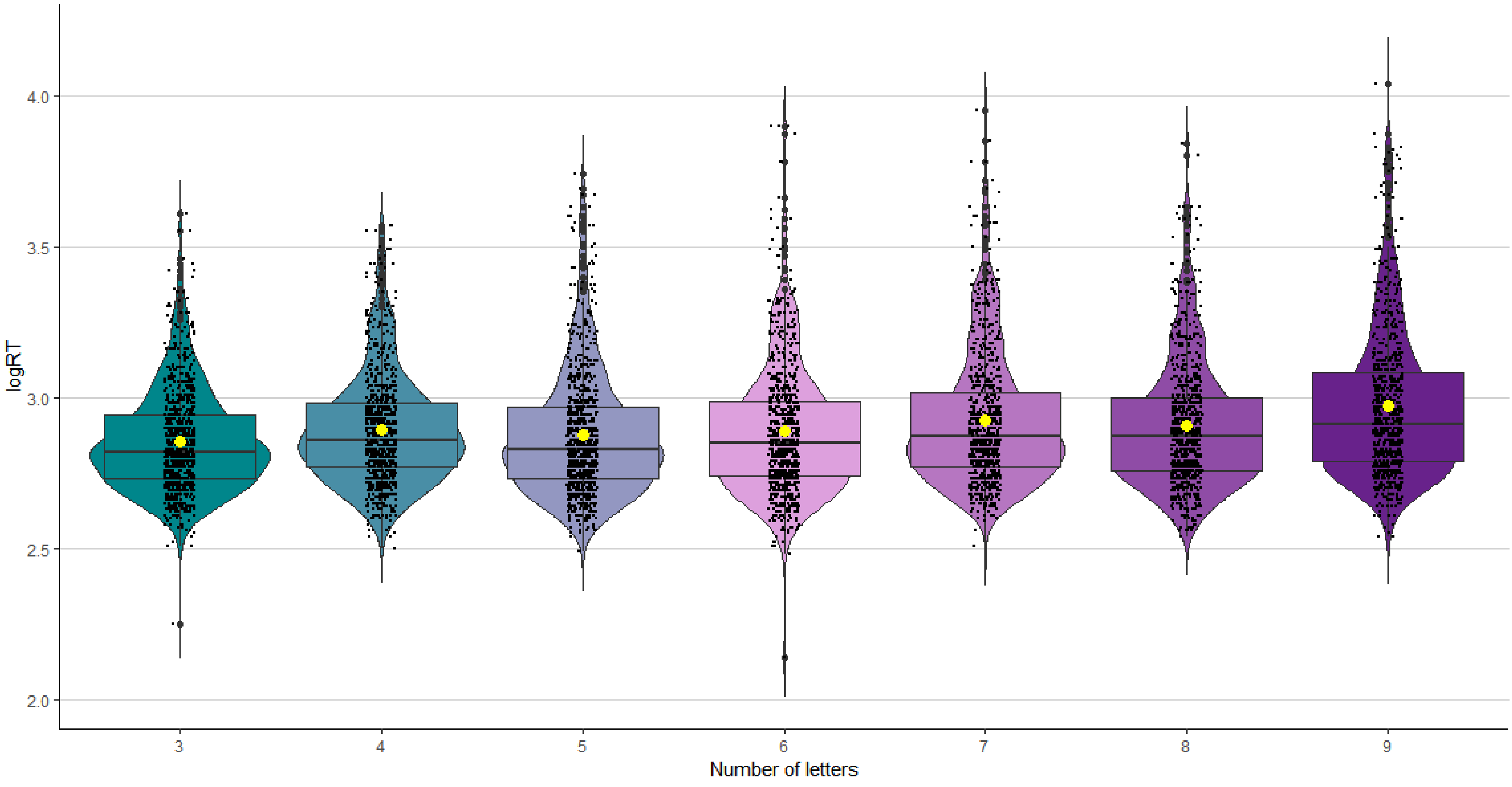


Figure 2.6 Linguistic Properties Effect Task: Mean logRT as a function of Frequency and Imageability. Individual black dots represent single trials, with group mean represented by the yellow dot.

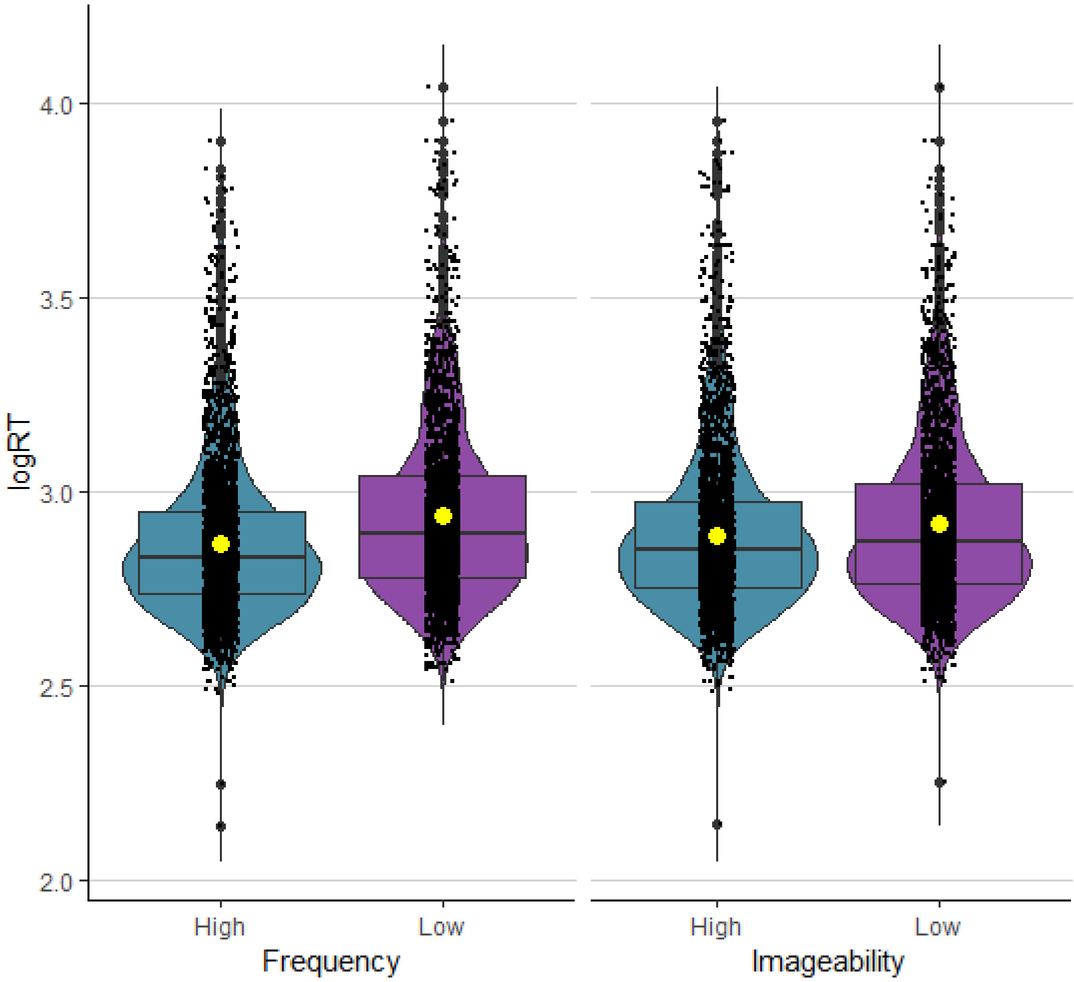


Table 2.3 Linguistic Properties Effect task: Mean logRT (standard deviations in parentheses) according to length and word properties.

| Length | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Overall word properties |
|---|---|---|---|---|---|---|---|---|
| Group | | | | | | | | |
| HIHF[a] | 2.81 (0.14) | 2.85 (0.16) | 2.84 (0.18) | 2.82 (0.19) | 2.89 (0.21) | 2.85 (0.19) | 2.92 (0.23) | 2.85 (0.19) |
| HILF[b] | 2.87 (0.17) | 2.91 (0.19) | 2.89 (0.22) | 2.91 (0.21) | 2.92 (0.22) | 2.95 (0.22) | 3.00 (0.26) | 2.92 (0.22) |
| LIHF[c] | 2.82 (0.15) | 2.88 (0.17) | 2.87 (0.22) | 2.87 (0.21) | 2.93 (0.21) | 2.89 (0.21) | 2.93 (0.22) | 2.88 (0.20) |
| LILF[d] | 2.94 (0.20) | 2.93 (0.21) | 2.90 (0.20) | 2.96 (0.22) | 2.97 (0.23) | 2.94 (0.21) | 3.07 (0.24) | 2.96 (0.22) |
| Frequency | | | | | | | | |
| High | 2.81 (0.14) | 2.87 (0.17) | 2.85 (0.20) | 2.84 (0.20) | 2.91 (0.21) | 2.87 (0.20) | 2.92 (0.23) | 2.87 (0.20) |
| Low | 2.90 (0.19) | 2.92 (0.20) | 2.90 (0.21) | 2.94 (0.22) | 2.94 (0.23) | 2.94 (0.21) | 3.03 (0.26) | 2.94 (0.22) |
| Imageability | | | | | | | | |
| High | 2.84 (0.16) | 2.88 (0.18) | 2.87 (0.20) | 2.87 (0.21) | 2.90 (0.22) | 2.90 (0.21) | 2.96 (0.25) | 2.89 (0.21) |
| Low | 2.87 (0.18) | 2.91 (0.19) | 2.88 (0.21) | 2.91 (0.22) | 2.95 (0.22) | 2.92 (0.21) | 2.99 (0.24) | 2.92 (0.21) |
| Overall length | 2.85 (0.17) | 2.89 (0.18) | 2.87 (0.21) | 2.89 (0.21) | 2.92 (0.22) | 2.91 (0.21) | 2.97 (0.25) | 2.90 (0.21) |

Note: a= High Imageability, High Frequency; b= High Imageability, Low Frequency; c= Low Imageability, High Frequency; d= Low Imageability, Low Frequency.

*Alternative Linear mixed-effects modelling analyses.* To explore the possible effect age of acquisition had on latencies, we ran two alternative models. The first was identical to the model described above, except that we replaced word frequency with age of acquisition. Again, all interactions among the fixed effects were non-significant and removed from the analysis. Within the model, there was no significant main effect for age of acquisition ($F(1, 132) = 0.901$, $p = .344$), but there were still significant main effects of length ($F(1, 132) = 29.05$, $p < .001$) and imageability ($F(1, 132) = 13.76$, $p < .001$). The fit of the model was significantly poorer than the original frequency-based model ($\chi 2 = 59.18$, $p < .001$). In the second alternative model, we included both frequency and age of acquisition as fixed effects, repeating the step-by-step removal of non-significant interactions. Once again, age of acquisition was not a significant fixed effect ($F(1, 131) = 1.42$, $p = .235$), but all other main effects remained significant (length: $F(1, 131) = 43.73$, $p < .001$; imageability: $F(1, 131) = 21.12$, $p < .001$; frequency: $F(1, 131) = 70.91$, $p < .001$). The fit of this model was significantly poorer than the original model that included frequency, length and imageability only ($\chi 2(1) = 8.98$, $p = .003$). Thus, the original frequency-based model was the most appropriate, and suggests that frequency, imageability and length are key word properties that affect reading latency.

**Effects of individual variables.** To assess how each individual's performance was influenced by different lexical and cognitive variables, we calculated several scores for each individual, based on their patterns of performance in the Linguistic Properties Effect task, the Partial Report task, the Woodcock Reading Mastery test and the Coltheart and Leahy task. Table 2.4 summarises the measures that were obtained for each individual.

Table 2.5 provides the individuals' scores, and Figures 2.7 and 2.8 show the violin plots for each of the measures. As in the analyses of Linguistic Properties Effect measures, trials that were missing due to technical problems were excluded. For this reason, 0.02% of total trials from the Coltheart and Leahy task were excluded from the accuracy analysis, and an additional 0.09% were excluded from the latency analyses. For the analyses of latency, inaccurate trials were also excluded. Considering the entire cohort, this resulted in the removal of 22.7% of all trials. Exclusion of such a high number of trials is potentially concerning, however somewhat unsurprising given the heavy recruitment of struggling readers. The implications of this are considered further in the discussion.

Table 2.4 Heterogeneity test scores calculated for each individual, how they were calculated and how they are interpreted.

| Score | Calculation of score | Description/ Interpretation |
| --- | --- | --- |
| **Woodcock Reading Mastery Test** | | |
| Reading Age | Standardised scores of the word identification task | High scores indicate a high Reading Age, low scores indicate a low Reading Age |
| **Coltheart and Leahy task** | | |
| RealWord Accuracy | Proportion correct on regular and irregular words combined | High scores indicate good real word reading, while low scores indicate poor real word reading |
| RealWord Latency | Mean latency on regular and irregular words combined | Low scores indicate fast real word reading, while high scores indicate slow real word reading |
| IrregularityDiff | Proportion correct on regular words minus proportion correct on irregular words | Positive scores indicate poorer irregular word reading, negative scores indicate poorer regular word reading. |
| Nonword AccuracyDiff | Proportion correct on real words minus proportion correct on nonwords | Positive scores indicate poorer nonword reading, negative scores indicate poorer real word reading. |
| Nonword LatencyDiff | Mean nonword latency minus mean real word latency | |

**Linguistic Properties Effect task**

| | | |
|---|---|---|
| FreqAccuracy Effect | Proportion correct on high frequency words minus proportion correct on low frequency words | Positive scores indicate better reading with high frequency compared to low frequency words, negative scores indicate better reading with low frequency compared to high frequency words. |
| FreqLatency Effect | Mean logRT for low frequency words minus mean logRT for high frequency words | |
| ImageAccuracy Effect | Proportion correct on high imageability words minus proportion correct on low imageability words | Positive scores indicate better reading with high imageability compared to low imageability words, negative scores indicate better reading with low imageability compares to high imageability words. |
| ImageLatency Effect | Mean logRT for low imageability words minus mean logRT on high imageability words | |
| WLE | Word length regressed onto logRT | High scores indicate logRT increased with longer words, low scores indicate little logRT change |

**Partial Report task**

| | | |
|---|---|---|
| VA span | Proportion correct on the partial report task | High scores indicate good span, low scores indicate limited span. Impairment categorised according to standardised age norms indicting normal, impaired or severely impaired performance |

Figure 2.7 . Violin plots of Reading Age (A), Real word Accuracy (B), Real word Latency (C), Irregularity Diff (D), Nonword Accuracy Diff (E), and Nonword Latency Diff (F). Black dots represent individual participants, yellow dots represent group mean, with 95% confidence intervals (black vertical line). Box plots show median (black central line), upper and lower quartiles.

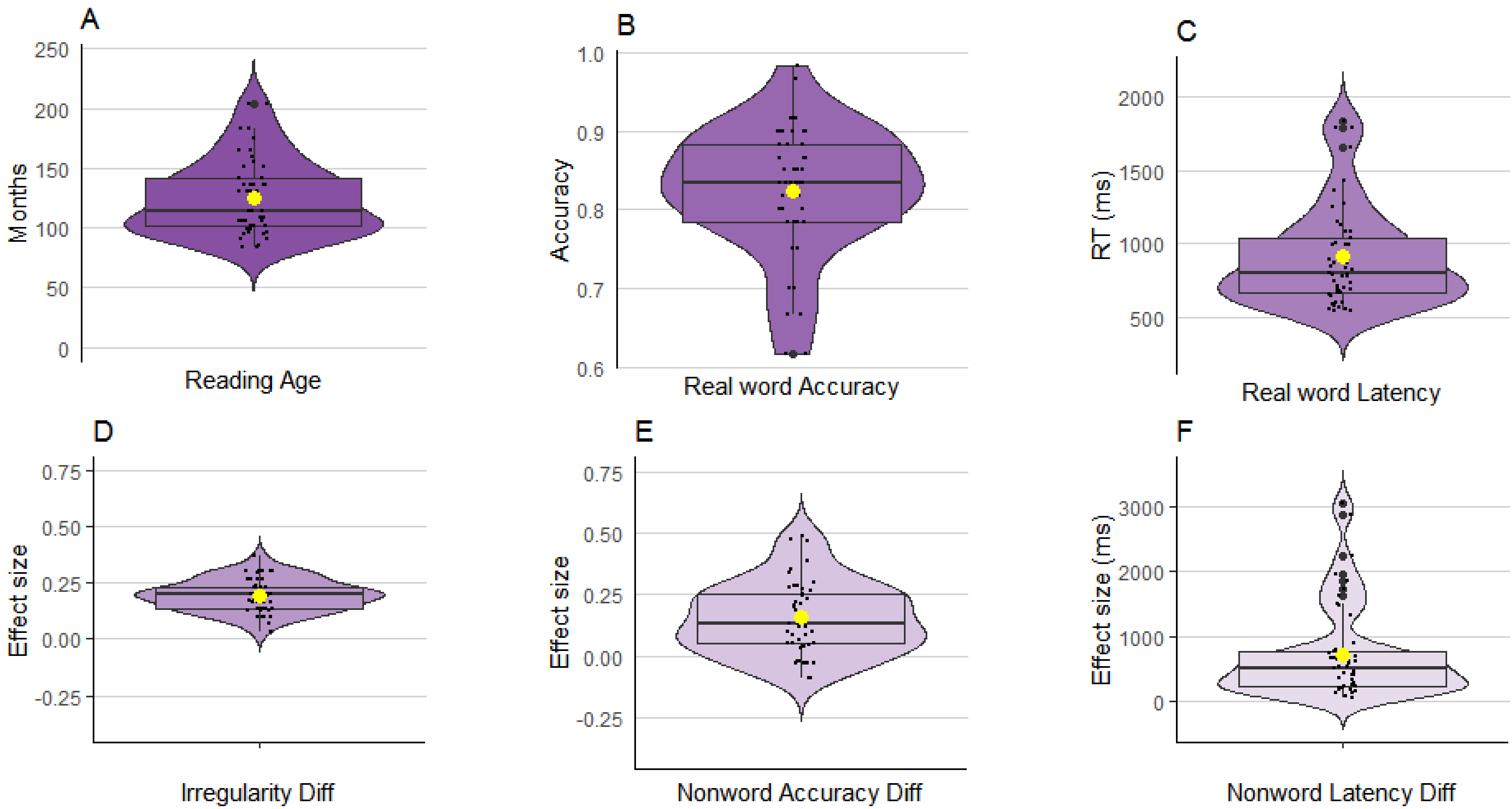


Figure 2.8 Violin plots of VA span (A), Frequency Accuracy Effect (B), Frequency Latency Effect (C), Imageability Accuracy Effect (D), Imageability Latency Effect (E), and WLE (F). Black dots represent individual participants, yellow dots represent group mean, with 95% intervals (black vertical line). Box plots show median (black central line), upper and lower quartiles.

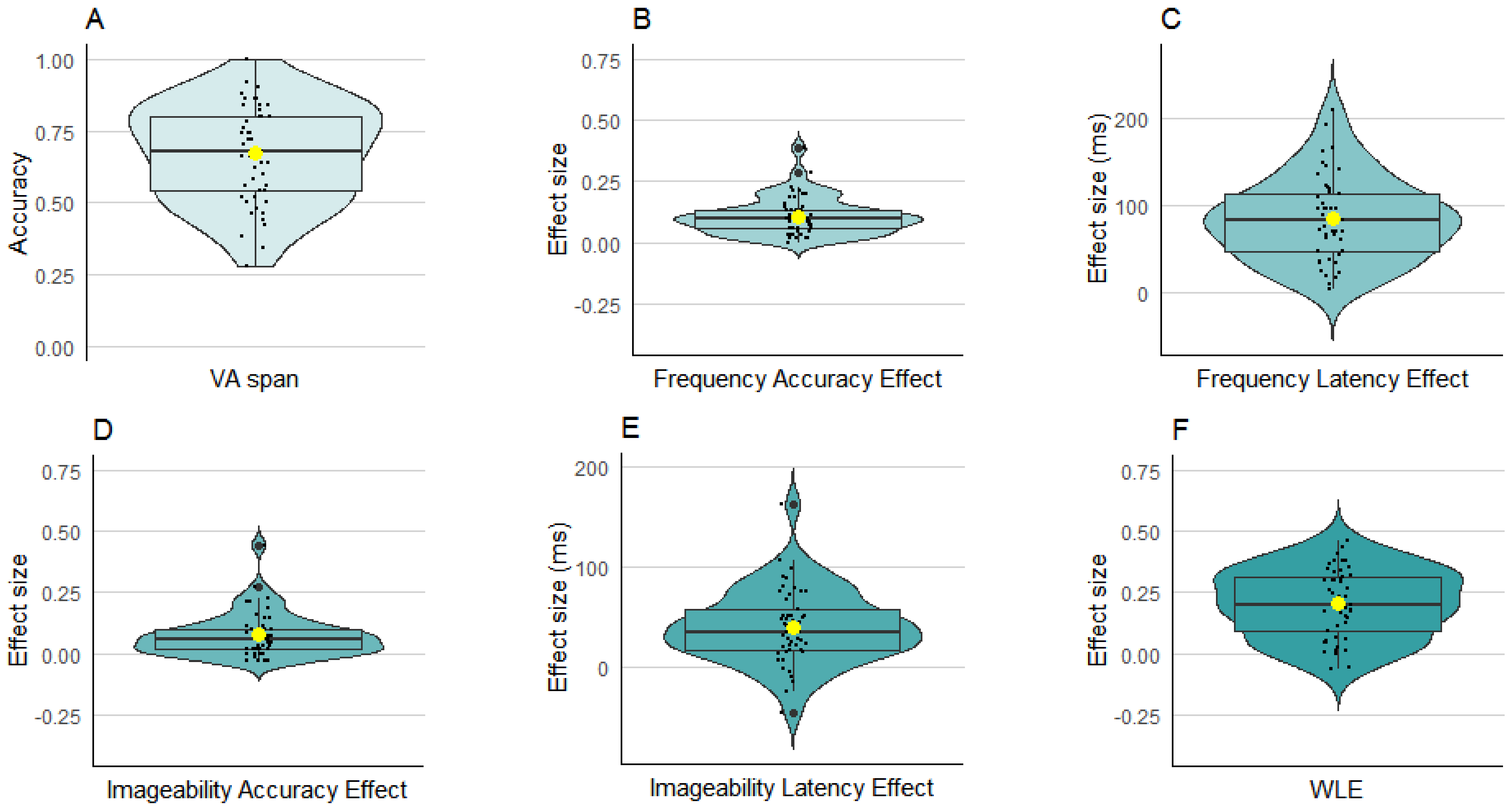


Table 2.5 Scores across all measures for each participant, and the overall group. * indicates Reading Age 2 years or more below their chronological age. For VA span; a score equal to above 35 indicate normal performance, a score between 34 and 17 indicate impaired performance (light grey), a score 17 or below indicate severely impaired performance (dark grey).

| ID[a] | CA[b] | RA[c] (std[d]) | FrACC Effect | FrRT[e] Effect | ImACC Effect | ImRT[e] Effect | WLE | VA span | Realword ACC | Realword RT[e] | Irreg Diff[f] | NonACC Diff[g] | NonRT[e] Diff |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 22 | 166 | 155 (99) | 0.01 | 46.19 | 0.01 | 50.51 | .007 | 0.86 | 0.90 | 671.52 | 0.14 | -0.03 | 420.62 |
| 23* | 161 | 105 (71) | 0.10 | 24.74 | 0.01 | 6.45 | .189 | 0.52 | 0.80 | 692.17 | 0.26 | 0.13 | 144.63 |
| 24* | 157 | 114 (79) | 0.07 | 61.91 | 0.10 | 40.21 | .027 | 0.74 | 0.82 | 701.88 | 0.23 | 0.05 | 409.38 |
| 25* | 161 | 99 (65) | 0.11 | 43.69 | 0.03 | 12.16 | .459 | 0.72 | 0.75 | 833.00 | 0.24 | 0.22 | 741.38 |
| 28* | 158 | 130 (88) | 0.10 | 96.63 | 0.07 | 47.40 | .054 | 0.90 | 0.75 | 558.84 | 0.16 | 0.05 | 599.40 |
| 29* | 167 | 94 (58) | 0.23 | 69.42 | 0.09 | 90.25 | .364 | 0.54 | 0.67 | 1034.95 | 0.20 | 0.30 | 884.14 |
| 31* | 166 | 99 (63) | 0.20 | 119.03 | 0.11 | 87.49 | .302 | 0.64 | 0.78 | 1652.60 | 0.30 | 0.28 | 1729.94 |
| 32* | 173 | 114 (72) | 0.10 | 74.17 | 0.07 | 78.57 | .185 | 0.86 | 0.80 | 777.17 | 0.26 | 0.27 | 626.58 |
| 33* | 161 | 136 (89) | 0.16 | 2.70 | 0.21 | 50.25 | .343 | 0.86 | 0.82 | 1079.33 | 0.10 | 0.25 | 613.67 |
| 34* | 165 | 96 (60) | 0.20 | 134.93 | 0.23 | -10.72 | .167 | 0.34 | 0.78 | 591.00 | 0.23 | 0.48 | 708.00 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 35* | 179 | 102 (63) | 0.09 | 122.74 | 0.11 | 107.05 | .355 | 0.84 | 0.83 | 893.58 | 0.20 | 0.26 | 1315.48 |
| 37 | 169 | 159 (93) | 0.04 | 33.78 | 0.01 | 24.69 | .089 | 0.70 | 0.92 | 690.22 | 0.17 | 0.09 | 246.50 |
| 38* | 168 | 85 (55) | 0.14 | 192.03 | 0.14 | 162.06 | .430 | 0.58 | 0.67 | 1245.23 | 0.13 | 0.34 | 3049.68 |
| 39* | 167 | 102 (66) | 0.00 | 145.78 | 0.09 | 75.86 | .378 | 0.74 | 0.85 | 997.78 | 0.30 | 0.12 | 780.49 |
| 40* | 163 | 136 (89) | 0.06 | 59.66 | 0.06 | 31.33 | .047 | 0.76 | 0.90 | 994.04 | 0.20 | 0.13 | 670.01 |
| 41* | 175 | 114 (72) | 0.13 | 112.33 | 0.01 | 34.02 | .171 | 0.52 | 0.83 | 1144.46 | 0.20 | 0.23 | 1620.43 |
| 43* | 165 | 90 (55) | 0.13 | 166.47 | 0.44 | 14.84 | .234 | 0.42 | 0.70 | 1786.74 | 0.20 | 0.25 | 1957.34 |
| 44 | 160 | 151 (94) | 0.03 | 69.81 | 0.00 | 33.97 | .205 | 0.68 | 0.88 | 797.96 | 0.23 | 0.05 | 527.52 |
| 45* | 172 | 114 (72) | 0.19 | 68.86 | 0.10 | 20.17 | .331 | 0.82 | 0.78 | 657.36 | 0.37 | -0.09 | 234.75 |
| 46* | 161 | 124 (83) | 0.11 | 109.62 | -0.03 | 75.24 | .244 | 0.50 | 0.83 | 1000.18 | 0.20 | 0.03 | 515.40 |
| 47* | 163 | 130 (86) | 0.13 | 101.22 | 0.10 | 74.94 | .171 | 0.46 | 0.85 | 1008.08 | 0.16 | 0.28 | 1496.39 |
| 48* | 163 | 105 (71) | 0.10 | 96.26 | 0.07 | 66.73 | -.005 | 0.38 | 0.83 | 1125.42 | 0.13 | 0.23 | 216.58 |
| 49* | 163 | 109 (74) | 0.10 | 46.12 | 0.10 | 32.74 | .317 | 0.80 | 0.70 | 874.50 | 0.26 | 0.10 | 659.17 |
| 50* | 154 | 102 (72) | 0.03 | 141.14 | 0.03 | 49.51 | .110 | 0.54 | 0.83 | 1080.20 | 0.13 | 0.06 | 273.80 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 51* | 163 | 130 (86) | 0.03 | 64.85 | 0.00 | 26.52 | -.058 | 0.72 | 0.87 | 682.52 | 0.20 | 0.17 | 171.72 |
| 52* | 160 | 94 (59) | 0.19 | 95.09 | 0.21 | 19.76 | .047 | 0.60 | 0.78 | 1831.57 | 0.17 | 0.35 | 441.58 |
| 54* | 173 | 84 (55) | 0.21 | 161.67 | 0.16 | -46.58 | .404 | 0.64 | 0.62 | 1785.73 | 0.30 | 0.39 | 1850.70 |
| 55* | 157 | 96 (67) | 0.10 | 120.21 | 0.07 | 44.83 | .176 | 0.46 | 0.78 | 1363.13 | 0.30 | 0.15 | 2218.45 |
| 57 | 159 | 136 (91) | 0.04 | 85.95 | -0.01 | 80.70 | .338 | 0.48 | 0.87 | 992.58 | 0.13 | 0.04 | 597.02 |
| 58* | 159 | 130 (88) | 0.01 | 46.20 | 0.07 | 25.31 | .013 | 0.84 | 0.82 | 826.69 | 0.17 | 0.12 | 197.07 |
| 59* | 164 | 105 (66) | 0.29 | 86.68 | 0.14 | 56.13 | .291 | 0.80 | 0.80 | 1425.79 | 0.14 | 0.47 | 2857.51 |
| 60 | 158 | 142 (93) | 0.13 | 81.63 | 0.04 | 15.85 | .199 | 0.74 | 0.88 | 790.94 | 0.23 | 0.05 | 231.78 |
| 61* | 157 | 109 (76) | 0.09 | 96.39 | 0.06 | 48.30 | .380 | 0.66 | 0.83 | 864.72 | 0.20 | 0.20 | 532.02 |
| 62 | 160 | 142 (91) | 0.06 | 15.64 | 0.00 | 50.15 | .144 | 0.78 | 0.90 | 775.69 | 0.20 | 0.20 | 346.89 |
| 63* | 159 | 102 (70) | 0.13 | 95.94 | 0.07 | 71.06 | .314 | 0.56 | 0.92 | 787.00 | 0.10 | 0.29 | 344.26 |
| 64* | 162 | 84 (55) | 0.39 | 209.67 | 0.27 | -15.11 | .267 | 0.44 | 0.62 | 1268.27 | 0.10 | 0.19 | 1469.04 |
| 65* | 158 | 90 (56) | 0.21 | 82.41 | 0.19 | -6.45 | .306 | 0.50 | 0.83 | 569.36 | 0.13 | 0.46 | 67.73 |
| 66* | 160 | 136 (89) | 0.09 | 65.71 | 0.03 | 20.25 | .261 | 0.28 | 0.90 | 735.04 | 0.20 | 0.10 | 675.92 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 67 | 160 | 183 (106) | 0.07 | 114.34 | 0.01 | 5.77 | .231 | 0.92 | 0.85 | 598.94 | 0.30 | -0.02 | 179.33 |
| 68 | 166 | 204 (110) | 0.01 | 21.76 | 0.01 | -2.18 | .056 | 0.88 | 0.88 | 549.68 | 0.23 | -0.09 | 224.46 |
| 69* | 165 | 105 (69) | 0.14 | 35.83 | 0.14 | 42.43 | .065 | 0.68 | 0.78 | 845.94 | 0.17 | 0.21 | 461.83 |
| 71 | 166 | 204 (110) | 0.06 | 37.69 | 0.06 | 13.26 | -.061 | 1.00 | 0.97 | 599.67 | 0.07 | 0.04 | 208.97 |
| 72 | 160 | 183 (106) | 0.03 | 17.23 | -0.03 | 47.42 | .148 | 0.62 | 0.92 | 750.40 | 0.10 | 0.09 | 204.56 |
| 73* | 160 | 136 (89) | 0.06 | 84.58 | -0.03 | 26.48 | .296 | 0.56 | 0.90 | 708.94 | 0.20 | -0.03 | 92.81 |
| 74 | 160 | 165 (100) | 0.01 | 8.17 | 0.04 | -24.84 | .009 | 0.80 | 0.98 | 548.15 | 0.03 | 0.15 | 123.09 |
| 75* | 167 | 136 (87) | 0.10 | 148.92 | 0.01 | 97.57 | .190 | 0.80 | 0.80 | 649.33 | 0.14 | -0.03 | 759.87 |
| 76 | 167 | 151 (93) | 0.06 | 87.69 | -0.03 | 44.19 | .381 | 0.66 | 0.85 | 674.39 | 0.30 | -0.02 | 415.26 |
| 77 | 159 | 165 (102) | 0.09 | 70.55 | 0.06 | 15.89 | .125 | 0.84 | 0.87 | 563.27 | 0.27 | 0.07 | 208.61 |
| 78 | 164 | 175 (101) | 0.09 | 34.20 | 0.00 | 23.03 | .296 | 0.92 | 0.90 | 580.76 | 0.20 | -0.03 | 137.67 |
| *M* | 163 | 126 | 0.11 | 83.84 | 0.08 | 39.50 | .204 | 0.67 | 0.82 | 911.36 | 0.20 | 0.72 | 723.66 |
| *SD* | 5 | 31 | 0.08 | 47.08 | 0.09 | 36.78 | .136 | 0.17 | 0.08 | 335.50 | 0.07 | 0.70 | 702.83 |

Note a) ID= Participant’s ID number, b) CA= Chronological age, c) RA= Reading Age, d) std = Standardised Scores on the Word Identification task of the WRMT, e) RT= Latency, f) IrregDiff= IrregularityDiff, g) NonACCDifff= NonwordAccuracyDiff

There are several points worth noting about these scores. First, the WLE scores (standardized B co-efficient; Panel F of Figure 2.8) from the Linguistic Properties Effect task showed good spread across individuals, with some people showing little or no evidence of any effect, and others showing quite large effects (see Appendix N for the full regression analysis calculating WLE scores). Second, although VA span scores (Panel A of Figure 2.8) also varied greatly across individuals, half of the participants fell within the impaired, or severely impaired ranges on this measure (see Table 2.6 for the impairment boundaries based on the oldest age norms, 10-12 years-old, from Coltheart and Leahy, 1996). This high incidence of VA span impairments may reflect our recruitment strategy, which could have favoured students likely to have difficulty reading. In both scores all participants were within the 95% confidence intervals.

Table 2.6 Impairment boundaries for accuracy scores on the sub-lists of the Coltheart and Leahy task, and the Partial Report task.

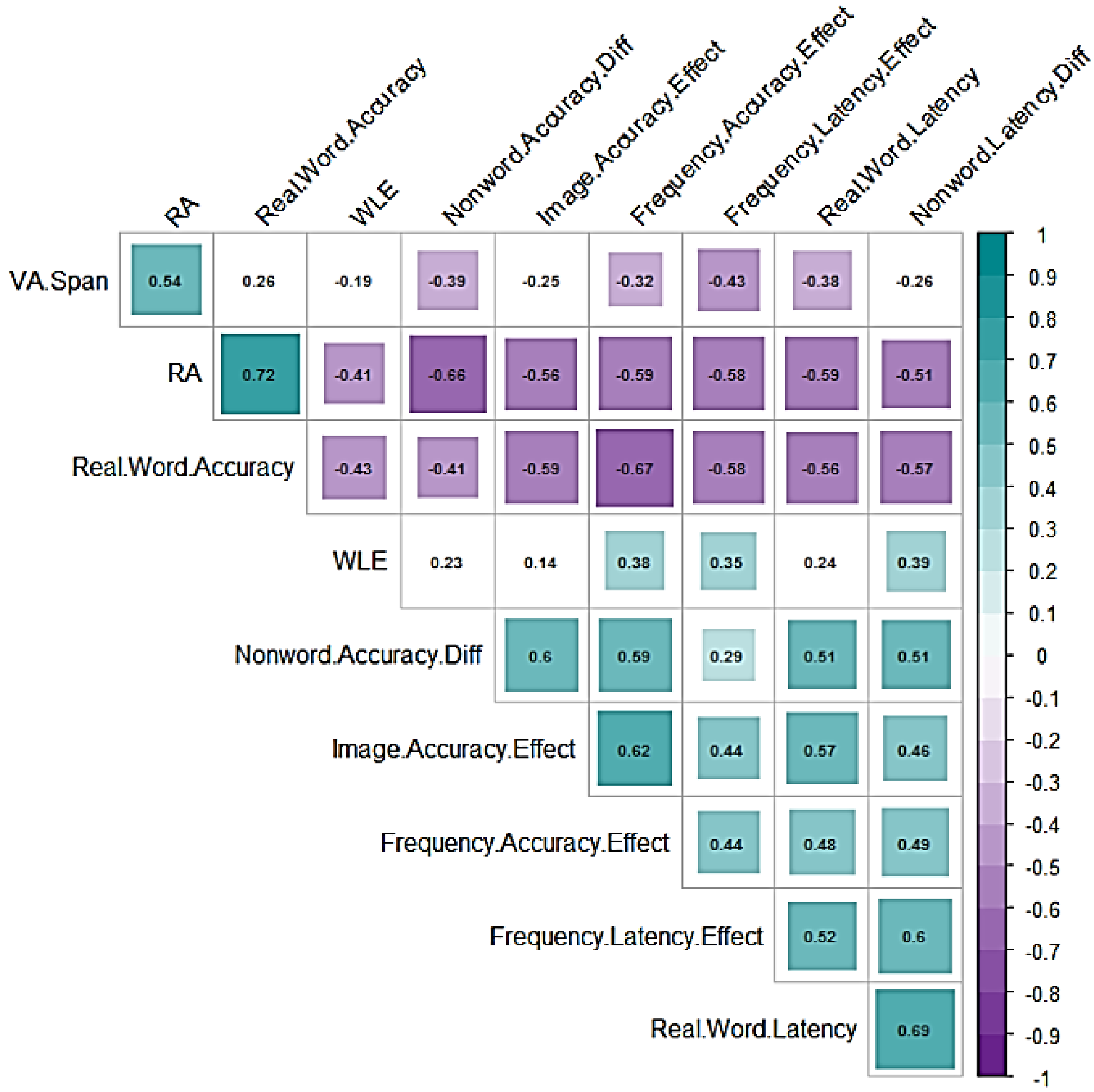
| Score | Unimpaired | Moderately impaired | Severely impaired |
|---|---|---|---|
| Regular Accuracy (/30) | 28+ | 26 – 27 | 0 – 25 |
| irregular Accuracy (/30) | 20+ | 16 – 19 | 0 – 18 |
| Nonword Accuracy (/30) | 22+ | 20 – 21 | 0 – 19 |
| VA span (/50) | 35+ | 29 – 34 | 0 – 28 |

A substantial number of our participants also scored poorly on the Coltheart and Leahy task. According to the age norms, 18 of the 49 participants were impaired on the regular word reading sub-list  (see Table 2.6 for the impairment boundaries, Appendix O provides all sub-list scores for each participant). On the irregular word list, 10 participants were impaired to some degree, and on the nonword list 27 participants were impaired. The real word accuracy violin plot (scores across regular and irregular words combined; Panel B of Figure 2.7) shows there was a good spread of scores, but a marked tail end (four individuals scored well below the 95% confidence interval). Reading latencies (Panel C of Figure 2.7) on this task were also highly variable across individuals, indeed a small number of individuals were extreme outliers ( five individuals scored well outside the 95% confidence intervals for the group).

The three additional difference scores that were obtained from the Coltheart and Leahy word task were: a) the supplementary whole-word processing index, the IrregularityDiff score (accuracy difference between irregular and regular words); b) the NonwordAccuracyDiff (accuracy difference between nonwords and real words); and c) the NonwordLatencyDiff (response latency difference between nonwords and real words). The IrregularityDiff scores (Panel D of Figure 2.7) had good spread (only two people scored outside the 95% confidence intervals). This figure shows that all participants read regular words more accurately than irregular words. The NonwordAccuracyDiff scores (Panel E of Figure 2.7) also had good spread, in fact, there was great variability on this measure: while the majority of participants tended to read real words more accurately than nonwords, a small cluster of participants (eight in total) actually performed marginally better on the nonwords compared to real words. The NonwordLatencyDiff scores (Panel F of Figure 2.7) had a large spread, with a marked upper-tail. All participants exhibited slower response times for nonwords than for real words.

**Correlation analysis.** Figure 2.9 presents the correlogram for the most important relationships between individual variables. Variables that yielded few or no significant correlations are not shown, namely: the IrregularityDiff, the supplementary whole-word index, and ImageLatencyEffect. See Appendix P for the complete correlation matrix.

Figure 2.9 Correlogram displaying the most important variables/relationships from the correlation analysis. Significant relationships ($p$ < .05) are represented by a coloured square, while insignificant relationships are left white. The strength of a relationship is depicted by the opaqueness and size of the coloured square. Positive relationships are green-scale, negative relationships are purple-scale.




Note: RA= Reading Age; WLE= word length effect

***Markers of phonologically-based reading.*** The primary measure of reading proficiency was Reading Age, with RealWord latency and Accuracy as secondary and tertiary measures respectively. Our first prediction was that, our key measure of phonological reading skills, NonwordAccuracyDiff, should be negatively associated with Reading Age and

RealWordAccuracy, and positively associated with RealWordLatency. This prediction was confirmed (see Figure 2.9). These finding indicates that the ability to utilise phonologically-based strategies is an important determinant of overall reading proficiency, even for adolescents. We did not predict any association between NonwordAccuracyDiff, our primary measure of phonological reading ability, and our primary index of whole-word reading ability, word length effect (WLE), or the supplementary index, IrregularityDiff. Indeed, no significant correlation between NonwordAccuracyDiff and either WLE or the IrregularityDiff was found, suggesting that both of the latter measures represent unique processes to that of the NonwordAccuracyDiff.

***Markers of whole-word reading.*** Our second prediction was that, our primary indicator of whole-word reading, WLE, should be negatively associated with Reading Age and RealWordAccuracy, and positively associated with RealWordLatency. This prediction was partially confirmed. WLE was negatively correlated with Reading Age and RealWordAccuracy; but the WLE and RealWordLatency relationship, although trending in the predicted direction, failed to reach significance. These findings are broadly consistent with the hypothesis that whole-word reading skills, as indexed by the WLE, are key determinants of overall reading proficiency in our sample of adolescent readers.

We did not make any formal predictions for the IrregularityDiff as it was merely supplementary measure. Nonetheless, given that this measure was a supplementary whole-word marker, we would have expected a negative association between this measure and both Reading Age and RealWordAccuracy, and a positive association with RealWordLatency. No significant correlations were found between IrregularityDiff and any of the reading proficiency measures. However, we did find a positive correlation between the IrregularityDiff and WLE, the two whole-word processing markers. In fact, WLE was the only measure to be significantly related to IrregularityDiff. Overall, these results suggest that IrregularityDiff was not the most suitable indicator of whole-word processes contributing to reading proficiency in our adolescent readers.

***VA Span.*** Our third prediction was that VA span would be positively associated with Reading Age and RealWordAccuracy, and negatively associated with RealWordLatency. This prediction was partially confirmed. VA span was positively correlated with Reading Age, and negatively correlated with RealWordLatency; but the VA span and RealWordAccuracy relationship, although trending in the predicted direction, did not reach statistical significance. These findings suggest that VA span is a determinant of overall reading proficiency.

Our forth prediction was that VA span would also be negatively associated with our primary measure of whole-word processing, the WLE, but not related to the phonological marker, NonwordAccuracyDiff. This prediction was not supported and the opposite pattern was found; VA span was not significantly correlated with WLE, however, there was a significant negative correlation between VA span and NonwordAccuracyDiff. The latter finding suggests that VA span may not index whole-word processing, nor is this measure entirely independent of those considered to index phonological reading ability.

*Considering the 'purity' of the VA span score.* One possible explanation for the correlation between NonwordAccuracyDiff and VA span is that VA span scores are influenced by the individual's phonological processing skills. For example, it may be that phonological processes are required for good performance on the VA span task, or alternatively, that excellent phonological processing skills somehow give a participant an "edge" on this task. To explore this possibility, we created a new measure, *PureSpan*, which expressed VA span scores as residuals, after regressing out variance shared with NonwordAccuracyDiff scores. Table 2.7 gives the key correlation findings. Overall, PureSpan was significantly and positively correlated with our primary measure of reading proficiency, Reading Age, but not our additional measure, RealWordLatency. PureSpan did not correlate with the WLE. These results suggest that the processes unique to the Partial Report task may be important to reading proficiency, however, to the extent that the WLE indexes parallel letter processing, these unique processes are unlikely to be directly related to whole-word processing skills as we have defined them here.

Table 2.7 Correlation co-efficient and significance scores (in parentheses) for the PureSpan score with the previous key scores. *Significant relationships at the $p<.05$ level.

| Measure | PureSpan |
|---|---|
| Reading Age | .304* (*p* =.034) |
| RealWordLatency | -.193 (*p* =.185) |
| WLE | -.112 (*p* =.443) |

***Supplementary analysis of Reading Age measure.*** Our primary measure of overall reading proficiency was absolute Reading Age, as assessed by the Woodcock Reading Mastery test. Thus, we used an *absolute* measure of proficiency, rather than a measure of proficiency *relative* to age expectations. This is a significant concern. Since there was considerable variability in our participants' chronological age, variability in the Reading Age measure could have simply reflected participants' chronological age, rather than proficiency relative to age expectations. To check this possibility, we calculated a difference score for each individual, by subtracting each participant's Reading Age from their chronological age and repeating the above analysis. The pattern of results obtained using this difference measure were almost identical to those reported above for the absolute Reading Age measure (this relative measure is included in the correlation matrix in Appendix P). Indeed, the new difference score and our original Reading Age measure were themselves extremely highly correlated with one another ($r$= .987). Therefore, in all subsequent analyses, we continue to present findings for the absolute Reading Age measure. The difference score will not be discussed or used further[6].

***Summary of correlation results.*** Both phonological and whole-word markers were associated with measures of general reading proficiency. Furthermore, there is evidence of some dissociation between indices of phonological and whole-word reading skills respectively. Namely, scores on the NonwordAccuracyDiff measure were not found to be significantly correlated with scores on the WLE measure. However, several key predictions relating to the VA span measure were not supported. Contrary to our prediction, there was no relationship between VA span and our WLE measure. This does not appear to be due to a statistical artefact: there was good variability in the distribution of scores on both the VA span and the WLE measures, and little evidence of significant skewness. In addition, and contrary to predictions again, scores on the NonwordAccuracyDiff *were* found to be negatively correlated with VA span scores. To further explore these possibilities, several regression models were created and are discussed below.

**Regression analysis.** To assess the contributions of phonological and whole-word processing to reading proficiency, several regression models were run. In the first set of regression models, we use the NonwordAccuracyDiff as a marker of phonological

[6] Standardised scores for the Word Identification task (presented in Table 2.5) were also entered as an alternative reading proficiency measure. Once again, the pattern of results were identical to those obtained when using Reading Age.

processing, and the WLE as the marker of whole-word processing. Reading Age and RealWordLatency, our key measures of overall reading proficiency, interchangeably served as the dependent variables. RealWordAccuracy was not used as a dependent variable, because this measure was not fully independent of NonwordAccuracyDiff. The former enters into the calculation of the latter. In the second set of regression models we again used the NonwordAccuracyDiff as a marker of phonological processing, but this time the VA span was the marker of whole-word processing, with the same dependent measures as before.

To assess collinearity among predictor variables, we calculated Variance Inflation Factor scores for each multiple regression analysis (James, Witten, Hastie, & Tibshirani, 2013). A score close to 1 indicates the predictor variables are fully independent; scores between 1 and 5 are considered acceptable; scores above 5 indicate substantial collinearity. Scores for all models were under 1.5 indicating good independence of predictor variables within the model.

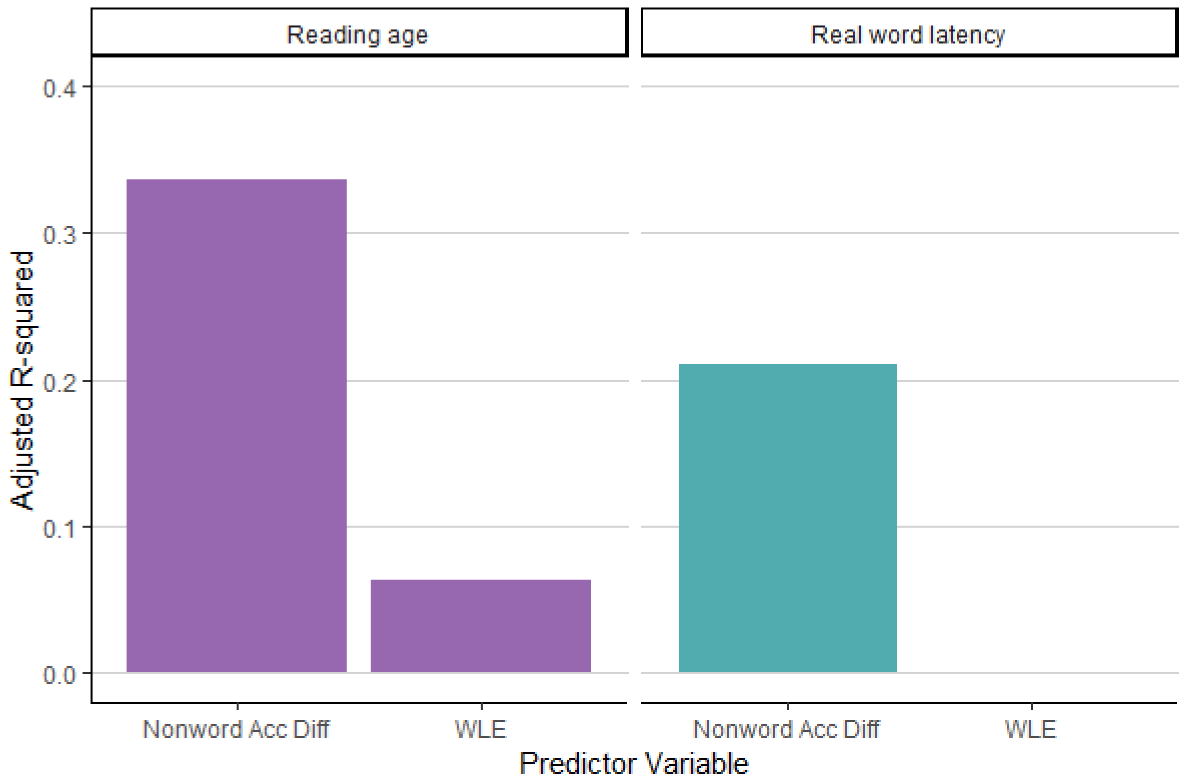
***The contributions of the NonwordAccuracyDiff and WLE to overall reading proficiency.*** Table 2.8 contains the results of the regression analysis for the contributions of NonwordAccuracyDiff and the WLE to overall reading proficiency, as indexed by Reading Age and RealWordLatency. The table presents adjusted $R^2$ scores for the combined model, and also for the respective ‘simple’ regression models (which incorporate only one predictor variable at a time). Figure 2.10 plots the difference in adjusted $R^2$ between the full regression model, and each of these simple models respectively. These values provide an indication of the unique contribution of each predictor to variance in Reading Age and RealWordLatency scores.

Table 2.8 Regression output for the NonwordAccuracyDiff / WLE model. **Significant at the *p* <.01 level. *Significant at the *p* <.05 level.

| Outcome | Model | Model Significance | AR[b] | 95% CI; |
|---|---|---|---|---|
| Reading Age | Combo model | < .001** (NAD[a] <.001**; WLE= .012*) | .488 | NAD[a] : (-175.39) - (-83.15); WLE: (-111.02) – (-14.28) |
| | NAD[a] | <.001 ** | .425 | (-190.40) – (-95.24) |
| | WLE | .003 ** | .152 | (-154.01) – (32.83) |
| Real word latency | Combo model | < .001 ** (NAD[a] <.001**; WLE= .307) | .249 | NAD[a] :(526.07) - (1739.42); WLE: (-309.56) – (96.05) |
| | NAD[a] | < .001 ** | .248 | (612.49) – (1794.37) |
| | WLE | .093 | .039 | (-104.18) – (1296.85) |

Note: a) NAD= NonwordAccuracyDiff. b) AR= Adjusted $R^2$

Figure 2.10 Adjusted R2 difference scores the NonwordAccuracyDiff / WLE model.



*Reading Age.* The model including the NonwordAccuracyDiff and WLE as joint predictors accounted for 48.8% of the variance in Reading Age scores. Difference scores showed that the NonwordAccuracyDiff uniquely accounted for 33.6% of the variance after taking into account the WLE. Conversely, the WLE uniquely accounted for 6.3% of the variance, after taking into account the effect of the NonwordAccuracyDiff. In sum, the model that contained *both* the NonwordAccuracyDiff and the WLE, provided a better fit for predicting reading proficiency, as indexed by our Reading Age measure, compared to simple models that contained each of these measures separately. This finding supports assertions that there are multiple determinants of reading proficiency, and that a model that considers only phonological processing skills does not provide a complete account of the variation in reading proficiency, at least not in the adolescent sample tested here. When considered independently, both the NonwordAccuracyDiff and the WLE contributed uniquely to Reading Age; however, the NonwordAccuracyDiff accounted for substantially more variance than the WLE. These results suggest that parallel letter processing, as indexed by the WLE, is important to accurate single word reading in adolescents, however, phonological processing, as indexed by the NonwordAccuracyDiff, is more important to this task.

*RealWordLatency.* In general, the pattern of results for RealWordLatency were the same as those for Reading Age. However, the full model accounted for considerably less of the variance in RealWordLatency scores (24.9%) than it had for the Reading Age scores. Difference scores show that, while the NonwordAccuracyDiff uniquely contributed to 21.0% of the variance, the unique contribution of WLE was negligible and did not reach statistical significance. This latter finding is somewhat surprising, given that the WLE is measure of parallel letter processing, which, by definition, should enable faster reading times. Our results suggest that, parallel letter processing as indexed by WLE, is not crucial to reading speed unlike phonological processing, as indexed by NonwordAccuracyDiff.

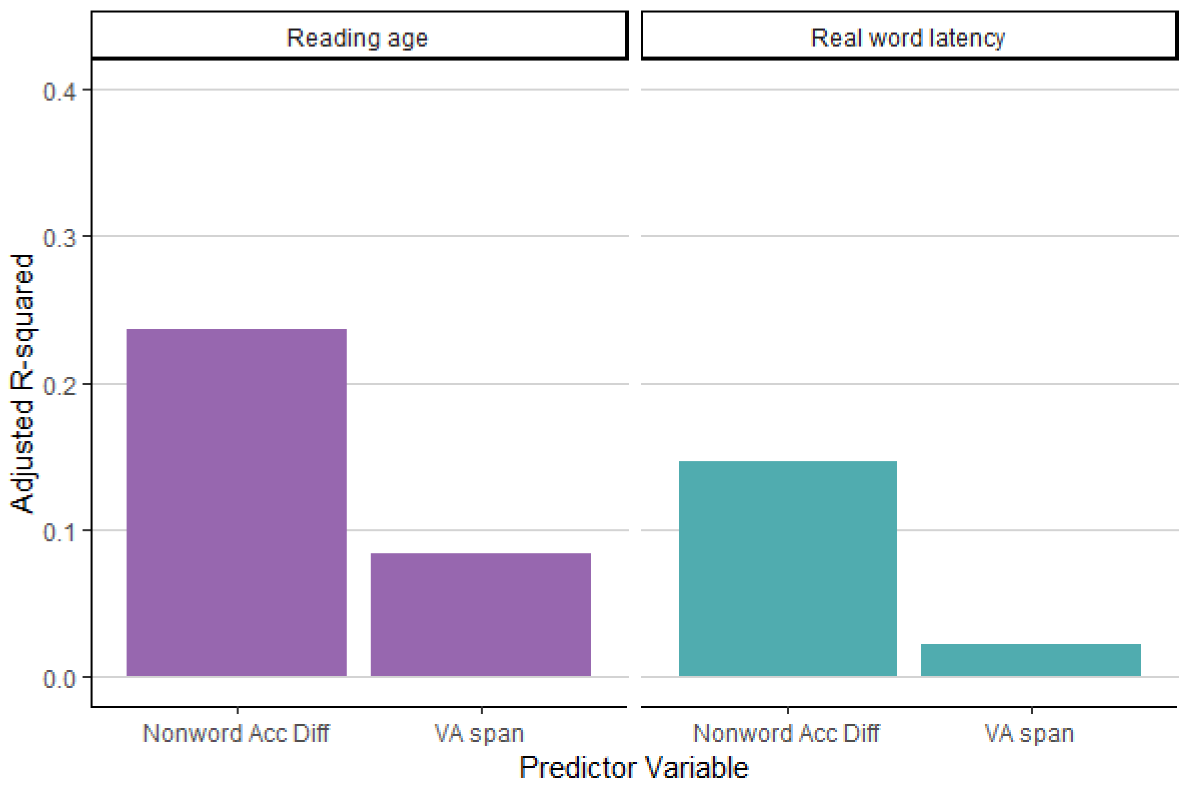
**The contributions of VA span and the NonwordAccuracyDiff to overall reading proficiency.** In this set of analyses, instead of using WLE scores as our primary index of lexical processing, we used VA span scores. We then repeated the analyses specified above. Table 2.9 contains the results of regression analyses looking at the contribution of these two measures to overall reading proficiency. Figure 2.11 shows the Adjusted $R^2$ difference scores.

Table 2.9 Regression output for NonwordAccuracyDiff / VA span model. **Significant at the *p* <.01 level. *Significant at the *p* <.05 level

| Outcome | Model | Model Significance | AR[b] | 95% CI; |
|---|---|---|---|---|
| Reading Age | Combo model | < .001 ** (VA= .004 **; NAD[a] < .001**) | .509 | VA: (19.69) – (19.93); NAD[a] : (-162.92) - (-67.39) |
| | NAD[a] | <.001 ** | .425 | (-190.40) – (-95.24) |
| | VA | < .001 ** | .273 | (52.45) – (142.35) |
| Real word latency | Combo model | < .001 ** (VA=.125; NAD[a] = .002**) | .270 | VA: (-942.81) – (119.26); NAD[a] : (380.82) - (1645.08) |
| | NAD[a] | < .001 ** | .248 | (612.49) – (1794.37) |
| | VA | .007** | .124 | (52.45) – (142.35) |

Note: a) NAD= NonwordAccuracyDiff. b) AR= Adjusted $R^2$

Figure 2.11 Adjusted R2 difference scores NonwordAccuracyDiff / VA span model.



*Reading Age.* For both analyses, the pattern of results was similar to those reported above. The full model that included both VA span and NonwordAccuracyDiff accounted for 50.9% of the total variance in Reading Age. Difference scores show that the NonwordAccuracyDiff uniquely contributed to 23.6% of the variance, after accounting for the variance shared with VA span. Conversely, VA span uniquely accounted for 8.4% of the variance, after accounting for the variance shared with NonwordAccuracyDiff. Once again, the large amount of variance explained by the predictor variables in the combined model is primarily due to the NonwordAccuracyDiff, however, the VA span by itself also explains a meaningful amount of variance.

*RealWordLatency.* As with the NonwordAccuracyDiff and WLE model, the full NonwordAccuracyDiff and VA span model accounted for substantially less variance in RealWordLatency (27.0%) than it had for Reading Age. However, unlike the NonwordAccuracyDiff and WLE model, both the NonwordAccuracyDiff and VA span were unique, significant predictors of RealWordLatency. Difference scores show that the NonwordAccuracyDiff uniquely contributed to 14.6% of the variance, while VA span uniquely contributed to 2.2% of the variance. Again, this result suggests that the large amount of variance explained by the predictor variables in the combined iteration of the model, is primarily due to the NonwordAccuracyDiff, but including VA span still has some benefit.

In sum then, a model that contained *both* the VA span and NonwordAccuracyDiff provided a better fit for predicting either Reading Age or RealWordLatency, than models that contained these measures separately. This supports assertions that reading involves multiple, diverging mechanisms. Altogether, these results support the assertion that measures of phonological skills alone do not provide a full explanation of the variability in reading proficiency in our adolescent sample. Instead VA span scores capture some additional skills that are crucial for accurate and efficient reading. This finding is consistent with multi-route models of reading, such as the VA span deficit hypothesis (Bosse et al ., 2007).

In general, the models that used Reading Age as the primary dependent measure account for more variance than those that used RealWordLatency as the dependent measure. This was an unexpected outcome. We had thought that RealWordLatency might be a particularly sensitive index of lexical reading skills, because low scores on this measure would suggest an ability to recognise familiar words at a glance. It is possible that other factors not considered here come into play in measures of reading latency. For example, some individuals may be generally faster responders, while others may adopt a more cautious approach to reduce the possibility of errors. There could also be some practice effects across the session, or within the individual tasks which affect the reliability of latency scores. Despite the slightly weaker outcomes for RealWordLatency, the measure is still beneficial to consider alongside Reading Age the primary indicator of reading proficiency. Together, they provide a more comprehensive indication of reading proficiency than any one alone.

**Analysis of Error Patterns.** One additional way we considered might offer insights into the cognitive mechanisms contributing to reading proficiency was to examine the types of errors a person makes while reading aloud. For example, a tendency to lexicalise nonsense words and read them as similar-sounding real words (i.e. *vun* > "van") may suggest a difficulty with phonological processes, and a reliance on whole-word processes. Conversely, a tendency to regularise irregularly-spelled words (e.g. *blood* > "blude") may suggest a difficulty with whole-word processing, and consequently, an overly heavy reliance on phonological processes to support reading. To assess this hypothesis, we calculated two scores for each individual: the *Lexicalisation Error Proportion* and the *Regularisation Error Proportion*. The Lexicalisation Error Proportion score was the proportion of nonwords that were lexicalised. The Regularisation Error Proportion score was the proportion of irregular words that were regularised. Both scores were created using data from the Coltheart and Leahy task (a full description of possible errors and the definitions used for coding are included in Appendix Q).

First, in order to assess the reliability of our error classification scheme, two independent judges coded the error output of each participant (i.e. incorrectly read trials), and an inter-rater reliability score was calculated via the "*kappa2*" function (from the *irr* package). A good agreement between the two judges was found ($\kappa = .686$, $p < .001$), however we were concerned that the inclusion of errors other than Lexicalisation and Regularisation was weakening the reliability, given that this kappa would also measure the reliability of agreement on additional errors of less interest. To address this issue we re-ran the analysis, looking at the agreement for Lexicalisation and Regularisation errors separately, having marked these errors as either present or absent, to try and reduce noise (i.e. every error not coded as a Lexicalisation error was considered an absence of Lexicalisation error). We found this manipulation did not have any meaningful effect on the inter-rater reliability (Lexicalisation error: $\kappa = .683$, $p < .001$, Regularisation error: $\kappa = .660$, $p < .001$). These results suggest that, despite our best efforts, the definition of these errors, taken from the literature, are somewhat open to interpretation. This subjectivity is important to remember as the subsequent individual analyses utilised coding data from only one judge (the "primary"). We selected the judge that coded less of the participants' errors as "Unknown", as we felt it may represent a better comprehension of the errors (Appendix R shows the breakdown of error coding for each of the two independent coders).

Figure 2.12 shows the distributions of scores for each of the two key error measures. There was a good distribution of scores for the Lexicalisation Error Proportion, with almost all participants producing at least some lexicalisation errors. For the Regularisation Error Proportion, scores were less broadly distributed, and a number of individuals scored at zero. This may be confounded by the lower level of group accuracy on the nonword sub-list (*M* = .667) compared to the irregular word sub-list (*M* = .727), which may afford more opportunity for lexicalisation.

Figure 2.12 Lexicalisation Error Proportion and the Regularisation Error Proportion violin plots. Black dots represent individual participants, yellow dots represent group mean, with 95% confidence intervals (black vertical line). Box plots show the median (black central line), upper and lower quartiles.

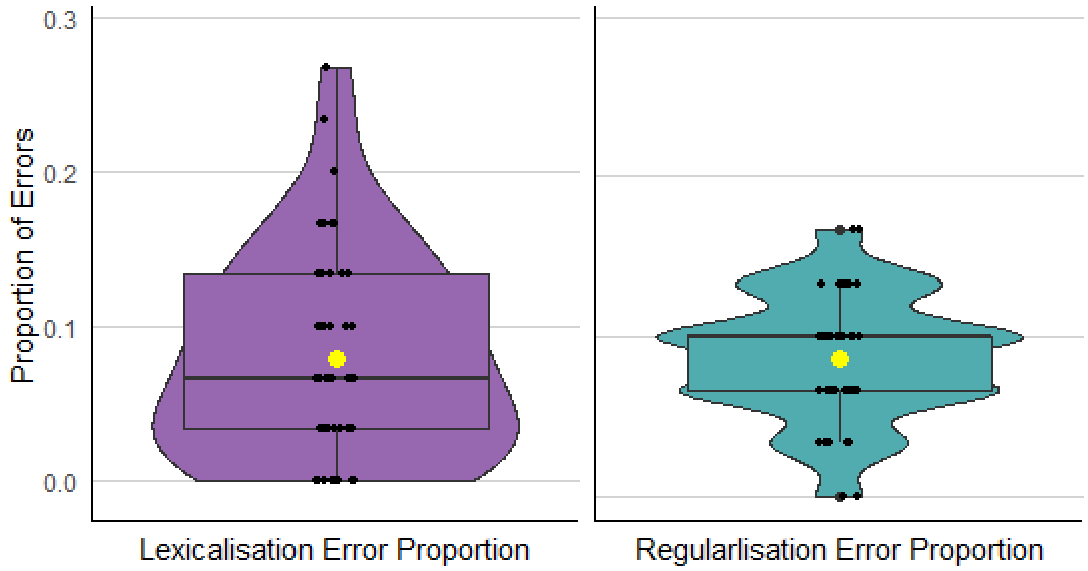



Table 2.10 shows the results of the correlation analyses for the two error measures. The Lexicalisation Error Proportion correlated with many of the other scores including, but not limited to, a negative correlation with Reading Age and a positive correlation with RealWordLatency, however, it did not correlate with either the WLE or the VA span (we did

not perform a correlation with NonwordAccuracyDiff, because this measure was derived from the same sub-list and was therefore not independent). Given that this measure is closely related to our key measure, NonwordAccuracyDiff, these findings are not surprising and suggest the Lexicalisation Error Proportion maybe a suitable marker of phonological processing.

Table 2.10 Correlation co-efficients (significance scores in parentheses) for the Lexicalisations and Regularisation Error Proportions, with each of the previous scores. **Significant relationships at the *p* <.01 level. *Significant relationships at the *p* <.05 level.

| | Lexicalisation Error Proportion | Regularisation Error Proportion |
|---|---|---|
| Reading Age | -.641** (*p* < .001) | .158 (*p* = .227) |
| ImageEffectACC | .646** (*p* < .001) | -.109 (*p* = .455) |
| FreqEffectACC | .716** (*p* < .001) | -.207 (*p* = .154) |
| ImageEffectRT | -.075 (.606) | .023 (.875) |
| FreqEffectRT | .454** (*p* = .001) | -.070 (*p* = .631) |
| WLE | .272 (*p* = .059) | .008 (*p* = .549) |
| VA span | -.280 (*p* = .051) | .288* (*p* = .045) |
| RealWordAccuracy | -.627 ** (*p* < .001) | *na* |
| RealWordLatency | .567 ** (*p* < .001) | -.134 (*p* = .358) |
| IrregularityAccuracyDiff | -.105 (.473) | *na* |
| NonwordAccuracyDiff | *na* | -.406** (*p* = .004) |
| NonwordLatencyDiff | .596** (*p* < .001) | -.045 (*p* = .759) |

The Regularisation Error Proportion did not correlate with either of the reading proficiency measures, Reading Age or RealWordLatency. The Regularisation Error Proportion was negatively correlated with NonwordAccuracyDiff (again we did not perform a correlation with RealWordAccuracy or IrregularityAccuracyDiff, because these measures were derived from the same sub-lists). These findings suggest that high rates of regularisation errors were associated with lower NonwordAccuracyDiff, and thus a tendency to regularise irregular words was associated with good nonword reading. This pattern suggests an increased dependence on phonological processing when irregular word reading is impaired. Curiously, the Regularisation Error Proportion was positively correlated with VA span, suggesting high rates of regularisation errors were actually associated with *higher* scores on the VA span, and thus a tendency to regularise was associated with VA span performance. This latter finding in particular seems counter intuitive.

Prior to performing regression analyses. We calculated Variance Inflation Factor scores to assess collinearity among predictor variables (James et al., 2013). Once again, the score was below 1.5. Table 2.11 contains the results of regression analyses looking at the contribution of these two measures to overall reading proficiency. Figure 2.13 shows the Adjusted $R^2$ difference scores.

*Reading Age.* The full model that included both Lexicalisation, and Regularisation Error Proportions, accounted for 38.6% of the total variance in Reading Age scores. Difference scores showed that only the Lexicalisation Error Proportion was a significant unique predictor, accounting for 38.2% of the variance in Reading Age, after the variance shared with Regularisation Error Proportion had been accounted for.

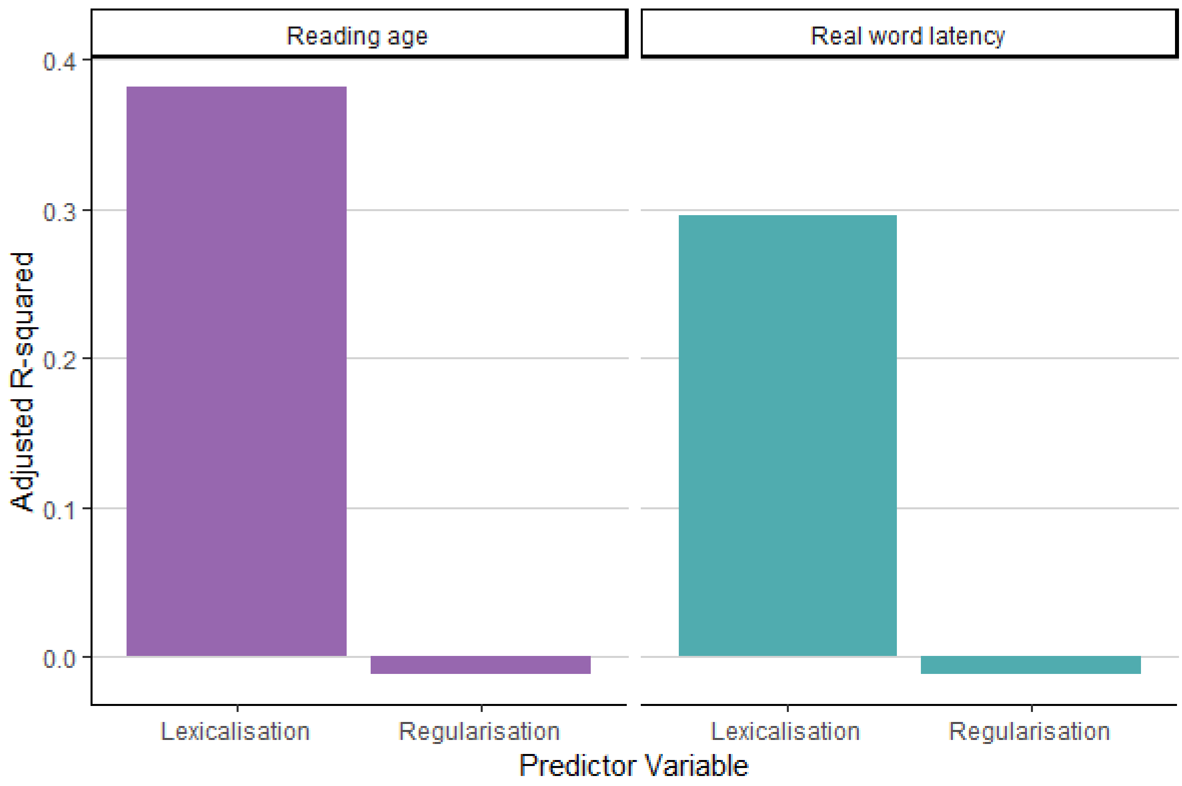
*RealWordLatency.* The full model accounted for 29.3% of the variance in RealWordLatency. Again, only the Lexicalisation Error Proportion was a unique, significant predictor of RealWordLatency. Difference scores show that Lexicalisation Error Proportion uniquely contributed to 29.6% of the variance. In sum then, when considered independently, only the Lexicalisation Error Proportion offered unique contributions to our key measures of reading proficiency.

Table 2.11 Regression output for LEPa / REPb. **Significant at the p<.01 level.

| Outcome | Model | Model Significance | AR[b] | 95% CI; |
|---|---|---|---|---|
| Reading Age | Combo model | < .001 ** (LEP[a] < .001**; REP[b] =.785) | .386 | LEP[a] : (-413.07) – (-191.59); REP[b] : (-160.73) - (211.52) |
| | LEP[a] | <.001 ** | .398 | (-412.73) – (-197.97) |
| | REP[b] | .277 | .004 | (-105.25) – (358.91) |
| Real word latency | Combo model | < .001 ** ( LEP[a] < .001**; REP[b] =.861) | .293 | LEP[a] : (1622.15) – (4202.83); REP[b] : (-2358.04) - (1979.21) |
| | LEP[a] | <.001 ** | .307 | (1684.44) – (4185.53) |
| | REP[b] | .358 | -.003 | (-3695.48) – (1362.31) |

Note: a) LEP= Lexicalisation Error Proportion. b) REP= Regularisation Error Proportion.

Figure 2.13 Adjusted R2 difference scores LEPa / REPb.



## Discussion

**Summary of key hypotheses and findings.** In this study, we looked at the cognitive processes contributing to reading proficiency in adolescents, and the relationships between markers of these different processes. Specifically, we looked at three key measures. The first was a measure of nonword accuracy (accuracy difference between nonwords and real words), which we used as a marker of phonological processing. The second was word length effect (the degree to which reading latencies increase as a function of word length), which we used as a marker of whole-word processing. The third was a measure of visual attention (VA span) that, according to the VA span deficit hypothesis, should also be a significant determinant of whole-word processing skills (Bosse et al., 2007).

Within the current sample there was a good spread of scores on each of these three key measures, including a good representation of lower scores (which was to be expected, given the initial recruitment targeted impaired readers). We also found several group patterns on these key tasks were consistent with previous findings. First, on the Partial Report task assessing VA span, we found the typical W-shaped serial response curve that has been widely reported for both unimpaired readers, and impaired readers with relatively intact whole-word skills (Antzaka et al., 2017; Awadh et al., 2016; Banfi et al., 2018; Lallier et al., 2018). That is, such identification accuracy is high for stimuli in the medial and most lateral positions, but is attenuated in positions just off-centre. Second, in the Linguistic Properties Effect task, we again replicated findings from studies that have found reading accuracy and latencies are determined by word length (Barton, Hashim, Björnström, & Hills, 2014; De Luca, Di Pace, Judica, Spinelli, & Zoccolotti, 1999; Marinus & de Jong, 2010; Spinelli et al., 2005; Ziegler et al., 2003; Zoccolotti et al., 2005), frequency (Balota et al., 2013; Cosky, 1976; Martens & de Jong, 2006; Weekes, 1997), and imageability (Gvion & Friedmann, 2013; Stadthagen-Gonzalez & Davis, 2006).

In the main analyses, we looked at the relationships between each individual's scores on the three key measures, and scores on various other measures of reading and other related cognitive skills. As predicted, all three key measures – NonwordAccuracyDiff, word length effect and VA span – correlated with our primary measure of reading proficiency, Reading Age. NonwordAccuracyDiff and VA span – but not word length effect – also correlated with our secondary measure of reading proficiency, RealWordLatency. However, contrary to our predictions, there was no correlation between VA span and our primary measure of whole-word processing, word length effect. Instead, the VA span measure correlated with the phonological processing marker, NonwordAccuracyDiff. Although we did not predict a relationship between these two variables, a couple of other studies found that accuracy and speed on phonemic awareness tasks were related to VA span (Lassus-Sangosse, N'guyen-Morel, & Valdois, 2008; Lewandowska, Milner, Ganc, Wlodarczyk, & Skarzynski, 2014; Saksida et al., 2016).

In the Regression analyses, we examined the combined and unique contributions of these key measures to overall reading proficiency. As predicted, all three key measures were unique predictors of reading proficiency. We found that a model containing both NonwordAccuracyDiff *and* the word length effect was a better predictor of Reading Age than models that contained only one of these predictors. The NonwordAccuracyDiff was the largest unique contributor to variance in Reading Age, but word length effect was also a significant unique contributor, and results were similar when the word length effect was replaced with VA span. When we repeated these analyses using RealWordLatency as the measure of reading proficiency, the results were similar to the outcomes for Reading Age, except that the unique contribution of word length effect was not significant while the contributions of the other measures (NonwordAccuracyDiff and VA span) were.

**Methodological limitations.** One possible concern with our findings is that scores for the word length effect, although well-distributed, did not reach the upper limits. That is, no participant showed a word length effect score high enough to indicate a complete inability to process letters in parallel. In contrast, VA span scores were widely distributed, with some participants showing a score low enough to indicate an almost complete inability to capture multiple elements in one "window". It could be interpreted that the word length effect was a less sensitive measure than VA span, however we often see that participants show moderate, rather than extreme, length effects (Balota et al., 2013; Cosky, 1976; Frederiksen & Kroll, 1976; Weekes, 1997). Even when an individual's parallel letter processing skills are severely limited, they can still capitalise on some letter redundancy to read highly familiar words. For example, the word *barbecue* will be highly familiar to all our adolescent participants, so it is unlikely they will need to scan every single letter to identify it. In designing this task, we wished to create a word list which systematically varied word length in letters, while controlling for other key variables known to influence reading latencies – specifically, word frequency, imageability and neighbourhood size. Nonetheless, as a group, our participants benefited from high frequency words compared to low frequency words, irrelevant of word length. Future studies could use nonwords similar to Juphard et al. (2004) to prevent compensation through familiarity and create an uncontaminated measure of parallel letter processing.

Future iterations of the word length effect measure could also be calculated slightly differently to: a) further avoid confound between length and other word properties (i.e. frequency and imageability); and b) ensure that the calculation of the effect considers individual variance in overall reading speed. First, three and nine letter words may be excluded from the analysis (or the test completely) in future, as these lengths have the most dramatic effect on frequency and imageability. Second, the increase in latency as a function of length could be calculated as a percentage. In doing so, two individuals, one with a short-word latency of 500ms and long-word latency of 1000ms, and another with a short-word latency of 1000ms and long-word latency of 2000ms, would have the same estimated word length effect –a 100% increase in latency as a function of length. As it is currently calculated, the word length effect for the latter individual would be greater than the former, even though the *relative* increase is identical.

Similarly, we should also consider that latency-based scores from the Coltheart and Leahy task may have been affected by the large amount of incorrect trials that were removed. There is a possibility that particularly poor readers would not have read enough words accurately to give a fair indication of their reading speeds. To address this issue, the task could be expanded using other standardised lists such as the Phonological Generalisation Test created by Rowse and Wilshire (2007), or particularly poor cases could be excluded from analysis. We are less inclined to employ this latter technique because: a) variation is beneficial for correlation analysis; and b) these extreme cases are still valuable in research of this nature.

It should also be borne in mind that our participant sample was likely to contain an over-representation of poor readers, an issue which may compound with the small sample size (which also increases the potential for a Type II error). The removal of unimpaired readers from the current analyses was considered. While we could not remove participants based on Reading Age or another measure of reading proficiency as this may have biased the results, however we could have excluded all those from non-reading remediation classes. However, one of the benefits of the non-reading remediation classes is they are completely unselected individuals who may or may not struggle with reading, thus increasing the variance in the sample. An ideal sample would include a representative number of readers of all proficiency levels, however, in our study, we found it difficult to recruit participants who did not feel that they struggled with reading. If more unimpaired readers could be recruited, future research could use a design that directly compared this group to impaired readers.

Our research also suggests that reading latency is a less sensitive measure of overall reading proficiency than accuracy, and, surprisingly, was not associated with whole-word reading processes, such as the word length effect. To the extent that the word length effect reflects a reader's reliance on serial processing, we would expect those with large word length effects to be slower readers overall. Meanwhile, someone who can rely on parallel letter processing and would not demonstrate a word length effect, should be generally fast at reading real words. Indeed, slow reading speed has often been affiliated with whole-word impairments (Woodhead et al., 2013; Woollams et al., 2007; Zabell & Everatt, 2002). It is possible that the shared variance between the word length effect and RealWordLatency was removed when we calculated the slope of the word length effect, meaning that general reading speed was effectively factored out. However, we specifically wanted to see whether a measure of parallel letter processing could predict reading latencies. Therefore, our calculation of word length effect slope created a score that enabled us to measure this processing without any noise from general reading speed (i.e. consistently slow readers were not biased against). Instead then, our findings could simply reflect the fact that there was not enough power in the current study (indeed, the correlation co-efficient for the word length effect and RealWordLatency trended towards significance).

**Considering the VA span.** We were concerned that there may be phonological demands within the Partial Report task that introduced noise into this measure. Despite efforts to ensure the independence of measures indexing the different cognitive functions, we are aware it is unlikely for these to be entirely separate of one another. We tried to explore this possibility by creating a new measure, *PureSpan,* which regressed out any shared variance between VA span and our key phonological measure, NonwordAccuracyDiff. As with the original VA span score, we found that the PureSpan score correlated with our primary measure of reading proficiency, Reading Age, but not our primary whole-word processing measure, the word length effect. Overall, in our study visual attention was consistently connected to reading proficiency, however, it these processes may not inherently linked to parallel letter processing. As we could not measure the association between

PureSpan and our measure of phonological processing, future studies should include additional indices of phonological processing, such as scores on phonemic awareness tasks.

We also administered a Span task which was less confounded by the phonological requirements of the Partial Report task, based off a task by Lobier, Zoubrinetzky, and Valdois (2012). We did not include this study in the formal analysis as only half of all participants completed the task due time constraints (a brief run-down of the methods and results can be found in Appendix S). The task had two conditions: Verbal and Non-verbal categorisation. The Verbal categorisation condition used letter or number stimuli, while the Non-verbal categorisation used non-orthographic stimuli (i.e. meaningless symbols and shapes), and in both conditions condition participants respond via keyboard. Lobier et al (2012) found a positive association between Verbal and Non-verbal categorisation scores, suggesting the two measures captured similar (VA) processes. Moreover, the Non-verbal categorisation scores were positively associated with measures of overall reading proficiency (reading age and reading speed). In our study, we examined whether the non-verbal span measure correlated with our key measures of phonological and whole-word processing. We also found a positive association between the Verbal and Non-verbal components of the task; however, the Non-verbal measure was not associated with our main measure of VA span, nor was it associated with any of our general or specific reading indices (Reading Age, NonwordAccuracyDiff or word length effect). These findings are consistent with previous research and critique (Catts, Compton, Tomblin, & Bridges, 2012; Hawelka & Wimmer, 2008; Leach, Scarborough, & Rescorla, 2003; Ziegler, Pech-Georgel, Dufau, & Grainger, 2010), and suggest that a measure of VA span may not index whole-word processes. We return to this point in the General Discussion.

**Practical implications.** The heterogeneity of reading impairments has important practical implications relevant to the development of reading intervention programmes. Many programmes aimed at children and adolescents that struggle with reading are focused on the acquisition and application of sight-sound rules (Ablinger et al., 2014; Ball & Blachman, 1988; Broom & Doctor, 1995a; Lovett et al., 1994; Torgesen, Morgan, & Davis, 1992; Torgesen, Wagner, Rashotte, Herron, & Lindamood, 2010; Williams, 1980, 1981). This approach would not only be unsuitable for those for whom phonological processing is not impaired (i.e. those with selective whole-word impairments), but may also limit the efficacy for the many who experience difficulties originating from *both* phonological and whole-word processing impairments. This will be covered in detail in the subsequent chapters outlining our research on intervention specificity. For now, the important conclusion is that heterogeneity is not simply a behavioural phenomenon, but also a cognitive one that must be taken into account throughout the development of reading models, empirical research and therapeutic approaches.

The current study had two primary aims. First, we investigated whether whole-word processing skills contribute to overall reading proficiency, above and beyond the contribution of phonological processing skills. We predicted that a key index of whole-word processing would account for a unique portion of variance in our measure of reading proficiency, independent of phonological processing demonstrating the importance of multiple cognitive processes in reading development. A second aim of the current study was to investigate whether a VA window could also explain overall reading proficiency, in particular, whole-word level reading skills. We predicted that an index of VA would account for a unique proportion of variance in our measure of reading proficiency and demonstrate an association with our index of whole-word processing. These hypotheses were partially upheld; our index of VA was found to account for a unique portion of variance in overall reading proficiency, independent of phonological processing, however VA was not associated with whole-word processing. Instead, VA was associated with phonological processing. Our findings are not consistent with the VA span deficit hypothesis, the implications for which will be discussed further in the General Discussion.

# Chapter 3 : Pre-intervention screening and profile building

In Chapter 2, we found that multiple cognitive factors were unique contributors to reading proficiency. Specifically, these factors were: a measure of nonword reading accuracy (a common index of phonological processing); and the degree of change in reading latencies as a function of word length, known as the word length effect (a measure that appears to capture whole-word level processing). These findings suggest that developmental dyslexia may have different cognitive underpinnings in different individuals and therefore ideally, reading interventions should be designed to target the specific problems identified in each case. Such tailoring maybe particularly important for whole-word cases, for whom an intervention targeting GPC skills may not be sufficient. In this study, we set out to test whether targeting an intervention tailored to an individual's specific cognitive difficulties would lead to better outcomes than applying a one-size-fits-all approach. To recruit individuals for this Intervention study (presented in Chapter 4), we first examined a large sample of adolescents (47 in total) who were known or suspected to be experiencing reading difficulties. In this Screening phase, we assessed some of the same aspects of participants' reading and other related cognitive skills as we did in the Heterogeneity study. These included: 1) a measure of reading proficiency, *Reading Age*, using the word identification sub-test of the Woodcock Reading Mastery Test III (Woodcock, 2011); 2) a measure of *VA span*, using the Partial Report task (Bosse et al., 2007); and 3) participants' regular, irregular and nonword reading skills using the Coltheart and Leahy (1996) item lists. Specifically, to index phonological-level processing we measured *nonword accuracy*, and to index whole-word level processing we measured *irregular word accuracy,* as well as the *nonword latency difference* between reading latencies on nonwords and real words (regular and irregular combined). Finally, to ensure that participants' difficulties could not be attributed to more global intellectual deficits, we also obtained an estimate of participants' general nonverbal intelligence, using the Block Design subtest of the Weschler Intelligence Scale for Children IV (Wechsler, 2003).

In the subsequent Profile Building phase, we identified a small number of participants with relatively "pure" profiles: difficulties at either the phonological, *or* at the whole-word level. We then assessed several further aspects of their cognitive profiles that have been previously associated with *either* phonological-level or whole-word level processing. These included phonemic awareness, estimated from phoneme segmentation and deletion tasks (and for some participants, also phoneme identification and Rhyme Detection tasks), and orthographic awareness, estimated from orthographic choice and homophone selection tasks. Finally, we also obtained a more comprehensive measure of general intelligence, estimated using the remaining standard sub-tests of the Weschler Intelligence Scale for Children IV, and two additional measures of reading proficiency, a word comprehension task and an oral paragraph reading task, both taken from the Woodcock Reading Mastery Test III.

**Predictions**

While neither of the two phases in the study were explicitly set up to test experimental hypotheses, we nonetheless made several predictions based on previous research and our findings from the Heterogeneity study. In the Screening phase, we predicted the following relationships: 1) nonword reading accuracy, a phonological processing marker, would be positively associated with Reading Age, our reading proficiency marker; 2) irregular word reading accuracy and VA span, as whole-word processing markers, would be positively associated with Reading Age; 3) our measure of nonword latency difference, a whole-word marker, would be negatively associated with Reading Age; 4) irregular word reading accuracy and VA span would be positively associated with one another; 5) our measure of nonword reading latency would be negatively associated with VA span; and 6) the phonological marker, nonword reading accuracy, would not be associated with any of the whole-word markers – irregular word reading accuracy, VA span or the nonword reading latency measure.

In the Profile Building phase, we predicted that individuals identified in the Screening phase as phonological cases – but not those identified as whole-word cases – would meet the criteria for impairment on the phonemic awareness tasks. Conversely, we predicted that whole-word cases – but not phonological cases – would meet the criteria for impairment on the orthographic awareness tasks and the single word comprehension task.

**Screening phase method**

Recruitment for this study occurred in two main phases: a) the Screening phase; and b) the Profile Building phase. Secondary schools within Wellington, New Zealand were asked if they would allow the primary researcher (EA) to describe the study to Year 9 students during school time and distribute letters calling for volunteers who were enrolled in reading remediation classes or otherwise suspected of having reading difficulties (see Appendix A and Appendix T for a copy of the information sheet sent to principals in the

2016 and 2017 recruitment respectively). EA verbally described the study to each eligible class, then distributed information letters and assent/consent forms for students and their parents to complete if they wished to participate (see Appendices U through W for copies of the forms during the 2016 recruitment, and Appendices B through D for copies of the forms during the 2017 recruitment; note, they were also asked to indicate if they were interested in participating in the subsequent phases if they were found to be suitable candidates). The only formal criteria for inclusion in the study was that the individual was currently completing Year 9 and had normal or corrected vision. There were no additional exclusion criteria. Participants who displayed relatively "pure" impairment profiles were invited to participate in the subsequent Profile Building phase and Intervention study. These will be discussed in more detail later (n. B the Intervention study will be discussed in Chapter 4).

**Screening phase: Participants.** Two cohorts were recruited for this study. The first cohort (recruited in 2016) were from School A, while the second cohort (recruited in 2017) included Schools B and C as well (all identical to Chapter 2). In the first screening recruitment, 21 participants consented to take part. In the second screening recruitment, 26 participants consented to take part. Out of the 47 total participants recruited in the screening of this study, one was excluded from the current analysis because they did not complete all tasks; a further two were excluded due to technical problems (the audio recorder failed during the session). Of the 44 remaining participants who completed all tasks, 17 were female and 27 were male, with a mean age of 13 years and 8 months (SD= 5 months). Small tokens of appreciation (USB sticks, water bottles, pens) were offered as a thank you for taking part and were not dependent on the number of tasks completed or performance during the session.

**Screening phase: Materials.** Each participant completed the following tasks: a) the Block Design subtest of the Wechsler Intelligence Scale for Children- IV (Wechsler, 2003); b) the Word Identification Task from the Woodcock Reading Mastery Test; c) the Coltheart and Leahy task; and d) the Partial Report task. The Wechsler Block Design sub-test was administered according to the standard procedure. The materials and procedure for administering and scoring the Woodcock Reading Mastery Test, Coltheart and Leahy task and Partial Report task are outlined in Chapter 2. No word was duplicated across any of the three reading tasks.

**Screening phase: Procedure.** Both the Coltheart and Leahy task, and the Partial Report task were created via OpenSesame (Mathôt et al., 2012). These tasks were administered on either a MacBook Pro (2016 cohort), or Samsung galaxy tab 10.1 (2017 cohort). All sessions were recorded using a digital audio recorder.

Each screening session lasted up to one hour, taking place in normal school hours. Sessions were conducted on a one-to-one basis; however, an observer (a research assistant) was also present throughout. All sessions were run by the lead researcher (EA). Each session proceeded in the same fashion as the sessions in described in Chapter 2 (the research project was summarised verbally; confidentiality was explained; questions were encouraged; they could take a break or stop participating at any point without penalty; participants verbally assented at the start of the session).

Tasks were administered in the same order for each participant: 1) the Block Design subtest from the Weschler Intelligence Scale for Children; 2) the Coltheart and Leahy task; 3) the Partial Report task; and 4) the word identification task from the Woodcock Reading Mastery Test. Some participants also completed additional tasks designed to assess other reading or cognitive skills discussed in Chapter 2. These will not be discussed at length here. For the 2016 cohort, task instructions were presented both on-screen in black size 32 font, and via an auditory pre-recording that was embedded within the programme. This was an attempt to standardise the procedure across participants while ensuring the experimenter did not come off “robot-like”, however it was found to be troublesome and so for the 2017 cohort instructions were given verbally by the experimenter.

As in Chapter 2, participants were told to try their best on all tasks, but not to panic if they did not know or were unsure of an answer at any point. Throughout all tasks, no feedback on performance was given, however occasionally if a student showed concern or appeared nervous, re-assurance was given without affirming performance. At the end of each session participants were offered the opportunity to ask any questions they might have. They were told that if they thought of any questions after they had left, their parents could email the lead researcher (EA) or the supervisor (CW).

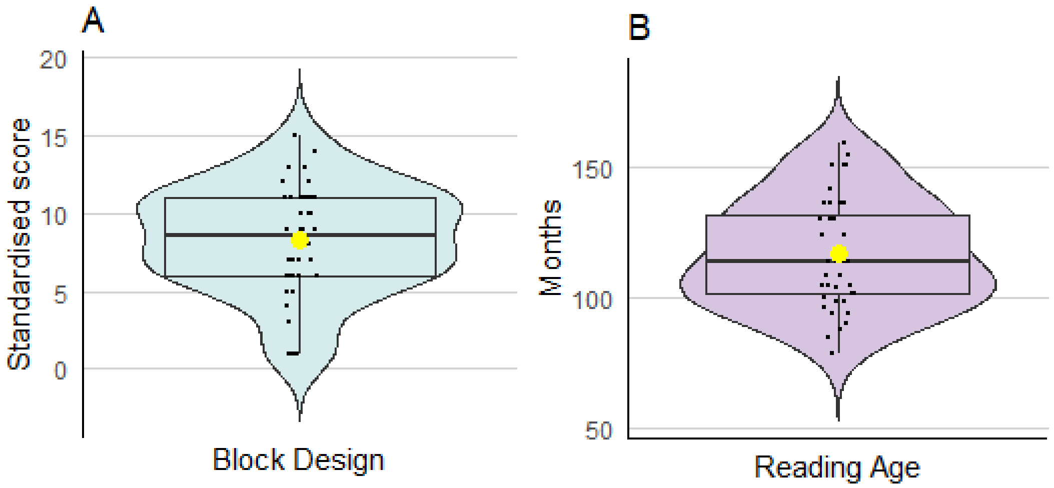
**Screening phase: Data Collection and Analysis.** For all tasks, reading accuracy was recorded during the session by EA, and where appropriate, cross checked against the audio-recording of the session. For the Coltheart and Leahy task, reading latencies were calculated manually from the digital recordings of the sessions using Audacity 2.1.2 (Audacity-Team, 2014): the experimenter measured the time from the onset of the beep that accompanied the stimulus, to the onset of the first correct response. Full details as to how these measures were pre-processed and analysed are provided in the Results section below. The criteria for identifying specific reading difficulties are also included in the Results section below.

**Results of Screening phase**

**Group trends.** Assumption checks were completed using SPSS, however, unless otherwise stated, all other analyses were completed in R[7]. This section looks at the overall group trends of reading remediation participants on the various tasks.

***Group trends: Presence of overall impairments.*** Estimates of general intelligence and Reading Age were calculated from the Block Design subtest (Wechsler, 2003) and word identification subtest (Woodcock, 2011) respectively. Figure 3.1 shows that both the Block Design and Reading Age scores show good spread, with no marked tail end on either. Overall, Block Design (general intelligence) scores indicated average intelligence, but Reading Age for the group was significantly lower (*M* = 117 months, *SD* = 21 months) than Chronological Age (*M* = 165 months, *SD* = 5 months; $t$ (43, 1) = 14.59, $p$ < .001), suggesting a significant number of reading impairments in the group.

Figure 3.1 Violin plots for Block Design scores (A) and Reading Age (B). Black dots represent individual participants, yellow dots represent group mean, with 95% confidence intervals (black vertical line). Box plots show median (black central line), upper and lower quartiles.

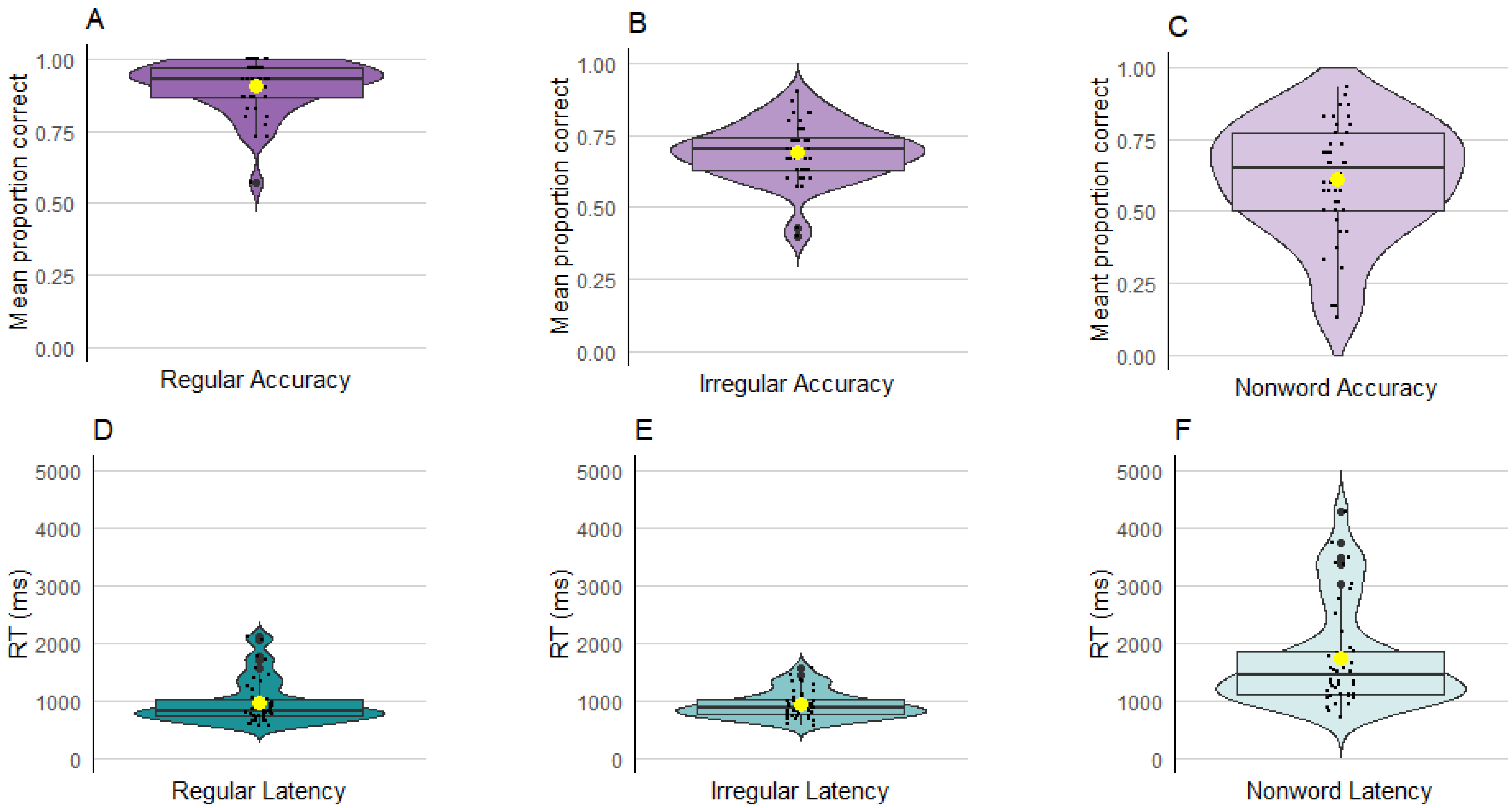



[7] References used: : Rstudio: Team (2015), lme4: Bates et al. (2014), lmerTest: Kuznetsova et al. (2017), lmTest: Hothorn et al. (2018), aod: Lesnoff and Lancelot (2012), Hmisc: Harrell Jr and Harrell Jr (2019), Tidyverse: Wickham (2017), corrplot Wei and Simko (2017), Irr: Gamer et al. (2012), Emmeans: Lenth (2018).

***Group trends: Coltheart and Leahy task.*** For the analyses of reading accuracy, one trial on the nonword reading task (participant 43: nonword "*brinth"*) could not be scored due to a technical problem. As can be seen in Figure 3.2, there was a good spread of scores on the regular and irregular word lists (Panel A and Panel B respectively), although there was a slight tail end for both (one individual on the regular word list and three individuals on the irregular word list scored below the 95% confidence interval). For the nonword list (Panel C of Figure 3.2), the spread was very good, with no marked tail end. Logistic regression was completed via the *glmer* function with accuracy as the dependent variable, word type as the fixed effect, and participant identify as a random effect. For the group as a whole, there a significant main effect of word type on overall reading accuracy ($F$ (87, 2) = 15.74, $p < .001$). Post-hoc analysis via the "*emmeans"* function (from the *emmeans* package) revealed significantly higher scores on the regular word list compared to both the irregular word list ($t$ = 3.41, $p$ = .002), and the nonword list ($t$ = 5.59, $p < .001$). There was no significant difference in accuracy between the irregular word and nonword lists ($t$ = 2.19, $p$ = .072).

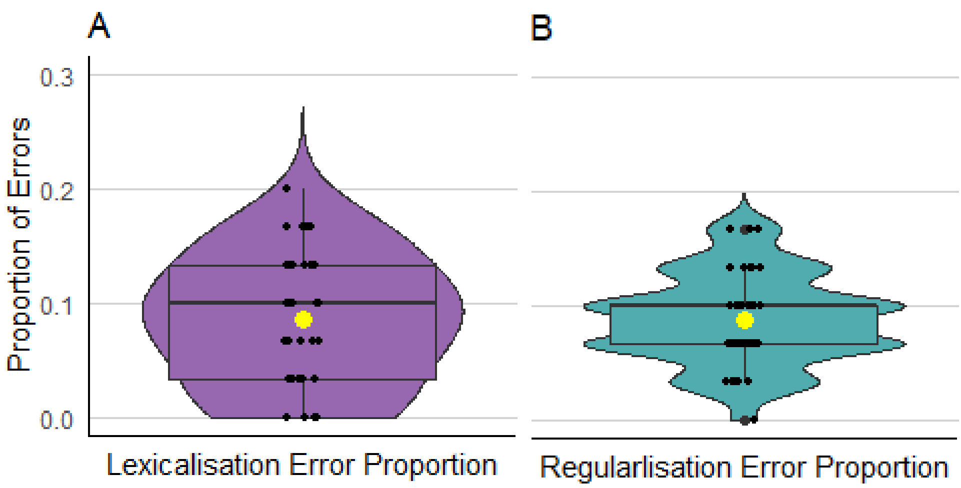
For the analyses of reading latencies, two additional trials were excluded due to technical problems (participant 28: irregular word "*eye";* participant 40: nonword "*peng")*. Inaccurate trials were also excluded. Considering the entire cohort, this resulted in the removal of 26.8% of the trials. Of these removed trials, 9.3% were regular words, 31.2% were irregular words, and 39.8% were nonwords. Exclusion of such a larger number of trials is potentially concerning, yet unsurprising given the explicit recruitment of struggling readers (we return to this issue in the discussion). As Figure 3.2 shows, for both the regular and irregular word lists (Panel D and Panel E respectively), there was relatively little spread, although there was an upper tail for the regular word list (eight individuals scored outside the 95% confidence interval). For the nonword list (Panel F of Figure 3.2), there was good spread, however there was a marked upper tail (six individual scores outside of the confidence intervals). Linear mixed effects modelling via the *lmer* function was completed with latency as the dependent variable, word type as the fixed effect, and participant identity as the random effect. Post-hoc analysis was completed via the *emmeans* function. The results showed that there was also a significant main effect of word type on reading latency ($F$ (75, 2) = 31.22, $p < .001$), with significantly longer latencies on nonwords relative to both irregular words ($t$ = 6.26, $p < .001$) and regular words ($t$ = 7.31, $p < .001$). There was no significant difference between the irregular and regular word lists ($t$ = 0.87, $p$ = .66).

Figure 3.2 Violin plots of regular word accuracy (A) and latency (D); irregular word accuracy (B) and latency (E); and nonword word accuracy (C) and latency (F). Black dots represent individual participants, yellow dots represent group mean, with 95% confidence intervals (black vertical line). Box plots show median (black central line), upper and lower quartiles.



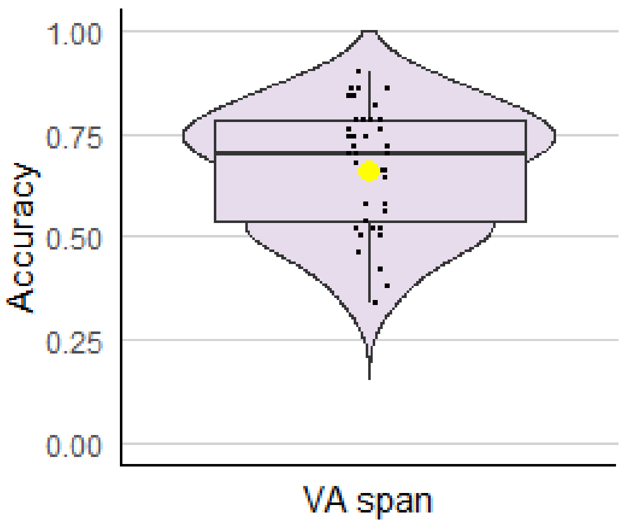
As in Chapter 2, we also examined the prevalence of two specific types of error on the Coltheart and Leahy task: lexicalisation errors on nonwords (e.g. *vun* > "van") which may suggest difficulty with phonological processes, and regularisation errors on irregular words (e.g. *blood* > "blude") which may suggest difficulty with whole-word processing. For each individual we calculated a Lexicalisation Error Proportion score, and a Regularisation Error Proportion score (a full description of possible errors is included in Appendix Q). The same independent judges from Chapter 2 coded the error output of each participant. As previously, we calculated the kappa statistic using the *Kappa2* function (via the *irr* package) for the Lexicalisation and Regularisation errors separately, treating each as a binary outcome (error coded as present vs. absent). For both error types, inter-rater reliability was moderate (Lexicalisation error: $\kappa = .527$, $p < .001$, Regularisation error: $\kappa = .545$, $p < .001$). These results again suggest that the definitions we used to identify error type were somewhat subjective. The subsequent individual analyses are based on data from the primary judge that scored fewer errors as "Unknown" (Appendix X shows the breakdown of error types for each of the two independent coders). Figure 3.3 shows the distributions of scores. There was a good distribution of scores for both measures, with almost all participants producing at least some lexicalisation errors, and some regularisation errors. Overall, there were comparable numbers of lexicalisation and regularisation errors, a finding that is consistent with the similar levels of reading accuracy for nonwords and irregular words discussed previously.

Figure 3.3 Violin plots of the Lexicalisation Error Proportion (A) and the Regularisation Error Proportion (B). Black dots represent individual participants, yellow dots represent group mean, with 95% confidence intervals (black vertical line). Box plots show the median (black central line), upper and lower quartiles.



***Group trends: VA span task.*** For the Partial Report task, the data showed no evidence of skew or kurtosis on the Q-Q plots[8] (Shapiro-Wilk's statistic was significant, $p < .001$, however this may not be surprising given the small sample size). Figure 3.4 shows that, overall, there was a good distribution of scores with no marked tail.

Figure 3.4 Violin plot of VA Span scores. Black dots represent individual participants, yellow dots represent group mean, with 95% confidence intervals (black vertical line). Box plots show median (black central line), upper and lower quartiles.




[8] Not included in this thesis

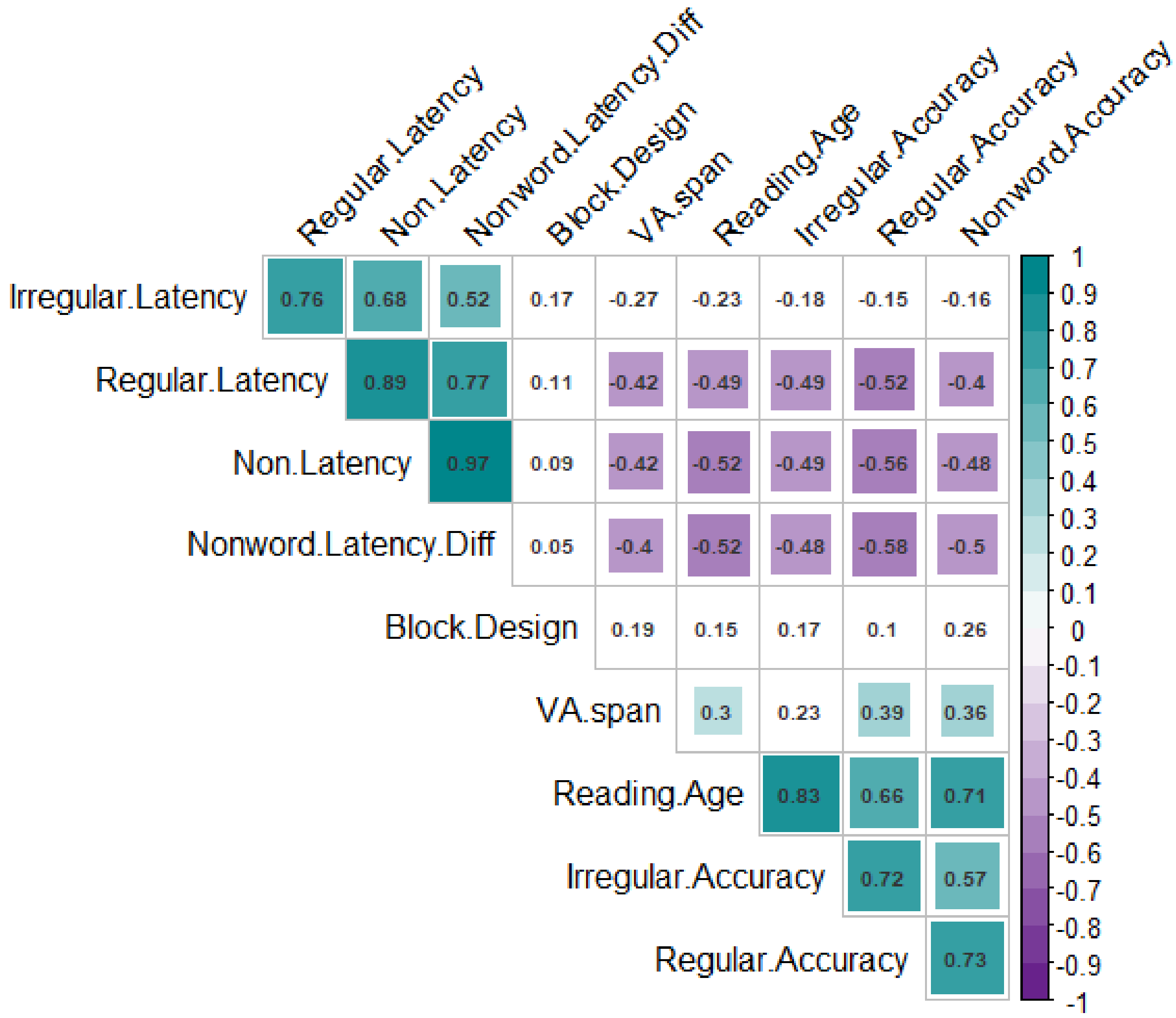
**Individual-level analyses.** Table 3.1 summarise the measures that were calculated for each individual. Figure 3.5 presents the correlogram for the relationships between these individual variables. First, it is worth noting that the Intelligence Estimate (Block Design) did not correlate with the measures of reading proficiency, Reading Age, or reading skills.

Table 3.1 Screening test scores calculated for each individual, how they were calculated and how they are interpreted.

| Score | Definition | Description/ Interpretation |
|---|---|---|
| Coltheart and Leahy task | | |
| Regular Accuracy<br>Irregular Accuracy<br>Nonword Accuracy | Proportion correct on the sub-list | Unimpaired, Impaired and Severely Impaired performance categorised according to standardised score boundaries (see table 3.3). |
| Regular Latency<br>Irregular Latency<br>Nonword Latency | Mean latency on the sub-list | Scores 1.5 standard deviations larger (i.e. slower) than group mean indicate reading speed impairments[9]. |
| Nonword LatencyDiff | Mean nonword latency minus mean real word (regular and irregular) log latency | Positive scores indicate poorer nonword reading, negative scores indicate poorer real word reading. |
| Partial Report task | | |
| VA span | Proportion correct on the Partial Report task | High scores indicate good span, low scores indicate limited span |
| Woodcock Reading Mastery Test | | |
| Reading Age | Word identification task standardised scores | Reading Age scores 2 years or more below chronological age indicate impairment. |
| Wechsler Intelligence Scale for Children | | |
| Block Design score | Standardised scores of the block sub-test | General intelligence impairments are indicated by scores less than 7. |

[9] Calculated separately for each cohort

Figure 3.5 Correlogram displaying the relationships between all the measures of the screening. Significant relationships ($p < .05$) are represented by a coloured square, while insignificant relationships are left white. The strength of a relationship is depicted by the opaqueness and size of the coloured square. Positive relationships are green-scale, negative relationships are purple-scale.




***Markers of phonological impairment.*** Our first prediction was that our primary measure of phonological reading skill, NonwordAccuracy, should be positively associated with our primary measure of reading proficiency, Reading Age. This prediction was confirmed. Our second prediction was that there should be no association between NonwordAccuracy and our three measures of whole-word skill: IrregularAccuracy, NonwordLatencyDiff, and VA span. This prediction was not supported. We found significant positive relationships between NonwordAccuracy and both IrregularAccuracy and VA span. We also found a significant negative relationship between NonwordAccuracy and

NonwordLatencyDiff; however, this finding should be treated with caution as the two measures are not truly independent.

***Markers of whole-word impairment.*** Our third prediction was that our key measures of whole-word reading skills, IrregularAccuracy, NonwordLatencyDiff, and VA span, would each be associated with the primary measure of reading proficiency, Reading Age. Specifically, we anticipated that both IrregularAccuracy and VA span would be positively associated with Reading Age, while NonwordLatencyDiff would be negatively associated with Reading Age. These predictions were supported.

Our fourth, and final, prediction was that our key measures of whole-word reading skills, IrregularAccuracy, NonwordLatencyDiff, and VA span, would all be associated with one another. Specifically, we predicted that IrregularAccuracy and VA span would be positively associated with one another, and both would in turn be negatively associated with NonwordLatencyDiff. These predictions were partially supported, both IrregularAccuracy and VA span were negatively correlated with NonwordLatencyDiff, however, IrregularAccuracy and VA span were not significantly correlated with one another. These results suggest that each of these key measures may be indexing a slightly different underlying mechanism[10].

***Error patterns as markers of phonological and whole-word impairment.*** Table 3.2 shows the results of the correlation analyses for the two error measures. The Lexicalisation Error Proportion correlated negatively with the key measure of reading proficiency, Reading Age, and RegularAccuracy (we did not perform a correlation with NonwordAccuracy, the key phonological measure, because these scores were derived from the same sub-lists). The Regularisation Error Proportion did not correlate with any measures (again we did not perform a correlation with IrregularAccuracy, a key whole-word measure, because these scores were derived from the same sub-lists). These results suggest that, similar to findings in Chapter 2, only the Lexicalisation Error Proportion could predict general reading proficiency, with neither score providing much insight into the reading profiles of the current sample.

[10] As shown in Chapter 2 it is unclear; a) what mechanisms the Partial Report task is really capturing, and b) what the contributions of the supposed ‘VA span’ mechanisms to reading are. The current intervention investigation was completed prior to the investigation in Chapter 2, and so the true complexity of VA span was not fully known. For the intervention study, we were highly influenced by other research differentiating reading impairments based, at least in part, on VA span. Hence, we used this as a key measure of whole-word impairments and anticipated it would be related to other key measures. This will be discussed in more detail in the General Discussion.

Table 3.2 Correlation co-efficients (significance scores in parentheses) for the Lexicalisations and Regularisation Error Proportions, with each of the previous scores. **Significant relationships at the p <.01 level. *Significant relationships at the p <.05 level.

| | Lexicalisation Error Proportion | Regularisation Error Proportion |
|---|---|---|
| Reading Age | -.401** (*p* = .007) | .077 (*p* = .618) |
| VA span | -.130 (*p* = .399) | .127 (*p* = .412) |
| RegularAccuracy | -.345** (*p* = .022) | -.003 (*p* = .984) |
| RegularLatency | .190 (*p* = .217) | -.121 (*p* = .435) |
| IrregularAccuracy | -.254 (.096) | *na* |
| IrregularLatency | .054 (*p* = .727) | -.039 (*p* = .801) |
| NonwordAccuracy | *na* | -.262 (*p* = .085) |
| NonwordLatency | .173 (*p* = .262) | -.111 (*p* = .473) |
| NonwordLatencyDiff | .168 (*p* = .274) | -.108 (*p* = .485) |

**Individual impairment patterns.** Since the screening described above served as a means of identifying candidates for the intervention study, we examined the various different profiles for each individual participant. We were primarily concerned with identifying two distinct profiles of reading impairment: phonological impairment and whole-word impairment. For both, we looked for participants that showed a Reading Age impairment (scores 2 years or more below their Chronological Age) but no evidence of below-average intelligence ( scores at 7 or above on the Block Design subtask). Anyone who did not meet these criteria was excluded from subsequent consideration for the Profile Building phase and Intervention study. Table 3.3 shows the cut-off scores that were used to define impaired performance on the Coltheart and Leahy Task, and the Partial Report task. For the Coltheart and Leahy task, cut-offs were based on the ranges recommended by Coltheart and Leahy (1996), using the oldest norms available (12 years-old). For the Partial Report task, VA span impairments were classified based on the typical age range associated with that participant's score according to the norms by Bosse and Valdois (2009). Those scoring within the 10-12 year age bracket (the oldest norms available) were classified as unimpaired; those scoring in the 8-9 age bracket were classified as moderately impaired; and those scoring in the 6-7 age bracket (or below) were classified as severely impaired.

Table 3.3 Impairment boundaries for accuracy scores on the sub-lists of the Coltheart and Leahy task, and the Partial Report task.

| Score | Unimpaired | Moderately impaired | Severely impaired |
|---|---|---|---|
| RegularAccuracy (/30) | 28+ | 26 – 27 | 0 – 25 |
| IrregularAccuracy (/30) | 20+ | 16 – 19 | 0 – 18 |
| NonwordAccuracy (/30) | 22+ | 20 – 21 | 0 – 19 |
| VA span (/50) | 35+ | 29 – 34 | 0 – 28 |

Table 3.4. illustrates how we identified the 'pure' phonological and 'pure' whole-word impairments. Table 3.5 provides the individuals' raw scores on each measure, and an indication of whether they fell below the cut-off for impairment on that measure. There are several points worth noting. First, Block Design subtest scores were widely distributed, with most participants (31) showing no evidence of impairment to general intelligence. Second, Reading Age was well distributed, and most participants met the criteria for impairment (39 of the total group; 28 of the 31 with average or above average general intelligence). Third, the number of participants meeting the criteria for impaired nonword reading (31 of the total group; 26 of the 28 meeting both reading impairment and general intelligence criteria) was larger than the number meeting the criteria for impaired irregular word reading (13 of the total group; eight of 28 meeting both reading impairment and general intelligence criteria).

Table 3.4 Criteria for identifying relatively "pure" phonological and "pure" whole-word reading impairments, including both the impaired and unaffected measures.

| Profile | Phonological | Whole-word |
|---|---|---|
| Nonword Accuracy | Moderately to severely impaired | Unimpaired (2016 cohort) to moderately impaired (2017 cohort) |
| Irregular Accuracy, VA span[11], NonwordLatencyDiff | Unimpaired on all three tasks | Moderately to severely impaired on at least one task |

[11] Screening commenced prior to the Heterogeneity study and as such the potential issues with VA span as a whole-word marker had not yet been identified. Following the completion of the heterogeneity study we removed VA span as a criterion for whole-word impairments and re-assessed the screening to establish if this would have affected the selection of participants for inclusion in the intervention, which it did not.

Table 3.5 Demographic information along with performance on scaled tests scores. * = Reading Age two or more years below Chronological Age. For the accuracy-based scores; light grey indicates the participant falls with the impaired band, dark grey indicates the participant falls within the severe impairment band. For NonwordLatencyDiff; light grey indicates the participants' mean difference is at least one SD larger than the group mean. Participants who were subsequently selected for and completed the intervention are bolded and underlined.

| ID[a] | Gender | Handed-ness | CA[b] | RA[c] (std[d]) | Block test | Accuracy | | | | Latencies | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Reg[e] | Irreg[f] | Non[g] | VA | Reg[e] | Irreg[f] | Non[g] | NonDiff[h] |
| 2016 | | | | | | | | | | | | | |
| **1*** | **F** | **R** | **170** | **142 (87)** | **10** | **0.93** | **0.87** | **0.67** | **0.74** | **790.00** | **870.00** | **930.00** | **94.97** |
| 2* | M | R | 164 | 109 (72) | 6 | 0.97 | 0.70 | 0.67 | 0.84 | 1010.00 | 940.00 | 1290.00 | 316.56 |
| 3* | F | R | 171 | 124 (78) | 8 | 0.90 | 0.77 | 0.17 | 0.72 | 810.00 | 990.00 | 1570.00 | 675.03 |
| 4* | F | R | 158 | 104 (79) | 6 | 0.93 | 0.63 | 0.53 | 0.78 | 770.00 | 740.00 | 1050.00 | 295.16 |
| 5* | F | L | 169 | 136 (84) | 7 | 0.93 | 0.73 | 0.70 | 0.78 | 720.00 | 760.00 | 1120.00 | 355.80 |
| 6* | M | R | 166 | 96 (60) | 9 | 0.87 | 0.67 | 0.57 | 0.66 | 820.00 | 830.00 | 1060.00 | 239.72 |
| 7* | M | R | 162 | 109 (74) | 13 | 0.93 | 0.70 | 0.63 | 0.7 | 1250.00 | 1370.00 | 1540.00 | 226.56 |
| 8* | F | R | 160 | 124 (83) | 11 | 1.00 | 0.73 | 0.77 | 0.78 | 750.00 | 680.00 | 1330.00 | 609.89 |
| 9* | M | R | 164 | 100 (66) | 11 | 0.87 | 0.63 | 0.50 | 0.76 | 770.00 | 930.00 | 1830.00 | 979.42 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10* | F | R | 167 | 105 (69) | 11 | 0.83 | 0.63 | 0.47 | 0.56 | 2100.00 | 1340.00 | 3390.00 | 1664.24 |
| **11*** | **F** | **R** | **157** | **130 (88)** | **11** | **1.00** | **0.70** | **0.80** | **0.52** | **1390.00** | **1270.00** | **2950.00** | **1618.55** |
| **12*** | **M** | **R** | **161** | **119 (80)** | **8** | **0.97** | **0.67** | **0.70** | **0.78** | **880.00** | **920.00** | **1370.00** | **467.80** |
| 13* | M | R | 167 | 105 (69) | 6 | 0.80 | 0.67 | 0.47 | 0.54 | 650.00 | 940.00 | 1270.00 | 474.28 |
| 14 | F | R | 159 | 151 (96) | 11 | 0.97 | 0.83 | 0.87 | 0.52 | 930.00 | 830.00 | 1100.00 | 221.06 |
| 15* | F | R | 165 | 136 (87) | 11 | 0.93 | 0.80 | 0.60 | 0.38 | 750.00 | 870.00 | 1230.00 | 423.90 |
| 16* | F | R | 165 | 94 (58) | 1 | 0.87 | 0.57 | 0.17 | 0.66 | 710.00 | 670.00 | 1060.00 | 371.76 |
| 17* | M | R | 168 | 130 (81) | 15 | 0.97 | 0.60 | 0.90 | 0.78 | 800.00 | 820.00 | 1070.00 | 262.10 |
| 18* | M | R | 165 | 79 (55) | 11 | 0.57 | 0.40 | 0.13 | 0.46 | 1760.00 | 920.00 | 3480.00 | 2136.31 |
| 19* | M | R | 167 | 114 (75) | 9 | 0.97 | 0.73 | 0.67 | 0.66 | 990.00 | 950.00 | 1840.00 | 868.18 |
| 20* | M | R | 164 | 114 (75) | 12 | 0.87 | 0.70 | 0.73 | 0.58 | 800.00 | 810.00 | 1520.00 | 723.25 |
| 21 | M | R | 168 | 151 (90) | 4 | 1.00 | 0.90 | 0.83 | 0.7 | 580.00 | 700.00 | 720.00 | 78.94 |
| | | | | | | 2017 | | | | | | | |
| 22 | M | R | 166 | 155 (99) | 11 | 0.97 | 0.83 | 0.93 | 0.86 | 600.00 | 754.48 | 1092.14 | 414.90 |
| 23* | M | L | 161 | 105 (71) | 7 | 0.93 | 0.67 | 0.67 | 0.52 | 660.32 | 736.75 | 836.80 | 138.26 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24* | M | L | 157 | 114 (79) | 10 | 0.93 | 0.70 | 0.77 | 0.74 | 741.25 | 649.38 | 1111.26 | 415.95 |
| 25* | M | R | 161 | 99 (65) | 9 | 0.87 | 0.63 | 0.53 | 0.72 | 897.88 | 744.21 | 1574.38 | 753.33 |
| 28* | M | R | 158 | 130 (88) | 7 | 0.83 | 0.67 | 0.70 | 0.9 | 552.52 | 567.16 | 1158.24 | 598.40 |
| 29* | M | R | 167 | 94 (58) | 1 | 0.77 | 0.57 | 0.37 | 0.54 | 1196.91 | 815.82 | 1919.09 | 912.72 |
| 30* | M | R | 162 | 136 (89) | 6 | 0.97 | 0.73 | 0.70 | 0.74 | 823.52 | 786.77 | 1240.38 | 435.24 |
| 31* | M | R | 166 | 99 (63) | 6 | 0.93 | 0.63 | 0.50 | 0.64 | 1702.21 | 1579.47 | 3382.53 | 1741.69 |
| 32* | M | L | 173 | 114 (72) | 6 | 0.93 | 0.67 | 0.50 | 0.86 | 784.56 | 831.64 | 1745.71 | 618.75 |
| **33*** | **M** | **R** | **161** | **136 (89)** | **14** | **0.87** | **0.77** | **0.57** | **0.86** | **1053.92** | **1108.04** | **1693.00** | **612.02** |
| 34* | F | R | 165 | 96 (60) | 5 | 0.90 | 0.67 | 0.30 | 0.34 | 588.56 | 594.30 | 1299.00 | 707.57 |
| **35*** | **F** | **R** | **179** | **102 (63)** | **11** | **0.93** | **0.73** | **0.57** | **0.84** | **934.71** | **841.23** | **2209.06** | **1321.09** |
| 37 | M | R | 169 | 159 (93) | 10 | 1.00 | 0.83 | 0.83 | 0.7 | 654.07 | 733.60 | 936.72 | 242.89 |
| 38* | F | R | 168 | 85 (55) | 9 | 0.73 | 0.60 | 0.33 | 0.58 | 1439.50 | 1007.78 | 4294.90 | 3071.26 |
| 39* | M | R | 167 | 102 (66) | 12 | 1.00 | 0.70 | 0.73 | 0.74 | 888.50 | 1153.90 | 1778.27 | 757.07 |
| 40* | M | R | 163 | 136 (89) | 13 | 1.00 | 0.80 | 0.77 | 0.76 | 851.20 | 1172.58 | 1664.05 | 652.15 |
| **41*** | **F** | **R** | **175** | **114 (72)** | **7** | **0.93** | **0.73** | **0.60** | **0.52** | **1327.36** | **911.68** | **2764.89** | **1645.37** |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 42* | F | R | 175 | 90 (55) | 5 | 0.73 | 0.43 | 0.43 | 0.5 | 1553.64 | 1349.77 | 3040.54 | 1588.84 |
| 43* | F | R | 165 | 88 (55) | 8 | 0.80 | 0.60 | 0.43 | 0.42 | 2047.54 | 1439.00 | 3744.08 | 2000.81 |
| 44 | F | R | 160 | 151 (94) | 1 | 1.00 | 0.77 | 0.83 | 0.68 | 655.50 | 983.78 | 1325.48 | 505.84 |
| 45* | M | R | 172 | 114 (72) | 7 | 0.97 | 0.60 | 0.87 | 0.82 | 651.34 | 667.06 | 892.12 | 232.92 |
| 46* | F | R | 161 | 124 (83) | 3 | 0.93 | 0.73 | 0.80 | 0.5 | 945.29 | 1070.05 | 1515.58 | 507.92 |
| 47* | M | R | 163 | 130 (86) | 8 | 0.93 | 0.77 | 0.57 | 0.46 | 974.11 | 1049.43 | 2504.47 | 1492.70 |
| *M* | | | 165.25 | 116.93 | 8.34 | 0.91 | 0.69 | 0.61 | 0.66 | 965.57 | 925.86 | 1729.15 | 783.44 |
| *SD* | | | 4.93 | 20.27 | 3.35 | 0.09 | 0.10 | 0.20 | 0.14 | 379.19 | 233.60 | 874.13 | 643.12 |

Note: a) ID = Participant, b) CA = Chronological Age, c) RA = Reading Age, d) std = Standardised Scores on the Word Identification task of the WRMT, e) Reg= Regular words, f) Irreg= Irregular words, g) Non= Nonwords, h) NonDiff= NonwordLatencyDiff.

Table 3.6 shows the number of participants that performed within the impaired range (moderately or severely impaired) on each the key tasks, and the number participants that fit the criteria for having a 'pure' or disproportionate phonological or whole-word impairment. Most impaired readers had mixed impairment profiles (15). For those who had pure profiles, the majority displayed a profile consistent with a phonological impairment. As can be seen in the table, only one individual demonstrated a pure whole-word impairment. Therefore, to obtain enough cases that contrasted adequately with our pure phonological cases, we loosened the criteria for whole-word impairment: individuals in this group could demonstrate moderate NonwordAccuracy impairments, as long as the VA span or NonwordLatencyDiff Impairments were still both present.

Table 3.6 The number of participants in each cohort, and the total sample, that fit the impairment criteria for each key measure; and the number of participants that fit the criteria for each impairment profile.

Total number scoring within the impaired range

| | 2016 cohort | 2017 cohort | Total |
|---|---|---|---|
| Nonword Accuracy | 16 | 15 | 31 |
| Irregular Accuracy | 6 | 7 | 13 |
| VA span | 10 | 11 | 21 |
| NonwordLatencyDiff | 3 | 4 | 7 |

Number meeting criteria for a 'pure' phonological or whole-word impairment

| | 2016 cohort | 2017 cohort | Total |
|---|---|---|---|
| Phonological | 4 | 6 | 10 |
| Whole-word | 1 | 0 | 1 |

Number meeting loosened criteria for a disproportionately whole-word impairment

| | 2016 cohort | 2017 cohort | Total |
|---|---|---|---|
| Whole-word | 1 | 2 | 3 |

**Screening phase comment.** The primary purpose of this Screening Phase was to identify participants with suitable reading impairment profiles for participation in the subsequent Profile Building phase and Intervention study. As such, this phase was not intended to provide any meaningful insight into the cognitive mechanisms of reading. Having said that, it is perhaps worth noting that many of the patterns observed were the same as those presented in Chapter 2. Again, our key indices of phonological and whole-word processing skill were associated with Reading Age, and many of the measures from these indices were also associated with one another, except for IrregularAccuracy and VA span, which showed no association. Of course, the sample tested in this study partially overlapped with the sample in Chapter 2. Nevertheless, it is interesting to note that the patterns found for a relatively unselected sample also apply to a sample that has been specifically selected based on the presence of reading difficulties.

Our examination of participants' individual profiles found that the majority of participants displayed neither pure phonological, nor pure whole-word impairments, but a mix of both. These results may reflect the complex way in which different types of underlying reading deficits manifest themselves during reading development, and the strategies or behaviours the individual adopts in response to them. They may also reflect the fact that none of our measures are likely to be truly pure indices of the specific processes we were aiming to measure.

There are also two potential issues with the Coltheart and Leahy task that we should acknowledge. First, as in Chapter 2, we removed a large amount of incorrect trials prior to the calculation of mean latency scores. Although the effect of this process is unclear, it is possible that particularly poor readers would be especially affected. Second, it is worth bearing in mind that our cut-off scores for impairment on each list were based on a normative sample of 10 to 12 year-olds. Given that our participant sample were aged between 13 and 14, these cut-off scores are likely to detect only the most severe impairments. In other words, our criteria may have been insensitive to milder levels of impairment. However, 30 of the 36 participants who demonstrated a low Reading Age relative to chronological age also qualified as impaired on at least one of the Coltheart and Leahy sub-lists. Fewer participants qualified as impaired on the irregular word list ($n$ = 12) compared to the nonword list ($n$ = 32). This is consistent with previous research which repeatedly demonstrates phonological impairments are more frequent than whole-word impairments (Peterson, Pennington, Olson, & Wadsworth, 2014; Tibi & Kirby, 2017).

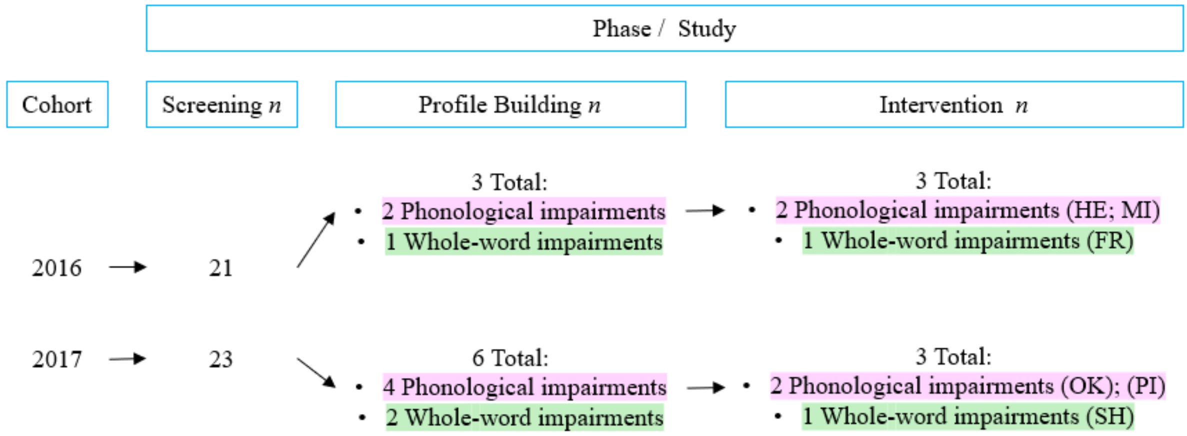
An alternative method for calculating accuracy-based impairments on this task would have been to use the Crawford t-test (Crawford & Howell, 1998). The Crawford t-test is popular in neuropsychology for assessing whether an individual is impaired when contrasted with a small sample of unselected, typical participants. We did not collect data from such a sample prior to the current study, however after the Screening and subsequent Intervention studies had been completed, the Heterogeneity study began and data from non-reading remediation classes was collected. Although this could not be used in the current research as participants had already been selected for the intervention, we were able to retrospectively check the potential impact of these values. Using the non-reading remediation participants (*n* = 12), the accuracy cut-offs for regular, irregular and nonwords did not change. This means that the identification of impairments on these three item lists was identical to the Coltheart and Leahy normed cut-offs.

**Profile Building phase method**

**Profile building phase: Participants.** Of the 13 individuals who demonstrated relatively pure profiles in the Screening Phase (either "pure" phonological impairments or relative whole-word impairments), we approached nine individuals who had shown particular interest in the study and invited them to participate in the Profile Building phase and Intervention study (see Appendix Y through AA for copies of the information-consent and assent forms for both cohorts). Of these nine individuals, six completed both the Profile Building phase and the Intervention study (the remaining three withdrew before any meaningful data had been collected, so their results will not be discussed further here[12]). The flow diagram in Figure 3.6 shows a breakdown of the number of participants at each stage of the screening and intervention. All participants discussed here attended School A. Four were female and two were male. Their mean age upon commencing the study was 13 years and 9 months (SD= 8 months).

[12] Since the Profile Building phase was primarily conducted to confirm suitability for the intervention, only participants who subsequently completed the intervention phase are discussed in this Chapter. Appendix AB. contains the data from two participants that completed the Profile Building but not the intervention.

Figure 3.6 Breakdown of participants who completed the screening in both 2016 and 2017 cohorts, and those identified with "pure" or relative impairment profiles.




**Profile building phase: Materials.** In this phase of the study, each participant completed several supplementary assessments designed to provide additional insights into their particular reading profile, and also completed a Problematic Letter Patterns task designed to identify suitable stimuli for use in the intervention phase. The supplementary assessments were: a) the Weschler Intelligence Scale for Children Version IV (all standard subtasks except the Block Design subtask) (Wechsler, 2003); b) a phonemic segmentation task (McBride-Chang, 1995); c) a phoneme deletion task (Hammill, Mather, & Roberts, 2001); c) a homophone selection task (Pexman, Lupker, & Jared, 2001); and d) an orthographic choice task (Manis et al., 1996). In addition, the three participants from the 2017 cohort (OK; PI; SH) also completed phoneme identification and Rhyme Detection tasks (Muter et al., 1998), as well as the Oral Reading Fluency and Word Comprehension tests from the Woodcock Reading Mastery test (Woodcock, 2011). Throughout all tasks, word stimuli were taken from the aforementioned research so that normed data and cut-offs could be used for comparison. Further details on materials and procedure for these tasks are provided below.

***Weschler Intelligence Scale for Children.*** The full Weschler Intelligence Scale for Children version IV was administered and scored according to the standard procedure.

***Phoneme segmentation task.*** This task, taken from McBride-Chang (1995), had two parts. In the first, participants listened to audio-recordings of 24 individual nonwords, repeating them back to the researcher one-at-a-time. For each word they were told if they repeated the word incorrectly and given two more chances. Performance on this portion of the task was not included in scoring. In the second part of the task, participants listened to each word again, and had to pronounce the individual phonemes of the word (e.g. *hib* > /h/ /ɪ/ /b/). There were three practice trials for this second part. Feedback was provided during the practice trials, but not during experimental trials. Accuracy was recorded. See Appendix AC for the stimuli list.

***Phoneme deletion task.*** In this task, taken from Hammill et al. (2001), participants heard an auditory recording of a real word (e.g. "*slope*"), accompanied by a visually presented letter or grapheme denoting one of the phonemes in the word (e.g. *l*). They were required to repeat the word back to the researcher, but this time omit the phoneme represented on-screen (i.e. /səʊp/). All of these latter repetitions were nonword homophones of other real words (i.e. "soap"). There were 20 experimental trials (plus two practice trials). Each trial commenced with a fixation dot presented in centre screen, followed 1000ms later by a blank screen where the audio-recording of the word was played. At the offset of the auditory recording, the letter or grapheme was displayed centrally in size 32 mono black font on a white background. Feedback was provided during the practice trials, but not during experimental trials. Accuracy was recorded. See Appendix AD for the stimuli list.

***Homophone selection task.*** This task, an adaptation from Castles and Coltheart (1996), consisted of 20 homophone pairs (e.g. *berry-bury*), taken from Pexman et al. (2001). The participant's task was to select the homophone that best matched a previously given semantic description (e.g. "*a fruit*"). Each trial commenced with a fixation dot presented in centre screen, followed 1000ms later by the pre-recorded semantic description of the target item. At the offset of this stimulus, two homophones were presented along the central y-axis of a computer screen in 32 mono black font – one in the centre left, one in the centre right of the screen. Participants were required to identify the correct homophone via keyboard ("z"= left word, "m"=right word). The target appeared in each side of the screen an equal number of times. There were three practice trials, during which feedback was provided; however, no feedback was given in experimental trials. Accuracy was recorded. See Appendix E for the stimuli list.

***Orthographic choice task.*** This task, taken from Manis et al. (1996), consisted of 57 phonologically identical word-nonword pairs (e.g. *blame-blaim*). The participant's task was to identify the real word (i.e. *blame*). Each trial commenced with a fixation dot presented in centre screen, followed 1000ms later by the word pairs presented in the same manner as for the homophone selection task. Again, participants responded via keyboard (“z”= left stimulus, “m”=right stimulus), and the target appeared in each side of the screen an equal number of times. There were five practice trials, during which feedback was provided; however, no feedback was given in experimental trials. Accuracy was recorded. See Appendix AF for the stimuli list.

***Phoneme identification task (2017 cohort only).*** In this task, taken from Muter et al. (1998), participants listened to an audio-recording of a word fragment, followed by an image[13] depicting the word (e.g. “*bi-*” accompanied by a picture of a bin). Participants were required to provide the finishing phoneme(s) (i.e. “-*n*”). There were eight experimental trials and two practice trials. Each trial commenced with a fixation dot presented in centre screen, followed 1000ms later by an image of the object, followed 100ms later by the audio-recording of the word fragment. The image remained on screen until participants responded. Accuracy was recorded. See Appendix AG for the stimuli list.

***Rhyme Detection task (2017 cohort only)***. In this task, taken from Muter et al. (1998), participants were presented with four images[12], one at the top of the screen (the probe item; e.g. a cat), and three at the bottom of the screen (e.g. a dog, a hat, and a shoe). They were required to select which of the bottom three images “rhymed” with the top. Participants responded by tapping on one of the images on-screen. Each trial commenced with a fixation dot presented in centre screen, followed 1000ms later by the array of four images. The probe item always appeared in the centre of the top half of the screen, and the remaining items were distributed evenly across the bottom half of the screen. The images remained on-screen until participants responded. There were 10 experimental trials in total. Accuracy was recorded. See Appendix AH for the stimuli for the stimuli list.

***Woodcock Reading Mastery Test (2017 cohort only).*** To assess reading at a paragraph-level, as well as semantic understanding, in the 2017 cohort we administered two additional sub-tests from the Woodcock Reading Mastery Test (Woodcock, 2011): The Oral Reading Fluency and Word Comprehension tests. Oral Reading Fluency is a measure of accuracy, reading speed and error rate on age-normed text passages (one-two paragraphs). Participants are asked to read the passage(s) out loud, errors and reading speed are noted and an overall score is calculated. Word Comprehension is a measure of single word semantic

[13] A small sample of coders (*n* = 13; female = 11, male = 2, age: *M*= 31 years 6 months, *SD* = 10 years 2 months) were asked to name the selected images with a single word. Average naming agreement across all images from the phoneme dentification and Rhyme Detection tasks was 90.1%.

understanding. Three sub-tests contribute to the overall accuracy-based score: Antonyms, Synonyms and Analogies. For each sub-test participants are asked to read a series of words one at a time, and then give a word that fulfils the brief of the subtest (i.e. in the Antonym sub-test participants were asked to give a word that meant the of the presented word; for example, if presented with *alive*, then the target response would be “dead”). Both the Oral Reading Fluency and the Word Comprehensions tests were administered according to the standard procedure, except participants started at the very beginning of the tests, rather than the recommended starting point for the age group. Age-based equivalence scores were calculated according to the standardised norms.

***Problematic Letter Patterns task.*** A number of the exercises planned for the Intervention study involved working with word *families* that shared common letter patterns. In the GPC training, these letter patterns consisted of graphemes that denoted a single phoneme (e.g. ‘oo’); in the visual training, they consisted of longer, easily identifiable letter sequences (e.g. ‘ation’), which we refer to as *clusters*. We assessed participants’ proficiency with a selection of both GPC and cluster redundancies, using two selection lists specially created for this study. The intervention was then tailored for each participant to include the GPC and cluster redundancies that the participant found most difficult.

The GPC selection list consisted of 156 words spanning 34 GPCs. Each GPC was represented in 5 words (some words contained two GPCs). The cluster selection list consisted of 165 words spanning 33 different letter clusters. These two-word lists were divided into two roughly equally sized blocked of words and presented to participants for oral reading using an ABAB design (GPC list Block 1 > cluster list Block 1 > GPC list Block 2 > cluster list Block 2). For all blocks, reading trials proceeded as follows: Each trial commenced with a fixation dot presented in centre screen for 1000ms, followed by a single word presented in size 32 mono black font, accompanied by a beep. The word stimulus remained on-screen until participants responded verbally. The order of items within blocks was pseudo-randomised and the same for all participants. No practice trials were completed given the similarity between this and the earlier single word reading tasks. No feedback on performance was given. See Appendix AI for the GPC stimuli list and Appendix AJ for the cluster stimuli list.

**Profile Building phase: session procedure.** Each Profile Building session was spread over no more than three testing sessions, each lasting up to one-hour. Sessions took were conducted in the same manner as the screening phase (within school normal hours, individually, in the presence of the researcher and an observer). As previously, in the first session, the researcher (EA) verbally summarised the project, explained how confidential data would be treated, and encouraged the participant to ask questions, or request a break/ to stop at any time if there wished. Once again, participants were told to try their best, but not to panic if they were unsure at any point. No specific feedback was given during any of the tasks, but if a student appeared nervous, general re-assurance was given.

Tasks were administered in the same order for each participant: 1) the full Weschler Intelligence Scale for Children; 2) the Problematic Letter Patterns test; 3) the homophone selection and orthographic choice tasks; 4) the phoneme segmentation and deletion tasks; and 5) where applicable, the additional profile building tasks (the phoneme identification, Rhyme Detection tasks, the Woodcock Oral Fluency and Word Comprehension tasks, in that order). All tasks, other than the Weschler Intelligence Scale for Children, and the Oral fluency and Word comprehension test from the Woodcock Reading Mastery Test, were created via OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). These tasks were administered on a Dell Inspiron laptop. All sessions were recorded using a digital audio recorder.

**Data Collection and Scoring.** The Wechsler Intelligence Scale for Children was scored according to the published procedures. Individuals' scores were classified into the bands specified in the scale, which were: extremely low, borderline, low normal, normal, high normal, superior, and extremely high. The Woodcock Reading Mastery Test subtests (where administered) were scored according to the published procedures. The test scores were then compared against the published ranges for all readers, which are expressed as age-equivalent scores. For all other tasks, accuracy was scored during the session by the researcher, and where appropriate, cross-checked against the audio recording. For the Problematic Letter Patterns test, reading latencies were calculated manually from the digital recordings of the sessions in the same manner as for the Heterogeneity study. These will be discussed further in the method of the subsequent Intervention phase.

For the phoneme segmentation test, we defined impairment as a score one standard deviation or more below the mean from McBride-Chang (1995), which was based on data from a group of eight to nine year-old unimpaired readers (there was no data available for the age group of the current participants, 13 to 14 year olds). For the phoneme deletion task, scores were converted to age-equivalents according to the standard procedure set out in Hammill et al. (2001). We defined impairment on this task as a score that was two or more years below chronological age. For the orthographic choice task we again defined impairment as a score one standard deviation or more below the mean from Manis et al. (1996). This definition was based on data from a group of with an average age of nine years-old. For the homophone selection task we also defined impairment as a score one standard deviation or more below the mean from Castles and Coltheart (1996) respectively. This definition was based on a group of nine to 10 year-olds. For the supplementary phoneme identification and rhyme tasks, we again defined impairment as a score one standard deviation or more below the mean for the oldest group (six year-olds) in Muter et al. (1998).

**Results of Profile Building**

Table 3.7 provides the individuals' scores on each of the Profile Building measures, and an indication of whether they fell below the cut-off for impairment on that measure. Regarding the Weschler Intelligence Scale for Children, four individuals had composite scores within the normal range for their age, one had a score within the lower borderline range (MI), and the remaining participant had a score that fell just within the extremely low range (SH). Participant SH was kept in this study because they represented one of only a few cases with the relative whole-word impairment profile. However, it is important to keep this finding in mind when we interpret other data for this participant.

Table 3.7 Performance on Profile Building tasks. Impairments are indicated in light grey. WISC[a] scores are calculated via the standard procedure. Phoneme, Homophone and Supplementary phoneme task scores measured as the mean proportion correct. Oral Fluency and Word Comprehension scores are measured in months.

| Participant | Case | WISC score[a] | Phoneme tasks | | Homophone tasks | | Supplementary Phoneme tasks | | Oral Fluency | Word Comprehension |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Segmentation | Deletion | Orthographic | Homophone | Identification | Rhyme | | |
| HE | P[b] | 112 | 0.42 | 0.80 | 1.0 | 0.94 | *na* | *na* | *na* | *na* |
| MI | P[b] | 72 | 0.50 | 0.85 | 0.85 | 0.81 | *na* | *na* | *na* | *na* |
| OK | P[b] | 116 | 0.54 | 0.85 | 0.95 | 0.75 | 0.88 | 1.0 | 126 | 165 |
| PI | P[b] | 96 | 0.50 | 0.80 | 0.95 | 0.81 | 1.0 | 1.0 | 137 | 153 |
| FR | WW[c] | 81 | 0.71 | 0.90 | 1.0 | 0.87 | *na* | *na* | *na* | *na* |
| SH | WW[c] | 67 | 0.21 | 0.65 | 0.95 | 0.93 | 0.88 | 0.70 | 117 | 99 |

Note: a) WISC score= Weschler Intelligence Scale for Children composite score, b) P= Phonological, c) WW= Whole-word

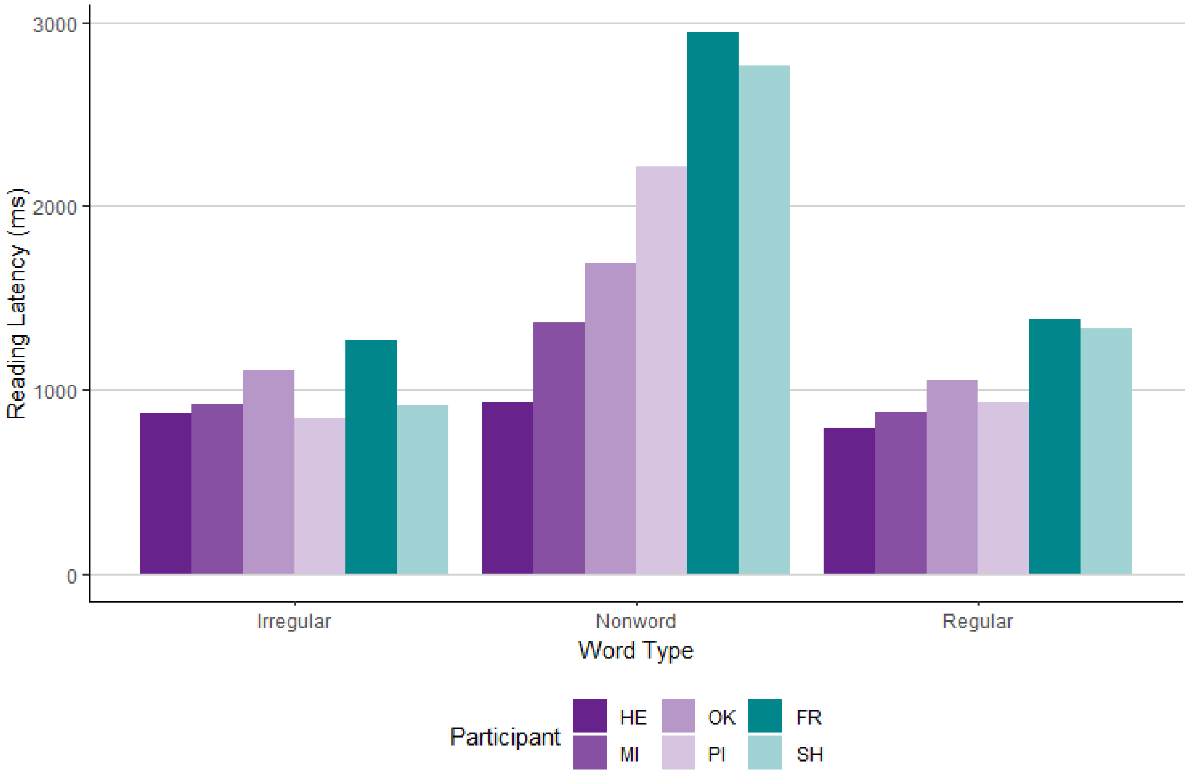
Considering the phonemic awareness tasks, none of our phonological cases were impaired on the phoneme segmentation task, but all four were impaired on the phoneme deletion task. The criteria for impairment on these two tests was derived from a younger control sample, which may have rendered them less sensitive at detecting impairments in a sample of adolescents. One of the whole-word cases, case FR, was not impaired on either task, as predicted; however, whole-word case SH was considerably impaired on both tasks. It is important to bear in mind that SH did not meet our original definition of a pure whole-word case (her nonword accuracy was not within the normal range), so she may share some features with phonological cases. Also, SH was the only case whose composite score on the Wechsler test fell within the extremely low range, so she may suffer from additional cognitive difficulties which may impact her performance across numerous tasks.

Considering the orthographic choice and homophone selection tasks, our prediction was that whole-word cases would exhibit deficits on these tasks, but phonological cases would not. Contrary to this expectation, none of our whole-word cases were impaired on either the homophone selection, or the orthographic choice task. Furthermore, while none of the phonological cases were impaired on the orthographic choice task, three out of the four cases were actually impaired on the homophone selection task. These findings contrast with previous literature (Castles & Coltheart, 1996; Law & Cupples, 2015; Manis et al., 1996).

For the supplementary phonemic awareness tasks, phoneme identification and Rhyme Detection (completed by the 2017 cohort only), no participant showed any impairment on either task. For the Woodcock Reading Mastery subtests (completed by the 2017 cohort only)**,** the age-equivalent estimates for Oral Fluency, as measured by the paragraph fluency task, were below the chronological age for all participants who were tested. These results suggest that paragraph-level reading was impaired, irrespective of the specific impairment profile. The age-equivalent estimates for Word Comprehension were below the chronological age for one phonological case (PI) and one whole-word case (SH), but the remaining phonological case OK scored slightly above his corresponding age-norm.

**Comprehensive Examination of Individual Profiles.** As part of the Profile Building phase, we investigated some of the individual data from the Screening phase in more detail. Considering all data available for each participant, from both of these sources, there are a number of additional observations worth noting. Considering first participants' reading latencies, Figure 3.7 shows the average latencies for each individual, for each of the three lists in the Coltheart and Leahy task. Contrasting phonological and whole-word cases we performed a Linear Mixed-Model (*lmer*) analysis with latency as the dependent variable, participant profile, word type and the interaction as fixed effects, and word identity as random effect. We found a significant interaction with profile and word type ($F$ (319, 2) = 16.6, $p$ < .001). Post-hoc analysis (*TukeyHSD*) revealed that, unsurprisingly, nonword latencies were longer for the whole-word cases than the phonological cases ($t$ (329, 1) = 8.48, $p$ < .001). Whole-word cases also showed longer reading latencies on regular words compared to phonological cases ($t$ (306, 1) = 3.23, $p$ = .017), however there was no difference in the reading latencies on irregular words ($t$ (321, 1) = 0.91, $p$ = .943). These results suggest reading latencies across multiple word types are affected in the current whole-word cases.

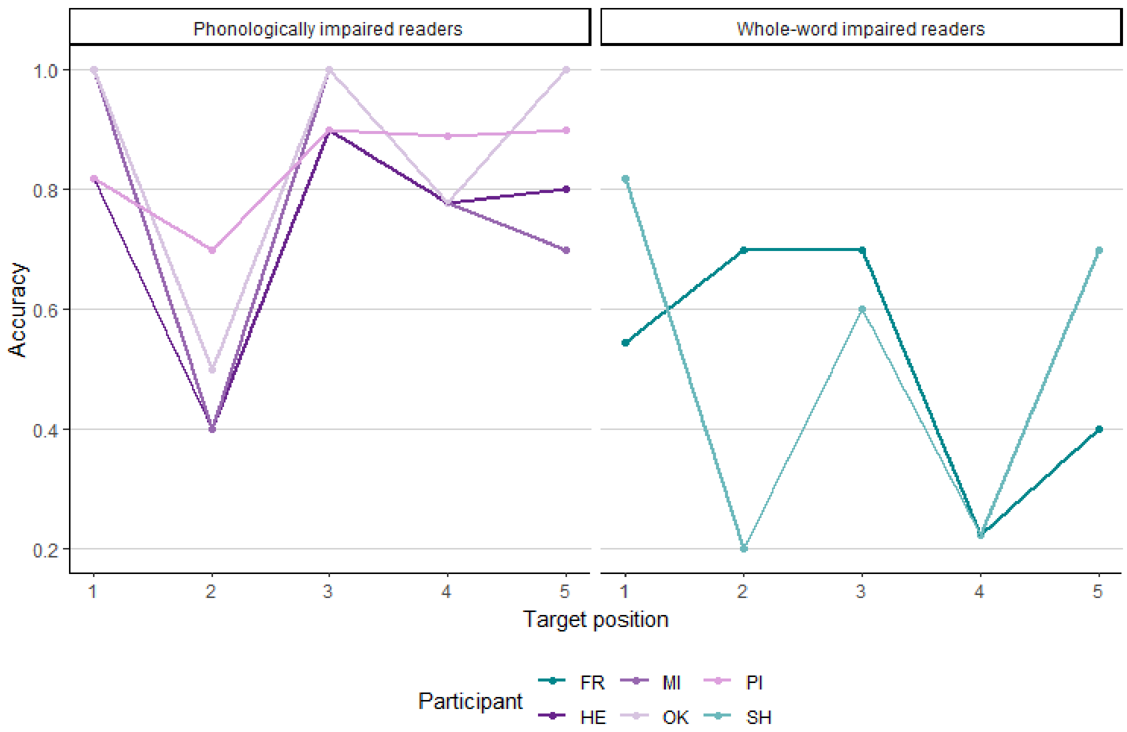
Figure 3.7 Coltheart and Leahy reading latencies according to word type. Phonological cases are represented in purple, whole-word cases are represented in green.




Turning now to the Partial Report task, aggregate scores for each individual are presented in the Screening phase (see Table 3.4), but given the suggestion that phonological and whole-word cases may have different serial position curves, it is worth examining these patterns for each of our cases individually. Figure 3.8 shows the serial position curves for each participant, created from their average accuracy at each position in the five-letter array.

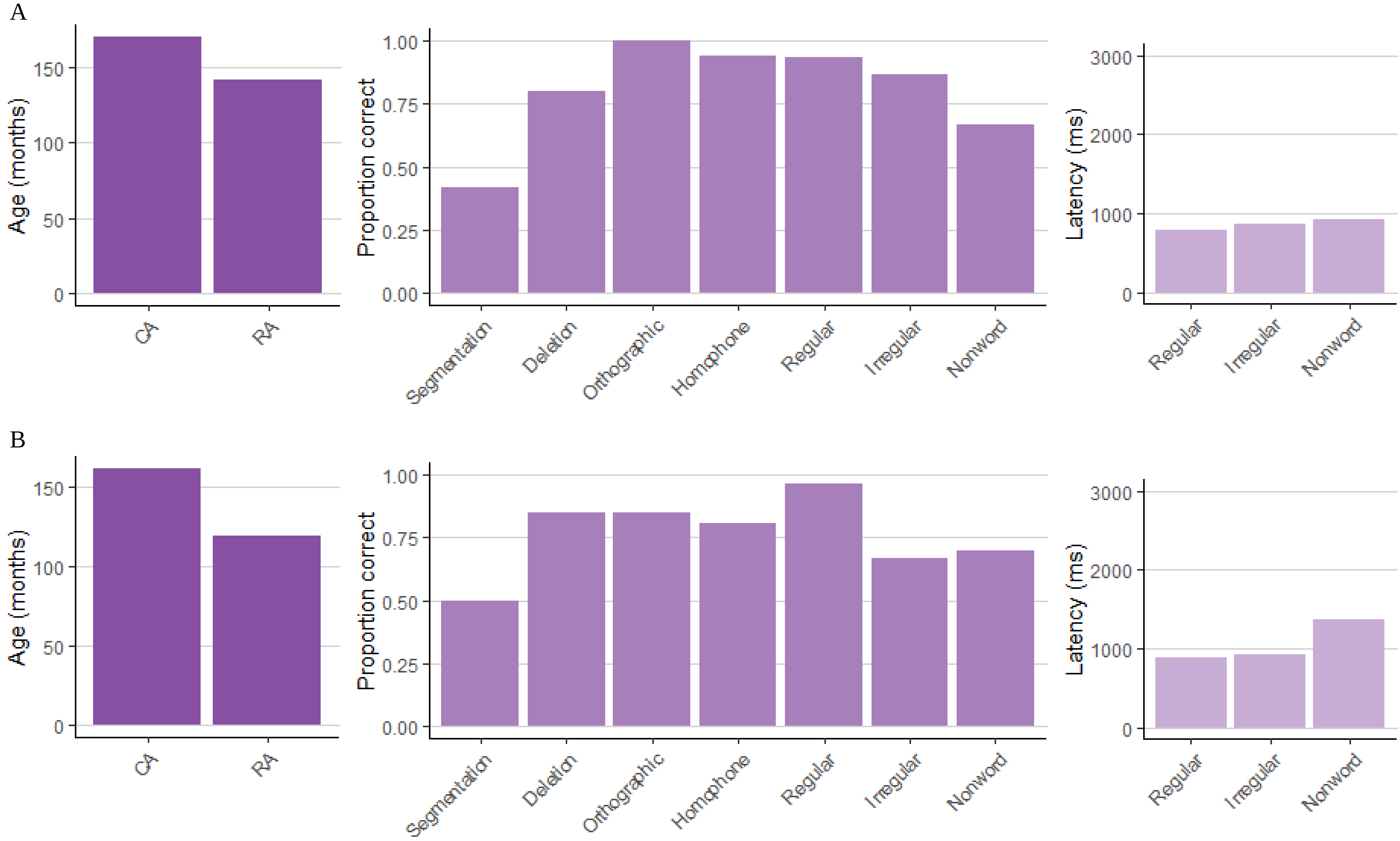
What is of interest here is not the overall accuracy (since overall accuracy was one of the criteria we used to define cases as phonological and whole-word in the first place), but rather the pattern across letter positions. Of the phonological cases, participants HE, MI and OK show a similar pattern: a typical W-function with relatively similar accuracy levels (although there is some variation at the fifth position). The remaining phonological case, participant PI, shows a flatter function, particularly across positions three to five. Of the two whole-word cases, participant SH shows a typical W-function. Once again, it is important to acknowledge that this individual did not fulfil the original criteria for pure whole-word cases, so similarities to phonological cases maybe expected. Finally, participant FR shows an unusual pattern. The resemblance of this pattern to the inverted U-shape that has been reported for previous whole-word impaired readers is interesting and worth keeping in mind when we analyse this person's response to the interventions (Valdois et al., 2003).

Figure 3.8 Serial position curves for each individual (VA span Partial Report task).

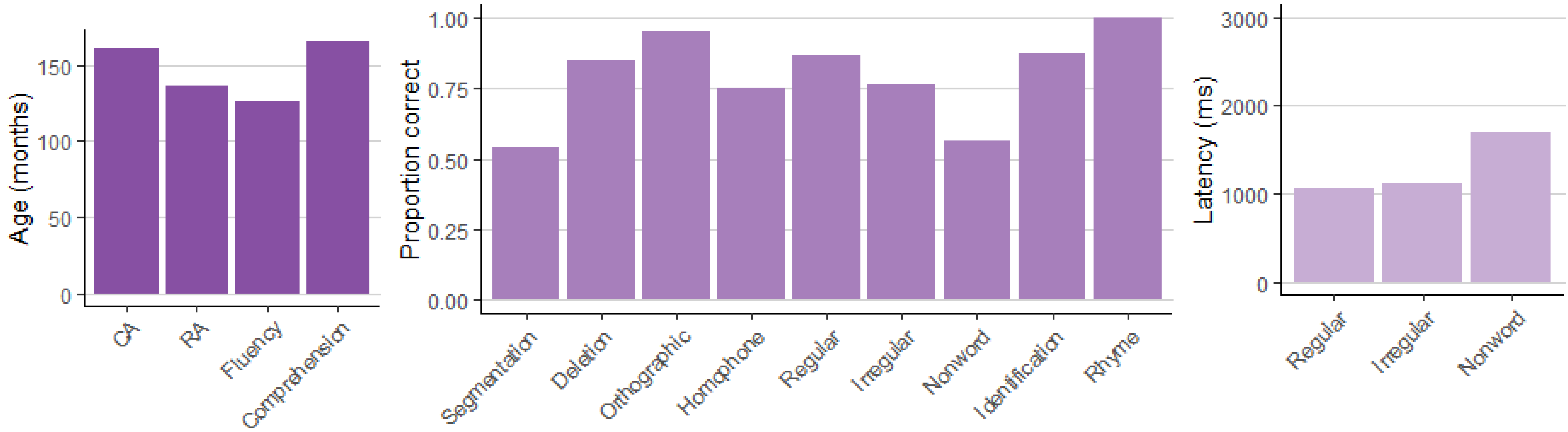



In order to summarise individual's performance across the various tasks and provide a useful reference for evaluating each individuals' results from the subsequent Intervention study, we created detailed graphical representation of their profiles. Phonological cases are presented in Figure 3.9, while whole-word cases are presented in Figure 3.10 These profiles included performance across all tasks for which we have data (across both the Screening and Profile Building phases).

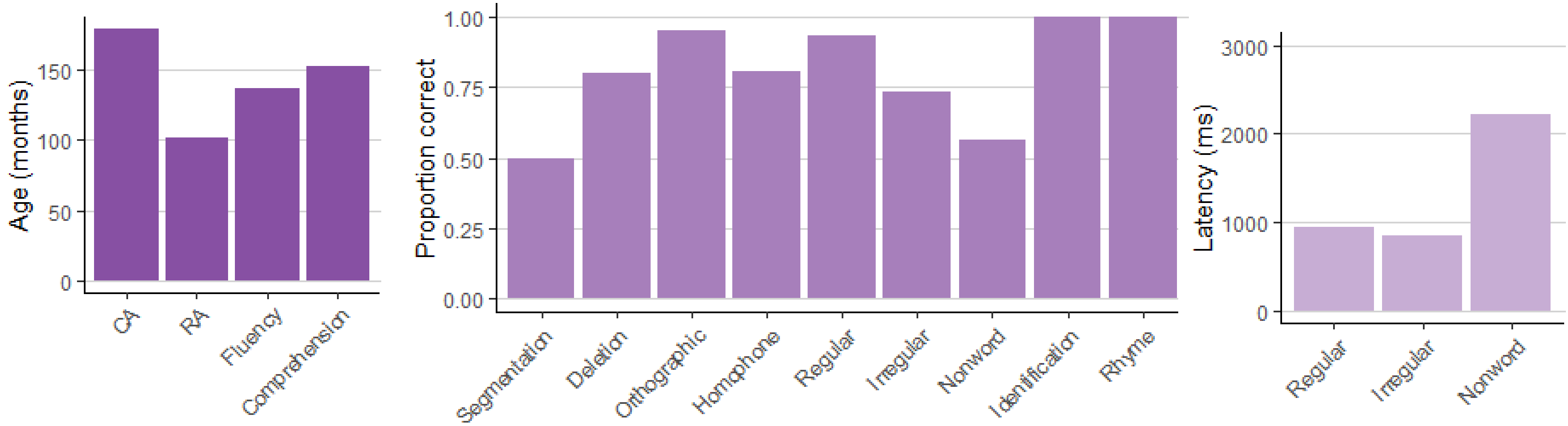
Figure 3.09 Extensive profiles for phonological cases HE (panel A), MI (panel B), OK (panel C) and PI (panel D).



C

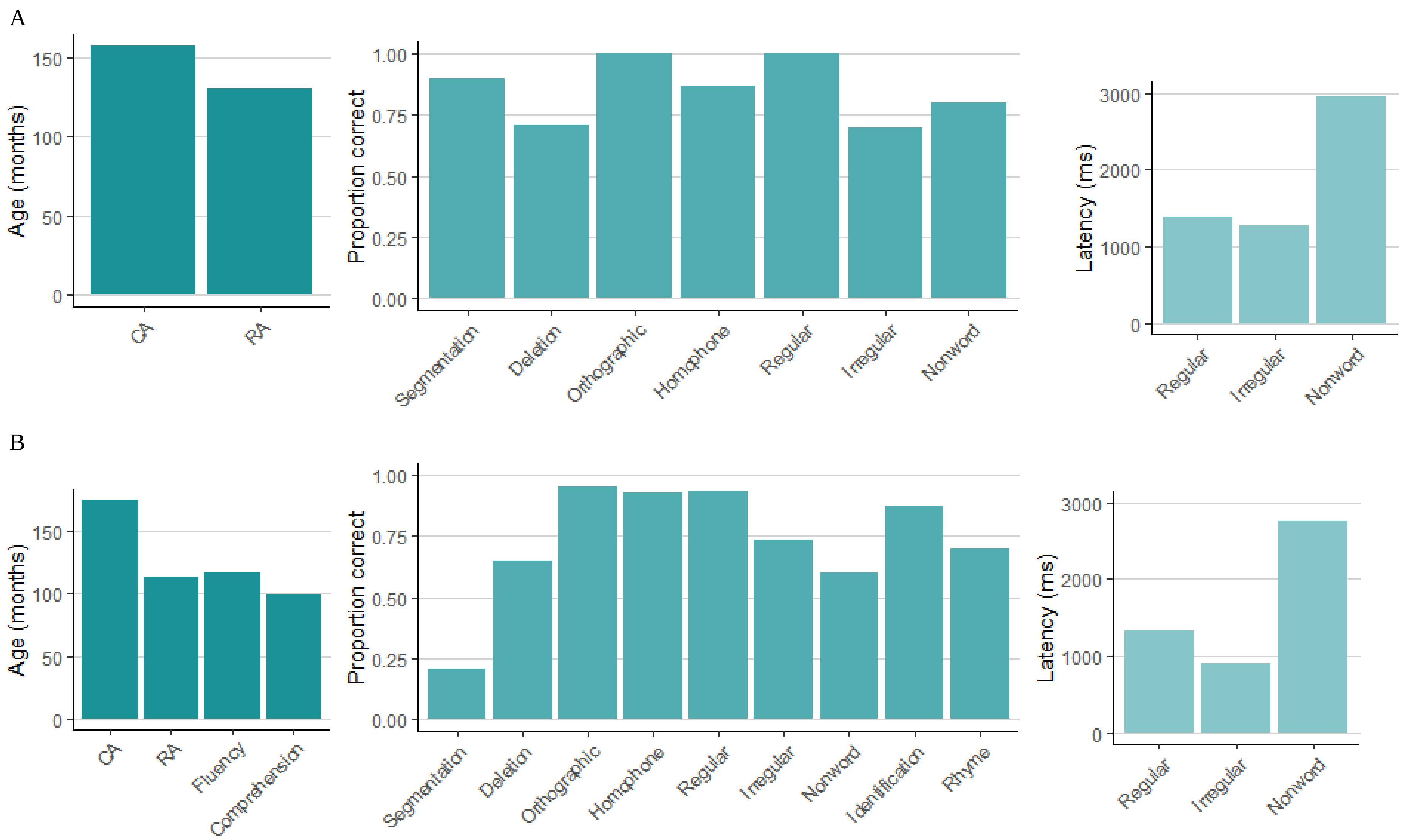



D




Note: CA= Chronological Age, RA= Reading Age, WISC= Weschler Intelligence Scale for Children

Figure 3.10 Extensive profiles for whole-word cases FR (panel A) and SH (panel B).



## Discussion

In many ways, the results for the Screening phase were comparable to those of the Heterogeneity study of Chapter 2. As predicted, our key measure of phonological processing (nonword reading accuracy) was again positively associated with reading proficiency, as indexed by Reading Age. Also, consistent with our predictions, all our key measures of whole-word processing – irregular word reading accuracy, VA span and the difference in reading latencies for real words and nonwords – were also associated with Reading Age. There were also significant associations between VA span and most of the other key measures, including an association with nonword accuracy, which is counter to our predictions but consistent with our findings from the heterogeneity. The exception was irregular word accuracy, which, counter to our predictions but again consistent with our findings from the heterogeneity study, was not reliably associated with VA span. The selective nature of these relationships in a sample of impaired readers who might otherwise be expected to show a pattern of inter-correlations between all measures is interesting. Also, the lack of association between VA span and irregular word accuracy (a key measure of whole-word reading) again suggests that VA span scores do not capture those processes that appear to be particularly critical to reading at a whole-word level. We return to this point in the General Discussion.

Several other observations were also consistent with those in the heterogeneity study. First, as a group, participants demonstrated the same W-shape serial response curve on the Partial Report task. Second, they showed similar distributions of lexicalisation and regularisation errors. However, the inter-rater reliability was much poorer than in the heterogeneity study, perhaps because the sample included a larger proportion of impaired readers, whose errors are likely to be more complex and maybe difficult to classify.

**Detailed profiles of selected cases.** Turning now to the primary objective of the screening phase – the identification of relatively pure whole-word and phonological cases. A majority of the individuals in the current study showed a mixed profile. Indeed, whole-word cases and phonological cases are rarely pure, most cases show a combination of both impairments (King, Giess, & Lombardino, 2007; Peterson, Pennington, & Olson, 2013; Peterson et al., 2014). We identified several pure cases of phonological impairment, however, identifying pure whole-word cases proved more difficult. Indeed, our original categorisation criteria only identified one participant (FR) that could be considered to have a selective impairment in whole-word processing. To try and increase the representation of whole-word impairments, we loosened the criteria for the categorisation of whole-word cases and included those who displayed moderate nonword accuracy deficits, as long as deficits on the whole-word measures were greater. With the loosened criteria we were able to recruit one additional case, participant SH.

In the Profile Building phase, participants exhibited immensely varied deficits. We predicted that phonological cases – but not whole-word cases – would show deficits on our key measures of phonological awareness (phoneme segmentation, phoneme deletion, and where applicable; phoneme identification and Rhyme Detection). However, while all four phonological cases met the criteria for impairment on the phoneme deletion task, none were impaired on any of the other tasks. The results for the whole-word cases were also mixed: one case (FR) was not impaired on any of the measures, but the other case (SH) met the criteria for impairment on both the phoneme segmentation and deletion tasks. Again, it should be borne in mind that SH was not a pure whole-word case (her nonword reading accuracy was also in the moderately impaired range).

We also predicted that the whole-word cases – but not the phonological cases – would show deficits on the orthographic and homophone measures used to assess a reader's ability to map the orthographic pattern to a lexical representation. We found that neither of the whole-word cases were impaired on either task. Also, and contrary to our prediction, three out of four phonological cases showed impaired homophone selection scores (MI, OK, PI). These findings are not consistent with previous case studies of similar cases, who have shown impairments on one or both tasks (Coltheart & Leahy, 1996; Law & Cupples, 2015; Manis et al., 1996). It is possible that whole-word cases did not show impairments on these tasks because the underlying processing deficits in these individuals were not severe enough. In addition, neither individual may have had a truly pure impairment profile (as indicated previously, with SH). Furthermore, phonological cases may have some whole-word processing impairments, and thus these four cases may not have been truly "pure" phonological cases. Alternatively, these results could reflect phonological demands of the tasks. The homophone selection task specifically presents two stimuli; a real word and a foil nonword. Phonological cases who struggle to read nonwords accurately maybe impaired on the task because it requires them to read nonwords, not because of any underlying whole-word processing deficits.

For participants from the 2017 cohort, we also examined single word comprehension and paragraph-level reading fluency (in the 2017 cohort only). On single word comprehension, we predicted that the whole-word case who completed this task (SH) would show an impairment, but the two phonological cases who completed it (OK and PI) would not. However, scores indicated impairment in one whole-word case (SH) and one phonological case (PI). It is possible that impaired readers, irrelevant of their profile, have limited exposure to vocabulary. On the paragraph-level reading fluency task, we predicted that all participants would show deficits irrelevant of their specific impairment profile, and this prediction was upheld, suggesting that all participants struggled with the integration of multiple reading processes. Overall, these results suggest that the reading impairments in the current sample were not limited to the identification of single words, and that a number of other processes not focused on in this study, or targeted in subsequent the intervention, were affected. We return to this latter point in the general discussion.

The results of the Screening and Profile Building study also speak to variation across participants sharing the same profile. For example, one participant, HE, reported during a Profile Building session that they were an avid reader, and in subsequent intervention sessions, they also discussed creative writing workshops they attend. Their enjoyment of and increased engagement in reading could explain why their impairments were less severe than their fellow phonological cases were. In contrast, participant PI, again a phonological case, reported a dislike of reading and overtly stated this was because they found it difficult to follow words across the page. This participant would not have had the same opportunity to compensate or partially overcome their reading difficulty as HE.

**Recommendations for the future.** Given the findings, from both the previous and current research, there are several minor changes or additions that should be considered for future screening attempts. To start, it may be worthwhile including a measure of word length effect, using tasks like the Linguistic Properties Effect task, as a criterion of the whole-word impairment profile. Estimating the degree to which an individual's reading speed is affected by word length may be a more specific measure of parallel letter processing difficulties than the current NonwordAccuracyDiff. VA span scores should also be omitted, until some of the issues raised in Chapter 2 can be addressed and a clearer understanding of what role, if any, visual attention has in the reading processes.

A measure of spelling accuracy and the assessment of error patterns may be useful in increasing the sensitivity of the screening, particularly given that reading and spelling or other writing deficits often accompany one another (Banfi et al., 2018; Daigle, Costerg, Plisson, Ruberto, & Varin, 2016; Georgiou et al., 2012; Romani, Olson, Ward, & Ercolani, 2002; Valdois et al., 2003; Zoubrinetzky et al., 2014). Specifically, previous research has found that typically, phonological cases write regular words as nonwords that are visually similar, but phonologically incorrect (Castles & Coltheart, 1996; Romani, Di Betta, Tsouknida, & Olson, 2008; Rowse, 2005). Conversely, whole-word cases regularise irregular words in the same manner as when reading them aloud., and that the nature of the spelling errors provide another useful index of the underlying deficit.

Future studies could also use a number of additional measures of phonological processing skills, such as: a) an alliteration task where the participant must list words starting with a target letter; b) an acronym or spoonerism task where they must combine the first phonemes of verbally-presented words; and c) a rhyme fluency task where the participant must list words that rhyme with a given target. Specifically, the reading sub-tasks of the WIAT may be useful (Wechsler, 2009).

Finally, future studies may benefit from having a single screening phase, containing all necessary tasks, that is all the screening and profile building tasks could be given to everyone participating in the screening phase. Doing so would allow for fine-grained analysis to be completed before selecting participants for the intervention, potentially reducing confounding or ‘non-pure’ profiles. However, as the current research suggests, finding adolescents with completely pure impairment profiles, phonological or whole-word, may prove difficult.

In the current study, we found that the overall association trends were largely repeated in this partially independent sample. Specifically, we found that the phonological index, whole-word indices and VA index were all associated with the index of reading proficiency. However, the primary aim of the current study was to identify suitable participants for the subsequent Intervention study. We successively identified and recruited six individuals with disproportionate phonological or whole-word profiles.

# Chapter 4 : Intervention study

In Chapter 2, we found that several different indices of cognitive processing were determinants of overall reading proficiency, and in Chapter 3, we found that poor readers vary with respect to their pattern of scores across these different indices. Such findings suggest that the cognitive factors underpinning developmental dyslexia may vary from person-to-person. If the underlying causes of reading impairments are heterogeneous, then reading improvements should be greater after a targeted intervention than after a more generic or non-target intervention. In this study, we worked with adolescents displaying two contrasting profiles of reading difficulty: one, a profile of phonological processing impairments, the other a profile of whole-word processing impairments (as defined in Chapter 3). These readers participated in two types of intervention: one targeted phonological processing skills, the other targeted whole-word processing skills. For each case we then examined whether the targeted intervention resulted in greater benefits than the alternative non-target intervention, and if so, whether the pattern of improvement was consistent with the rationale underlying the training approach. In the following section, we review studies that have used interventions designed to target phonological and/or whole-word reading strategies, with particular focus on single-case research and other investigations that have identified relatively pure impairment profiles. We then outline the approach that will be taken in the current study.

## Interventions designed to target phonologically-based reading

A common method used to improve reading skills is to encourage part-based processing by breaking words up, often placing a heavy focus on phonemic processing (Ball & Blachman, 1988; Lovett et al., 1994; Swanson & Hoskyn, 1998; Torgesen et al., 2010). These methods may use some, or all of the following approaches: 1) explicit training identifying graphemes (e.g. *ph*) and converting them into their corresponding phonemes; 2) implicit training knowledge of grapheme-phoneme correspondences (GPCs), by working with “families” of words that share the same GPC (e.g. *graph, photo*); and 3) improving phonemic awareness skills through tasks that require participants to segment, blend or otherwise transform spoken words. For example, Williams (1980, 1981) used a range of tasks that included breaking a word into its component graphemes and converting each one into a phoneme, blending separated phonemes together, and spelling words out. A large group (63) of undifferentiated readers (seven to 12 years-old) showed improvements on both trained and untrained word reading accuracy following this intervention. Another study combined a grapheme-phoneme analysis task similar to that of Williams (1980, 1981), with a phoneme blending task, and found that children (five to six years-old) with phonemic awareness deficits showed significantly larger improvements when exposed to both tasks, as opposed to just the blending task (Torgesen et al., 1992). Several more recent studies using similar designs have replicated these findings (see Schneider, Roth, and Ennemoser, 2000; for a metanalysis see Bus & Van IJzendoorn, 1999), suggesting that training solely with phonemes is insufficient; effective phonological treatments also need to focus on graphemes.

Of course, studies focusing on groups of undifferentiated participants cannot tell us whether these interventions provide merely general reading practice, or whether they operate in a more specific way to address phonologically-based reading difficulties. To resolve this matter, we need to turn to studies of more narrowly delineated groups, or individual cases. Broom and Doctor (1995a) used a combination of explicit training with difficult GPCs and implicit training with exemplar words that contained these GPCs via word reading and spelling. Training was given to one case, an 11 year-old delayed reader with prominent phonological difficulties. This individual showed improved reading accuracy, not only on the trained words, but also on an untrained set of regular words and nonwords. Several subsequent studies with similar cases have also adopted this method and again, observed improvements to both trained and untrained word stimuli (Jones, 2013; Rowse & Wilshire, 2007). The observed improvement on untrained nonwords in these studies is consistent with rationale underlying the intervention; that is, that training phonological, part-based reading skills should produce benefits that are generalisable to completely novel words.

While these findings appear promising, some of the studies described did not examine whether the techniques used were *more* effective than whole-word based approaches, nor whether they were better suited to those with documented phonological impairments, than to those with an alternative pattern of impairments. There is some evidence that GPC training programmes are less effective for individuals whose impairments either extend beyond or are completely separate of phonological difficulties (Brunsdon, Hannan, Nickels, & Coltheart, 2002; McArthur et al., 2013; O'Shaughnessy & Swanson, 2000). For these individuals, programmes targeting reading at the whole-word level may be more effective.

**Interventions designed to target whole-word based reading**

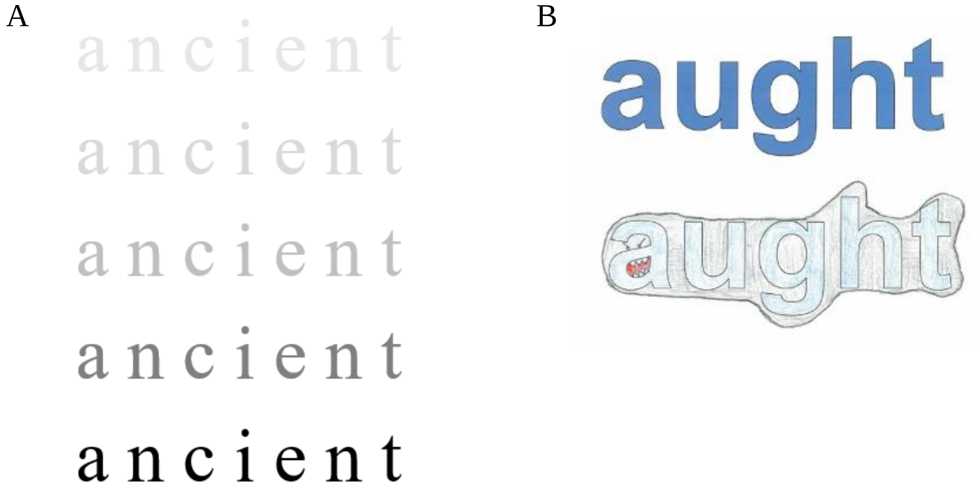
Fewer studies have focused on improving whole-word reading strategies, and for those that have, the methods used, and their effectiveness have varied greatly. One whole-word specific approach involves focusing on words that cannot be pronounced via common sight-sound rules. Broom and Doctor (1995b) worked with another 11-year-old case, this time displaying whole-word reading difficulties. They developed an intervention which focused on irregular word reading and spelling practice, explicitly demonstrating to participants how these words break GPC rules. They found some significant improvement to trained word reading accuracy; however, this improvement did not extend to untrained regular or irregular words. One reason may be that this intervention did not target any of the underlying causes of whole-word impairments, but rather the superficial symptoms.

More recently studies have focused on encouraging the visual recognition of words as whole units, irrespective of their spelling pattern. These studies fall roughly into four broad categories. The first involves using brief presentation in order to facilitate fast visual recognition of a word as a whole unit. Judica et al. (2002) presented words on-screen for a very short amount of time (60-150ms), and then asked participants to either read aloud, read silently or type the word back as a response. When administered to an undifferentiated group of impaired readers (11 years-old), this intervention produced generalised learning benefits: significant accuracy and latency improvements to untrained real words and nonwords.

The second type of approach focuses on training visual attention and visual identification skills more generally. For example, Valdois et al. (2014) used a combination of visual search, discrimination, matching and parsing tasks with a seven-year-old whole-word impaired reader. There were significant accuracy and latency improvements across untrained regular, irregular and nonwords. However, this study did not compare the outcomes of this visual training to outcomes of a phonologically-based intervention, or recruit a phonological case for comparison.

The third approach places greater emphasis on meaning and the global form of the written word. For example, Rowse and Wilshire (2007) presented words to participants on visual-mnemonic flashcards, in which the words were overlaid onto semantically-related pictures (e.g. the word *sew* was printed over the image of spool of cotton). Subsequently, they presented the same word sets as degraded images, which gradually became more distinct (Figure 4.1 Panel A). Participants were required to read the word out loud as soon as they were able to identify it. A whole-word impaired reader (10 year-old NS) demonstrated reading accuracy improvements on both trained words and untrained words, specifically irregular words. In contrast, a phonologically impaired reader (10 year-old WB) who completed this programme showed improvements that were limited to trained words: there was no significant generalisation to untrained words. Also, for WB, this programme was not as effective for trained words as a phonological programme. However, one inconsistent outcome of this study was that NS, the whole-word case, also benefitted significantly from a more conventional phonological intervention, so it could not be concluded that the visual intervention was *superior* to the phonological intervention in targeting whole-word impairments.

Figure 4.1 Demonstration of stimuli used in prior visual interventions. Panel A shows the degraded images from Rowse and Wilshire (2007); Panel B shows the cluster flashcards from Jones (2013).




A fourth visual approach involves working with redundant letter sequences – that is, sequences of letters that repeatedly occur in the same order in multiple different words (e.g. *ation*: *vacation, creation, national, alienation, stationery*). Jones (2013) used the degraded images described in Rowse and Wilshire (2007) but supplemented this with a task focusing on these redundant letter clusters. They used printed flashcards containing a four-to-five letter redundancy, incorporated into an illustration that emphasised the shape of the cluster (Figure 4.1 Panel B). The participant then rehearsed a short list of words containing this cluster. One 11-yeard-old participant with a whole-word deficit showed improved reading accuracy on untrained words following this intervention. However, again, this intervention did not produce reliably greater benefits than a standard, phonological intervention for this individual.

A similar approach, which also focused on cluster-style letter redundancies, was adopted by Law and Cupples (2015). Participants were shown a cluster and asked to generate words containing that letter pattern. After discussing the different ways that cluster could be pronounced based on the word (i.e. *ough* in "*cough*" vs "*dough*"), participants were asked to imagine a scene that incorporated the meanings of those words (e.g. a baker *coughing* as he rolled out his *dough*). They then practiced reading other words that also contained the cluster. Following this intervention, two cases with relatively pure whole-word impairments (seven and nine years-old upon commencing the study) showed improved reading accuracy on trained words, untrained words and redundant words (untrained words containing the trained letter patterns). However, in this study, only whole-word cases were tested, and these individuals only completed training that targeted whole-word based reading. Consequently, we cannot determine whether the benefits of this intervention were general or more specific to the whole-word processes targeted by the treatment. Also, some of the clusters that were trained consisted of only-two to-three letters (e.g. *ai*), and in these instances, the training stimuli may have been superficially similar to those used in standard GPC training programmes.

**Study outline and hypotheses**

In sum, more research into intervention-specific improvements is needed to address whether different underlying cognitive difficulties can be selectively targeted, especially for those whose profile suggests a difficulty at the whole-word level. In the current research, we sought to create a novel intervention for these individuals, based on the rationale that efficient whole-word recognition relies upon effective parallel letter processing. As we discussed in Chapter 1, a number of theories and models emphasise the importance of parallel letter processing in whole-word recognition (Ans et al., 1998; Bosse et al., 2007; Coltheart et al., 2001; Pritchard et al., 2018). As our starting point we took the clustered letter patterns used by Jones (2013) and Law and Cupples (2015). We reasoned that training participants to identify common and highly redundant letter patterns within words (e.g. *ogue*: *rogue, synagogue, dialogue*), might help reduce the load on parallel letter processing. That is, by repeatedly exposing participants to the same clusters, we aimed to encourage them to process words more holistically. By using letter clusters, we hoped to maximise the potential for generalised skill learning, enabling readers to transfer those techniques to other words that were either: a) completely unrelated to the cluster and not included in training (known as *untrained words*); or b) untrained but related to the cluster (i.e. words containing the trained letter pattern, known as *redundant words*). Throughout the programme, we included activities that reinforced the meaning of the words being trained, and also sought to vary the phonemic properties of the cluster in the trained words (that is the cluster did not sound identical in all training words). A moving word task was also used to encourage rapid, whole-word recognition. Our aim was to ensure that the focus was on the visual form of the cluster and/or word, and that training was distinct from the alternative GPC-based training.

The primary aim of the current research then, was to test whether a cluster-based *visual* programme of this kind could offer specific benefits to whole-word impaired readers, above and beyond the more typical phonologically-based *GPC* intervention. In addition, we set out to look at whether phonological cases would show the converse pattern, this is greater benefits following a conventional GPC training than following a visual training. To address these aims, we worked with secondary school students who had been identified in the Screening and Profile Building chapter as having distinct and disproportionate profiles, administering a typical GPC intervention, as well as our novel visual intervention. Participants completed both interventions, irrelevant of their impairment profile. For each individual we used measures obtained before and after each intervention to estimate programme efficacy, directly comparing the visual intervention to the phonological intervention.

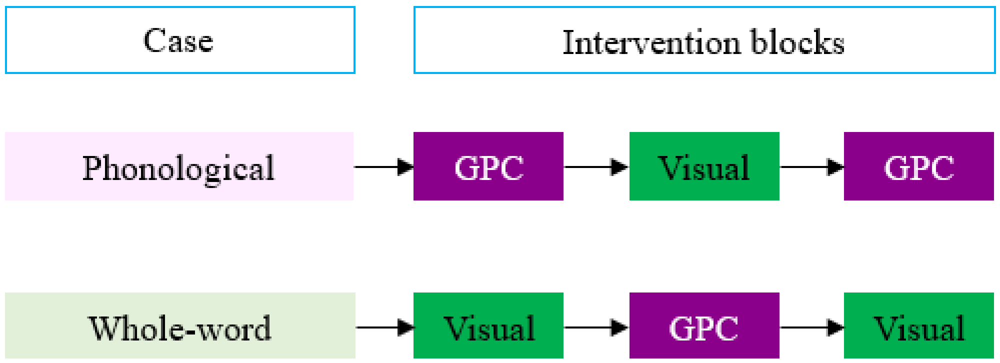
We predicted that whole-word impaired cases would show significantly greater benefit from visual training, when compared to phonological training. Based on the prominence of irregular word reading accuracy and general reading latency impairments in these individuals, we also predicted that performance on these measures in particular would improve, with greater improvements resulting from visual training. Conversely, we predicted that phonologically impaired cases would show significantly greater benefit from GPC training, when compared to visual training. Based on the prominence of nonword reading accuracy impairments in these individuals, we also predicted that performance on this measure in particular would improve, with greater improvements resulting from the GPC training.

**Method**

**Participants.** Participants were the six individuals described in the Profile Building phase of Chapter 3 (phonological cases: HE, MI, OK and PI; whole-word cases: FR and SH). For a full outline of recruitment, demographic details and reading impairment profile, please see Chapter 3. Participants were offered a total of eight movie vouchers for their participation across this study, given in pairs at roughly one-month intervals.

**Design.** The current study utilised an ABA alternating treatment design, with three intervention blocks as illustrated in Figure 4.2. The first and last blocks comprised the "target" intervention, which was the intervention designed to target the individual's specific reading difficulties. For phonological cases, the target intervention was the GPC intervention, and for whole-word cases, the target was the visual intervention. Finally, the middle block comprised the "non-target" intervention (for phonological impaired cases, this was the visual training, and for whole-word cases, it was the GPC training. Note: due to the practical constraints, not all participants completed all blocks; we return to this point later).

Figure 4.2 Outline for intervention for phonological and whole-word cases.




During each intervention block, participants completed 14 individual training tasks, spread across eight sessions, and also completed several reading tests designed to assess their performance at various stages in the training. These included: a) a Pre-intervention reading test, containing lists of regular, irregular and nonwords that were not trained in the intervention; b) a Letter Pattern Acquisition test, administered prior to the intervention, which consisted of words containing the letter patterns to be targeted in the intervention (more detail on these patterns given shortly); c) a Post-intervention reading test, which was identical to the Pre-intervention test but administered at the end of the intervention blocks; and d) a series of Review reading tests, consisting of words containing the letter patterns trained in the immediately preceding session.

**Materials.** Below we describe the materials used in the two interventions. We also provide details of the materials used in the reading assessments described above (the Pre-intervention reading test, a Letter Pattern Acquisition test, a Post-intervention reading test and a series of Review reading tests).

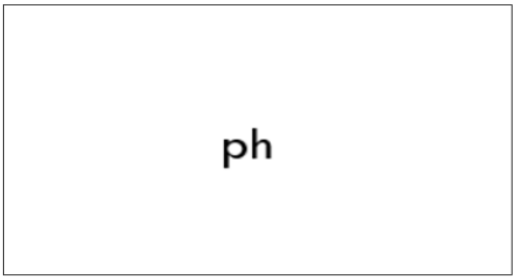
***GPC Intervention.*** The GPC intervention focused on training each participant to sound out and spell words containing target GPCs (e.g. "*ph*"). Rather than merely isolating and training each specific GPC, we took an implicit approach emphasising *families* of words that all shared the target GPC; the exercises were designed to orient the participant's attention to the shared GPC (e.g. *emphasis, paragraph*). The procedure adopted was based on that described by Broom and Doctor (1995a) and is described in further detail below (see also Jones, 2013; Rowse & Wilshire, 2007). Although phonologically impaired readers typically show greater difficulty with novel or non-words, we used real words in the training for ecological validity; an effective intervention for any type of dyslexia should increase real word reading proficiency or else it cannot be considered to have targeted reading processes successfully.

The word families were tailored to the individual and their most problematic GPCs. To identify problematic GPCs, we used data from the GPC list of the Problematic Letter Patterns test, administered as part of the Profile Building phase (see Appendix AI for the stimuli list). Specifically, we sorted all words according to their target GPCs, and then ranked the sets from least-to-most accurate (since there were five exemplars for each GPC, scores were marked out of 5). If two or more GPCs were tied using this method, they were ordered according to the average reading latency for correct trials (see Appendix AK for the full rank-ordered list of GPCs assessed for each participant).

For the intervention materials we then generated 10 exemplar-families for each GPC. These words were taken from the English Lexicon Project (Balota et al., 2007) and the MCR Psycholinguistic database (Coltheart, 1981). We attempted to minimise any variance in difficulty across the families using age of acquisition scores from Gilhooly and Logie (1980), and ensuring the average scores for each family were within a one-and-a-half-year "age" bracket (the average scores for each family fell between 101-117 months). Across a single block, a total of 100 words (110 for the 2017 cohort) were trained See Appendix AN for all possible training families. The training for each GPC family was as follows in Figure 4.3.

Figure 4.3 Outline of GPC intervention.

1. The target GPC was first displayed centrally on-screen. Participants were required to sound the GPC out, and were corrected if necessary. They were then asked if they could think of any words that contained this letter pattern (again, they were corrected if necessary).

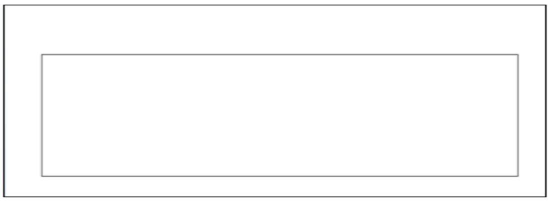



2. The first word in the target GPC family was presented centrally on-screen, and participants were asked to read it out loud. If they made a mistake, they were corrected and encouraged to have another go. The word was then removed from the screen.


3. The participant was asked if they could define the word they had just read, and if they could not, the researcher assisted.

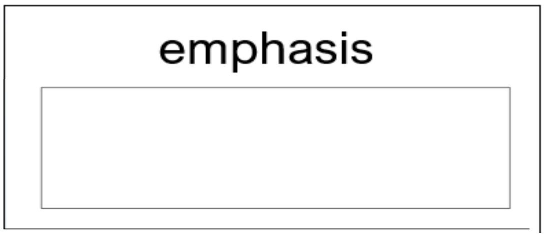
4. The participant and researcher discussed how the word could be used in a sentence.

5. The participant was asked to spell the word. They were required to type their response into a large black rectangle that appeared on-screen. If they spelled the word correctly, they moved onto the next word.



   If not, they were asked to try again using a cue. The cue consisted of the words' letters presented atop the entry box, in a pseudo-randomised order. If they spelled the word correctly, they moved onto the next word.


   If not, they were asked to try one final time with another cue. This time the cue consisted of the letters in the correct order. At this point, participants would proceed to the next word irrelevant of their response.



This process was repeated for all 10 words in the family. Once this was completed, participants practiced spelling the whole family by hand. They were given extra-wide lined paper and asked to write each of the training words as the researcher read them out. Once the participants had attempted all 10 words, they were asked to read them out one-by-one. Their reading and spelling were corrected as needed. The training of the GPC for that part of the session ended with a final writing task – sentence writing. Participants completed this task on the computer. They saw the same blank screen with a black rectangular outline as in the previous spelling practice and were required to type out the sentence as it was read aloud by the researcher (see Appendix AO for a list of stimuli). The sentence contained at least five of the trained words – for example, "*The woman's nephew is good at physics, plays the saxophone, likes pheasants and dolphins and collects autographs."* Participants were told they could ask the researcher to slow down or repeat any words at any point, and that only the trained words were of interest – we were not interested in the spelling of the remaining words in the sentence. The researcher and participant went over the sentence at the end of this task.

***Visual intervention***. The visual intervention focused on training faster whole-word recognition and semantic processing of word families containing the cluster letter strings (e.g. *ation*). Rather than concentrating on sound consistency as we had in the GPC training, we emphasised how, despite phonological variance, these letter patterns repeated throughout a number of words (e.g. *national*, *creation*). We created two novel tasks: a) the *Make n' Break* task; and b) the *Movers n' Shakers* task.

Word families used in the training were tailored to the individual's most problematic letter clusters, as identified in the Problematic Letter Patterns test, using the method outlined for the GPC intervention (see Appendix AJ for a full stimuli list and Appendix AL for the full rank-ordered list of clusters assessed for each participant). Again, for each cluster, we generated a family of ten words containing that cluster, using the same resources as for the GPC intervention. Once again, we attempted to minimise any variance in difficulty across the families using age of acquisition scores (Gilhooly & Logie, 1980). The length of clusters restricted the words available. As a result, there was greater variability across the word families than there was for the GPC families; average scores for each family were within a four-year "age" bracket (the average scores for each family fell between 86-141 months). Across a single block, a total of 100 words were trained (90 for the 2017 cohort). See Appendix AP for all possible training families. The training for each cluster family was as follows in Figure 4.4.

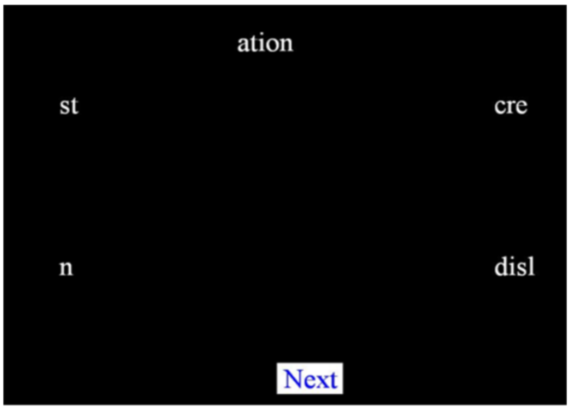
Figure 4.4 Outline of visual intervention.

1. As with the GPC training, the focal cluster for that part of the session was introduced – displayed centrally on-screen. The participant was asked if they could think of any words that contained this letter pattern and corrected if they made any errors.


The participant then proceeded through Steps 2 to 4 outlined for the GPC training (reading the word aloud, discussing meaning and using the word in a sentence). The participant was then asked to spell the word as outlined in the GPC training. However, in this intervention, if they made any errors, they were not offered any further prompts, such as scrambled letters. The above process was repeated for each of the 10 words in the family, one at a time.

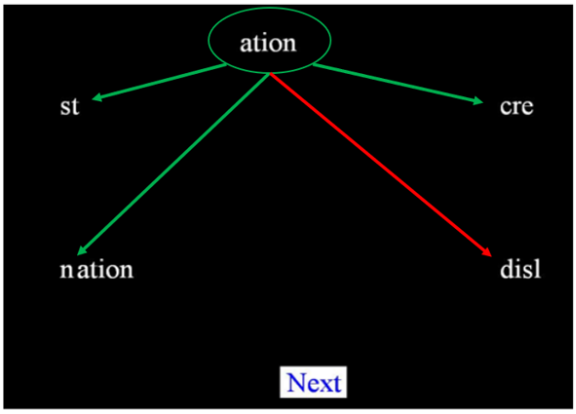
Participants then completed the Make n' Break task for the current word family, which proceeded as follows:

2. For the first part of the task (the "Make" part), the cluster was presented centrally along the x-axis atop the screen, and four different letter strings appeared in the four corners. Some letter strings could be combined with the cluster to make the trained words introduced earlier, while others would make nonwords.




Participants were required to make the real training words by the dragging the cluster down to the target strings using the mouse, while avoid making nonwords.

When the cluster was dragged into a correct position, the word would form and turn green. When the cluster was dragged to an incorrect position, the nonword

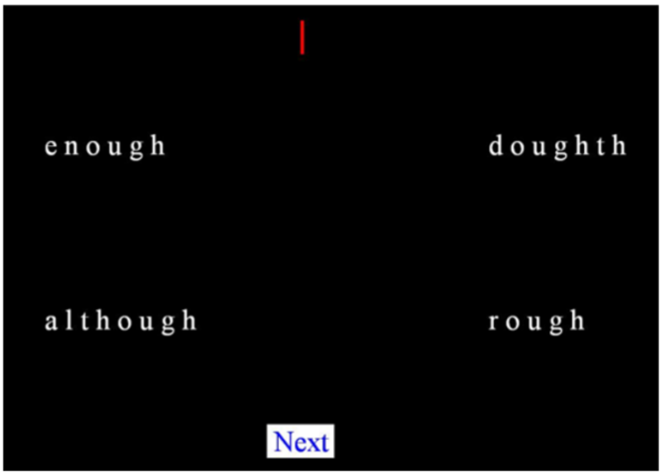


would form and turn red; the cluster then re-appeared in its original position, so another word could be made.

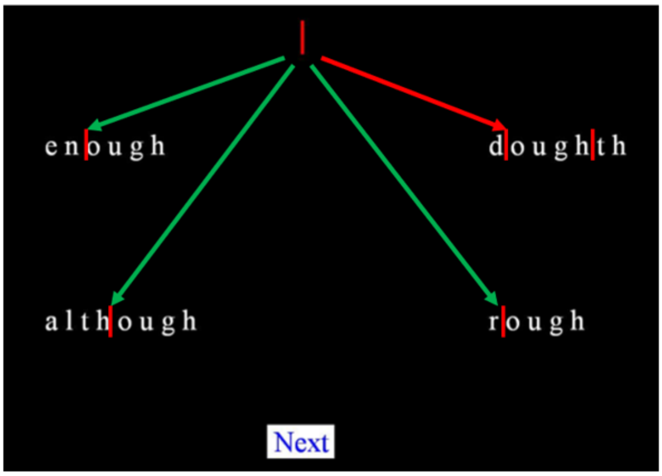
Each trial contained between two to three "real" strings, while the remaining strings formed nonwords. The strings could precede (e.g. *"n"> nation*), proceed (e.g. *"eigh"> eighteen*) or surround the cluster (e.g. *"ne___ary"> necessary*). Once participants made all the real words they believed possible, they clicked the "Next" button, which recorded a screenshot of their responses allowing for scoring later.

3. The researcher then asked the participant to read out the real words they had made and recall their definitions. Assistance was given on words they struggled to read or define, and any words they had not successfully made were also discussed. There were four trials in this part of the task.

4. In the second part of the task (the "Break"), a red vertical bar appeared centrally at the top of the screen, and the four words / nonwords from the "Make" part of the task appeared in the four corners.




Participants were required to drag the bars and separate out the cluster from the rest of the word, but only for real, training words and not nonwords.




After the line had been dragged to a word, the line re-appeared in its original position.

Each trial contained between two-to-three real, training words, with the remaining nonwords. Once participants broke all the real words they believed possible, they clicked the "Next" button, which recorded a screenshot of their responses allowing for scoring later on.

5. The researcher then asked the participant to read out the real words they had broken and recall their definitions. Assistance was given on words they struggled to read or define, and any words they had not successfully broken were also discussed. There were four trials in this part of the task.

   The second task performed for each word family was the Movers n' Shakers task. For each word in the family, the procedure was as follows.

6. The training word appeared on the left of the screen within a white banner running along the central x-axis (superimposed on a black background), and then immediately moved rightwards along the x-axis at a rate of 0.12 mm/ms, disappearing when it had completed its journey. The initial appearance of the word was accompanied by a beep (not audible to participants). The participant's task was to read the word aloud as quickly and accurately as possible, before it disappeared from the screen.

   If the participant was unable to read the word correctly before if disappeared, the word was presented again, however the movement speed was decreased to 0.07 mm/ms; then if they still could not respond correctly, the speed was further reduced in increments of 0.17 mm/ms, with 0.03 mm/ms the slowest moving speed. If the word could still not be accurately read at this slowest speed, it was presented in a fixed position in the centre of the banner.

7. After completing these tasks with the entire word family, a sentence reading and writing task was administered following the same format as in the GPC intervention, for example: "*The boy used the stationary to draw an illustration of the combination of old and new national flags from his imagination"* (see Appendix AQ for a stimuli list).

   The above process was repeated for each cluster family.

***Pre-intervention and post-intervention reading tests.*** These tests consisted of 20 regular words, 20 irregular words, and 20 nonwords, from Coltheart and Leahy (1996), and supplemented by Rowse and Wilshire's (2007) phonological generalisation test which included words of each type, matched to the original Coltheart and Leahy list for frequency, letter and syllable length. They were administered in the same manner as the Coltheart and Leahy task in the previous Heterogeneity study, and the Screening study (words appeared centrally, onset was accompanied by a short beep, words remained on-screen until participants responded, both and accuracy and reading latency recorded). These words were administered prior to each intervention block (the pre-intervention reading test) and again at the completion of the block (the post-intervention reading test). To ensure that improvement over the course of the study could not be attributed to practice effects, we created three versions of this test – Version one, Version two and Version three – each consisting of an entirely different set of words. Version one was always administered in the first intervention block, Version two in the second block and Version three in the third block, irrespective of the nature of the training undertaken in that block (See Appendix AR for the full word lists).

***Letter Pattern Acquisition test (2017 cohort only).*** This test consisted of three words for each of the letter patterns that were to be targeted in the intervention block, totalling 33 words for the GPC test and 27 for the cluster test. These words were not explicitly included in training, but they did contain the (as-yet-to-be) trained letter patterns. The administration was the identical to the pre- and post- intervention reading tests. The Letter Pattern Acquisition test was administered prior to each intervention block along with the pre-intervention reading test, and later merged with the review test (detailed below). See Appendix AS for the stimuli used in the GPC test, and Appendix AT for stimuli used in the cluster test.

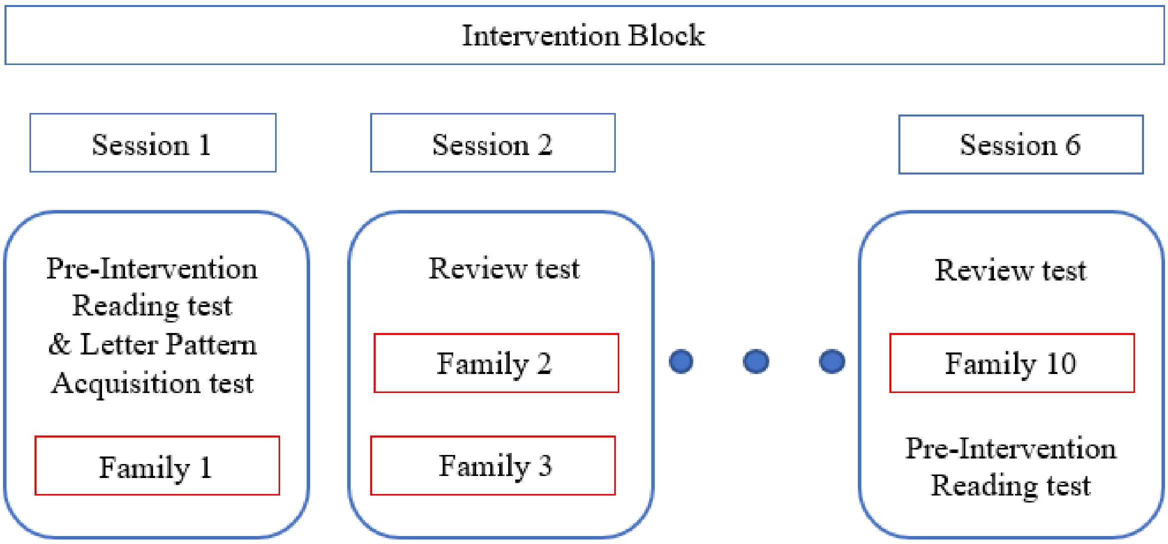
***Review test.*** This test consisted of the 10 words from the families trained in the previous session, as well as the three redundant words (untrained words that contained the trained letter patterns) from the Letter Pattern Acquisition test. Administration was identical to the Letter Pattern Acquisition and pre-post intervention reading tests. There were up to two blocks; each consisting of 13 words containing the trained letter pattern. Feedback was provided once the test was completed to help students with words they still found difficult.

**Overall procedure.** All tasks, other than selected visual training exercises as noted below, were created via OpenSesame (Mathôt et al., 2012). The Make n' Break and Movers n' Shakers tasks of the visual intervention were created and administered in Matlab

(MathWorks, 2005). All tasks were administered on a Dell Inspiron laptop. All sessions were recorded using a digital audio recorder, and a Logitech webcam (picture not recorded).

Each session lasted up to one hour, with two-to-three sessions a week in normal school hours. Sessions were conducted in an identical manner to the previous phases (one-to-one; observer present; questions were encouraged; they could take a break or stop participating at any point without penalty). Figure 4.5 shows how an intervention block was organised. Each block contained six sessions, each roughly divided into two halves. In the first half of Session one, participants completed the appropriate pre-intervention reading test and, (where applicable) the Letter Pattern Acquisition test. In the second half of Session 1, the participant began their first training session, which focussed on a specific letter pattern (a GPC or cluster, depending on the intervention block). Before the training commenced, they were first taken through demonstrations of the various exercises they would complete in that block and encouraged to ask questions.

Figure 4.5 Outline for intervention including what measures were used to assess performance throughout the intervention blocks.




In Sessions two to five, the participant completed two new letter patterns per session, one in the first half of the session, and the other in second half. In Session six, they completed their final letter pattern in the first half of the session, and then in the second half of that session, they completed the Post-intervention reading test, which was identical to the Pre-intervention test. In addition, at the commencement of Sessions two to six, participants completed the Review reading test, which comprised a mix of trained and redundant words containing the letter patterns targeted in the previous session (words trained in the final session of a block were reviewed at the start of the new block). The review test was completed prior to the commencement of the first part of that session. See Appendix AM for an intervention outline of GPCs/clusters in each session.

**Data analysis.** Reading accuracy for all reading tests was recorded during the session by the lead researcher, and cross checked against the audio-recording of the session. Reading latencies were calculated manually from the digital recordings of the sessions using Audacity 2.1.2 (Audacity-Team, 2014), measuring the time from the onset of the beep that accompanied the stimulus, to the onset of the first correct response. The two primary outcome measures were: a) the change in accuracy across the pre- and post- intervention reading tests for each respective intervention block; and b) the change in latency across the pre- and post-intervention reading tests for each respective intervention block.

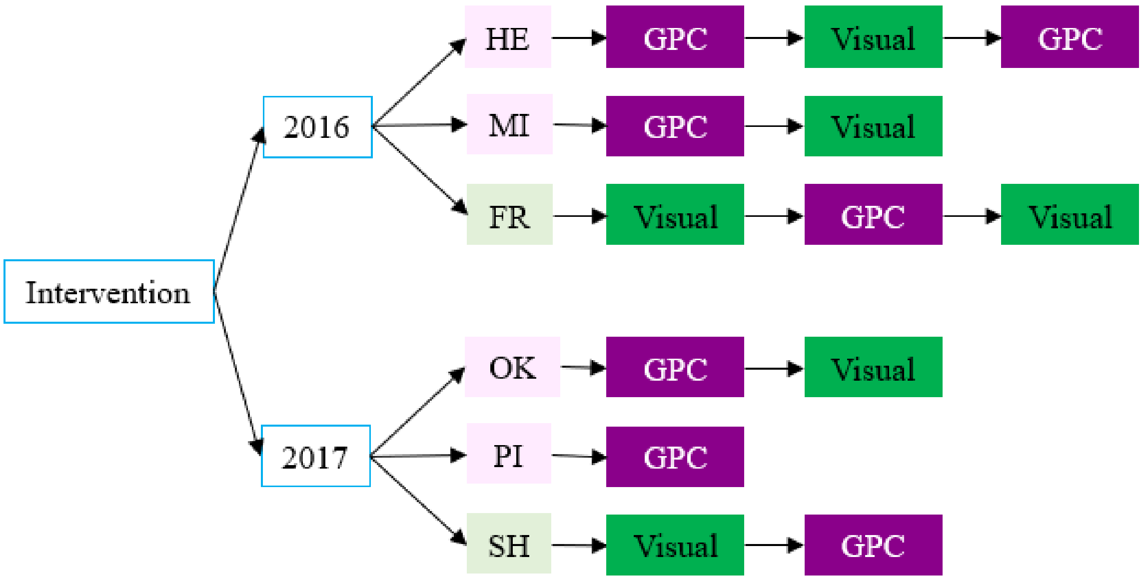
The current set of analyses were performed to assess whether impaired readers benefited from undergoing a tailored intervention programme above and beyond any benefit from a programme that was designed for an alternative profile of impairment. Specifically, we wanted to test whether whole-word cases would benefit from a visual-based intervention programme, more so than from a phonological-based intervention programme. Conversely, we wanted to test whether phonological cases would benefit from a phonological-based intervention programme, more so than from a visual-based intervention programme. Given the sample size, we were primarily concerned with within-subject comparisons, however some group analyses were also completed.

## Results

All analyses were completed in R[14]. Throughout all analyses, the alpha level was $p < .05$. Due to practical constraints, not all participants completed all intervention blocks. Figure 4.6 outlines the blocks actually completed by each participant.

[14] Packages used: Rstuido: R-Studio-Team (2015), lme4: Bates et al. (2014), lmerTest: Kuznetsova et al. (2017), lmTest: Hothorn et al. (2018), Hmisc: Harrell Jr and Harrell Jr (2019), Tidyverse: Wickham (2017), emmeans: Lenth (2018)

Figure 4.6 Structural outline of the intervention as completed by each participant. Phonological cases/GPC intervention blocks are shaded purple, whole-word cases/ Visual training blocks are shaded green.




**Primary Outcome measures.** As in previous analyses, trials that were missing due to technical problems were excluded (this accounted for 0.12% of all trials across all participants combined). For the analyses of latency, trials that were incorrect during the pre-intervention test and/or its corresponding post-intervention were also excluded. Considering the entire cohort for this study, this process eliminated 21.1% of all trials. Once again, the exclusion of so many trials is potentially concerning, however unsurprising given these were impaired readers. Visual inspection of Q-Q plots[15] revealed substantial skew and kurtosis in the reading latencies, so we applied a base 10 log transformation to these scores. Visual inspection of the Q-Q plots[16] for the transformed latencies revealed approximately normally distributed after transformation.

To assess whether there were any significant differences according to word type or participant that may have affected subsequent improvement analysis, we performed logistic regression (*glmer*) with all data points combined. Word type, participant and the interaction between these variables were all entered as fixed effects. Word identity was entered as a random effect. There was no significant interaction between word type and participant, so

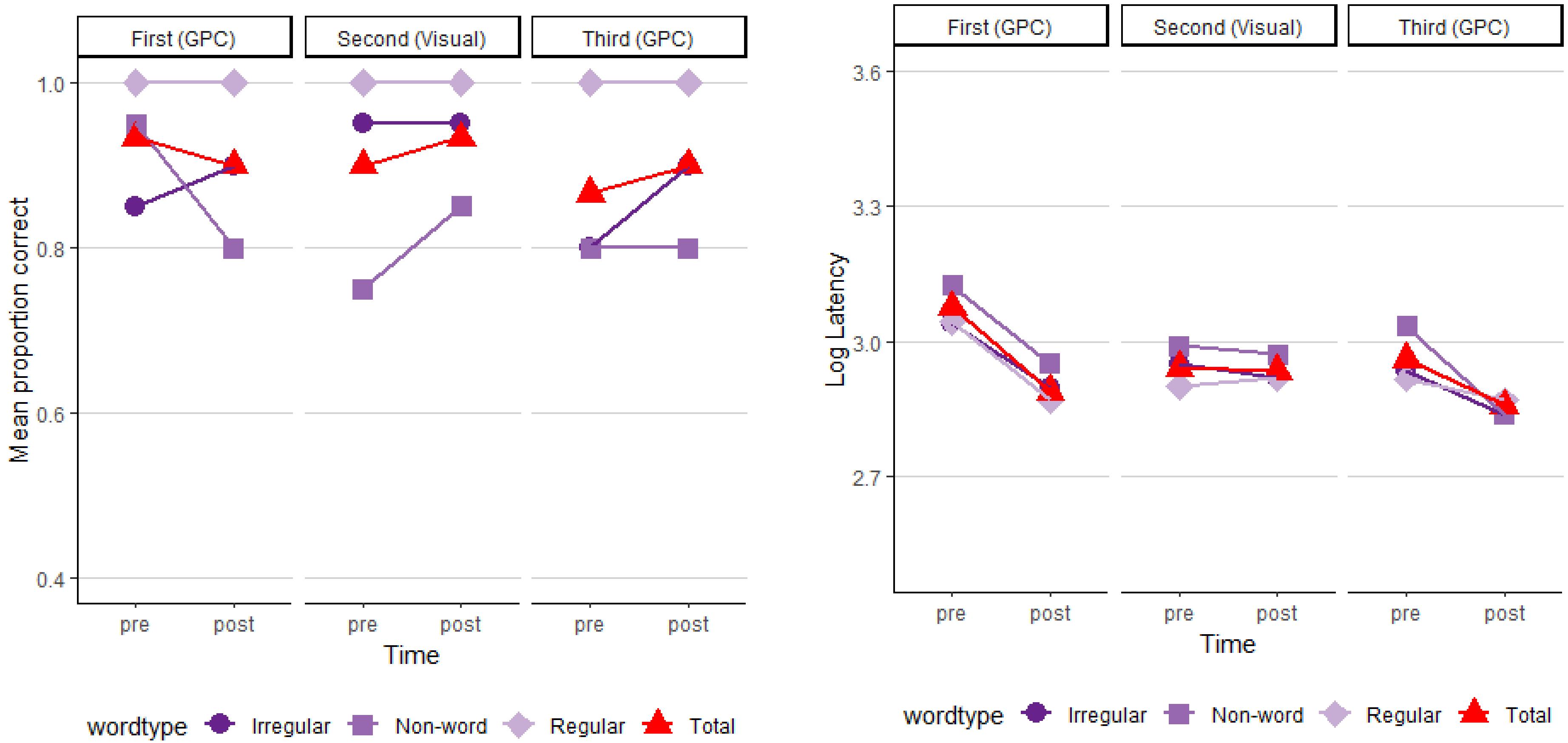
[15] Not included in thesis
[16] Not included in thesis

this was removed, and the analysis was rerun. The final model revealed main effects of word type ($F = 10.53$, $p < .001$) and participant ($F = 3.44$, $p < .001$). Specifically, nonword reading accuracy was significantly poorer than either irregular ($t = 6.57$, $p < .001$) or regular ($t = 4.59$, $p < .001$) word reading accuracy, and irregular word reading accuracy was significantly poorer than regular word reading accuracy ($t = 9.06$, $p < .001$). Thus, it is important to bear in mind that there were considerably less data points for nonwords, and irregular words. For the main effect of participant, phonological case HE was significantly better than a number of participants (MI: $t = 2.97$, $p = .035$; OK: $t = 3.63$, $p = .004$; SH: $t = 3.38$, $p < .010$). Overall, these results suggest that some words and/or participants might have less power within their analyses of reading latency than others.

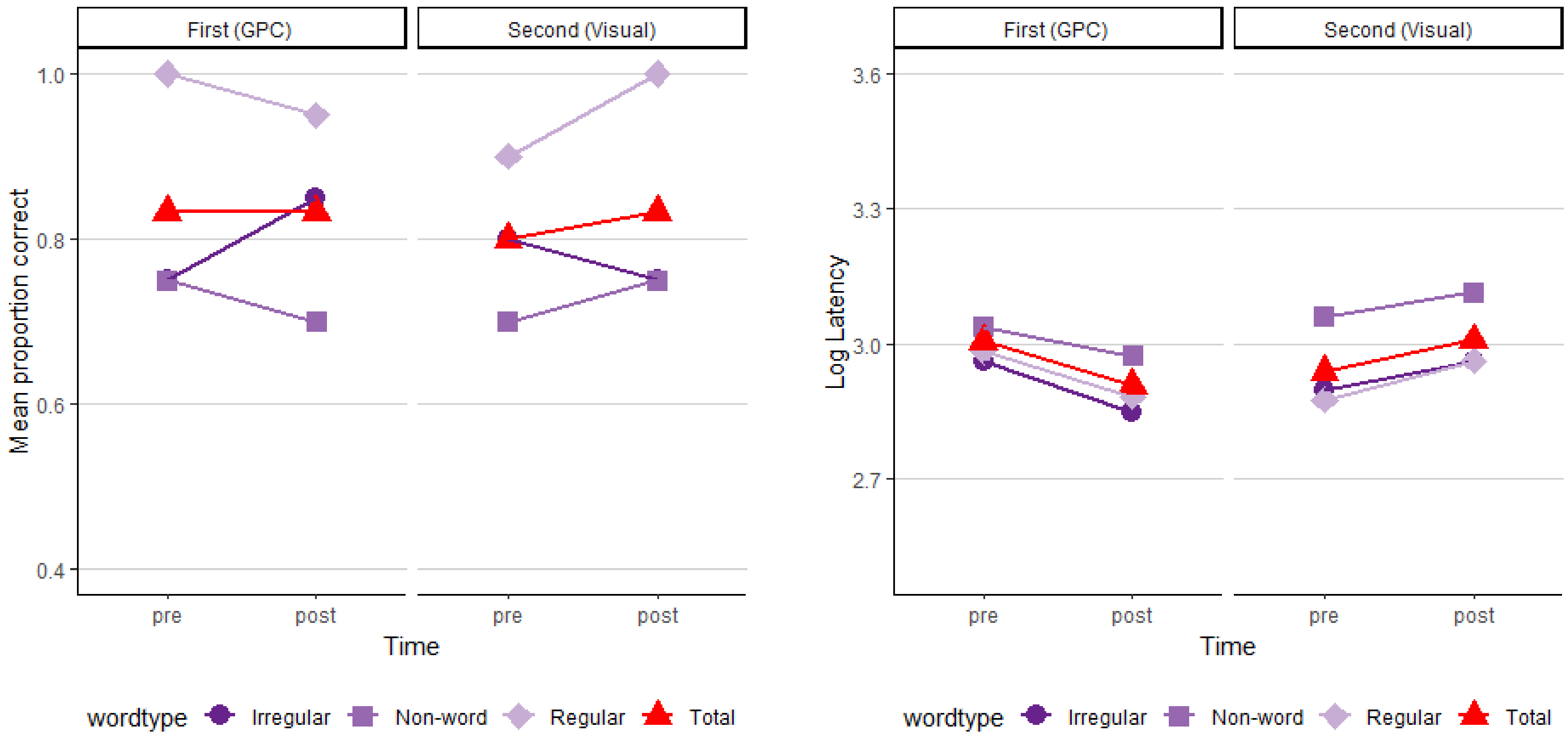
**Primary outcome measure: Phonological cases.** Figure 4.7 presents accuracy and latency changes for each of the phonological cases across each intervention block. The figure shows data for the regular, irregular, and nonword lists separately, as well as an overall accuracy score (total). We start by summarising the trends evident in the figures; in the following section we provide the statistical analyses of each individual's data.

Figure 4.7 Accuracy and latency changes for each of the phonological cases – HE (A), MI (B), OK (C), and PI (D) – according to word type (regular, irregular, nonword) and Intervention block (first (GPC), second (visual) and, where applicable, third (GPC)). Total score is collapsed across word type.

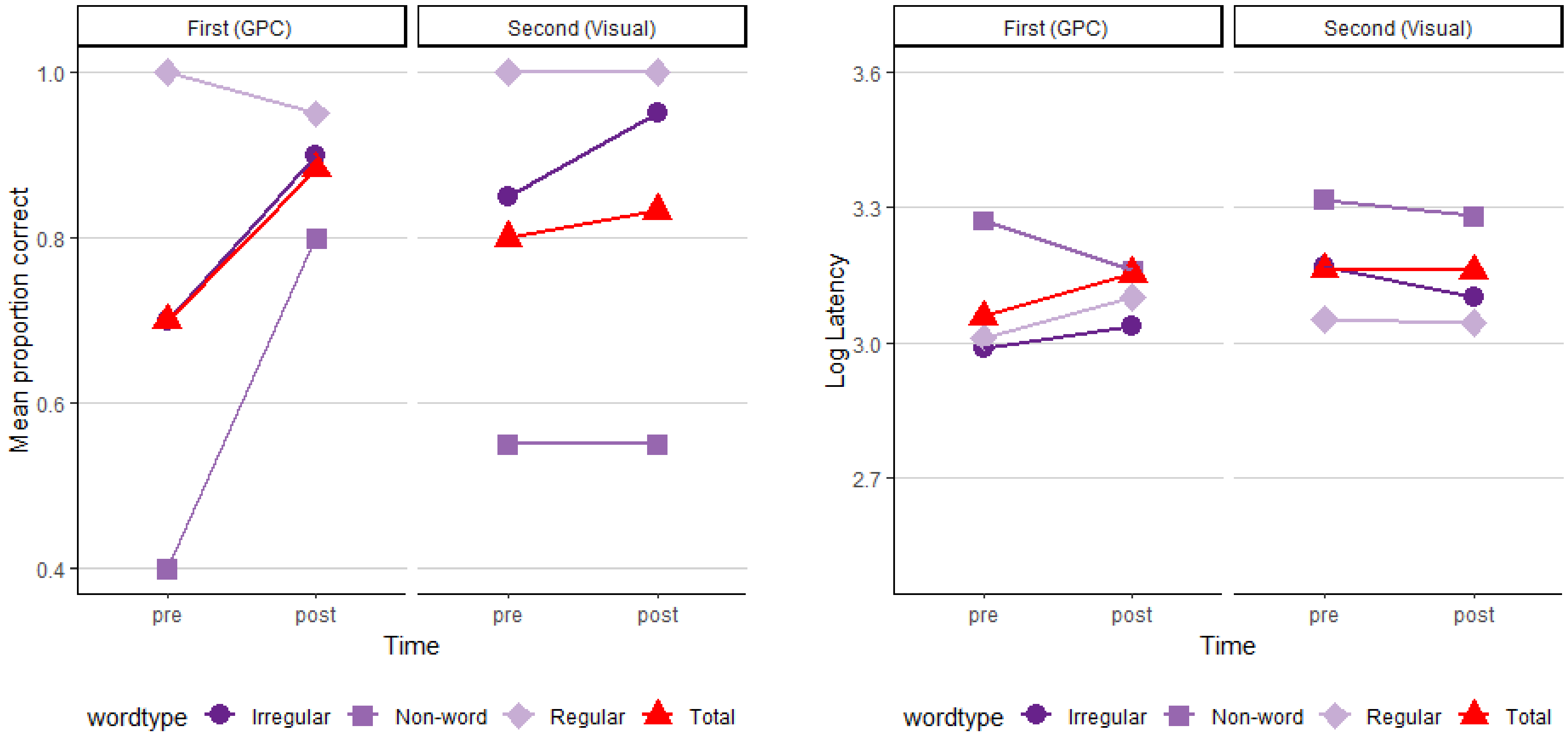
A)



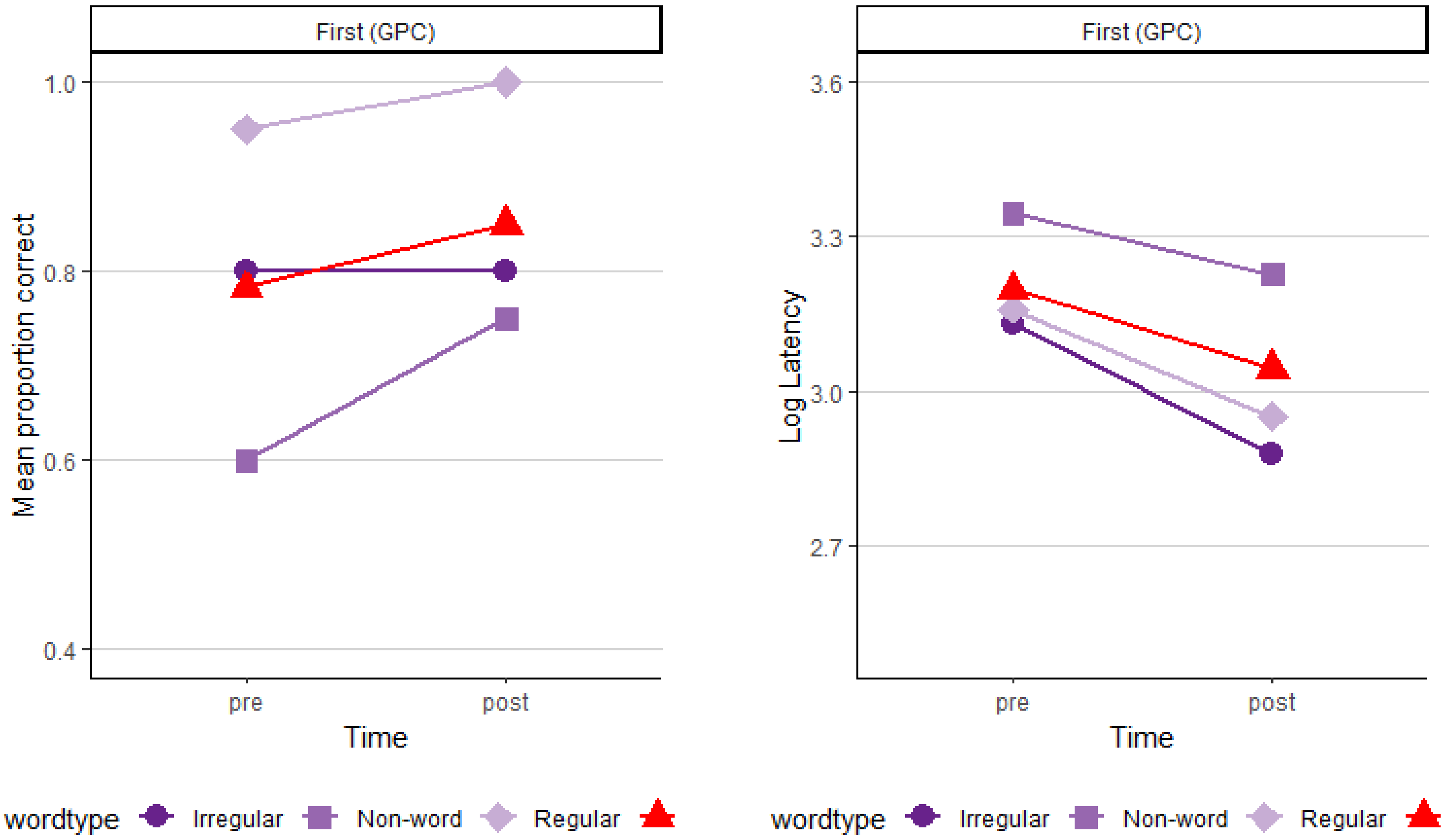
B)


First (GPC)
Second (Visual)
Mean proportion correct
1.0
0.8
0.6
0.4
pre
post
Time
wordtype
Irregular
Non-word
Regular
Total
Log Latency
3.6
3.3
3.0
2.7

C)


First (GPC)
Second (Visual)
Mean proportion correct
1.0
0.8
0.6
0.4
pre
post
Time
wordtype
Irregular
Non-word
Regular
Total
First (GPC)
Second (Visual)
Log Latency
3.6
3.3
3.0
2.7
pre
post
Time
wordtype
Irregular
Non-word
Regular
Total

D)


First (GPC)
Mean proportion correct
1.0
0.8
0.6
0.4
pre
post
Time
wordtype
Irregular
Non-word
Regular
First (GPC)
Log Latency
3.6
3.3
3.0
2.7
pre
post
Time
wordtype
Irregular
Non-word
Regular

The first individual, HE (Figure 4.7 Panel A), completed all three intervention blocks. As is evident in the figure, and contrary to our predictions, there was little effect of intervention on reading accuracy. Regular word accuracy was consistently at ceiling in all pre- and post- intervention tests. Irregular word accuracy shows a trend towards improvement in the two GPC training blocks, but no change for the visual training block, where performance was already near ceiling. Contrary to our specific predictions, intervention type did not appear to have any consistent effect on accuracy change for nonwords considered alone. Although we did not have specific predictions as to how reading latency might change for phonological participants, for HE there is an overall trend towards shorter latencies irrespective of word type (total) following the GPC training blocks, which suggests this intervention might have been particularly beneficial for this individual.

The next phonological case, MI, only completed one target GPC intervention block, and one non-target visual intervention block. They did not complete the third target GPC intervention block. As Panel B of Figure 4.7 shows, there was some effect of intervention on reading accuracy, however again, the patterns do not align in any consistent way with our prediction for phonological cases. It is interesting to note that some trends in this data were similar to HE: for example, after the first (GPC) training block, nonword accuracy appeared to decrease, while irregular word accuracy increased, and in the second (visual) training block, nonword accuracy increased. With regard to reading latencies, in the GPC block, there appears to be an overall trend towards improvement across all word types, while there is a slight trend towards increased latencies in the visual block.

The next phonological case, OK, again only completed one target GPC intervention block, and one non-target visual intervention block. As Panel C of Figure 4.7 shows, there were some marked trends with respect to the effect of intervention on reading accuracy. Although regular word reading accuracy was close to ceiling at all timepoints, irregular word accuracy showed a marked trend towards improvement in both training blocks. For nonwords, there was a dramatic improvement across pre-post intervention testing as a result of the GPC intervention, but no change as a result of the visual training. This pattern is consistent with our predictions for phonological cases. Regarding reading latencies, there is a less discernible effect of intervention than with the other two phonological cases, HE and MI. In the GPC block, there appears to be a small trend towards improvement in nonword reading latencies, but not in the other two word types. In the visual block, there is little improvement on any of the word lists.

The remaining phonological case, PI, only completed one intervention block, the first target GPC intervention block. Non-target visual and the second target GPC blocks were not completed. As Panel D of Figure 4.7 shows, there was some evidence of a trend towards improved reading accuracy following the intervention, particularly on nonwords, and to a lesser extent, on regular words. Irregular word accuracy on the other hand showed no change. With regard to reading latencies, there were the same trends towards general improvement across all word types as was observed with participants HE and MI in the GPC training blocks.

To investigate whether any of the trends described above were statistically significant, we analysed each individual's data. We first calculated whether accuracy changes were statistically significant using a logistic regression (*glmer*). In these models, the dependent variable was accuracy and the fixed effects were intervention (GPC or visual), testing time (pre- vs post- intervention reading performance), and the two-way interactions involving those factors (n. B. as participant PI only completed one block, only testing time was entered as a fixed effect for this individual). We did not enter word type as a main *or* random effect in the individual analyses because there was not enough power for the model(s) to run. However, word identity was entered as a random effect. For these analyses, z-values obtained were the Wald-test statistics.

Our key predictions rest on the interaction between intervention type (GPC or visual) and testing time (pre- or post- intervention). A significant interaction between these two variables is required to demonstrate differential improvement in accuracy according to intervention type. Only results relating to this interaction will be presented here. Appendix AU gives the full models for both accuracy and subsequent latency measures, including all main effects and any other nonsignificant interactions.

With respect to accuracy, only one participant, OK, showed a significant interaction between intervention and testing time ($z$= 2.90, $p$ = .004), with larger accuracy improvements following GPC training, than following visual training. An interaction could not be calculated for participant PI as they only completed one of the three interventions blocks, however, it is worth noting that PI showed a main effect of testing time ($z$= 5.51, $p$ < .001), such that their reading accuracy improved after a block of GPC training. For the remaining phonological participants, HE and MI, the key interaction failed to reach significance.

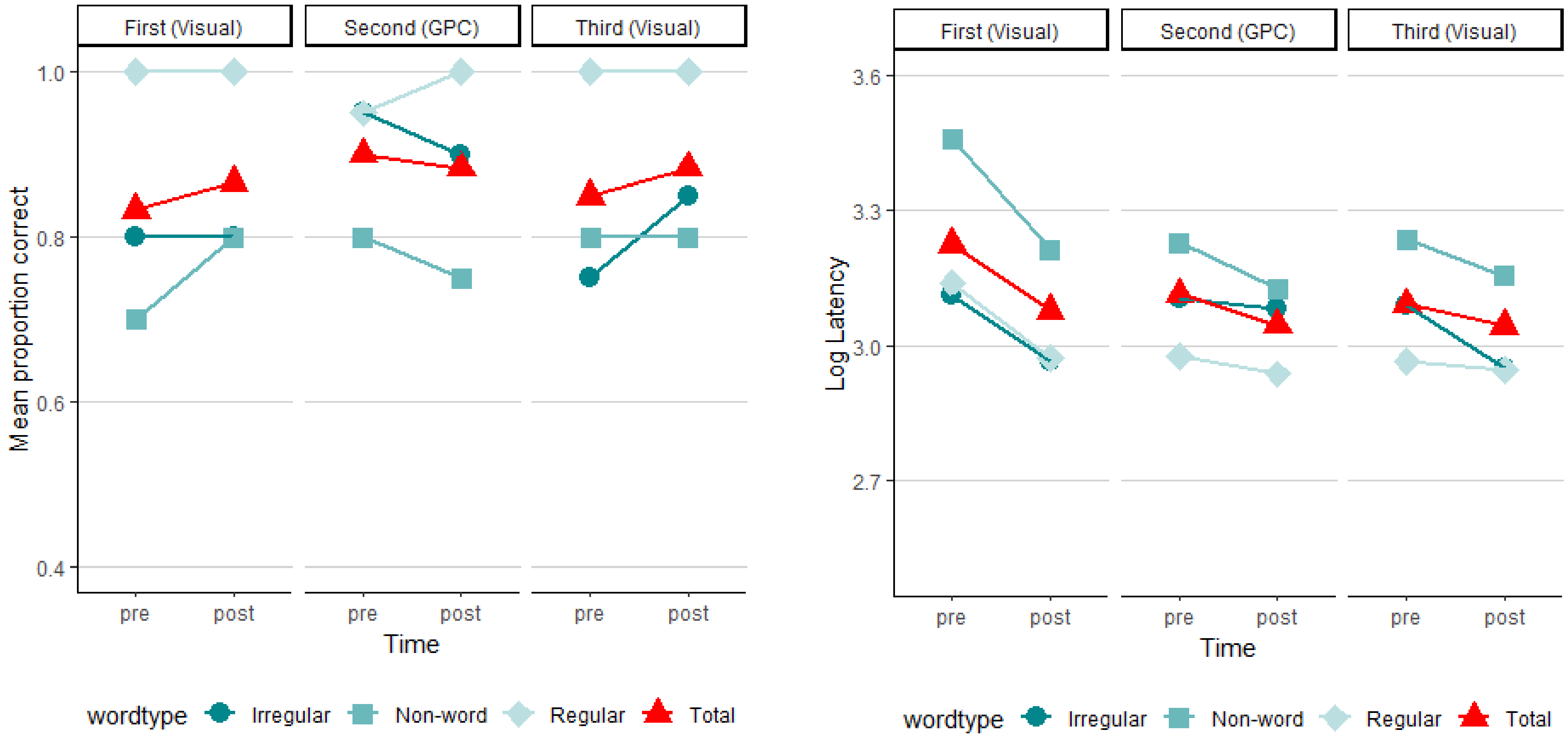
Turning now to latency, we analysed these data in the same way as described for accuracy, except that, since these data are continuous, we applied a linear mixed-effects model (*lmer*). In all models, the dependent variable was log latency and the fixed effects were again intervention, testing time, and the two-way interaction involving those factors (once again, for participant PI only testing time was entered). Again, the random effect was word identity. F-values were obtained from these analyses were based on the Type III hypotheses tests for the fixed effects. Degrees of freedom were estimated via Satterwaite's approximation (rounded to the nearest integer). Below we discuss any significant interactions (see Appendix AU for the full model).

Two of the three phonological cases for whom data was available, HE and MI, showed a significant interaction between intervention and testing time (HE: $F$ (1, 153)= 20.65, $p < .001$; MI: ($F$ (1, 91)= 18.56, $p < .001$; OK*: ns*). In both cases, reading latency improvements were larger after GPC training than after visual training. Once again, interactions were not calculable for participant PI, however a main effect of testing time was found ($F$ (1, 46)= 35.00, $p < .001$) such that their reading latencies improved after the GPC training.

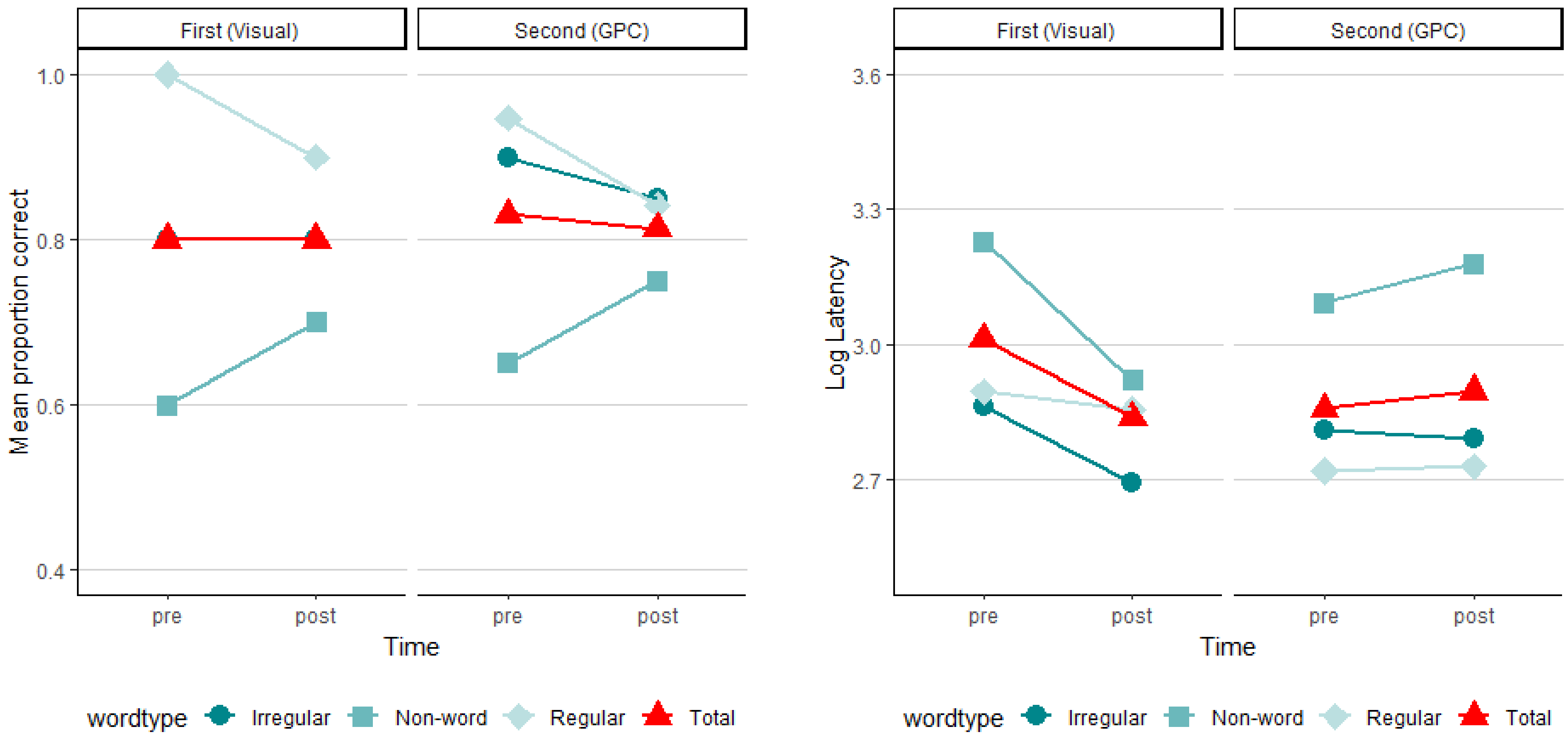
**Primary Outcome Measures: whole-word cases**. Figure 4.8 presents accuracy and latency changes across intervention block for each of the whole-word cases. The general trends in these figures are noted below and statistical analyses are discussed subsequent to that.

Figure 4.8 Accuracy and latency changes for each of the whole-word cases – FR (A) and SH (B) –according to word type (regular, irregular, nonword) and Intervention block (first (visual), second (GPC) and, where applicable, third (visual)). Total score is collapsed across word type.

A)



B)


First (Visual)
Second (GPC)
Mean proportion correct
1.0
0.8
0.6
0.4
pre
post
Time
wordtype
Irregular
Non-word
Regular
Total
First (Visual)
Second (GPC)
Log Latency
3.6
3.3
3.0
2.7
pre
post
Time
wordtype
Irregular
Non-word
Regular
Total

Participant FR (Panel A of Figure 4.8), completed all three intervention blocks. As the figure shows, regular word accuracy was close to ceiling across all timepoints. For nonwords, there was a trend towards improved accuracy in the first visual intervention block, but no evidence of improvement following the GPC block or the second visual block. For irregular words, there was a trend towards improved accuracy following the last, but not the first visual intervention block, and a trend following the opposite direction in the GPC intervention block. These trends are partially consistent with the predictions. With respect to reading latencies, FR shows some general improvement across all blocks and word types. Improvement appears more marked following the visual intervention blocks (most particularly the first visual block), than it does following the GPC intervention block. These improvements appeared to be evident across all three word-types (total), although may be slightly more prominent for nonwords.

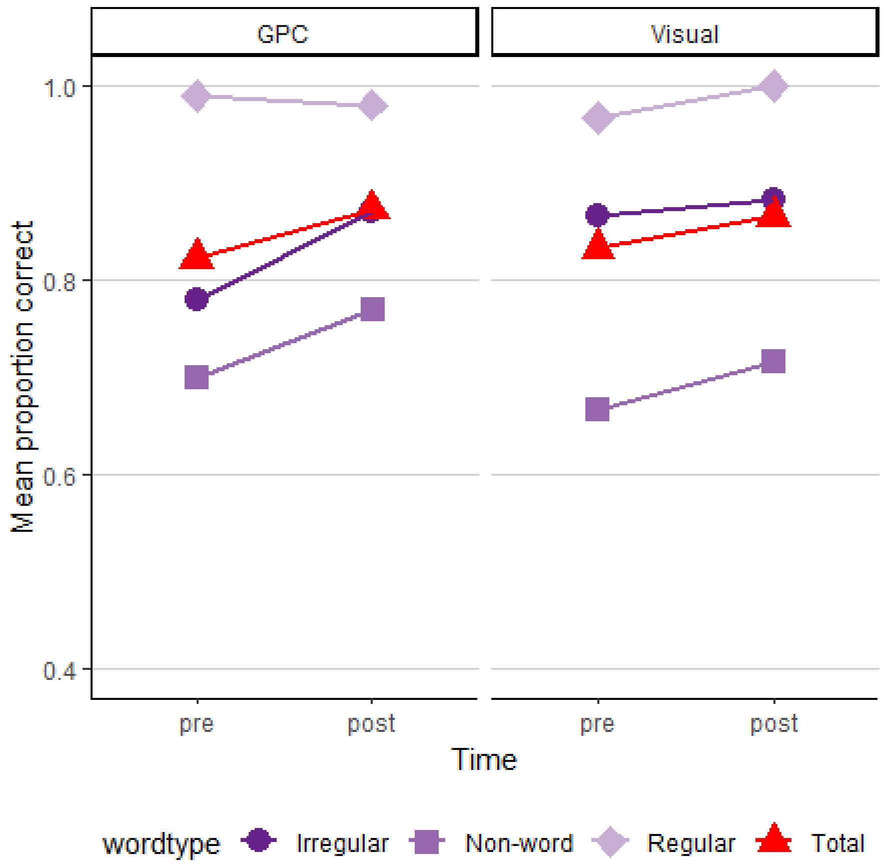
Participant SH, the remaining whole-word case, completed one visual intervention block, and one GPC intervention block. They did not complete the third visual intervention block. As shown in Panel B of Figure 4.8, there were some interesting trends in SH's data. There was no evidence of improvement in regular word accuracy across either of the intervention blocks (in fact, there was a slight trend in the opposite direction). For nonword accuracy, there were patterns of improvement following both the visual and the GPC training block. However, contrary to the hypothesis, irregular word accuracy did not change following the visual training, and there was a trend towards a small decrease in accuracy across following the GPC training. In SH's reading latencies, improvement appears more marked following the visual intervention blocks (most particularly the first visual block), than it does following the GPC intervention block (similar to FR). These improvements appeared to be evident across all three word-types (total), although may be slightly more prominent for nonwords.

To investigate whether the patterns we observed for the whole-word cases were statistically significant, we analysed each individual's accuracy and latency data in the same manner as for the phonological cases. Again, we discuss any significant interactions between intervention type (GPC or visual) and testing time (pre- or post- intervention), and Appendix AU gives the full models for both accuracy and latency. In the analyses of accuracy, neither of the whole-word cases exhibited a significant interaction between intervention and testing time. In the analyses of reading latency, both FR and SH showed significant interactions between intervention and testing time (FR: $F$ (1, 145)= 3.92, $p$ = .050; SH: $F$ (1, 86)= 11.14, *p* = .001), such that reading latency improvements were larger after visual training than after GPC training.

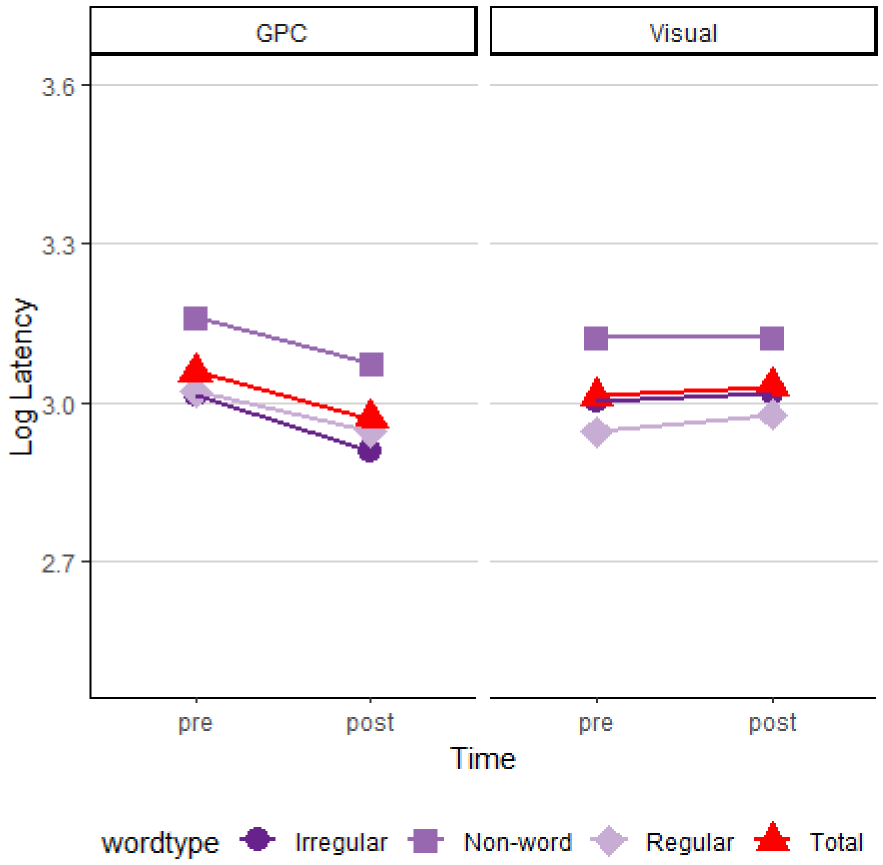
**Group analyses.** To assess whether there were any significant group-level interactions between intervention and testing time we first collapsed across impairment profile, creating two groups: a phonological group (four cases) and a whole-word group (two cases). These "groups" are of course very restricted and cannot be considered wholly representative of those with these particular profiles; nevertheless, it is useful to consider the patterns at this broader level. Starting with the phonological group, Panel A of Figure 4.9. presents accuracy changes, and panel B presents the latency data. The patterns observed at an individual level are clearly evident in this grouped data. That is, the phonological group exhibit a marked trend towards improved accuracy on both nonwords and irregular words following GPC training, but there is less evidence for improvement following the visual training. There also appears to be a marked trend towards improved latency all word types following the GPC training, but there is no evidence of improvement following the visual training.

Figure 4.9 Accuracy (A) and latency (B) changes for the phonological group according to word type (regular, irregular, nonword) and Intervention block (GPC, visual). Total represents the overall mean across all word types.

A)




B)



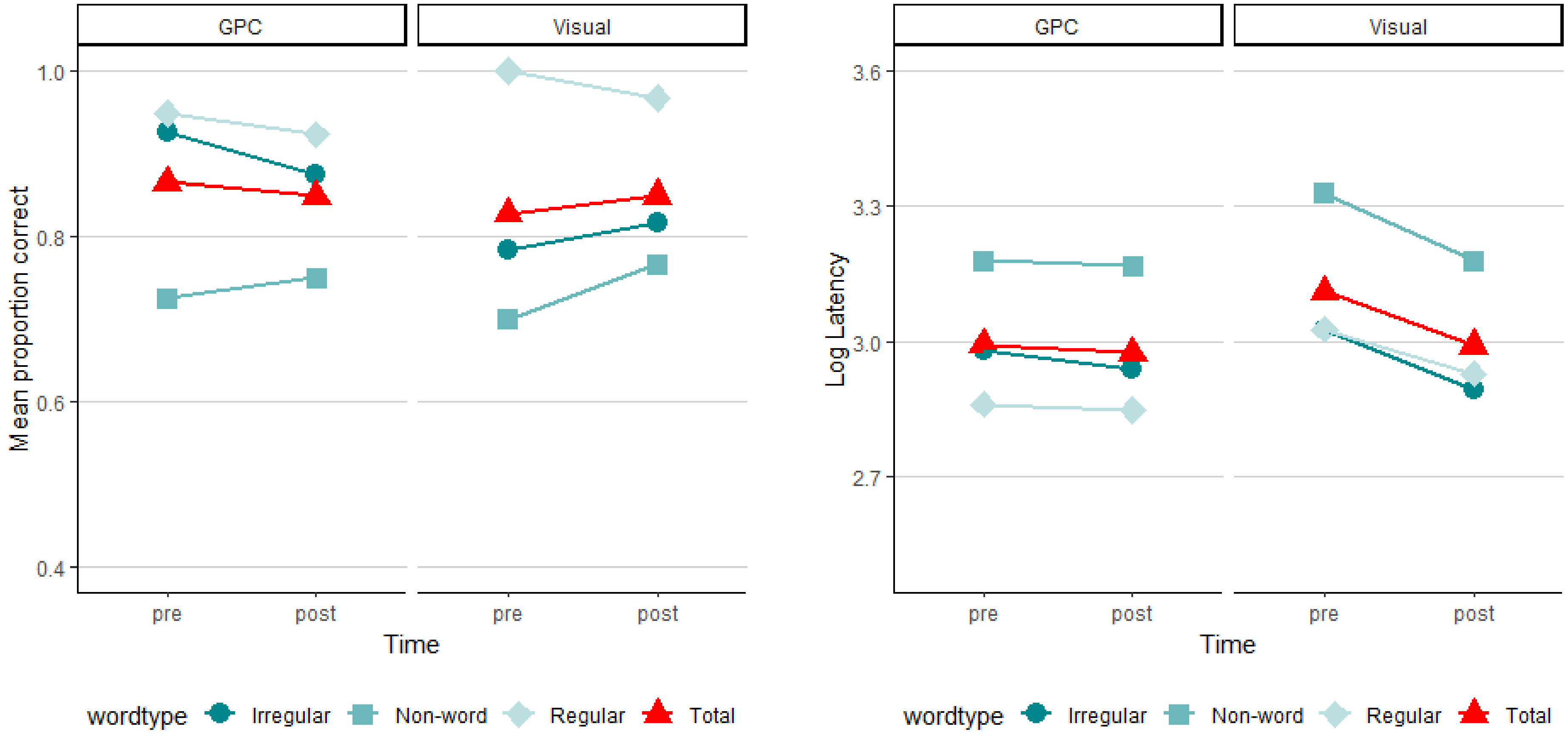
To calculate whether accuracy changes were statistically significant for the phonological group as whole, logistic regression was used. The model was the same as for the individual analyses, except that it also included participant identity as a random effect (the dependent variable was accuracy, the fixed effects were intervention, testing time, and the two-way interactions involving those factors, word identity was entered as a random effect). There were no significant interactions between intervention type and testing time. The full model is present in Appendix AV. We also ran a more complex model which broke the accuracy data down according to word type (regular word, irregular word, or nonword). This model was the same as the previous one, expect that it included word type as a fixed effect, as well as all two-way and three-way interactions. Once again participant and word identity were entered as random effects. Again, there was no significant interactions between any of the variables. The full model is presented in Appendix AV.

To calculate whether latency changes were statistically significant for the phonological group, linear mixed-effects modelling was used. The fixed and random effects were the same as for the accuracy analyses, however, the dependent variable was log latency (the fixed effects were intervention, testing time, and the two-way interactions involving those factors, and the random effects were word identity and participant). We found that the phonological group showed a significant interaction between intervention and testing time ($F$ (1, 611)= 25.03, $p < .001$), such that reading latency improvements were larger after the GPC training – their target intervention – than after visual training. The full model is present in Appendix AV. We also ran the more complex model which included word type and all possible interactions as fixed effects. We found that that there were no significant interactions involving word type. The full model is presented in Appendix AV.

Now we will turn to whole-word cases considered as a group. Panel A of Figure 4.10 presents accuracy changes for the whole-word cases, considered as a group, and panel B presents the latency data. Some of the patterns observed at an individual level are evident in this grouped data, however, for this group, the patterns across word type were less consistent across participant. Nonetheless, the figure suggests slightly greater improvement on nonword and irregular word accuracy following visual training, when compared to GPC training. For latency, there appears to be a marked trend towards improved latency on all word types following the visual training, but there is no evidence of improvement following the GPC training.

Figure 4.10 Accuracy (A) and latency (B) changes for the whole-word group according to word type (regular, irregular, nonword) and Intervention block (GPC, visual). Total represents the overall mean across all word types.

A) B)



To calculate whether accuracy changes were statistically significant for the whole-word group, logistic regression was used. The model was the same as for the phonological group. There were no significant interactions between intervention type and testing time. This mimics the patterns from the individual analyses. The full model is given in Appendix AV. We also ran the more complex model that included word type, and all possible interactions as fixed effects. Again, there were no significant interactions between any of the variables. The full model is given in Appendix AW.

To calculate whether latency changes were statistically significant for the whole-word group, linear mixed-effects modelling was used. The model was the same as for the phonological group. The whole-word group also showed a significant interaction between intervention and testing time ($F$ (1, 297)= 13.84, $p < .001$), such that reading latency improvements were larger after the visual training – their target training – than after GPC training. The full model is given in Appendix AV. Again, we also ran the more complex model looking at word type interactions. We found that that there were no significant interactions involving word type. The full model is given in Appendix AW.

We subsequently completed a combined analysis which included all participants, with impairment profile (phonological versus whole-word) included as a fixed effect. To analyse the accuracy data, we performed logistic regression with accuracy as the dependent variable. The fixed effects were intervention, testing time, impairment profile, and all two-way and three-way interactions. Once again participant and word identity were entered as random effects. We found that there were no significant interactions in this model. The full model is presented in Appendix AX. We then repeated the above analysis, this time we included word type, and all possible interactions involving this variable. Again, there were no significant interactions in this model. The full model is presented in Appendix AY.

To analyse the combined log latency data, we used a linear mixed-effects model, keeping the fixed and random effects identical to the accuracy analysis. In the simple model (i.e. not including word type), the three-way interaction between intervention, time and impairment profile was significant ($F$ (1,1041)= 34.05, $p < .001$). The full model is presented in Appendix AX. For the complex model (i.e. including word type and all interactions involving this variable), we again observed the three-way interaction obtained in the previous model between intervention, testing time and impairment group ($F$ (1,1202)= 31.57, $p <$ .001). Also, perhaps unsurprisingly, there was a significant interaction between word type and impairment group (*F* (1,202)= 13.91, $p < .001$). However, we did not find the predicted interaction between word type, testing time, impairment group and intervention Appendix AY contains the full model.

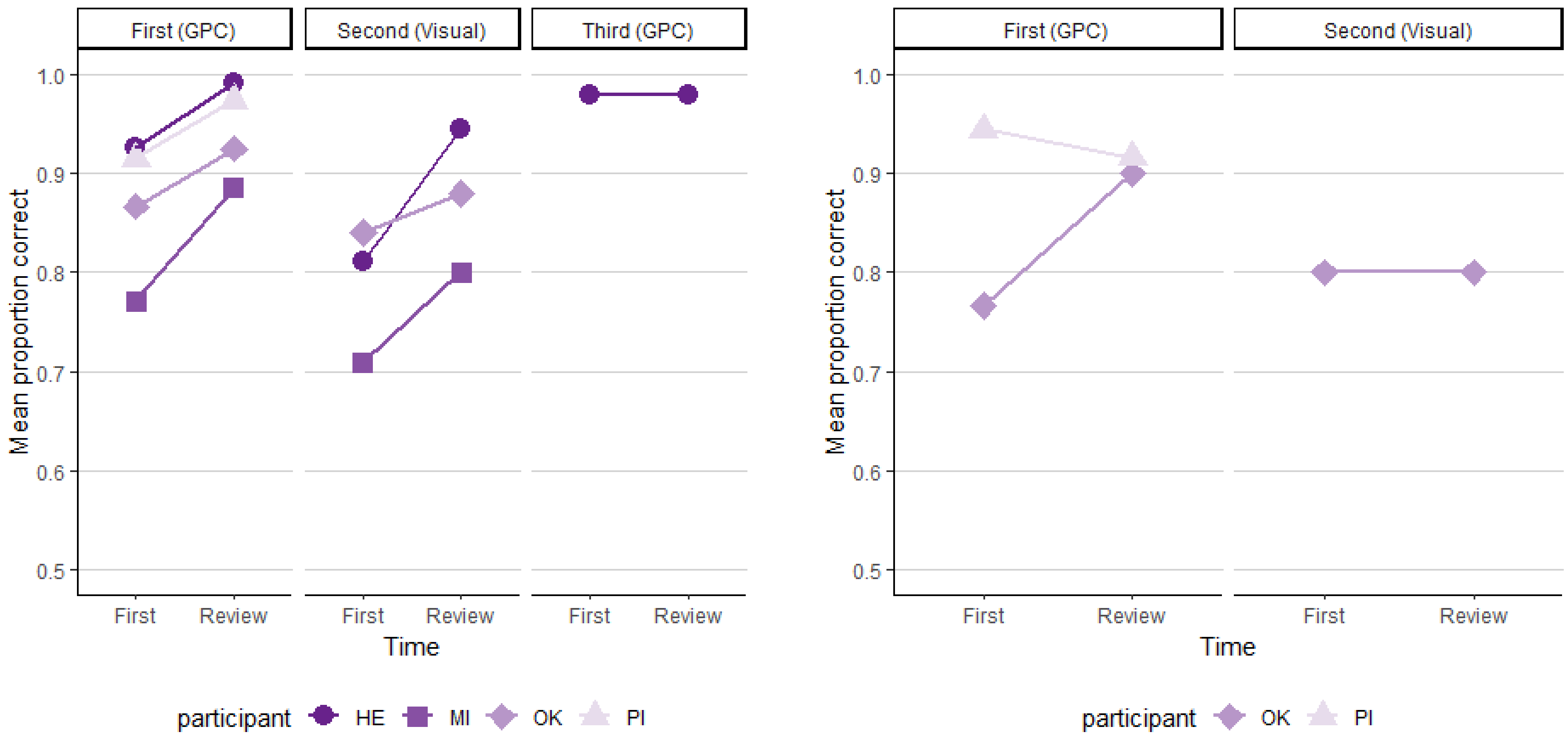
**Summary of Primary Outcome Results.** The results of our analyses of the primary outcome measures – accuracy and reading latencies on the pre- and post- intervention reading tests – provide partial support for our hypotheses. The analyses of reading accuracy yielded few reliable results; however, one individual, phonological case OK, demonstrated a significantly larger improvement in reading accuracy following their target intervention – GPC training – than they did following the non-target visual intervention. There were no other significant intervention-specific.

Reading latency on the other hand, yielded several significant intervention effects. In the individual analyses, two of the three phonological cases for whom we had data from two or more blocks (MI and HE) demonstrated significantly greater latency improvements following the target intervention – GPC training – than they did following the non-target visual intervention. The same outcome was observed in the group-level analyses of phonological cases. These effects were not predicted for the phonological cases, for whom our predictions were limited to the reading accuracy data. Further, both of the two whole-word cases we tested (FR and SH) demonstrated significantly greater improvements following their target intervention – visual training – than they did following the non-target GPC intervention. Again, this outcome was also observed in the group analysis of whole-word cases. Word type did not significantly modulate any of the intervention-related improvements reported above, in either individual or group analyses.

**Secondary Outcome measures.** For the secondary outcome measures, we looked at whether participants' reading accuracy improved on words that contained the letter patterns that were trained during the interventions (GPCs in the GPC intervention and clusters in the visual intervention). Starting with the phonological cases, Panel A of Figure 4.11 shows reading accuracy for trained words at two timepoints: a) the first reading attempt during the relevant training session (henceforth known as first attempt); and b) during the Review test completed in the following session (henceforth known as review attempt). For the 2017 cohort (OK and PI), we were also able to obtain data for a set of words that contained the trained letter patterns but were not themselves explicitly trained (i.e. redundant words). Panel B of Figure 4.11 shows reading accuracy for these words at first and review timepoints.

Figure 4.11 Proportion of words read correct in the first reading attempt compared to the review attempt in the subsequent session for Phonological cases. Panel A shows trained word accuracy changes, Panel B shows redundant word accuracy changes (2017 cohort only).

A) B)



For the first individual, HE (who completed all three intervention blocks) the graph shows that, contrary to our predictions, words trained in the visual intervention appeared to exhibit greater improvements than words trained in the GPC intervention. The next case, MI (who only completed one GPC intervention block, and one visual intervention block), shows some small effect of intervention on trained word reading accuracy, however this trend does not appear to be specific to either block. For OK, (who again only completed one target block and one non-target intervention block), there was little evidence of improvement in either intervention. The final phonological case, PI (who only completed one target intervention block) shows little improvement in GPC training.

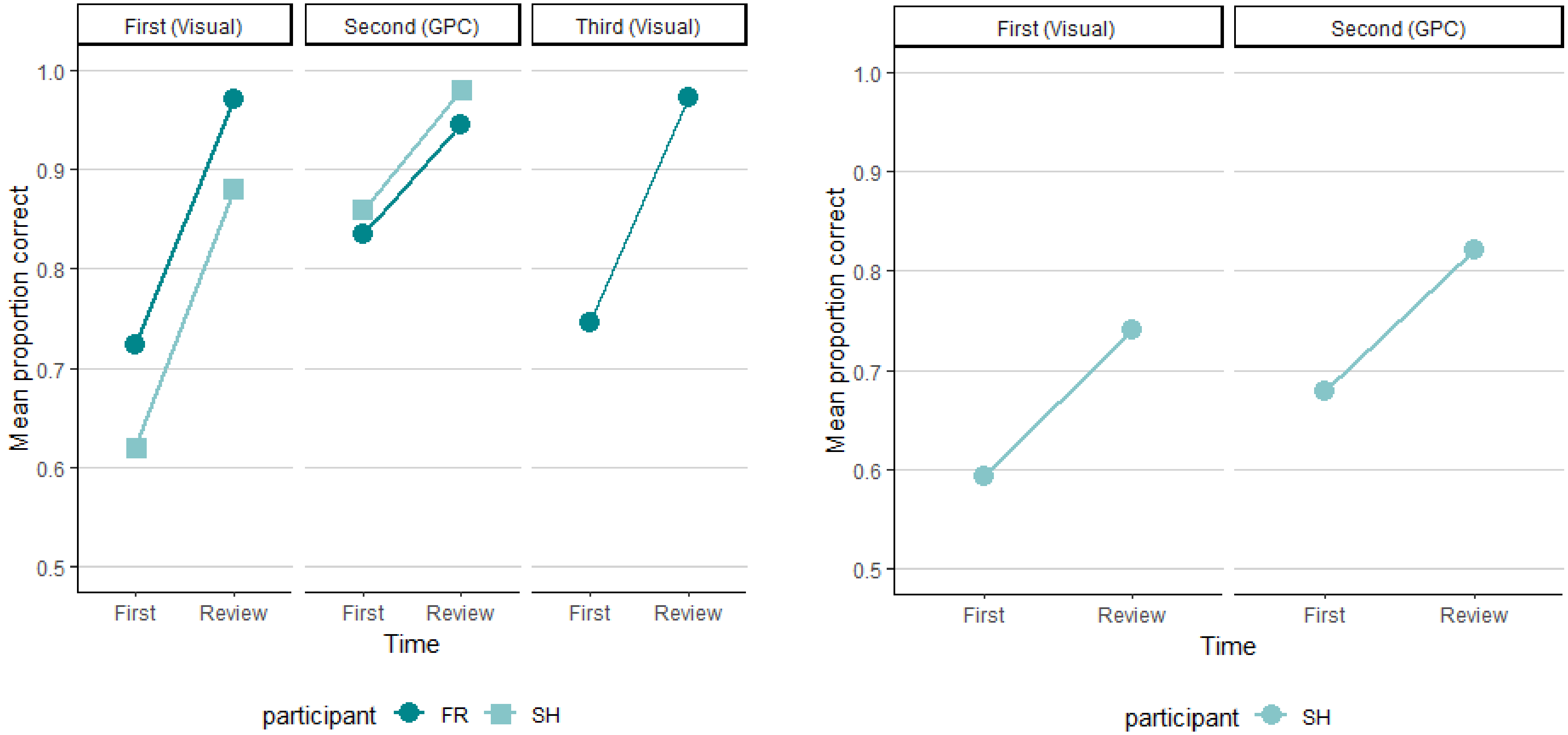
For each phonological case, we analysed these data using logistic regression. In these models, the dependent variable was accuracy and the fixed effects were intervention (GPC vs visual), timepoint (first attempt vs review attempt), and the two-way interaction between those factors (n.B. as participant PI only completed one block, only timepoint was entered). Word identity was entered as a random effect. Appendix AZ gives the full models, including all main effects and any other nonsignificant interactions – here we present significant interactions. Only one participant, MI, showed a significant interaction between intervention and timepoint ($z$= 2.17, $p$ = .030), such that larger accuracy improvements were found in the trained words from GPC training, compared to trained words from the visual training. An interaction could not be calculated for participant PI as they only completed one intervention block, however, it is worth noting that they showed a main effect of timepoint ($z$= 3.11, $p$ = .002), such that their reading accuracy improved with GPC training. For the remaining phonological participants, HE and OK, the key interaction failed to reach significance.

For participant OK, there appeared to be a small improvement in accuracy on redundant words following GPC training, but no change following the visual training. For PI, there was a small trend of decline in redundant word reading accuracy following GPC training (PI did not complete the visual training). We analysed these data in the same way as described for trained words. Case OK, who was the only phonological case for whom we had data available for testing an interaction between intervention and timepoint, showed no significant interaction. Appendix AZ gives the full model.

**Secondary outcome measure: Whole-word cases**. Turning now to the whole-word cases, Panel A of Figure 4.12 shows reading accuracy for trained words containing the target letter patterns and (where applicable) Panel B shows reading accuracy for redundant words.

Figure 4.12 Proportion of trained words read correct in the first reading attempt compared to the review attempt in the subsequent session for whole-word cases. Panel A shows trained word accuracy changes, Panel B redundant word accuracy changes (2017 cohort only).

A) B)



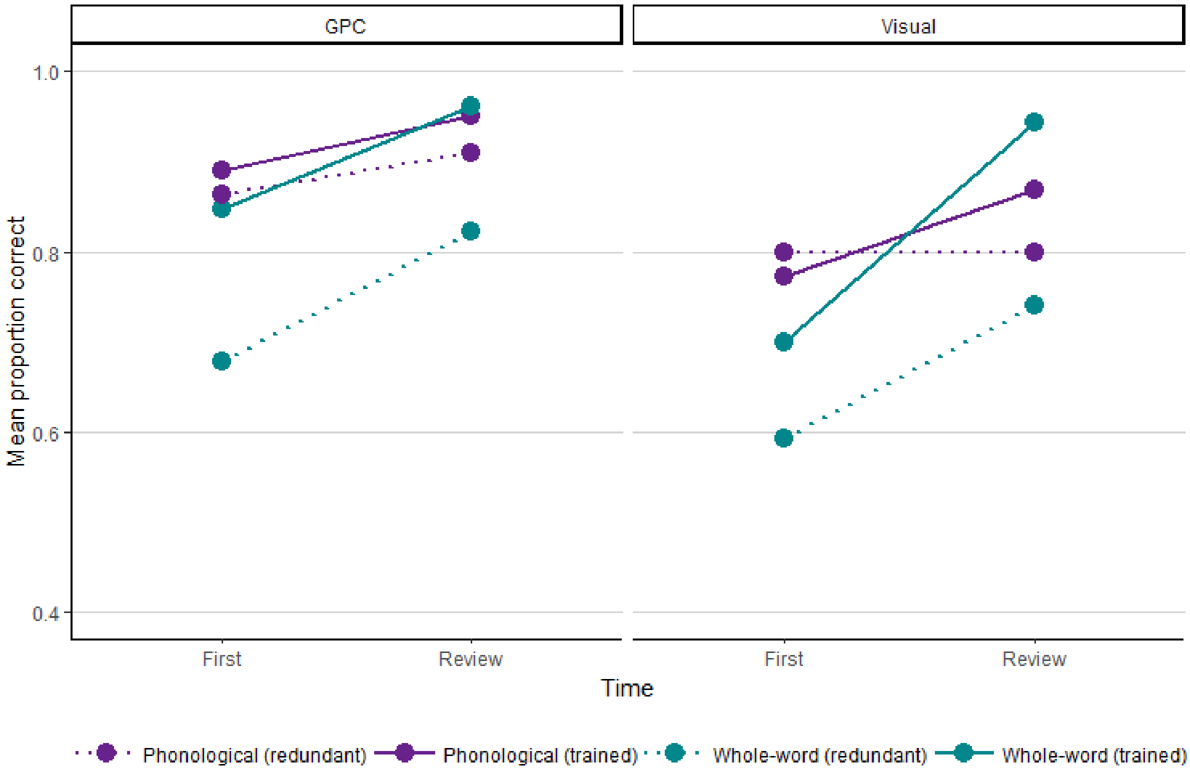
For the first individual, FR (who completed all three intervention blocks), the graph shows a trend relatively consistent with our predictions, with a sizeable improvement for explicitly trained words following both visual training blocks and a moderate improvement following the GPC block. The second case, SH (who only completed one visual intervention block, and one GPC intervention block) shows a sizable pattern intervention-specific effect on reading accuracy, particularly following the visual training block, although there is also reasonable improvement following the GPC training block

We analysed the data for each whole-word case using the same logistic regression as for the phonological cases. Only one participant (FR) exhibited a significant interaction between intervention and timepoint ($z$= 1.98, $p$ = .048), with larger improvements as a function of the visual intervention, compared to GPC training. For SH, no significant interaction was found. Appendix AZ gives the full model.

For the participant who completed redundant word accuracy measures (SH), there was a trend towards improvement following both the visual and GPC training blocks. We again analysed the data as we did for the phonological cases and found no significant interaction. Appendix AZ contains the full model.

**Secondary outcome measure: Group analyses.** Figure 4.13 presents trained word accuracy changes for all phonological cases (for whom the data was available) considered as a group, and the same for the whole-word cases as another group. For the phonological group there appears to be little improvement in trained or redundant word accuracy following either GPC or visual training; it seems that the individual improvement pattern of participant MI is attenuated when considered as part of grouped data. The whole-word group on the other hand exhibit a relatively large trend of improvement in trained and redundant word reading accuracy following both GPC training and visual training. It is perhaps unsurprising that this group trend is more akin to the trends found in the individual data considering that there were only two participants within the latter group, both of whom showed relatively prominent and similar trends of improvement.

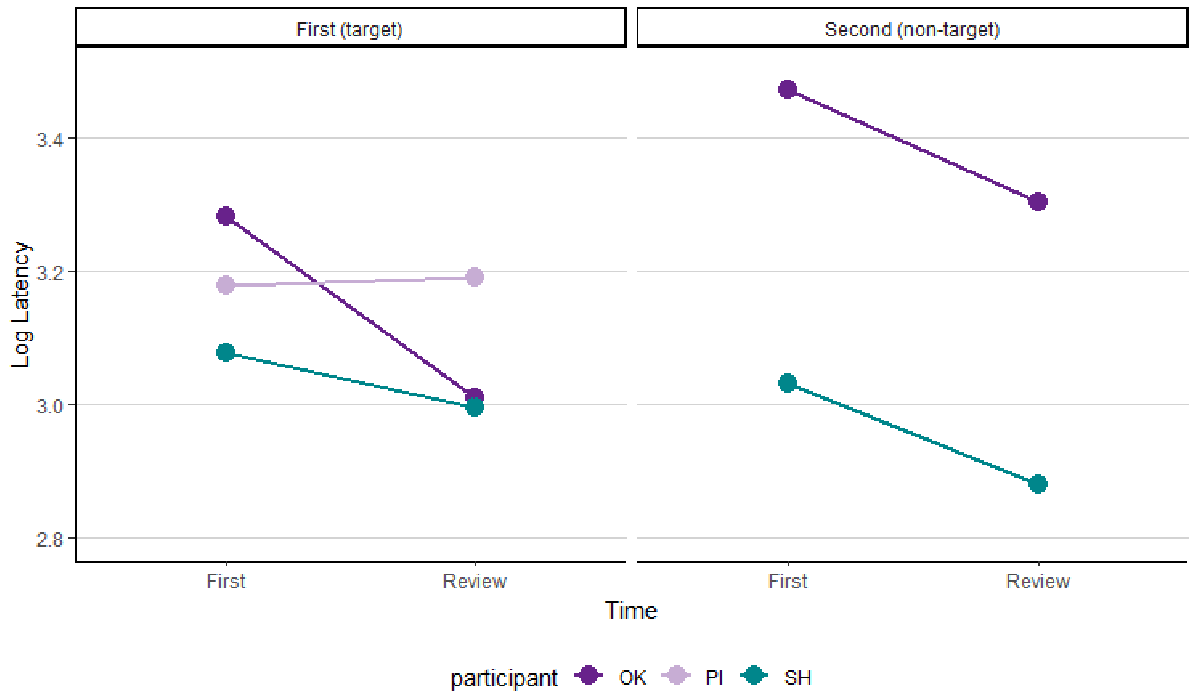
Figure 4.13 Accuracy changes on trained (solid lines) and redundant (dotted lines) words for the phonological (purple) and whole-word (green) presented groups according to Intervention block (GPC vs visual).




We analysed the data for each group using logistic regression. We used the same model as with the individual case analyses, except participant was added as a random effect. For trained words, there were no significant interactions between intervention type and timepoint for either group. Appendix BA gives the full model. This mimics some, but not all the patterns from the individual analyses. For redundant words, as there were only two phonological cases and one whole-word case, we decided not to complete this analysis. We subsequently completed a combined logistic regression analysis, which included all participants, with impairment profile (phonological vs whole-word), and all interactions as fixed effects. For trained words we found that there were no significant interactions in this model. For redundant words we again found that there were no significant interactions in this model. Appendix BB gives the full models.

**Secondary outcome measure: redundant word latency (2017 cohort).** For the analysis of latency in redundant words, incorrectly read trials from the first attempt and/or review attempt were excluded. Considering the entire cohort for which data was available, this process eliminated 19.5% of all trials. As in the primary outcome measures, reading latencies for all individuals were substantially positively skewed, so we applied a base log transformation to these scores. Figure 4.14 shows the resulting mean latencies for the two phonological cases (OK and PI), and the one whole-word case (SH) for whom this data was collected. For phonological case OK, reading latencies showed a trend towards improvement across both GPC and visual training blocks, with improvement appearing to be more marked following the GPC training block than the visual training block. For phonological case PI, there appears to be a very slight increase in reading latency (i.e. slower) in the GPC block (PI did not complete the visual training block). For whole-word case SH there appears to be some improvement following both the GPC and visual training blocks; however, this trend appears to be more marked following the GPC training block than the visual block which is counter to our predictions.

Figure 4.14 Mean log latencies for redundant words in the first and review exposures for the 2017 cohort according to first (target) and second (non-target) intervention blocks.




We analysed the data for each case using a linear mixed-effects model. The dependent variable was log latency, and the fixed effects were timepoint, intervention and the interaction between these factors (once again, for participant PI only timepoint was entered). Word identity was entered as a random effect. In these analyses, no significant interactions were found (for PI no main effect was found). Appendix BB contains the full models.

As in previous analyses, we also investigated group patterns. Again, as there were only two phonological cases and one whole-word case, we decided to complete a single combined analysis, which included all participants, with intervention, testing time, impairment profile (phonological vs whole-word) and all interactions included as fixed effects, and participant as a random effect. The only significant interaction was between testing time and impairment profile ($F$ (1,136)= 2.07, $p$ = .040), however the key three-way interaction between intervention, testing time and impairment profile was not significant. This mimics the patterns from the individual analyses. Appendix BB gives the full models for all analyses on redundant word reading latencies.

**Summary of Secondary Outcome results.** The results from the secondary outcome measures – accuracy and latency changes on trained words and redundant words containing the target letter pattern – provide partial support for our hypotheses. One phonological case, MI, demonstrated a significantly larger improvement in trained word reading accuracy following their target intervention – GPC training – than they did following the visual training. One whole-word case, FR, also demonstrated a significantly larger improvement in trained word reading accuracy following their target intervention – visual training – than they did following GPC training. However, there were no other significant intervention-specific accuracy or latency improvements in any of the other individual analyses, or in the group analyses, for either the trained or the redundant words.

## Discussion

We shall briefly recap the outcomes of the primary measures, reading accuracy and latency on untrained words, starting with whole-word cases. For these cases it was hypothesised that the visual intervention would be the most beneficial intervention. We predicted significantly larger improvements to reading accuracy in general – and irregular word accuracy more specifically – following visual training, than following GPC training. Our results did not support this hypothesis. Neither individual nor group analyses showed any advantage of visual training over GPC training for general reading accuracy, or irregular word accuracy more specifically. We also predicted significantly larger improvements in overall reading latencies for whole-word cases following the visual training, than following GPC training. Our results were consistent with this hypothesis, showing that visual training provided greater improvements in reading speed than GPC training. This finding was also replicated at the group level.

Next for whole-word cases, the secondary outcome measures. It was hypothesised that accuracy on trained words would mirror the patterns from the primary measures. Specifically, we predicted that there would be significantly larger improvements in accuracy following the visual training, than following the GPC training. For trained words, our results were mixed, providing some support for this hypothesis. Participant FR, but not SH, showed a significant advantage of visual training; but there was no overall effect in the group-level analyses. Data for redundant words (words not included in the training session but containing trained letter patterns) were only available for participant SH. Again, the hypotheses mirrored the primary outcomes measures, specifically we predicted that there would be significantly larger improvements in accuracy and latency following the visual training than following the GPC training. Our findings did not support this prediction for redundant words.

Turning now to the outcomes of the primary measures for the phonological cases, it was hypothesised that the GPC intervention would be the most beneficial training. We predicted significantly larger improvements to reading accuracy in general – and nonword accuracy more specifically – following the target GPC training, than following the visual training. Overall our results did not provide strong support for this hypothesis; only one participant (OK) showed significantly larger accuracy improvements across words of all types following the GPC training, than following the visual training. However, for nonword accuracy specifically, none of the other individual or group, analyses showed any significant advantage of GPC training over visual training. We did not make any explicit predictions for latency improvements, however, we found that two of the three participants for whom enough data was available (HE and MI) showed significantly larger improvements across words of all types following the GPC training than following the visual training. An intervention-specific benefit was also found at the group level.

For the secondary outcome measures with phonological cases, again the results did not provide strong support for our hypotheses. Only participant MI showed a significant advantage of GPC training on trained word accuracy but there was no overall effect in the group-level analyses. For participant OK, who also completed a redundant word assessment, our results did not support the hypothesis: again, there was no advantage of GPC training over visual training for either reading accuracy or latency.

In the current research, we set out to target the underlying cognitive difficulties in developmental dyslexia. Overall, the findings suggest that we were able to target these differential processes. We felt that an intervention tailored for whole-word cases would be most effective if the load on parallel letter processing could be reduced. We reasoned that implicitly training these participants to recognise a common letter combination (e.g. *ation*) as a single visual pattern as opposed to individual letter patterns might encourage them to process words more holistically. Our results suggest we were successful with this aim. We return to this point in the General Discussion.

**Methodological considerations.** One possibility is that the interventions did successfully target the precise difficulties in each impairment profile, but that these gains did not translate into *differential* changes in reading accuracy. Considering words that were explicitly trained during the sessions, there were general accuracy improvements on those words, irrespective of intervention type. The very simple act of explicitly and repeatedly engaging with a word may engender benefits that are quite independent of the specific skills being trained during the intervention. If so, it would not be unexpected to see similar rates of improvement in these words irrespective of intervention type.

Given this potential issue, we might expect redundant word accuracy and latency to be more sensitive measures, however our results do not support this. It is possible that the techniques and skills learned in the training sessions cannot be extended to redundant words. This explanation may be particularly salient for the visual intervention, where the orthographic similarity for some word families may be less salient than the phonological differences. That is, although a portion of the visual letter pattern across a word family will be shared, the pronunciation of the letter pattern may vary from word-to-word. For example, for a cluster like *ation* the orthographic similarity across training words (e.g. *nation*, *stationary*, *creation*) may be just as salient as the phonological similarity, given that the latter is relatively consistent from word-to-word (e.g. /ˈn**eɪʃ(ə)n**/, /ˈst**eɪʃ(ə)n**(ə)ɹi/, /kriːˈ**eɪʃ(ə)n**/ – corresponding phonetic notation of "ation" bolded ). Conversely, for a cluster like *ough* the orthographic similarity across training words (e.g. *cough*, *through*, *dough*) may be more salient than the phonological differences, given that the latter varies from word-to-word extensively (e.g. /k**ɒf**/, /θr**uː**/, /d**əʊ**/ – corresponding phonetic notation of *ough* bolded).

There are a number of other important issues to bear in mind when evaluating these results. First and foremost, the role of order effects in our study. We used an alternative treatment design, so that participants could be compared with themselves and direct comparisons could be made between the two different training programmes. These comparisons prevent the problems that arise when we combine data across multiple, individuals which may differ in crucial, but undetectable ways. Broom and Doctor (1995a, 1995b), for example, tested only one intervention per participant (a GPC intervention for their phonological case, and a word-based intervention for their comparison, whole-word case), and were consequently unable to make strong claims about the superiority of either intervention for either case. Conversely, Law and Cupples (2015) were able to contrast the effectiveness of two different interventions, but since their participants were all whole-word cases, they were unable to make strong claims as to the specificity of the interventions in targeting underlying processes. By having impairment profile as a between-subject variable, and intervention as a within-subject variable, we were able to focus our hypotheses not on *absolute* rates of improvement, but *relative* rates of improvement across the two interventions.

However, using an ABA alternating treatment design also means that up to twice the number of data points were available from intervention A, than from intervention B. If improvement rates are *strongly* dependent on the order of intervention block, then the target intervention, being block ‘A’, may have been inherently favoured using this design. The patterns of data for participants who completed the full ABA design (HE and FR) suggest this favouring was unlikely to have occurred, as improvements observed in the first intervention block were also evident in the third at least to some extent, but of course we cannot know this for individuals who did not complete all three blocks. In the current research, we could not adequately look at or test order effects because of our ABA design. A more ideal treatment design might have been ABBA, counter-balanced across participants so that equal numbers of participants completed ABBA, and its complementary structure, BAAB.

Another crucial issue to consider is the individual variability amongst participants within the same profile group. Given that the number of cases studied here was very small, it is difficult to estimate the degree of variability between cases. However, the differences between our two whole-word cases, FR and SH, appeared to be particularly marked. FR’s profile was pure (severe VA span impairments and marked real-nonword latency difference; no evidence of nonword accuracy or phonemic awareness impairments). Conversely, SH had the most complex profile of any participant, with impairments across tasks indexing both whole-word, *and* phonological processing (severe VA span impairments, marked real-nonword latency difference and orthographic processing impairments; but also, evidence of nonword accuracy and phonemic awareness impairments). While the phonological cases exhibited less variation, there were still some differences amongst individuals, particularly with respect to their performance on the orthographic awareness tasks. Finally, in addition to these sources of cognitive variation, there are likely to be differences in participants' levels of motivation and interest in improving their reading. Indeed, we mentioned some differences with respect to participants HE and PI in Chapter 3. To ensure that these somewhat uncontrollable variations have minimal influence in future studies, at least at the group level of analysis, every effort should be made to maximise the sample size, with particular emphasis on the increased recruitment of whole-word cases.

Finally, another key aspect of the design was that the stimuli used in the pre- and post- intervention tests were the same for all participants, and assigned to block number, rather than intervention type. This means that the between-subject comparisons made on these measures cannot be attributed to differences in the tests themselves. However, it also means that if the stimuli in these lists were not of equal difficulty, there may be greater opportunity for improvement in some blocks than in others.

**Intervention summary.** Overall the results provide some support for a targeted-intervention approach. Importantly, the typical GPC training was not the most suitable approach for all delayed readers, instead impaired readers with non-phonological deficits benefit from a non-GPC training. These results are consistent with previous intervention research that has found larger improvements following target intervention (Jones, 2013; Law & Cupples, 2015; Rowse & Wilshire, 2007). Instead of applying a one-size-fits-all approach, training should first seek to identify the specific deficits, and then work to build techniques that would be particularly useful in overcoming these deficits.

# Chapter 5 : General Discussion

There were two major parts to this research. In the first part, we explored individual variation in reading and reading-related skills in an undifferentiated adolescent sample. Specifically, we examined two key measures: one that indexes phonological or part-based processing during reading (specifically, nonword reading accuracy), and one that indexes whole-word based processing (specifically, the effect of word length on reading latencies). We found that both measures were associated with overall reading proficiency, but not with each other. Further, each measure explained a unique proportion of the variance in overall reading proficiency. We also examined the relationship between these key reading measures and a commonly-used measure of VA span, based on a visual partial report task (Bosse et al., 2007). While VA span scores were associated with overall reading proficiency, contrary to our predictions, they were not associated with our primary measure of whole-word reading ability (word length effects).

In the second part we developed and assessed a novel, theoretically-driven intervention that targeted deficits in whole-word reading skills (called the visual training) and contrasted its effectiveness with an intervention that targeted deficits in phonological reading skills (called the GPC training). We first identified individuals with relatively selective deficits in either phonological or whole-word reading skills and invited a sample of these cases to participate in an intervention. We contrasted the effectiveness of the visual intervention to the GPC intervention. For reading accuracy we found no intervention-specific improvements, however, for reading latency, we did find intervention-specific improvements. That is, following the visual intervention, whole-word cases showed greater improvements in overall word reading speed than they did following the GPC intervention. Conversely, phonological cases showed greater improvements in overall reading speed following the GPC intervention than they did following the visual intervention.

We will now consider the implications of the results from these studies.

## Unique contributions of whole-word processes to reading proficiency

As outlined in Chapter 1, several recent theories propose that proficient reading relies upon two distinctly different sets of cognitive skills: one set utilises part-based analysis of words (*phonological* process), and the other handles words as a single coherent unit (*lexical or whole word* processes; see Ans et al., 1998; Coltheart et al., 2001; Pritchard et al., 2018).

According to this perspective, reading proficiency will depend upon an individual's skills across both these domains. However, not all theories take this view. One competing view is that phonological skills are central to the development of reading, and variability in these skills is responsible for most of the variation in reading proficiency amongst individuals (Griffiths & Snowling, 2002; Snowling, 1995). In support of this second view, Griffiths and Snowling (2002) found that children's performance on tasks that require phonological processing (e.g. nonword reading and phonemic awareness tasks) were positively associated with their performance on tasks that require whole-word skills (e.g. irregular word reading accuracy).

We found that, in an adolescent sample containing a large proportion of delayed readers, a measure of whole-word reading ability (specifically, word length effects) was a unique and independent predictor of reading proficiency, even after factoring out variance in phonological processing ability (assessed using a measure of nonword reading accuracy). These findings suggest that variation in phonological processing skills alone does not sufficiently account for variability in reading development, at least not in individuals who have several years of reading experience. Instead, we suggest that parallel letter processing – that is, the ability to process the letters of a familiar word rapidly and in parallel – is also a critical component. Previous research with impaired readers also provides evidence that word length effects index specific aspects of reading proficiency. In a sample of adolescent readers (11 to 14 years-old) with developmental dyslexia, Spinelli et al. (2005) were able to identify one sub-group that demonstrated marked word length effects suggesting a heavy reliance on serial letter processing, and another subgroup for whom these effects were minimal suggesting greater reliance on parallel letter processing. This pattern suggested that some readers may struggle due to parallel letter processing difficulties, while others may struggle for different reasons. Further, in our study, we found that the length effect and a measure of irregular word accuracy were associated, while the length effect and a measure of nonword accuracy were not. These findings along with the outcomes of the regression analyses suggest that, as per our prediction, parallel letter processing contributes to whole-word reading skills specifically.

In addition, the screening phase for our intervention study identified several cases meeting the criteria for dyslexia and demonstrating a relatively selective difficulty with one of these sets of skills. Specifically, out of a sample of 39 dyslexic individuals, we identified ten whose nonword accuracy was below the normal cut-off, but other whole-word cognitive skills (irregular word reading, reading speed, VA span) were within the normal range, and only one individual whose performance on these whole-word measures was below normal, but nonword accuracy was not. For example, case FR displayed impairments on measures of VA span and reading speed, but not nonword reading accuracy or phonemic awareness. This pattern suggests that FR's phonological processing skills were typical for their age, but their whole-word reading skills were well below normal. Such cases cannot be readily explained by the phonological deficit severity hypothesis (Griffiths & Snowling, 2002; Snowling, 1995).

The present research focused on reading performance in an adolescent sample. It is likely that whole-word skills become increasingly important as the reader develops and begins to transition from a largely part-based, bottom-up reading strategy to a more efficient, whole-word, top-down strategy. This proposal is consistent with a recent study of adults with developmental dyslexia that identified a sizeable subgroup with normal phonological processing skills (e.g. accurate nonword reading), but poor whole-word processing skills (e.g. inaccurate irregular word reading; Romani et al., 2008). Castles and Coltheart (1993) looked at irregular and nonword reading patterns in typically developing children (eight to 14 years-old) and found that age predicted substantially more of the variance in irregular word reading than it did for nonword reading. As a group, the participants in the Heterogeneity study performed poorly on our measure of irregular word accuracy and demonstrated a clear word length effect. Further, in the Screening study, we identified one case with pure whole-word reading difficulties, and another couple with disproportionate whole-word difficulties. Many more displayed complex profiles including both phonological and whole-word impairments. Altogether, these results suggest that older readers may enable us to measure or test whole-word reading processes.

However, our inferences regarding whole-word processing in reading rely on the assumption that the measures we used indexed the processes we purported them to. In this research, we have used the word length effect as our key index of whole-word processing. Our reasoning was that a larger effect of word length on reading latencies would indicate greater reliance on sequential letter processing, and conversely less reliance on whole-word based recognition strategies, which enable letters to be processed in parallel. However, an alternative explanation is possible. The SERIOL model proposes that length effects do not reflect the degree to which letters are processed in series, but rather the familiarity and neighbourhood density of the word (Whitney, 2001; Whitney & Cornelissen, 2005, 2008;

Whitney & Marton, 2013). The SERIOL model proposes that a sequential reading strategy is engaged during all types of word reading, even when the word is highly familiar. Consequently, more time is needed to process the visual information in a long word than in a short word. However, in highly skilled adult readers, the time cost involved in processing longer words is offset by their generally lower neighbourhood density. That is, long words have fewer neighbours than equally familiar short words and therefore, more discriminable activation patterns. However, this effect of neighbourhood density only emerges when readers have acquired a rich vocabulary of orthographic representations. Consequently, if a reader demonstrates strong word length effects, this would not indicate an unusually heavy reliance on serial letter processing, but rather that they have not developed high levels of familiarity with the stimulus words. To test this hypothesis, a nonword training task could be used. Participants could be trained with nonwords that have no close lexical neighbours (e.g. *cilognet*, *calognet*, *holognet*, *colognem*). The number of neighbours (i.e. the number of words differing by one letter) could be manipulated, as could familiarity through training exposure.

Further, it is important to keep in mind that our key measure of whole-word processing (word length effects) generally accounted for less of the variance in overall reading proficiency than measures associated with phonological processing (e.g. nonword accuracy). Also, in the dyslexic subsample identified in the Screening phase, more cases qualified as having a selective profile of phonological impairment (10 in total) than qualified as having a selective impairment in whole-word processing (three in total). These findings are consistent with several recent studies examining measures of phonological and whole-word processing respectively. For example, one study of undifferentiated readers (seven to eight years-old) found that measures of phonemic awareness, a phonological measure, accounted for more variance in reading proficiency than naming speed, a whole-word measure (Tibi & Kirby, 2017). Finding a similar pattern, Peterson et al. (2014) profiled dyslexic readers (eight to 13 years-old) based on their performance on a series of phonemic and orthographic choice tasks. A majority of participants performed disproportionately poorly on the phonemic choice task, while only a small number showed such a pattern on the orthographic choice task. Furthermore, in a five-year follow up, only one individual demonstrated a “switch” from phonological difficulties to whole-word difficulties, demonstrating the pervasiveness of these profiles over time. Together, these and the current findings suggest that, even in adolescent readers phonological processing skills are still crucial for successful reading. It is possible that, readers still rely heavily on phonological processing skills during this stage of development, even when reading familiar words. However, the effect of poor phonological processing skills may be more indirect: individuals with poorer phonological processing skills may be less inclined to read frequently, and consequently, acquire less reading experience than their peers.

We propose that reading proficiency in adolescence depends upon the co-operation of different cognitive processes that are at least partially independent of one another. The development of these processes, and ultimately reading proficiency, might be likened to the growth of a tree (see Figure 5.1 for an illustration of normal "tree" growth and Figure 5.2 for an illustration of delayed growth). For the tree to grow, there must be a seedling; the capacity to process and draw meaning from a wide range of complex visual stimuli. From here, the seedling requires different resources to grow: we must be exposed to words both explicitly (i.e. we must be told that particular letters or letter patterns correspond to specific sound patterns), and implicitly (i.e. we observe how others read letter patterns and words, and we are corrected when we mispronounce a word). A solid base or trunk of the tree develops; sight-sound rules become highly familiar and can be applied with relative ease. As the tree continues to grow, orthographic representations start to form, allowing for words to be represented and recognised as a whole unit (that is, the leaves start to appear on the tree). As the tree continues to grow, these newly learned mappings between orthography and phonology can then be generalised, leading to greater efficiency when reading new words, that is, this learning at the whole-word level also enhances the efficiency of part word processing. Within our tree analogy, the leaves enhance the overall health of the tree, including its trunk. Finally, the Reading Proficien-tree is fully developed!

Figure 5.1 The Reading Proficien-Tree

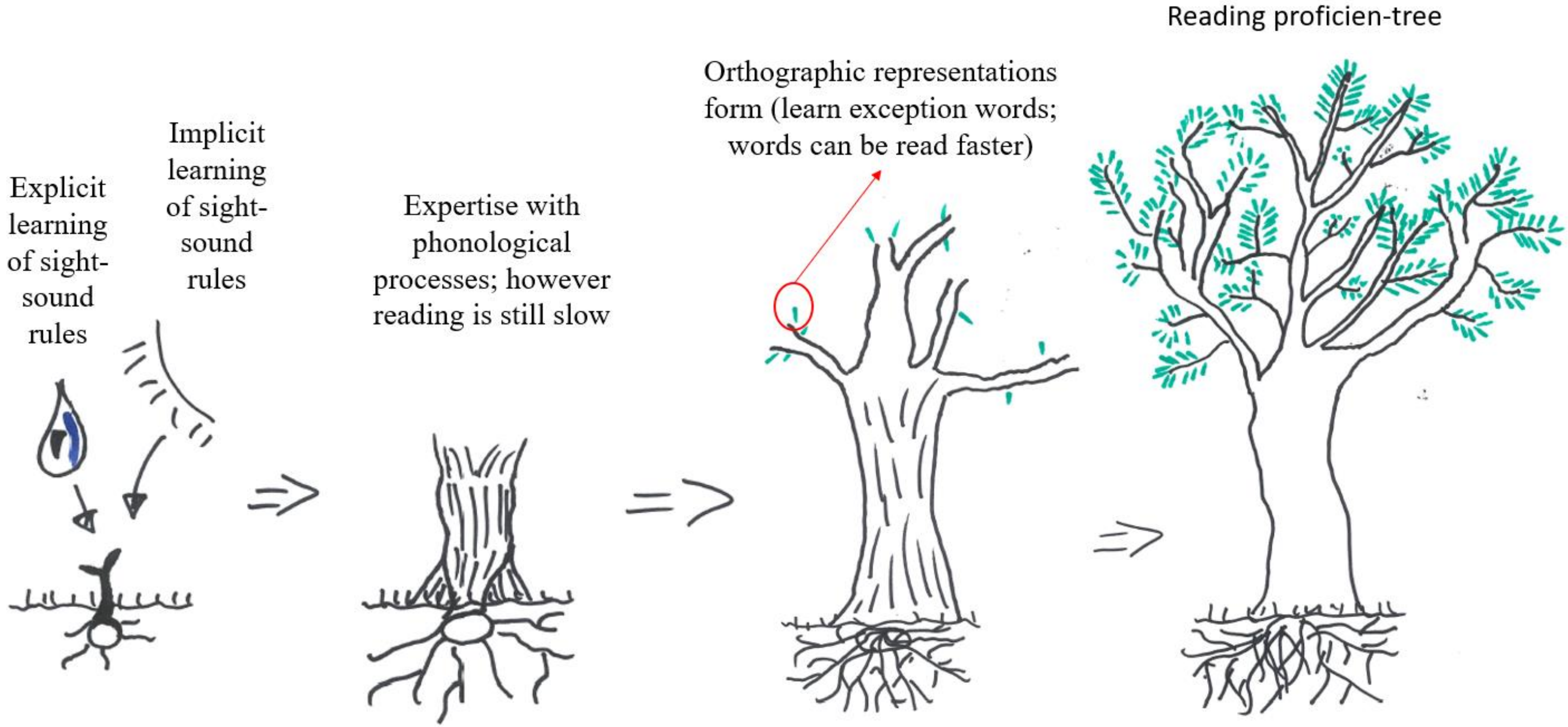
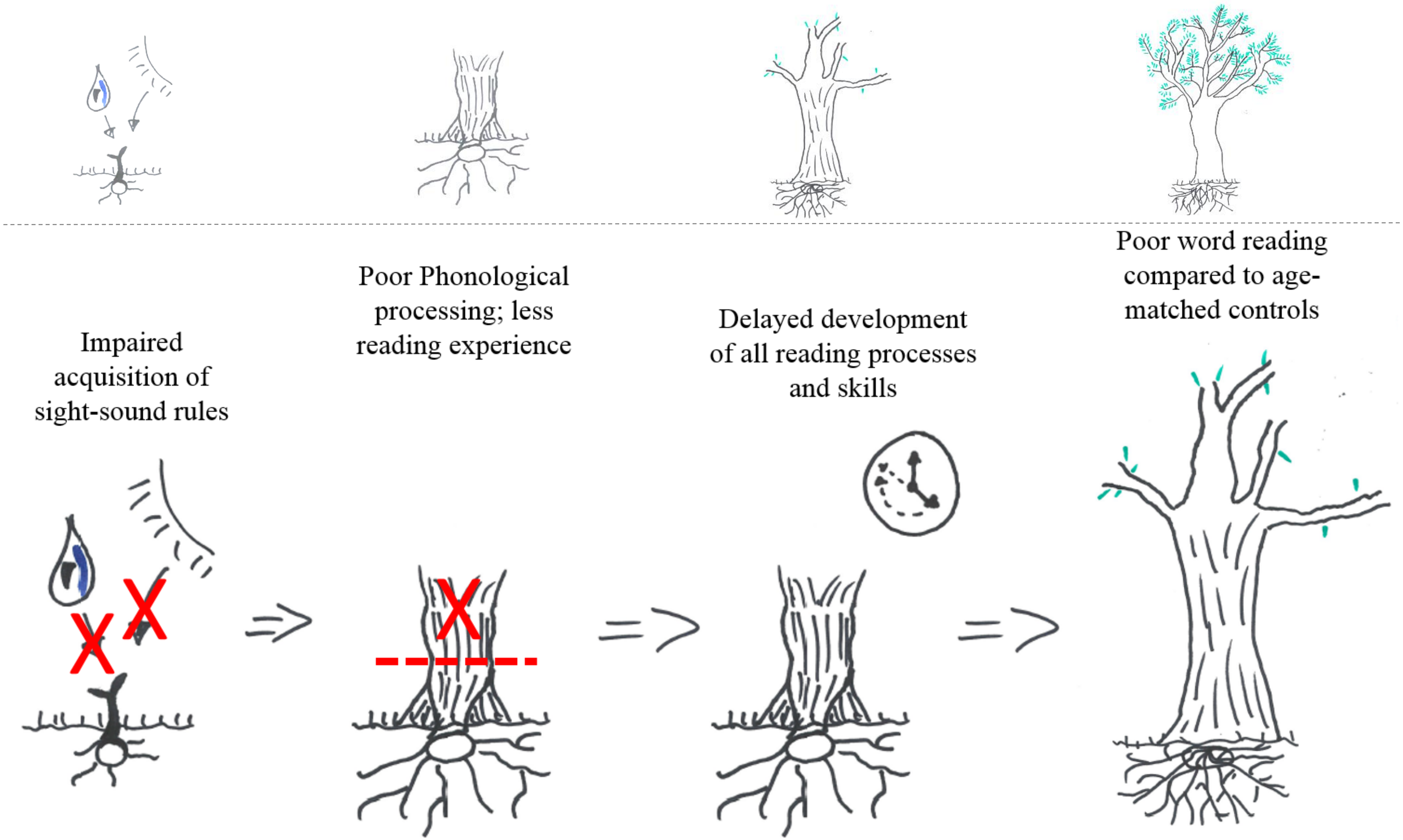


Figure 5.2 Delayed reading according to the reading proficien-tree



**Evaluating the VA span deficit hypothesis**

We turn now to our findings concerning the VA span deficit hypothesis (Bosse et al., 2007). As outlined in Chapter 1, the Multi-Trace Memory model proposes that readers possess a visual "window" which determines how many letters they can process in parallel (Ans et al., 1998). In a simulation of this model, the window was found to be crucial for accurate irregular word reading (Ans et al., 1998). The VA span deficit hypothesis (Bosse et al., 2007) proposes that certain types of developmental difficulties, such as those primarily affecting whole-word reading skills, are caused by a restriction to this visual window. It further proposes that this window can be measured using visual partial report tasks, which involve probed recall of a briefly presented array of visual stimuli. Those exhibiting a selective deficit on measures of whole-word reading skills (e.g. disproportionately poor irregular word reading) are predicted to perform poorly on these partial report tasks.

If the VA span deficit hypothesis is correct, the size of the individual's VA window will determine how many letters they will be able to process simultaneously. So, the wider the window, the less their reading latencies for familiar words will be impacted by word length. Empirically, VA span scores should therefore be negatively associated with word length effects. We did not find this to be the case. In our sample, scores on a visual partial report task were not associated with our primary measures of whole-word reading ability, including the word length effect measure, nor with a supplementary whole-word measure – irregular word accuracy. These findings are inconsistent with previous studies which have reported negative associations between VA span and measures of whole-word reading, such as irregular word accuracy or word length effects, across both children and adolescents (Bosse et al., 2007; Bosse & Valdois, 2009; van den Boer et al., 2013).

Surprisingly, while we did not observe the predicted association between VA span scores and the word length effect, we did observe a positive correlation between VA span scores and a measure of nonword reading accuracy. This finding suggests that there might be a phonological component to performance on the partial report task. The task requires participants to respond verbally by identifying the name of the target letter. Indeed, the Partial Report task has previously been criticised for its use of orthographic stimuli and response style (verbal), which means that participants may be able to use phonological strategies to enhance their performance on the task (Hawelka & Wimmer, 2008; Ziegler et al., 2010). If this is the case, then it should be possible to partial out the variance that VA span scores share with nonword accuracy scores and obtain a new span measure that is relatively uncontaminated by phonological processes. This residualised measure might be better at capturing the visual attentional processes that are critical to whole-word reading. Contrary to this hypothesis, we found that a residualised measure of this kind was not associated with the word length effect. Of course, residualizing a measure in this way has one particular disadvantage – we may be taking out variance that is genuinely attributable to the measure of interest. In future studies, a better approach may be to use a nonverbal version of the partial report task, consisting of meaningless symbols and shapes, requiring only a button press response (see Hawelka & Wimmer, 2008; Lobier et al., 2012; Ziegler et al., 2010).

It is also possible that the Partial Report task is an analytic and not a global measure, because it relies on the ability to recall a specific letter, rather than the whole letter string. However, during encoding, the target letter is unknown and the presentation time is short, meaning that performance may be enhanced when all letters and their positions are coded in a single capture. As such, it should not matter how many letters the participant is asked to recall. Instead, a greater issue may be the use of consonant strings. It may be that these strings cannot be interpreted as words, real or otherwise, preventing whole-word strategies entirely.

Additionally, and putting aside the issue of phonology, it is important to remember that the Partial Report task requires other mental processes not accounted for in this study. It requires the participant to not only attend to but also to retain visual information post-presentation, that is, the participant must hold the identity and position of the letters once they disappear and mentally overlay this representation onto the subsequent probe (for discussion, see Kuo, Stokes, & Nobre, 2012; Schneider, Mertes, & Wascher, 2016). It is difficult to devise a task that would measure visual attention span that doesn't make demands on working memory processes. However, one way to minimise the load is by assessing stimulus recognition in a more implicit way. For example, participants could be presented with stimuli arrays which they identify as either familiar (i.e. the array has been presented previously) or unfamiliar (i.e. a completely novel array). In such a task, both accuracy and latency measures could be used to index the span of visual attention. Another possibility is to use a matching task in which participants view two arrays in close succession (on different parts of the screen) and must say if they were the same or different. If the delay is kept short, the task may place fewer demands on working memory. One final possibility may be to use a Navon-style task, which can be used to assess two levels of attention: global and local (Navon, 1977). In this task, the outline of a large visual stimulus, such as a triangle, is made up of smaller stimuli that are either congruent (i.e. triangles), or incongruent (i.e. squares), and the participant is required to identify the large stimulus.

A second possibility is that VA span is crucial to reading, and that the partial report task is a good measure of this ability, but that its *contribution* to whole-word processing is less evident in our older readers than it would be in a younger sample. By adolescence, it is possible that numerous other factors such as reading experience contribute to overall reading ability, specifically whole-word based skills. These factors may be responsible for most of the individual variability in reading performance at this age. Instead then, VA span may be more fundamental to the initial development of whole-word processing skills, and less crucial for sustaining these skills.

The third possible explanation for our findings is that the visual attentional capacities indexed by a measure of VA span do not enhance whole-word reading skills selectively, but rather support reading proficiency in a more generic manner. Visual attention may play a more general role that is not specific to either phonological or whole-word processing. This proposal is certainly consistent with the observation that VA span scores did index some unique processes utilised during reading. Above, we suggested that there may be a two-way interplay between whole-word and phonological processes during reading development. Specifically, as an individual's body of orthographic knowledge develops, they have a richer database of orthographic-phonological correspondences to draw upon (implicitly) when they need to read using part-based strategies.

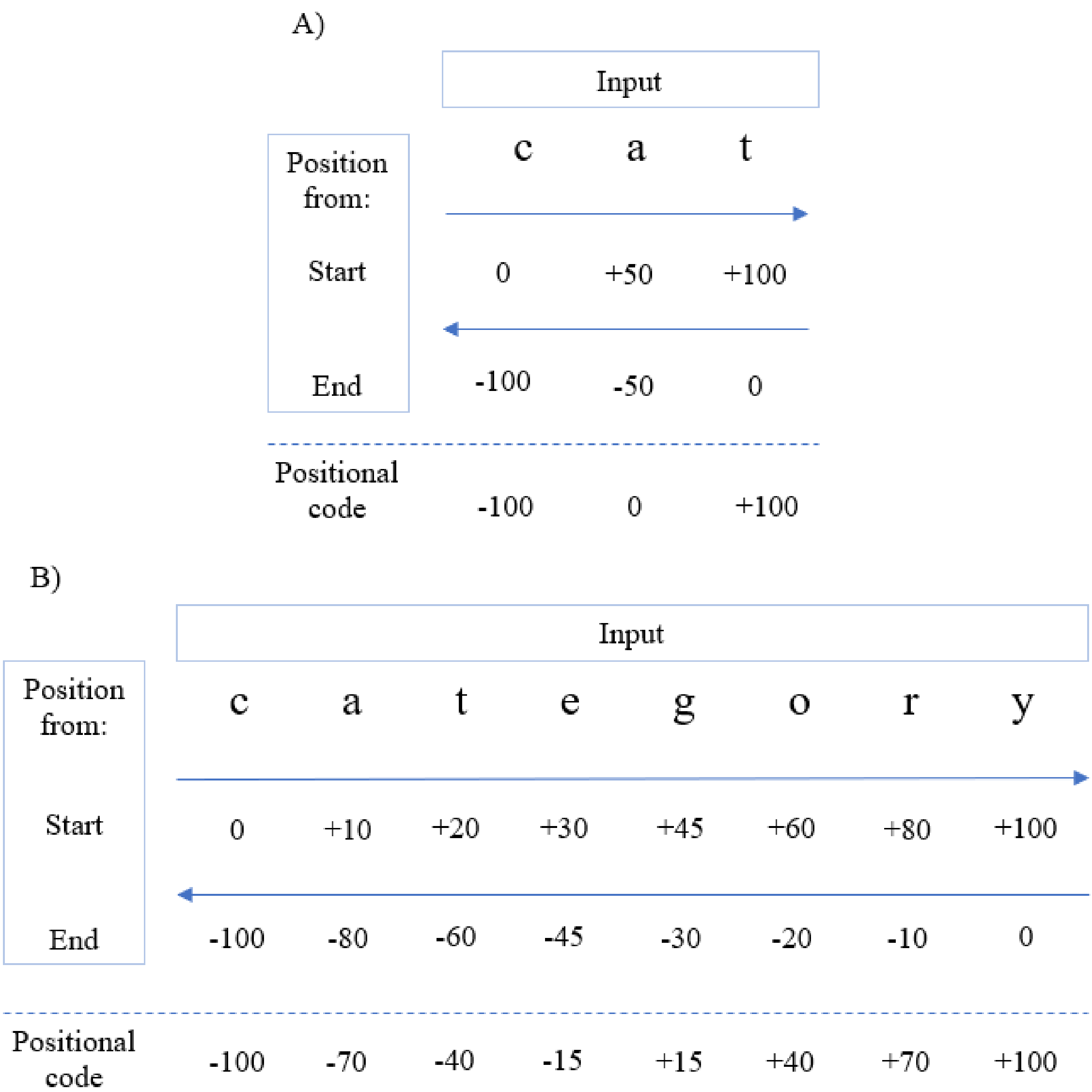
A final possibility we need to consider is that the VA window may not be the locus of variability at all. That is, the VA span deficit hypothesis is not an accurate model of the underlying deficit in individuals with whole-word processing difficulties. The Multi-Trace Memory model predicts that readers vary their processing style – and hence, the "size" of the VA window – according to whether the word they are reading is familiar (Ans et al., 1998). Therefore, apparent limitations in VA span might not be driven by a fixed limitation in the span itself, but rather a lack of familiarity with the word(s) being read. The strength of a word's lexical representation within the episodic memory layer depends upon how familiar the words is. If a word is relatively unfamiliar, there may be limited feedback from the lexical layer to the second ("echo") orthographic layer. As a result, the VA window must focus on the stimulus again, this time in analytic mode, and the predicted over-reliance on analytic or part-based processing occurs. Within this framework, it is not possible to tease apart these two competing explanations – a limited VA window *or* abnormally weak lexical representations. Indeed, it may be possible to account for the variability in whole word processing without recourse to the notion of a VA window at all. For example, SERIOL does not incorporate the notion of a VA window. Instead, it proposes that familiarity influences a reader's ability to predict the upcoming letters in a word (Whitney, 2001; Whitney & Cornelissen, 2008; Whitney & Marton, 2013). A lack of familiarity with words will compromise this capacity, and, as we mentioned above, longer words will be particularly disadvantaged here. To differentiate between the predictions of the VA span deficit hypothesis and the SERIOL model we could compare how VA span and familiarity uniquely predict overall reading proficiency.

To further understand what might be happening in whole-word reading, SERIOL's notion of *letter position tagging* is a useful starting point. In this model, letter position tagging is a crucial first step in reading via both phonological and whole word processes (Whitney, 2001; Whitney & Cornelissen, 2008). However, rather than being equally important for *both* phonological and whole word reading, we propose that position tagging plays a particularly important role in whole-word reading. Although not addressed directly in the current research, this proposal is consistent with previous literature on 'migration' errors (Kezilas, Kohnen, McKague & Castles, 2014; Kohnen & Castles, 2013; Kohnen, Nickels, Castles, Friedmann and McArthur, 2012). Migration errors occur when the letters of a word or non-word are switched (i.e. *carton* vs *catron*), negatively impacting reading accuracy. Previous research with … found higher error rates were associated with poor orthographic lexicon…. Interestingly, and in conflict with the expectation of both proposals for VA, migration errors were not often made with non-words, and there was no relationship between the number of errors on either real or non-words and performance on phonemic awareness tasks. These results suggest that letter position tagging may be more important to whole-word reading.

Specifically, we suggest that the representations of letter positions and their respective letter content are activated in parallel. As shown in Panel A of Figure 5.3, the *position tagging units* code the position of each letter relative to both the beginning and the end of the stimulus. This relative positioning is reminiscent of Henson's (1998, 1999) *start-end model* in short-term memory. For example, in the word *cat*, "*c*" at the grapheme layer would be tagged with the lowest positional code, because 0% of the stimulus letters appear to its left, and 100% of the remaining letters appear to its right. Conversely, the letter "t" will receive the highest positional code, because 100% of the other stimulus letters appear to its left. The middle letter "a" will be tagged with an intermediate positional code, because half the other letters in the stimulus appear to its left and half to its right.

This relative tagging will also allow differentiation in cases where an input stimulus could potentially activate multiple lexical units. For example, *cat* could activate both the "*cat*" unit and the "*category*" unit. As Panel B in Figure 5.3 shows, in both words, "c" has the same positional code (that is -100), however, the codes for subsequent letters "a" and "t" are different.

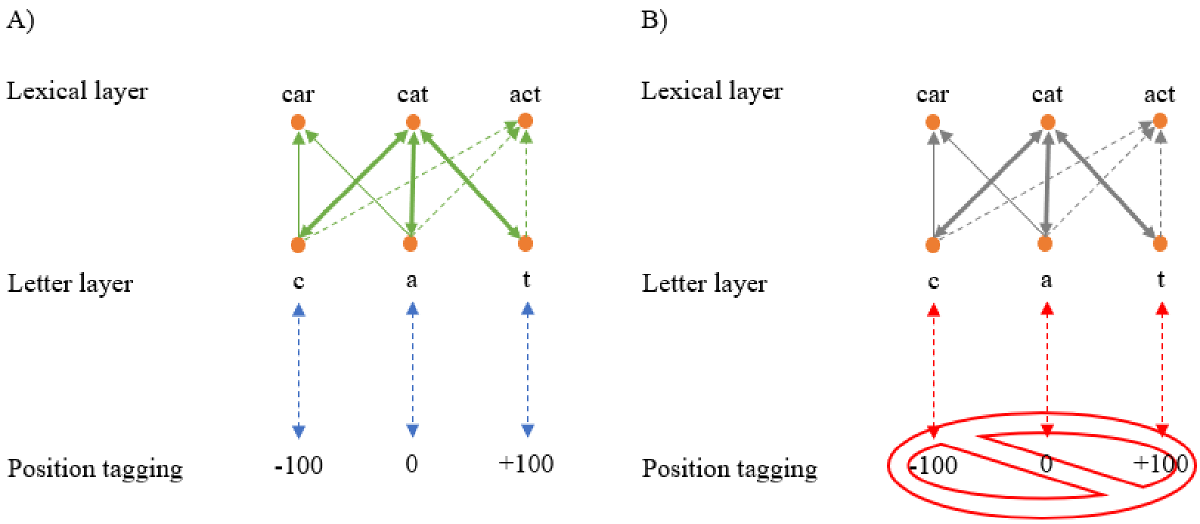
Figure 5.3 Proposed functioning of relative position tagging for whole-word reading



An overall schematic of this "Letter Tag" model is provided in Panel A of Figure 5.4. The relationship between each letter's identity and its position is maintained via bidirectional, resonating activation between the respective letter and position units. Each specific letter-position combination will activate all words containing that combination; the lexical unit of the word containing all the activated letter-position combinations will therefore receive the most activation. The direct mapping from letter units to lexical units allows for rapid whole-word recognition, which does not require direct grapheme-phoneme conversion.

Within this type of model, whole-word impairments would be conceptualised as difficulties with the position tagging processes, and/or mapping between letters and their positions (that is establishing effective resonance between letter and position units). This idea is illustrated in Panel B of Figure 5.4. We believe this offers a more specific account of attentional mechanisms in the reading processes, but we are certainly not proposing this as a conclusive account of underlying reading mechanisms. It illustrates just one of the ways we could account for the current findings without reference to any sort of visual attentional window.

Figure 5.4 Structural outline of the "Letter Tag" model




## Targeted intervention in a sample of adolescent readers

In the first part of this research, we found evidence to suggest that there are multiple predictors of reading proficiency, including phonological processing skills and whole-word processing skills. Therefore, we reasoned that marked difficulties with either one – or both – of these sets of skills could be responsible for the patterns of reading performance we see in developmental dyslexia. If this is the case, then the effectiveness of a reading intervention for these individuals will depend on how well it targets the skill set(s) most affected in that person. In the second part of this research, we tested this hypothesis.

This research question has both practical and theoretical implications. From a practical perspective, it is important that we seek to maximise the beneficial outcomes following reading intervention. To do this, we need to know whether different patterns or profiles of reading difficulty are better suited to different training programmes. From a theoretical perspective, the results have implications for our conceptual explanations of developmental dyslexia. Single deficit models of developmental dyslexia would predict that, although some treatments may be more effective than others overall, there should be little variability amongst individuals in which treatments are most effective. For example, the phonological deficit severity hypothesis (Griffiths & Snowling, 2002) would predict that all individuals with developmental dyslexia would benefit from an intervention that targets phonological processes (e.g. GPC training). In contrast, dual (and multiple) deficit models predict that tailoring a treatment to the individual's specific impairment will be more effective than a one-size-fits-all treatment. For example, the DRC model and the Multi-trace Memory model both predict that phonological cases will benefit more from a GPC intervention than a programme that focuses on the visual or other global properties of words. Whole word cases will exhibit the contrasting pattern (Ans et al., 1998; Coltheart et al., 2001).

Although the reading benefits we observed in this study were largely confined to reading latency measures, the results are consistent with the multiple deficit view. Four of the five cases we tested for whom data was available – two phonological: HE and MI, two whole-word: FR and SH – showed evidence of intervention-specific benefits following the target training. For phonological cases this means greater improvement following GPC training, and for whole-word cases this means greater improvement following visual training. These results suggest that disproportionately poor phonological and whole-word processing skills are best targeted using different methods.

These findings are particularly strong, because our study directly contrasted the effectiveness of two different reading interventions in the same individual. Previous research has shown that interventions that target an individual's specific reading profile can lead to improvements on reading measures (Broom & Doctor, 1995a, 1995b; Judica et al., 2002;

Law & Cupples, 2015; Valdois et al., 2014). However, these studies only examined one form of treatment, so we cannot know whether the reported improvements were greater than what one might expect using a different, less tailored form of treatment. In our study, we demonstrated that individuals' improvements on their tailored treatment *significantly exceeded* the improvements they showed following a non-target intervention.

Other previous investigations examining similar predictions have found mixed results. In their study, Rowse and Wilshire (2007) administered two different interventions, one a standard phonological training and the other a visual training that focused on word form. Two reading-delayed children (11 years-old) displaying contrasting profiles participate. The phonological case showed generalised improvements to untrained stimuli that were intervention specific – that is, the phonological intervention resulted in significantly greater improvements than the visual intervention. However, the whole-word case showed improvements that were not intervention-specific – that is, the phonological and visual interventions were equally effective. Similarly, Jones (2013) used a standard phonological training and a visual training that focused on redundant letter patterns (i.e. clusters), with another two contrasting cases. However, unlike Rowse and Wilshire (2007), they did not find any intervention-specific effects for either the phonological, or the whole-word case. Our research, which also used a standard phonological training, and a visual training employing clusters, found significant intervention-specific benefits similar to Rowse and Wilshire (2007), except that these findings were not limited to phonological cases; whole-word also cases showed greater benefits from a targeted intervention. Therefore, the results of our study suggest that redundant letter patterns may be useful for targeting whole-word impairments.

The results from our primary reading latency measure support the hypothesis that an intervention targeted to the participant's impairment profile will result in greater gains in reading efficiency than the alternative intervention. Given the small number of participants tested here, and the variability across their profiles, this result is particularly promising. Considering first the whole-word cases, the specific benefits conferred by the visual intervention suggest that the whole-word cases became more efficient at mapping orthographic representations following this training. Latency improvements were not specific to word type, suggesting that nonword improvements were equivalent to regular and irregular word improvements. This may be surprising given that one of the goals of the visual intervention was to encourage greater reliance on *whole-word level* processes, which of course by definition, do not exist for novel words. However, we have also noted that whole-word cases tend to show unusually longer reading latencies for novel or nonwords (Valdois et al., 2003; Zabell & Everatt, 2002; Zoccolotti et al., 1999). Indeed, this was a criterion for the whole-word profile in the current study. These results suggest that our visual intervention works by encouraging participants to process *all* orthographic materials in larger letter chunks, and this skill extends to nonwords as well. For phonological cases, the beneficial effects of GPC training on reading latencies suggests that reading speed may be determined by both phonological and whole-word processes. Considering these results alongside the findings from primary and secondary measures of accuracy, it may be that the GPC intervention worked to increase the efficiency of sight-sound rule *application*, as opposed to rule *acquisition*.

An important feature of our study is that our participants were older than most of those previously tested (Broom & Doctor, 1995a, 1995b; Jones, 2013; Law & Cupples, 2015; Rowse & Wilshire, 2007). Indeed, a recent meta-analyses found particularly strong benefits for older readers, suggesting that adolescents not only benefit from reading interventions but may do so to a great extent than younger readers (Suggate, 2016). There are a number of advantages to working with older children and adolescents when trying to assess the efficacy of interventions for reading. First, whole-word impairments may not become apparent until adolescence. Young readers may not be expected to read complex words (i.e. irregular words, long words), or read particularly fast as they are still learning. Only as they get older and the disparities between them and their peers grow, do difficulties with these stimuli or tasks become apparent. Second, older readers, even those with dyslexia, will have developed some of the most basic skills required for reading, such as the recognition of letter form and shape. Interventions with these older cases can build upon these basic skills and focus on higher-level processes that involve mapping this visual form to sound and meaning.

Another key feature of our design was that our measures of improvement were based on sets of words that were not explicitly trained in the intervention. The reason we focused on these untrained words was to assess generalisation. Generalisation is an important goal for any intervention study, because it signifies that an intervention successfully targeted underlying processes, rather than merely increasing familiarity with the specific training stimuli. We reasoned that each participant's tailored treatment would encourage the use of the trained skills and techniques, even outside of the training blocks, leading to greater improvements following an intervention tailored to that person. Overall then, our criteria for the current study were incredibly strict, making the intervention-specific improvements even more impressive.

Nonetheless, it is important to remember that the benefits we observed were limited to measures of reading latency; they did not extend to reading accuracy. There are a number of possibilities for this. One is that the interventions did not sufficiently target the *precise* difficulties hypothesised for each impairment profile. Our visual training used a novel method, inspired by previous studies that used cluster-style letter patterns (Jones, 2013; Law & Cupples, 2015). It was based on the clear theoretical rationale, that encouraging participants to package commonly occurring letter strings into a unit (i.e. a cluster) would facilitate parallel letter processing. As such, it was based on a distinctly different concept than the GPC training, which aimed to facilitate grapheme-to-phoneme conversion. However, at a superficial level, some elements of the interventions may have appeared similar to the participant. For example, both visual and GPC interventions require participants to read, define and spell word families, and both involve working with families of words that share common letter patterns. Also, although we made a clear distinction between GPCs and clusters at a theoretical level, at a more practical level, they may not have appeared distinct. For example, some of the clusters used in the visual training consist of easily identifiable and common letter-sound correspondences, so they may have similar properties to GPCs (i.e. the cluster *ation* has a consistent pronunciation across words). Also, it is possible that different participants may have adopted different strategies to read the reading and test stimuli. Some participants might tend to capitalise on their strengths across both types of interventions. For example, in whole-word cases, a useful strategy might be to process the cluster as if they were GPCs, and in phonological cases, the converse strategy – focusing on larger letter patterns – might be advantageous. If some of the participants did adopt such a strategy, then the techniques learned or rehearsed in the two interventions might not have been that different, and the opportunity for intervention-specific improvements would be diminished.

A second possibility is that it may be unreasonable to expect that targeting the kinds of skills we did would lead to improved accuracy on words that weren't explicitly trained, particularly irregular words. Instead, it may be more reasonable to expect improvements to be evident in reading latency measures as readers become more efficient in mapping across orthographic-phonemic representations. Irregular words, by definition, must be learned individually. Therefore, it is unclear as to how an intervention targeting general reading strategies could lead to improved accuracy on irregular words.

A third possible explanation is that the training stimuli we used were not challenging enough to elicit substantial change in this older, adolescent sample. The average age of acquisition of the words trained in both interventions was well below the chronological ages of all participants in this study, and for most, they were also below Reading Age scores (SH is the exception). Consequently, the opportunity for accuracy improvements using these words may have been limited because they were not novel or challenging enough. Accuracy may benefit more from a more error-driven approach, which focuses on words that are more challenging or unfamiliar to the participant.

One final possibility is that our accuracy measure was simply not sensitive enough to inspire any genuine improvements following treatment. It is more difficult to detect training-related improvements on measures of accuracy, because the statistical comparisons have low power. Also, in the cases studied here, ceiling effects may constrain the opportunity for improvement. While our participants did produce errors on the irregular word and/or nonword lists during the pre-intervention tests, all performed at or near ceiling on regular words at this timepoint. Clearly, if a participant achieves an accuracy rate of 25 or more out of 30 on a specific word list before treatment, the opportunity for change is extremely limited.

**Limitations and Suggestions for Future Research.** The current study focused entirely on reading at the single-word level. Of course, as this thesis is evidence of, reading is much more than simply recognising and verbally reading a word in isolation. Proficient reading requires the integration of meaning across multiple words in a sentence, and the retention of meaning or concepts for integration with later sentences in the same paragraph. Paragraph level reading is commonly assessed using either a fluency task that involves reading a short paragraph aloud, or a passage comprehension task that involves reading a short paragraph and then answering a series of questions about the content. There is considerable evidence that skills at the single word level are important predictors of competency on these kinds of passage reading tasks. For example, Shankweiler et al. (1999) found that young readers (seven to nine years-old) defined as impaired on single-word reading tasks also performed below the age-norm in a text comprehension task. Similarly, such children demonstrate an increase in the number and duration of fixations in passage reading (De Luca et al., 1999; Prado, Dubois, & Valdois, 2007). In our study, as part of the profile building phase, we assessed passage reading accuracy and speed in a subset of our intervention participants. Unsurprisingly, these individuals, who all performed well below age norms on standard single word tests, also showed poor passage level reading. However, that is not to say that single word reading is the only determinant of passage fluency and comprehension. Cutting, Materek, Cole, Levine, and Mahone (2009) looked at the predictors of reading difficulty in children (nine to 14 years-old) who scored within the normal range on single word reading tasks, but below normal on a measure of passage reading fluency. They found that this group showed impairment on measures of syntax skills and verbal semantic processing. Therefore, it is paramount that future research investigates how reading is affected at a paragraph-level in developmental dyslexia, both from an aetiological and therapeutic perspective.

An important methodological issue in our research that needs consideration is that of examiner bias. In this study, the intervention sessions were delivered by a researcher who was non-blinded to the study hypothesis (EA). While every effort was made to present both interventions with equal enthusiasm and rigour, it remains possible that the researcher may have inadvertently favoured the individual's target intervention. It would be therefore be worthwhile in future, larger studies of this type to utilise researchers who are blind to the study hypotheses.

Turning now to recommendations for future extensions of this research, we recommend a much larger replication, using a larger sample, and if possible a full ABBA design (see Chapter 4 for discussion). Ideally, such a study would also include more training sessions, and base its outcome measures on larger stimulus sets. If the effects found here are successfully replicated, then the next step would be to isolate the specific elements of the interventions that are responsible for change. For example, it would be useful to determine whether the differential effects of the two treatments arise as a result of the particular types of stimuli that are trained (i.e. GPCs vs clusters), or as a result of the specific tasks that the participant is asked to perform, irrespective of the stimuli. This could be investigated by independently varying the training stimuli and tasks performed, examining the effect of each manipulation in isolation. If we were able to find an effect of stimulus type independent of the task, this would suggest that participants are (at least implicitly) aware of the difference between letter pairs and clusters, and also that these letter patterns successfully target the hypothesised underlying deficits. Conversely, if we found an effect of task independent of stimulus type, then this would suggest that the activities within the tasks are driving the effects. In the case of the visual training, this would suggest that there is something about the actions of moving, manipulating words and recalling meaning that is beneficial for whole-word processing.

Another crucial objective for future research is to examine the long-term benefits of these interventions, and also whether they translate into meaningful improvements in an everyday environment. Indeed, the litmus test for any reading intervention is whether it leads to enduring changes in reading proficiency. To test for such changes, both screening and intervention cohorts could complete an additional measure of reading proficiency that could be administered prior to, immediately after, and six-months following the intervention phase of the study. Alternatively, the two different interventions could be administered at two different timepoints with an intervening delay of several months. In this way, each intervention could be assessed not only immediately after training, but also several months post-training. The ecological validity of improvements during reading intervention research is also important. No matter how effective a reading intervention is within the context of a study setting, the ultimate test would be whether improvements were found in reading outside of the research sessions. Future research could assess whether those who undertake the intervention perform better in class tests or assessments than their reading delayed peers that do not undergo any training.

Finally, in future studies it may be useful to consider not only reading, but also spelling. Previous research has found that spelling and other writing deficits often accompany developmental dyslexia, and that the nature of the spelling errors provide another useful index of the underlying deficit (Banfi et al., 2018; Daigle, Costerg, Plisson, Ruberto, & Varin, 2016; Georgiou et al., 2012; Romani, Olson, Ward, & Ercolani, 2002; Valdois et al., 2003; Zoubrinetzky et al., 2014). Further, in intervention studies, it would be useful to assess spelling as an outcome in its own right. Improvements on spelling measures would suggest that the trained skills are not limited to written word identification but extend to the conversion of verbal stimuli to written words.

**Concluding comments.** In sum, the results of the current research are largely consistent with existing theoretical hypotheses and empirical research. Specifically, our first study explored individual variation in reading and reading-related skills in an undifferentiated adolescent sample. We found that indices of both phonological and whole-word processing explained a unique proportion of the variance in overall reading proficiency. We also examined the relationship between these key reading measures and a commonly-used measure of VA span. While VA span scores were associated with overall reading proficiency, contrary to our predictions, they were not associated with our primary measure of whole-word reading ability. It remains unclear whether our negative results are due to limitations in the measures used to assess visual attentional span, or whether the concept itself is problematic. This is a question for future research.

Our second study focused on the development and assessment of a novel, visual intervention targeting whole-word reading skills and contrasted it against a typical GPC intervention targeting phonological reading skills. We found no intervention-specific improvements in reading accuracy. However, we did find intervention-specific improvements in reading latencies. That is, whole-word cases showed greater reading speed improvements following the visual intervention, while phonological cases showed greater reading speed improvements following the GPC intervention. These results suggest that an intervention tailored to an individual's impairment profile is more effective than a generic approach that does not consider the nuance in each individual case.

Typically developing adolescent readers are robust to variations in the characteristics of words. They are able to identify a word irrespective of length, familiarity, or regularity. We would expect these readers to be able to read typo-laden texts without stumbing. Adolescents with developmental dyslexia on the other hand may find themselves stumbling in the sentence above. Even though they are not looking for the errors (like you may have just gone back and done), the errors are likely to affect their reading efficiency, particularly if they struggle with recognising words at a single glance. For these individuals, there are still many situations in in day-to-day life where reading is crucial. Understanding the instructions for fixing the kitchen sink, following directions to the forest, and enjoying all seven Harry Potter books, all require a level of reading proficiency. It is paramount that the interventions we offer those with dyslexia target the specific cognitive difficulties that are hampering this proficiency.

"If you are going to get anywhere in life you have to read a lot of books."

Roald Dahl

**References**

Ablinger, I., von Heyden, K., Vorstius, C., Halm, K., Huber, W., & Radach, R. (2014). An eye movement based reading intervention in lexical and segmental readers with acquired dyslexia. *Neuropsychological Rehabilitation, 24*(6), 833-867. doi:10.1080/09602011.2014.913530

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review, 105*(4), 678-723.

Antzaka, A., Lallier, M., Meyer, S., Diard, J., Carreiras, M., & Valdois, S. (2017). Enhancing reading performance through action video games: the role of visual attention span. *Scientific Reports, 7*(1), 14563-14563. doi:10.1038/s41598-017-15119-9

Audacity-Team. (2014). Audacity (Version 2.0. 2). Retrieved from https://audacityteam.org/

Awadh, F. H. R., Phénix, T., Antzaka, A., Lallier, M., Carreiras, M., & Valdois, S. (2016). Cross-language modulation of visual attention span: An Arabic-French-Spanish comparison in skilled adult readers. *Frontiers in Psychology, 7*, 1-10. doi:10.3389/fpsyg.2016.00307

Baddeley, A. D., Logie, R. H., & Ellis, N. C. (1988). Characteristics of developmental dyslexia. *Cognition, 29*(3), 197-228. doi:10.1016/0010-0277(88)90024-8

Ball, E. W., & Blachman, B. A. (1988). Phoneme Segmentation Training: Effect on Reading Readiness. *Annals Dyslexia, 38*(1), 208-225. doi:10.1007/bf02648257

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1563-1571. doi:10.1037/a0032186

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of experimental psychology: General, 133*(2), 283-316. doi:10.1037/0096-3445.133.2.283

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*(3), 445-459. doi:10.3758/BF03193014

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior research methods, 39*(3), 445-459.

Banfi, C., Kemény, F., Gangl, M., Schulte-Körne, G., Moll, K., & Landerl, K. (2018). Visual attention span performance in German-speaking children with differential reading and spelling profiles: No evidence of group differences. *PLoS One, 13*(6), e0198903. doi:10.1371/journal.pone.0198903

Barton, J. J. S., Hashim, M. H., Björnström, L. E., & Hills, C. (2014). The word-length effect in reading: A review. *Cognitive Neuropsychology, 31*(5-6), 378-412. doi:10.1080/02643294.2014.895314

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Bijeljac-Babic, R., Millogo, V., Farioli, F., & Grainger, J. (2004). A developmental investigation of word length effects in reading using a new on-line word identification paradigm. *Reading and Writing, 17*(4), 411-431. doi:10.1023/B:READ.0000032664.20755.af

Björnström, L. E., Hills, C., Hanif, H., & Barton, J. J. S. (2014). Visual Word Expertise: A Study of Inversion and the Word-Length Effect, with Perceptual Transforms. *Perception, 43*(5), 438-450. doi:10.1068/p7698

Bogon, J., Finke, K., Schulte-Körne, G., Müller, H. J., Schneider, W. X., & Stenneken, P. (2014). Parameter-based assessment of disturbed and intact components of visual attention in children with developmental dyslexia. *Developmental Science, 17*(5), 697-713. doi:10.1111/desc.12150

Bosse, M. L., Chaves, N., Largy, P., & Valdois, S. (2015). Orthographic learning during reading: the role of whole-word visual processing. *Journal of Research in Reading, 38*(2), 141-158.

Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: the visual attention span deficit hypothesis. *Cognition, 104*(2), 198-230. doi:10.1016/j.cognition.2006.05.009

Bosse, M. L., & Valdois, S. (2009). Influence of the visual attention span on child reading performance: a cross-sectional study. *Journal of Research in Reading, 32*(2), 230-253. doi:0.1111/j.1467-9817.2008.01387.x

Broom, Y. M., & Doctor, E. A. (1995a). Developmental phonological dyslexia: A case study of the efficacy of a remediation programme. *Cognitive Neuropsychology, 12*(7), 725-766. doi:10.1080/02643299508251400

Broom, Y. M., & Doctor, E. A. (1995b). Developmental surface dyslexia: A case study of the efficacy of a remediation programme. *Cognitive Neuropsychology, 12*(1), 69-110. doi:10.1080/02643299508251992

Brunsdon, R. K., Hannan, T. J., Nickels, L., & Coltheart, M. (2002). Successful treatment of sublexical reading deficits in a child with dyslexia of the mixed type. *Neuropsychological Rehabilitation, 12*(3), 199-229. doi:10.1080/09602010244000048

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Journal of Behavior research methods, 41*(4), 977-990.

Bus, A. G., & Van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*(3), 403-414.

Campbell, R., & Butterworth, B. (1985). Phonological dyslexia and dysgraphia in a highly literate subject: A developmental case with associated deficits of phonemic processing and awareness. *The quarterly journal of experimental psychology, 37*(3), 435-475. doi:10.1080/14640748508400944

Castles, A., & Coltheart, M. (1993). Varieties of Developmental Dyslexia. *Cognition, 47*(1), 149-180. doi:10.1016/0010-0277(93)90003-e

Castles, A., & Coltheart, M. (1996). Cognitive Correlates of Developmental Surface Dyslexia: A Single Case Study. *Cognitive Neuropsychology, 13*(1), 25-50. doi:10.1080/026432996382051

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and Nature of Late-Emerging Poor Readers. *Journal of Educational Psychology, 104*(1), 166-181. doi:10.1037/a0025323

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A, 33*(4), 497-505.

Coltheart, M., & Leahy, J. (1996). Assessment of Lexical and Nonlexical Reading Abilities in Children: Some Normative Data. *Australian Journal of Psychology, 48*(3), 136-140. doi:10.1080/00049539608259520

Coltheart, M., Masterson, J., Byng, S., Prior, M., & Riddoch, J. (1983). Surface dyslexia. *The Quarterly Journal of Experimental Psychology Section A, 35*(3), 469-495. doi:10.1080/14640748308402483

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256. doi:10.1037/0033-295X.108.1.204

Coltheart, M., Woollams, A., Kinoshita, S., & Perry, C. (1999). A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin & Review, 6*(3), 456-463. doi:10.3758/BF03210835

Conlon, E., Lovegrove, W., Barker, S., & Chekaluk, E. (2001). Visual Discomfort: The Influence of Spatial Frequency. *Perception, 30*(5), 571-581. doi:10.1068/p2954

Cornelissen, P., Richardson, A., Mason, A., Fowler, S., & Stein, J. (1995). Contrast sensitivity and coherent motion detection measured at photopic luminance levels in dyslexics and controls. *Vision Research, 35*(10), 1483-1494. doi:10.1016/0042-6989(95)98728-R

Cosky, M. J. (1976). The role of letter recognition in word recognition. *Memory & Cognition, 4*(2), 207-214. doi:10.3758/bf03213165

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*(4), 482-486.

Cutting, L. E., Materek, A., Cole, C. A., Levine, T. M., & Mahone, E. M. J. A. o. d. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *59*(1), 34-54.

Daigle, D., Costerg, A., Plisson, A., Ruberto, N., & Varin, J. (2016). Spelling Errors in French-speaking Children with Dyslexia: Phonology May Not Provide the Best Evidence. *Dyslexia, 22*(2), 137-157. doi:10.1002/dys.1524

De Groot, F., Huettig, F., & Olivers, C. N. (2016). When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 42*(2), 180-196. doi:10.1037/xhp0000102

De Luca, M., Barca, L., Burani, C., & Zoccolotti, P. (2008). The effect of word length and other sublexical, lexical, and semantic variables on developmental reading deficits. *Cognitive and Behavioral Neurology, 21*(4), 227-235. doi:10.1097/WNN.0b013e318190d162

De Luca, M., Di Pace, E., Judica, A., Spinelli, D., & Zoccolotti, P. (1999). Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologia, 37*(12), 1407-1420. doi:http://dx.doi.org/10.1016/S0028-3932(99)00038-X

Denckla, M. B., & Rudel, R. G. (1976). Rapid ‘automatized’ naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14*(4), 471-479. doi:https://doi.org/10.1016/0028-3932(76)90075-0

DFNZ. (2017). Recognising Dyslexia. Retrieved from http://www.dyslexiafoundation.org.nz/info.html

Edwards, V., & Hogben, J. (1999). New Norms for Comparing Children’s Lexical and Nonlexical Reading: A Further Look at Subtyping Dyslexia. *Australian Journal of Psychology, 51*(1), 37–49. https://doi.org/10.1080/00049539908255333

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2*(3), 361-379. doi:10.1037/0096-1523.2.3.361

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology, 74*(2), 159-180. doi:10.1111/j.2044-8295.1983.tb01851.x

Gagl, B., Hawelka, S., & Wimmer, H. (2015). On Sources of the Word Length Effect in Young Readers. *Scientific Studies of Reading, 19*(4), 289-306. doi:10.1080/10888438.2015.1026969

Gamer, M., Lemon, J., Gamer, M. M., Robinson, A., & Kendall's, W. (2012). Various coefficients of interrater reliability agreement (package 'irr'). Retrieved from http://idg.pl/mirrors/CRAN/web/packages/irr/irr.pdf

Georgiou, G. K., Papadopoulos, T. C., Zarouna, E., & Parrila, R. (2012). Are auditory and visual processing deficits related to developmental dyslexia? *Dyslexia: An International Journal of Research and Practice, 18*(2), 110-129. doi:http://dx.doi.org/10.1002/dys.1439

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395-427. doi:10.3758/bf03201693

Gordon, P. C., Plummer, P., & Choi, W. (2013). See before you jump: Full recognition of parafoveal words precedes skips during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 633-641. doi:10.1037/a0028881

Griffiths, Y. M., & Snowling, M. J. (2002). Predictors of exception word and nonword reading in dyslexic children: The severity hypothesis. *Journal of Educational Psychology, 94*(1), 34-43. doi:10.1037/0022-0663.94.1.34

Gvion, A., & Friedmann, N. (2013). A selective deficit in imageable concepts: a window to the organization of the conceptual system. *Frontiers in Human Neuroscience, 7*. doi:10.3389/fnhum.2013.00226

Hammill, D. D., Mather, N., & Roberts, R. (2001). *ITPA-3: Illinois Test of Psycholinguistic Abilities*: Pro-ed.

Harrell Jr, F. E., & Harrell Jr, M. F. E. (2019). Package 'Hmisc': CRAN2018.

Hawelka, S., & Wimmer, H. (2008). Visual target detection is not impaired in dyslexic readers. *Vision Research, 48*(6), 850-852. doi:10.1016/j.visres.2007.11.003

Henson, R. N. A. (1998). Short-Term Memory for Serial Order: The Start-End Model. *Cognitive Psychology, 36*(2), 73-137. doi:10.1006/cogp.1998.0685

Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition, 27*(5), 915-927. doi:10.3758/BF03198544

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*(1), 74-95. doi:10.1037/0033-295X.98.1.74

Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G., & Mitchell, D. (2018). Package 'lmtest': Testing linear regression models.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with Application in R* (Vol. 112). New York: Springer.

Jones, B. (2013). *Treatment of Developmental Reading Disorders: A Visual Attention Approach.* (Masters Thesis), Victoria University of Wellington, Wellington, New Zealand. Retrieved from http://researcharchive.vuw.ac.nz/xmlui/handle/10063/2754

Joo, S. J., White, A. L., Strodtman, D. J., & Yeatman, J. D. (2018). Optimizing text for an individual's visual system: The contribution of visual crowding to reading difficulties. *Cortex, 103*, 291-301. doi:https://doi.org/10.1016/j.cortex.2018.03.013

Jorm, A. F., & Share, D. L. (1983). An invited article: Phonological recoding and reading acquisition. *Journal of Applied psycholinguistics, 4*(2), 103-147. doi:10.1017/S0142716400004380

Judica, A., De Luca, M., Spinelli, D., & Zoccolotti, P. (2002). Training of developmental surface dyslexia improves reading performance and shortens eye fixation duration in reading. *Neuropsychological Rehabilitation, 12*(3), 177-197. doi:10.1080/09602010244000002

Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision: Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition, 55*(2), 332-340. doi:10.1016/j.bandc.2004.02.035

Kezilas, Y., Kohnen, S., McKague, M., & Castles, A. (2014). The locus of impairment in English developmental letter position dyslexia. Frontiers in human neuroscience, 8, 356.

King, W. M., Giess, S. A., & Lombardino, L. J. (2007). Subtyping of children with developmental dyslexia via bootstrap aggregated clustering and the gap statistic: comparison with the double-deficit hypothesis. *International Journal of Language and Communication Disorders, 42*(1), 77-95. doi:10.1080/13682820600806680

Kirk, J., & Reid, G. (2001). An examination of the relationship between dyslexia and offending in young people and the implications for the training system. *Dyslexia, 7*(2), 77-84. doi:10.1002/dys.184

Kohnen, S., & Castles, A. (2013). Pirates at parties: Letter position processing in developing readers. Journal of Experimental Child Psychology, 115(1), 91-107.

Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When ‘slime’becomes ‘smile’: developmental letter position dyslexia in English. Neuropsychologia, 50(14), 3681-3692.

Kohnen, S., Nickels, L., Geigis, L., Coltheart, M., McArthur, G., & Castles, A. (2018). Variations within a subtype: Developmental surface dyslexias in English. *Cortex, 106*, 151-163. doi:https://doi.org/10.1016/j.cortex.2018.04.008

Kuo, B.-C., Stokes, M. G., & Nobre, A. C. (2012). Attention modulates maintenance of representations in visual short-term memory. *Journal of Cognitive Neuroscience, 24*(1), 51-60. doi:10.1162/jocn_a_00087

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13). doi:10.18637/jss.v082.i13

Kwok, R. K. W., Cuetos, F., Avdyli, R., & Ellis, A. W. (2017). Reading and Lexicalization in Opaque and Transparent Orthographies: Word Naming and Word Learning in English and Spanish. *Quarterly Journal of Experimental Psychology, 70*(10), 2105-2129. doi:10.1080/17470218.2016.1223705

Lallier, M., Abu Mallouh, R., Mohammed, A. M., Khalifa, B., Perea, M., & Carreiras, M. (2018). Does the Visual Attention Span Play a Role in Reading in Arabic? *Scientific Studies of Reading, 22*(2), 181-190. doi:10.1080/10888438.2017.1421958

Lallier, M., Donnadieu, S., & Valdois, S. (2013). Developmental dyslexia: Exploring how much phonological and visual attention span disorders are linked to simultaneous auditory processing deficits. *Annals Dyslexia, 63*(2), 97-116. doi:10.1016/ j.neuron.2011.11.002.

Lassus-Sangosse, D., N'guyen-Morel, M.-A., & Valdois, S. (2008). Sequential or simultaneous visual processing deficit in developmental dyslexia? *Vision Research, 48*(8), 979-988. doi:10.1016/j.visres.2008.01.025

Law, C., & Cupples, L. (2015). Thinking outside the boxes: Using current reading models to assess and treat developmental surface dyslexia. *Neuropsychological Rehabilitation, 27*(2), 149-195. doi:10.1080/09602011.2015.1064453

Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*(2), 211-234. doi:10.1037/0022-0663.95.2.211

Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-squares means: R package version.

Lesnoff, M., & Lancelot, R. (2012). aod: analysis of overdispersed data: R package version 1.

Lewandowska, M., Milner, R., Ganc, M., Wlodarczyk, E., & Skarzynski, H. (2014). Attention dysfunction subtypes of developmental dyslexia. *Medical Science Monitor, 20*, 2256-2268. doi:10.12659/MSM.890969

Lobier, M., Zoubrinetzky, R., & Valdois, S. (2012). The visual attention span deficit in dyslexia is visual and not verbal. *Cortex, 48*(6), 768-773. doi:10.1016/j.cortex.2011.09.003

Lovegrove, W. (1993). Weakness in the Transient Visual System: A Causal Factor in Dyslexia? *Annals of the New York Academy of Sciences, 682*(1), 57-69. doi:10.1111/j.1749-6632.1993.tb22959.x

Lovett, M. M., Borden, S. L., DeLuca, T., Lacerenza, L., Benson, N. J., & Brackstone, D. (1994). Treating the Core Deficits of Developmental Dyslexia: Evidence of Transfer of Learning After Phonologically- and Strategy-Based Reading Training Programs. *Developmental Psychology, 30*(6), 805-822. doi:10.1037/0012-1649.30.6.805

Manis, F. R., Seidenberg, M. S., Doi, L. M., McBride-Chang, C., & Petersen, A. (1996). On the bases of two subtypes of development dyslexia. *Cognition, 58*(2), 157-195. doi:10.1016/0010-0277(95)00679-6

Marinus, E., & de Jong, P. F. (2010). Variability in the word-reading performance of dyslexic readers: Effects of letter length, phoneme length and digraph presence. *Cortex, 46*(10), 1259-1271. doi:10.1016/j.cortex.2010.06.005

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2*(3), 175-199. doi:10.1007/bf01067101

Martens, V. E., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language, 98*(2), 140-149. doi:10.1016/j.bandl.2006.04.003

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324. doi:10.3758/s13428-011-0168-7

MathWorks, I. (2005). MATLAB: the language of technical computing. Desktop tools and development environment, version 7: MathWorks.

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., & Banales, E. (2013). Sight Word and Phonics Training in Children With Dyslexia. *Journal of learning disabilities*, *48*(4), 391-407. doi:10.1177/0022219413504996

McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology*, *87*(2), 179-192. doi:10.1037/0022-0663.87.2.179

Moores, E., Cassim, R., & Talcott, J. B. (2011). Adults with dyslexia exhibit large effects of crowding, increased dependence on cues, and detrimental effects of distractors in visual search tasks. *Neuropsychologia*, *49*(14), 3881-3890. doi:10.1016/j.neuropsychologia.2011.10.005

Moores, E., Tsouknida, E., & Romani, C. (2015). Adults with dyslexia can use cues to orient and constrain attention but have a smaller and weaker attention spotlight. *Vision Research*, *111*, 55-65. doi:10.1016/j.visres.2015.03.019

Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1998). Segmentation, Not Rhyming, Predicts Early Progress in Learning to Read. *Journal of Experimental Child Psychology*, *71*(1), 3-27. doi:10.1006/jecp.1998.2453

Nation, K., & Snowling, M. J. (1998). Semantic Processing and the Development of Word-Recognition Skills: Evidence from Children with Reading Comprehension Difficulties. *Journal of Memory and Language, 39*(1), 85-101. doi:http://dx.doi.org/10.1006/jmla.1998.2564

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353-383. doi:10.1016/0010-0285(77)90012-3

Nickels, L., Biedermann, B., Coltheart, M., Saunders, S., & Tree, J. J. (2008). Computational modelling of phonological dyslexia: How does the DRC model fare? *Cognitive Neuropsychology, 25*(2), 165-193. doi:10.1080/02643290701514479

O'Shaughnessy, T. E., & Swanson, H. L. (2000). A Comparison of Two Reading Interventions for Children with Reading Disabilities. *Journal of learning disabilities, 33*(3), 257-277. doi:10.1177/002221940003300304

Patterson, K. E. (1978). Phonemic dyslexia: Errors of meaning and the meaning of errors. *Journal of Quarterly Journal of Experimental Psychology, 30*(4), 587-607. doi:10.1093/neucas/6.2.166-a

Patterson, K. E., & Marcel, A. J. (1977). Aphasia, dyslexia and the phonological coding of written words. *Quarterly Journal of Experimental Psychology, 29*(2), 307-318. doi:10.1080/14640747708400606

Peterson, R. L., Pennington, B. F., & Olson, R. K. (2013). Subtypes of developmental dyslexia: Testing the predictions of the dual-route and connectionist frameworks. *Cognition, 126*(1), 20-38. doi:10.1016/j.cognition.2012.08.007

Peterson, R. L., Pennington, B. F., Olson, R. K., & Wadsworth, S. J. (2014). Longitudinal Stability of Phonological and Surface Subtypes of Developmental Dyslexia. *Scientific Studies of Reading, 18*(5), 347-362. doi:10.1080/10888438.2014.904870

Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 139-156. doi:10.1037/0278-7393.27.1.139

Prado, C., Dubois, M., & Valdois, S. (2007). The eye movements of dyslexic children during reading and visual search: Impact of the visual attention span. *Vision Research, 47*(19), 2521-2530. doi:10.1016/j.visres.2007.06.001

Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A Computational Model of the Self-Teaching Hypothesis Based on the Dual-Route Cascaded Model of Reading. *Cognitive Science, 42*(3), 722-770. doi:doi:10.1111/cogs.12571

Pynte, J., Kennedy, A., & Ducrot, S. (2004). The influence of parafoveal typographical errors on eye movements in reading. *European Journal of Cognitive Psychology, 16*(1-2), 178-202. doi:10.1080/09541440340000169

R-Studio-Team. (2015). RStudio: Integrated Development for R . Boston, MA: RStudio, Inc. Retrieved from https://www.rstudio.com/products/rstudio

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology, 62*(8), 1457-1506. doi:10.1080/17470210902816461

Romani, C., Di Betta, A. M., Tsouknida, E., & Olson, A. (2008). Lexical and nonlexical processing in developmental dyslexia: A case for different resources and different impairments. *Cognitive Neuropsychology, 25*(6), 798-830. doi:10.1080/02643290802347183

Romani, C., Olson, A., Ward, J., & Ercolani, M. G. (2002). Formal lexical paragraphias in a single case study: how “masterpiece” can become “misterpieman” and “curiosity” “suretoy”. *Brain and Language, 83*(2), 300-334. doi:10.1016/S0093-934X(02)000226

Rowse, H. J. (2005). *Comparison of Effectiveness of Phonological and Whole-Word Treatment Programmes Within Two Dyslexia Subtypes.* (Doctoral thesis), Victoria University of Wellington, Wellington, New Zealand. Retrieved from http://researcharchive.vuw.ac.nz/handle/10063/591

Rowse, H. J., & Wilshire, C. E. (2007). Comparison of phonological and whole-word treatments for two contrasting cases of developmental dyslexia. *Cognitive Neuropsychology, 24*(8), 817-842. doi:10.1080/02643290701764207

Ruiz, A., Ansaldo, A. I., & Lecours, A. R. (1994). Two Cases of Deep Dyslexia in Unilingual Hispanophone Aphasics. *Brain and Language*, *46*(2), 245-256. doi:Lesioning an attractor network: Investigations of acquired dyslexia.

Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J.-F., Bricout, L., . . . Ziegler, J. C. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology*, *52*(10), 1503-1516. doi:10.1037/dev0000184

Savin, N. E., & White, K. J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica: Journal of the Econometric Society*, *45*(8), 1989-1996. doi:10.2307/1914122

Schneider, D., Mertes, C., & Wascher, E. (2016). The time course of visuo-spatial working memory updating revealed by a retro-cuing paradigm. *Scientific Reports, 6*. doi:10.1038/srep21442

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology*, *92*(2), 284-295. doi:10.1O37//0O22-O663.92.2.2B4

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568. doi:10.1037/0033-295X.96.4.523

Shallice, T., & Coughlan, A. K. (1980). Modality specific word comprehension deficits in deep dyslexia. *Journal of Neurology, Neurosurgery & Psychiatry, 43*(10), 866-872. doi:10.1136/jnnp.43.10.86

Shallice, T., & Warrington, E. K. (1975). Word recognition in a phonemic dyslexic patient. *Quarterly Journal of Experimental Psychology, 27*(2), 187-199. doi:10.1080/14640747508400479

Shallice, T., Warrington, E. K., & McCarthy, R. (1983). Reading without semantics. *The quarterly journal of experimental psychology, 35A*(1), 111-138. doi:10.1080/14640748308402120

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., . . . Shaywitz, B. A. (1999). Comprehension and Decoding: Patterns of Association in Children With Reading Difficulties. *Scientific Studies of Reading, 3*(1), 69-94. doi:10.1207/s1532799xssr0301_4

Share, D. L. (1995). Phonological recoding and self-teaching: *Sine quanon* of reading acquisition. *Journal of Cognition, 55*(2), 151-218. doi:10.1016/0010-0277(94)006452

Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological Psychiatry, 57*(11), 1301-1309. doi:10.1016/j.biopsych.2005.01.043

Shih, S.-I., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review, 109*(2), 260-305. doi:10.1037//0033-295X.109.2.260

Siegel, L. S. (1988). Evidence that IQ scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology, 42*(2), 201-215. doi:10.1037/h0084184

Snowling, M. J. (1995). Phonological processing and developmental dyslexia. *Journal of Research in Reading, 18*(2), 132-138. doi:10.1111/j.1467-9817.1995.tb00079.x

Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length Effect in Word Naming in Reading: Role of Reading Experience and Reading Deficit in Italian Readers. *Developmental neuropsychology, 27*(2), 217-235. doi:10.1207/s15326942dn2702_2

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*(4), 598-605. doi:10.3758/BF03193891

Stenneken, P., Egetemeir, J., Schulte-Korne, G., Muller, H. J., Schneider, W. X., & Finke, K. (2011). Slow perceptual processing at the core of developmental dyslexia: a parameter-based assessment of visual attention. *Neuropsychologia, 49*(12), 3454-3465. doi:10.1016/j.neuropsychologia.2011.08.021

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods, 42*(2), 497-506. doi:10.3758/brm.42.2.497

Suggate, S. P. (2016). A Meta-Analysis of the Long-Term Effects of Phonemic Awareness, Phonics, Fluency, and Reading Comprehension Interventions. *Journal of learning disabilities, 49*(1), 77-96. doi:10.1177/0022219414528540

Swanson, H. L., & Hoskyn, M. (1998). Experimental Intervention Research on Students with Learning Disabilities: A Meta-Analysis of Treatment Outcomes. *Review of Educational Research, 68*(3), 277-321. doi:10.3102/00346543068003277

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language, 9*(2), 182-198. doi:10.1016/0093-934X(80)90139-X

Team, R. J. R., Inc., Boston, MA URL http://www. rstudio. com. (2015). RStudio: integrated development for R. *42*, 14.

Temple, C. M., & Marshall, J. C. (1983). A case study of developmental phonological dyslexia. *British Journal of Psychology, 74*(4), 517-533. doi:10.1111/j.2044-8295.1983.tb01883.x

Tibi, S., & Kirby, J. R. (2017). Investigating Phonological Awareness and Naming Speed as Predictors of Reading in Arabic. *Scientific Studies of Reading*, 1-15. doi:10.1080/10888438.2017.1340948

Torgesen, J. K., Morgan, S. T., & Davis, C. J. J. o. E. p. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. *84*(3), 364-370. doi:10.1037/0022-0663.84.3.364

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals Dyslexia, 60*(1), 40-56. doi:10.1007/s11881-009-0032-y

Tree, J. J., & Kay, J. (2006). Phonological dyslexia and phonological impairment: An exception to the rule? *Neuropsychologia, 44*(14), 2861-2873. doi:10.1016/j.neuropsychologia.2006.06.006

Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia:

Evidence from two case studies. *Reading and Writing, 16*(6), 541-572. doi:10.1023/A:1025501406971

Valdois, S., Peyrin, C., Lassus-Sangosse, D., Lallier, M., Demonet, J. F., & Kandel, S. (2014). Dyslexia in a French-Spanish bilingual girl: behavioural and neural modulations following a visual attention span intervention. *Cortex, 53*, 120-145. doi:10.1016/j.cortex.2013.11.006

van den Boer, M., de Jong, P. F., & Haentjens-van Meeteren, M. M. (2013). Modeling the Length Effect: Specifying the Relation With Visual and Phonological Correlates of Reading. *Scientific Studies of Reading, 17*(4), 243-256. doi:10.1080/10888438.2012.683222

Wechsler, D. (2003). Wechsler Intelligence Scale for Children–Fourth Edition: Australian Standardised Edition (WISC-IV Australian). In. San Antonio, TX: Harcourt Assessment.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology, 50*(2), 439-456.

Wei, T., & Simko, V. (2017). R package “corrplot”: Visualization of a correlation matrix: R package version 0.73. Retrieved from https://github.com/taiyun/corrplot

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review, 8*(2), 221-243. doi:10.3758/BF03196158

Whitney, C. (2008). Supporting the serial in the SERIOL model AU - Whitney, Carol. *Language and Cognitive Processes, 23*(6), 824-865. doi:10.1080/01690960701828202

Whitney, C., & Cornelissen, P. (2005). Letter-position encoding and dyslexia. *Journal in Research in Reading, 28*(3), 274-301. doi:10.1111/j.1467-9817.2005.00270.x

Whitney, C., & Cornelissen, P. (2008). SERIOL Reading. *Language and Cognitive Processes, 23*(1), 143-164. doi:10.1080/01690960701579771

Whitney, C., & Marton, Y. (2013). The SERIOL2 Model of Orthographic Processing. In: Education Resources Information Center.

Wechsler, D. (2009). *Wechsler Individual Achievement Test (3rd ed.)*. San Antonio, TX: Pearson.

Wickham, H. (2017). Tidyverse: Easily install and load’tidyverse’packages. *R package version 1.1.1*.

Williams, J. P. (1980). Teaching decoding with an emphasis on phoneme analysis and phoneme blending. *Journal of Educational Psychology, 72*(1), 1-15. doi:10.1037/0022-0663.72.1.1

Williams, J. P. (1981). "Teaching Decoding With an Emphasis on Phoneme Analysis and Phoneme Blending": Correction to Williams. *Journal of Educational Psychology, 73*(1), 16. doi:10.1037/h0078092

Wilmer, J. B., Richardson, A. J., Chen, Y., & Stein, J. F. (2004). Two Visual Motion Processing Deficits in Developmental Dyslexia Associated with Different Reading Skills Deficits. *Journal of Cognitive Neuroscience, 16*(4), 528-540. doi:10.1162/089892904323057272

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, Retrieval Processes, and Reading: A Longitudinal Study in Average and Impaired Readers. *Child Development, 57*(4), 988-1000. doi:10.2307/1130373

Woodcock, R. W. (2011). Woodcock reading mastery tests: WRMT-III. In. Bloomington, Minnesota: Pearson.

Woodhead, Z. V. J., Penny, W., Barnes, G. R., Crewes, H., Wise, R. J. S., Price, C. J., & Leff, A. P. (2013). Reading therapy strengthens top–down connectivity in patients with pure alexia. *Brain, 136*(8), 2579-2591. doi:10.1093/brain/awt186

Woollams, A. M., Ralph, M. A. L., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review, 114*(2), 316-339. doi:10.1037/0033-295X.114.2.316

Wybrow, D. P., & Hanley, J. R. (2015). Surface developmental dyslexia is as prevalent as phonological dyslexia when appropriate control groups are employed. *Cognitive Neuropsychology, 32*(1), 1-13. doi:10.1080/02643294.2014.998185

Yanosky, D., Anthony, L., Jackson, K., Mirto, J., Crotts, J., & Bender, W. (2004). Test Review: Hammill, D.D., Mather, N., & Roberts, R. (2001). The Illinois Test of Psycholinguistic Abilities (3rd ed.) (ITPA-3). Austin, TX: PRO-ED. Assessment for Effective Intervention, 29(4), 71–76. https://doi.org/10.1177/073724770402900409

Zabell, C., & Everatt, J. (2002). Surface and phonological subtypes of adult developmental dyslexia. *Dyslexia, 8*(3), 160-177. doi:10.1002/dys.223

Ziegler, J. C., Pech-Georgel, C., Dufau, S., & Grainger, J. (2010). Rapid processing of letters, digits and symbols: what purely visual-attentional deficit in developmental dyslexia? *Developmental Science, 13*(4), 8-14. doi:10.1111/j.1467-7687.2010.00983.x

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology, 86*(3), 169-193. doi:10.1016/S0022-0965(03)00139-5

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language, 93*(3), 369-373. doi:10.1016/j.bandl.2004.10.010

Zoccolotti, P., De Luca, M., Lami, L., Pizzoli, C., Pontillo, M., & Spinelli, D. (2013). Multiple stimulus presentation yields larger deficits in children with developmental dyslexia: a study with reading and RAN-type tasks. *Child Neuropsychology, 19*(6), 639-647. doi:10.1080/09297049.2012.718325

Zoccolotti, P., De Luca, M., Pace, E. D., Judica, A., Orlandi, M., & Spinelli, D. (1999). Markers of developmental surface dyslexia in a language (Italian) with high grapheme-phoneme correspondence. *Applied Psycholinguistics*, *20*(2), 191-216.

Zoubrinetzky, R., Bielle, F., & Valdois, S. (2014). New Insights on Developmental Dyslexia Subtypes: Heterogeneity of Mixed Reading Profiles. *PLoS One, 9*(6). doi:10.1371/journal.pone.0099337

**Appendix A.** Heterogeneity and Screening study letter to principal (2017 Screening cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON

Emma Ashcroft
**PhD Student**
Email: *[email]*

Dr. Carolyn Wilshire
**Senior Lecturer, School of Psychology**
Email: *[email]*

**A comparison of two different cognitive interventions for school students with reading difficulties.**
**Ethics approval number: 0000022713**

Dear **Principal,**

My name is Emma Ashcroft and I am a PhD student working under the supervision of Dr Carolyn Wilshire in the School of Psychology at Victoria University. My thesis involves investigating dyslexia and developing suitable interventions to aid the development of reading skills. The aim of my research is to assess the efficacy of reading interventions specially designed to target different reading difficulties. I am particularly interested in trialling a new intervention for atypical "visual" forms of dyslexia, characterised by problems developing an efficient sight vocabulary, despite normal phonological skills. However, I will be trialling my methods on a range of individuals with different kinds of reading difficulties.

I am writing to enquire as to whether your school may be willing to participate in this research. I have chosen to write to you because [school name] has a reputation within the community as a centre with specialised resources and a proactive approach to addressing reading difficulties. Our target group of participants ranges from Year 3 to 10 students who are currently enrolled in remediation classes or otherwise identified as struggling readers. If you were to take part, your role would be simply to disseminate the information packs to students in these classes, so they can share them with their families. This could be done in paper form (in a sealed envelope) or electronically, whichever works best for you. The information packs would provide our contact details, and parents who are interested could then respond to us directly. I have attached a copy of the information sheet to parents for your viewing.

The study is organised into two parts. In the first part, all students in special reading classes willing to participate will complete some simple reading and cognitive tasks. This will allow us to create a brief profile of their reading difficulties. This part takes around one hour.

In the second part of the study, we will be asking 2-4 participants from the first part if they wish to take part in a reading intervention trial, which is designed to target two very different cognitive skills crucial for reading: a) phonological awareness or the ability to convert text to sounds, and b) visual attention. We will then contrast the effectiveness of these two approaches for the different participants. This intervention would consist of a series of weekly sessions over a period of approximately four months.

These sessions would be available for completion on school premise, with your permission, or at students' homes outside school hours. A secondary supervisor will be present throughout all sessions, this can either be another researcher or a teacher. The secondary supervisor will not partake in the research.

Full consent will be obtained from all students and their families before they participate in any part of this research. Parents will also be asked for their permission to give students movie vouchers as appreciation for their participation. This research has been approved by the School of Psychology Ethics Committee (approval number 00000 22713).

After the research is complete, I will also be available to visit your school and discuss the overall findings with any interested staff. If students and their families wish, I will also provide teachers with feedback on individual children which may be useful for setting future goals.

Thank you for your time. I will be in touch with you within the next fortnight to hear your views on this proposal, or please feel free to email me on the address provided above.

Yours sincerely,

Emma Ashcroft

**Appendix B.** Heterogeneity and Screening study parent information sheet (2017 Screening cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft | Dr. Carolyn Wilshire |
| **PhD Student** | **Senior Lecturer, School of Psychology** |
| Email: *[email]* | Email: *[email]* |


**A comparison of two different cognitive interventions for school students with reading difficulties.**

**Ethics approval number: 0000022713**

Dear **Parent,**

My name is Emma Ashcroft and I am a PhD student working under the supervision of Dr. Carolyn Wilshire in the School of Psychology at Victoria University. My PhD thesis involves investigating reading difficulties, with the aim of developing new interventions to improve students' reading skills.

I am therefore looking for school students who are currently in reading remediation classes or have been otherwise identified as struggling readers, and would be interested in participating in an assessment of their reading. If you agree, your child will be given several tasks to examine their reading and visual skills. We estimate the tasks will take no longer than two one hour sessions. Testing will take place at the school. Students are welcome to stop at any time. A digital audio recorder will be used during the reading task so that your child's responses can be transcribed later on.

Subsequent to this, we will be looking to recruit 2-4 students whose reading styles fit the profiles we are focusing on. This will consist of one-on-one sessions to target each student's reading problems. We will be inviting students to take part in the intervention phase of the research at a later date. You and your child do not need to decide whether or not you want to take part now.

Finally, the sessions are strictly confidential. Your child's name will never be released to anyone outside of the research team, and in written reports and publications of the research, they will be referred to only by a random combination of letters.

Attached is a more detailed information sheet and consent form. You are welcome to contact myself or my supervisor, Dr. Carolyn Wilshire, to find out more. Our details are given above. If you have any queries regarding ethics, please contact [Ethics administrator] ([email/ contact number]) who is chair of the Human Ethics Committee at Victoria University.

Yours Sincerely,

Emma Ashcroft

**Appendix C.** Heterogeneity and Screening study parent consent form (2017 Screening cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft | Dr. Carolyn Wilshire |
| **PhD Student** | **Senior Lecturer, School of Psychology** |
| Email: *[email]* | Email: *[email]* |


**A comparison of two different cognitive interventions for school students with reading difficulties.**
**Ethics approval number: 00000 22713**

**What is the purpose of this research?**
Reading difficulties are highly prevalent in primary and secondary school students, with some estimating up to 20% of children in New Zealand schools struggling to learn fluent reading skills. Given how important reading is for everyday communication and academic study, it is essential that intervention are developed for those who find reading particularly difficult. Importantly, these intervention need to target the specific needs of individuals'.

The goal of this study is to develop interventions which target different patterns of reading impairment.

**Who is conducting the research?**
Dr Carolyn Wilshire and PhD student Emma Ashcroft from the School of Psychology at Victoria University will be conducting the research. Emma will subsequently write a PhD thesis from the results of the research. This research has been approved under the delegated authority of the School of Psychology Ethics Committee (SoPHEC) at Victoria University.

**What is involved if you agree for your child to participate?**
If you and your child agree to participate in this study, Emma will visit your child's school complete a couple of sessions with them. Another researcher or teacher will be present in these sessions but they will not be actively involved. These sessions will last no longer than an hour each and will comprise of three tasks.

The first task will ask students' to mimic a series of visual patterns using plastics blocks. The second task will ask students' to read words and sentences aloud. The third task will ask students to manipulate word sounds, according to instructions from the experimenter. The latter two tasks will be recorded using a digital tape recorded and later transcribed. Students reading accuracy will then be used to calculate an estimated reading age.

During the research you are free to withdraw your child (or any information they have provided) at any time up until the session has been completed, without having to give any reasons and without any penalty. Approximately 2-4 students will subsequently be invited to take part in the intervention phase of this study. In the event that we are able to tailor an intervention to suit your child, please indicate whether you wish to be contacted regarding this phase of the study below.

**What happens to the information you provide?**
Your child's information will be kept confidential. All data files will be coded using a random selection of letters. Any files which identify you will only be accessible by the investigators. Data *without identifying names* may appear in journal publications, conference presentations and will form part of a thesis submitted for assessment. Your child's name will never be used in any presentation or publication.

**Thank you for your participation.**
As a token of our appreciation, we will offer each child a small reward. This is in no way dependent on their performance.

If you would like to know the results of this study they can be emailed, or sent to you in another capacity, if you wish. Individual student's performance data will not be available for this phase of the study, as no formal diagnostic inferences may be made from this research.

If you have any further questions regarding this study, please contact one of us above.

**Statement of Consent**

I have been given a full explanation of this project, and have had an opportunity to ask questions and have had them answered to my satisfaction. I understand that my child will be given a number of tasks to examine his or her reading and other related cognitive skills. Following this, my child may be invited to take part in further sessions designed to improve their reading skills, which they may also take part in if we so wish.

I understand that the data obtained may appear in graduate student reports, poster presentations, conference presentations and peer-reviewed publications, but that in all cases, my child will be referred to only by a random stream of letters. All paper work and electronical files will be kept by Dr. Carolyn Wilshire in a secure place. Summaries of the data will be kept by Emma Ashcroft and Dr Wilshire and may be shared with other members of the research team.

I understand that I may withdraw my child (or any information we have provided) at any time up until the session has been completed, without having to give reasons and without penalty of any sort.

I agree that ___________________________, who is under my guardianship, may take part in this research.

Name: ________________________________________________

Signature: ________________________________________________

Date: ________________________________________________

If you would like to receive a summary of the results of this study once it is completed, please tick here: ☐ (Please note, individual data will not be available).

If we are able to tailor the intervention tasks to your child's needs, would you like to be contacted regarding this phase of the research? If so please tick here: ☐

Copy to:

(a) Parent
(b) Researcher (initial both copies below)

**Appendix D.** Heterogeneity and Screening student info and assent form (2017 Screening cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA
UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft<br>**PhD Student** | Dr. Carolyn Wilshire<br>**Senior Lecturer** |


**Reading study**

**Ethics approval number: 00000 22713**

Dear **Student,**

My name is Emma Ashcroft and I am a PhD student at Victoria University. For my project I am looking at reading difficulties, and how I can help students improve.

I would like to invite you to take part in some reading and Lego-style tasks. This would take up two lessons. Another adult would also be present for this session, but they will not be involved. If you do decide to take part, you can change your mind and stop at any time, even during the session.

I will tape what you say in the tasks on an audio recorder and write them out later on. The details about what you say and do is private. After the study, I will work out how everyone performed as a group. I will send everyone a copy of the group results. No names will be mentioned in this report. I will also write my PhD thesis using the results added together. Your name will never mentioned.

After these sessions, I will be asking a small number of students (2-4) if they would like to take part in some new exercises designed to improve reading. If I do ask you to take part, you can decide later on whether you want to do this.

If you wish to take part, please sign the form below. If you have any questions, your parents can email me on the address at the top of the previous page.

Yours Sincerely,

Emma Ashcroft

**Statement of Assent**

I, ___________________________, agree to take part in this research.

Signed: ___________________________________

Name: ____________________________________

Date: _________________________

**Appendix E.** Heterogeneity study principal letter (2018 cohort)

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON

| | |
|---|---|
| Emma Ashcroft | Dr. Carolyn Wilshire |
| **PhD Student** | **Senior Lecturer, School of Psychology** |
| Email: *[email]* | Email: *[email]* |

**A comparison of two different cognitive interventions for school students with reading difficulties.**
**Emma Ashcroft**

Dear **Principal [name],**

My name is Emma Ashcroft and I am writing to you regarding a research project I am currently running. I am a PhD student working under the supervision of Dr Carolyn Wilshire in the School of Psychology at Victoria University. My thesis involves investigating dyslexia and other forms of reading difficulties. The aim of my research is to investigate the legitimacy of recent ideas and suggestions about what dyslexia looks like, and what the underlying causes of reading difficulties might be.

I am writing to enquire as to whether your school may want to participate in this research. I have chosen to write to you because [reason]. For these reasons, I think [school name] would fit in nicely. Specifically, my target group of participants are Year 9 to 10 students. For my research, I like to work with a range of different students to get a more complete idea of what is and isn't considered a reading difficulty. For that reason, I work with students both formally diagnosed with dyslexia and those who have not been diagnosed but identified by teachers as in need of reading remediation, as well as those who show no apparent signs of reading difficulties for comparison. If you were to take part, I would ask teachers to disseminate the information packs to students in these classes, so they can share them with their families. This could be done in paper form (in a sealed envelope) or electronically, whichever works best for you. The information packs would provide our contact details, and parents who are interested could then respond to us directly, or to your teachers, whichever is preferred. I have attached a copy of the information sheet to parents for your viewing.

The study is comprised of one-off, one-hour long sessions where I work with one student at a time, completing a few different tasks that assess their cognitive and reading styles. This will allow me to create a brief profile of their reading difficulties, if there are any. These sessions would be available for completion on school premise, with your permission, or at students' homes outside school hours. A secondary supervisor will be present throughout all sessions, this can either be another researcher, or a teacher if any are interested. The secondary supervisor will not partake in the research.

Full consent will be obtained from all students and their families before they participate in any part of this research. This research has been approved by the School of Psychology Ethics Committee (approval number 00000 22713).

After the research is complete, I will also be available to visit your school and discuss the overall findings with any interested staff. If students and their families wish, I will also provide teachers with feedback on individual children which may be useful for setting future goals.

Thank you for your time. I will be in touch with you within the next fortnight to hear your views on this proposal, or please feel free to email me on the address provided above.

Yours sincerely,

Emma Ashcroft
BSc/Ba, PGDip Cognitive and Behavioural neuroscience
PhD Candidate
School of Psychology | Te Kura Mātai Hinengaro
Victoria University of Wellington | Te Whare Wānanga o Te Ūpoko o te Ika a Māui

**Appendix F.** Heterogeneity study parent information sheet (2018 cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft<br>**PhD Student**<br>Email: *[email]* | Dr. Carolyn Wilshire<br>**Senior Lecturer, School of Psychology**<br>Email: *[email]* |


**A comparison of two different cognitive interventions for school students with reading difficulties.**

**Ethics approval number: 0000022713**

Dear **Parent,**

My name is Emma Ashcroft and I am a Doctoral student working under the supervision of Dr. Carolyn Wilshire in the School of Psychology at Victoria University. My PhD thesis involves investigating reading difficulties, with the aim of understanding and developing new interventions to improve students' reading skills.

So far in my research I have been working with students enrolled in reading remediation classes to look at the potential benefits of new intervention techniques. Now I am looking to gain a greater understanding of what is happening in the mind when we try to read. For this I am looking for students across secondary schools to take part in a one-off session looking at their reading skills. For this study anyone may take part. The sessions will be one-on-one and last no longer than one hour during school time. Students are welcome to stop at any time. A digital audio recorder will be used during the reading task so that your child's responses can be transcribed later on.

Finally, the sessions are strictly confidential. Your child's name will never be released to anyone outside of the research team, and in written reports and publications of the research, all data will be combined so that no individual will be discussed separately. The findings may also be used for teaching purposes in secondary school and university psychology classes however no names or other identifying information will be used. Should the study identify potential reading difficulties in your child they will be asked for permission to inform their teacher.

Attached is a more detailed information sheet and consent form. You are welcome to contact myself or my supervisor, Dr. Carolyn Wilshire, to find out more. Our details are given above. If you have any queries regarding ethics, please contact [Ethics administrator] ([email/ contact number]) who is chair of the Human Ethics Committee at Victoria University.

Yours Sincerely,
Emma Ashcroft

**Appendix G.** Heterogeneity study parent consent form (2018 cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON

Emma Ashcroft
**PhD Student**
Email: *[email]*

Dr. Carolyn Wilshire
**Senior Lecturer, School of Psychology**
Email: *[email]*

**A comparison of two different cognitive interventions for school students with reading difficulties.**
**Ethics approval number: 00000 22713**

**What is the purpose of this research?**
Reading difficulties are highly prevalent in primary and secondary school students, with some estimating up to 20% of children in New Zealand schools struggle to learn fluent reading skills. Given how important reading is for everyday communication and academic study, it is essential that the process of reading and reading difficulties are well understood. The aim of this research is to look at the mental processes involved in reading and how they affect one another.

**Who is conducting the research?**
Dr Carolyn Wilshire and PhD student Emma Ashcroft from the School of Psychology at Victoria University will be conducting the research. Emma will subsequently write a PhD thesis from the results of the research. This research has been approved under the delegated authority of the School of Psychology Ethics Committee (SoPHEC) at Victoria University.

**What is involved if you agree for your child to participate?**
If you and your child agree to participate in this study, Emma will visit your child's school complete a one-on-one session with them. Another researcher or teacher will be present in these sessions but they will not be actively involved. These sessions will last no longer than an hour each and will comprise of several short tasks. In these tasks students will be asked to read words aloud, manipulate word sounds, and identify images on screen for a very short amount of time. The tasks will be recorded using a digital tape recorded and later transcribed.

During the research you are free to withdraw your child (or any information they have provided) at any time up until the session has been completed, without having to give any reasons and without any penalty.

**What happens to the information you provide?**
Your child's information will be kept confidential. All data files will be coded using a random selection of letters. Any files which identify you will only be accessible by the investigators. Data *without identifying names* may appear in journal publications, conference presentations

and will form part of a thesis submitted for assessment. The findings may also be used for teaching purposes in secondary school and university psychology classes however no names or other identifying information will be used. Should the study identify potential reading difficulties in your child they will be asked for permission to inform their teacher.

**Thank you for your participation.**
As a token of our appreciation, we will offer each child a small reward. This is in no way dependent on their performance.

If you would like to know the results of this study they can be emailed, or sent to you in another capacity, if you wish. Individual student's performance data will not be available for this phase of the study, as no formal diagnostic inferences may be made from this research.

If you have any further questions regarding this study, please contact one of us above.

You can return a physical copy of the consent forms to your child's teacher, or you may fill it in electronically and email it directly to Emma.

**Statement of Consent**

I have been given a full explanation of this project, and have had an opportunity to ask questions and have had them answered to my satisfaction. I understand that my child will be given a number of tasks to examine their reading skills.

I understand that the data obtained may appear in graduate student reports, poster presentations, conference presentations and peer-reviewed publications, but that in all cases, no child will be identified. All paper work and electronical files will be kept by Dr. Carolyn Wilshire in a secure place. Summaries of the data will be kept by Emma Ashcroft and Dr Wilshire and may be shared with other members of the research team.

I understand that I may withdraw my child (or any information we have provided) at any time up until the session has been completed, without having to give reasons and without penalty of any sort.

I agree that ___________________________, who is under my guardianship, may take part in this research.

Name: __________________________________________________

Signature: __________________________________________________

Date: __________________________________________________

If you would like to receive a summary of the results of this study once it is completed, please tick here (please include email address below): ☐ (Please note, individual data will not be available).

Email: __________________________________________________

Copy to:

(c) Parent
(d) Researcher (initial both copies below)

**Appendix H.** Heterogeneity study student information and assent form (2018 cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft | Dr. Carolyn Wilshire |
| **PhD Student** | **Senior Lecturer** |


**Reading study**

**Ethics approval number: 00000 22713**

Dear **Student,**

My name is Emma Ashcroft and I am a PhD student at Victoria University. For my project I am looking at how the mind reads and what happens when someone struggles with reading.

I would like to invite you to take part in some reading tasks that would take up one lesson.

Another adult would also be present for this session, but they will not be involved. If you do decide to take part, you can change your mind and stop at any time, even during the session.

I will tape what you say in the tasks on an audio recorder and write them out later on. The details about what you say and do is private. You will also be asked if you want Emma to tell your teacher if it might be good for you to have some extra attention or help with reading.

After the study I will work out how everyone performed as a group and write my PhD thesis using the results added together. Your name will never be mentioned. The group results may also be used in a psychology class at your school, however no one’s name will ever be used.

If you wish to take part, please sign the form below. If you have any questions, your parents can email me at any stage.

Yours Sincerely,

Emma Ashcroft

**Statement of Assent**

I, ___________________________(first name), agree to take part in this research.

Signed: ___________________________________

First and last name: ___________________________________

Date: ____________________________________

If the reading tasks show you might benefit from some extra attention or help with reading do you want Emma to tell your teacher? (please tick one)

☐ **Yes**

☐ **No**

**Appendix I.** Additional demographic information about participants in the Heterogeneity study.

| Participant | Gender | Class | Handedness | Eyesight |
|---|---|---|---|---|
| 22 | Male | Reading Remediation | Right | Normal |
| 23 | Male | Reading Remediation | Left | Normal |
| 24 | Male | Reading Remediation | Left | Normal |
| 25 | Male | Reading Remediation | Right | Glasses |
| 28 | Male | Reading Remediation | Right | Normal |
| 29 | Male | Reading Remediation | Right | Unclear |
| 31 | Male | Reading Remediation | Right | Normal |
| 32 | Male | Reading Remediation | Left | Unclear |
| 33 | Male | Reading Remediation | Right | Glasses |
| 34 | Female | Reading Remediation | Right | Unclear |
| 35 | Female | Reading Remediation | Right | Normal |
| 37 | Male | Reading Remediation | Right | Normal |
| 38 | Female | Reading Remediation | Right | Normal |
| 39 | Male | Reading Remediation | Right | Unclear |
| 40 | Male | Reading Remediation | Right | Glasses |
| 41 | Female | Reading Remediation | Right | Normal |
| 43 | Female | Reading Remediation | Right | Normal |
| 44 | Female | Reading Remediation | Right | Unclear |
| 45 | Male | Reading Remediation | Right | Glasses |

| | | | | |
|---|---|---|---|---|
| 46 | Female | Reading Remediation | Right | Glasses |
| 47 | Male | Reading Remediation | Right | Normal |
| 48 | Male | Reading Remediation | Right | Unclear |
| 49 | Male | Reading Remediation | Right | Normal |
| 50 | Female | Reading Remediation | Left | Normal |
| 51 | Male | Reading Remediation | Right | Normal |
| 52 | Female | Reading Remediation | Right | Normal |
| 54 | Male | Reading Remediation | Right | Normal |
| 55 | Male | Reading Remediation | Right | Normal |
| 57 | Male | Reading Remediation | Right | Normal |
| 58 | Female | Reading Remediation | Right | Normal |
| 59 | Female | Reading Remediation | Left | Normal |
| 60 | Female | Reading Remediation | Right | Normal |
| 61 | Female | Reading Remediation | Right | Normal |
| 62 | Male | Reading Remediation | Right | Normal |
| 63 | Female | Reading Remediation | Right | Unclear |
| 64 | Male | Reading Remediation | Left | Normal |
| 65 | Female | Reading Remediation | Right | Normal |
| 66 | Male | Non-Reading Remediation | Right | Unclear |
| 67 | Male | Non-Reading Remediation | Right | Normal |
| 68 | Male | Non-Reading Remediation | Right | Unclear |
| 69 | Female | Non-Reading Remediation | Right | Normal |

| | | | | |
|---|---|---|---|---|
| 71 | Female | Non-Reading Remediation | Right | Normal |
| 72 | Female | Non-Reading Remediation | Right | Normal |
| 73 | Male | Non-Reading Remediation | Right | Normal |
| 74 | Male | Non-Reading Remediation | Right | Normal |
| 75 | Male | Non-Reading Remediation | Right | Normal |
| 76 | Female | Non-Reading Remediation | Right | Normal |
| 77 | Male | Non-Reading Remediation | Right | Normal |
| 78 | Male | Non-Reading Remediation | Right | Normal |

**Appendix J.** Test reliability and validity information

Coltheart and Leahy item list (Coltheart and Leahy, 1996). Control data from X participants (age range: X) was obtained and used to calculated age-equivalent brackets. Data is comparable to previous word lists by Castles and Coltheart (1993) which has been tested for, and shown to have, good reliability and validity (Edwards & Hogben, 1999).

Homophone Selection test (Pexman et al, 2001; control data from Castles and Coltheart, 1996). Control data from 10 male participants was collected (gender and age-matched to dyslexic case-study). For Regular/Regular words: *M*=52.9 (*SD*=3.51); Regular/Irregular words: (with the Regular word as the target) *M*= 27.0 (*SD*= 2.45), (with the irregular word as the target) *M*= 25.1 (*SD*= 1.91); and Irregular/nonwords: *M*= 27.4 (*SD*= 1.95).

Orthographic Choice task (Manis et al, 1996). Control data from 51 participants (gender and age matched to dyslexic sample, ranging from 9-to-15-years) was collected, with averaged score as *M*= 90.0% (*SD*= 4.9%).

Partial Report task (Bosse et al, 2007). Control data from 55 participants (gender and age matched to dyslexia sample, ranging from 9-to-13-years) was collected, with average score as *M*= 87.0% (*SD*= 0.08%). Data from another sample of 14 controls was concurrent with this (M= 90.2%, SD= 0.05%; Prado et al, 2007).

Phonemic Deletion task (Illinois Test of Psycholinguistic Abilities: 3rd Edition; Hammil et al, 2001). Age-equivalent scores for the phonemic deletion task are determined from raw scores. Reliability and validity data are presented within the manual. It has been found that the tasks included in the Illinois Test battery, including the Phonemic Deletion task, show good validity and low error rate for measuring of spoken and written language (Yanosky et al, 2001).

Phoneme Identification task and Rhyme Detection task (Muter et al, 1998). Longitudinal data from 38 control participants was obtained at three time points. Phoneme Identification was scored out of 8; Rhyme Detection was scored out of 10. At time 1 (Age: *M*= 4 years, 3 months), Phoneme Identification: *M*= 2.0 (*SD*= 2.5); and Rhyme Detection: *M*= 5.4 (*SD*= 2.9). At time 2 (Age: *M*= 5 years, 3 months), Phoneme Identification: *M*= 3.4 (*SD*= 3.2); and Rhyme Detection: *M*= 8.1 (*SD*= 2.4). At time 3 (Age: *M*= 6 years, 3 months), Phoneme Identification: *M*= 6.3 (*SD*= 2.6); and Rhyme Detection: *M*= 8.3 (*SD*= 2.5).

Phoneme Segmentation task (McBride-Chang, 1995). Control data from 136 participants (Age: *M*= 12 years, 3 months) was collected with an average score of *M*= 13.9 (*SD*= 5.3).

Woodcock Reading Mastery test battery (Woodcock, 2011). Standardised and age-equivalent scores for the Word Identification, Oral Fluency and Comprehension tasks are determined from raw scores. Reliability and validity are presented in the manual.

**Appendix K.** Coltheart and Leahy task wordlist.

| Regular words | Irregular words | Non-words |
|---|---|---|
| Bed | blood | aspy |
| Brandy | bouquet | baft |
| Chance | bowl | bick |
| Check | break | bleaner |
| Chicken | brooch | boril |
| Context | ceiling | borp |
| Cord | choir | brennet |
| curb | colonel | brinth |
| drop | come | crat |
| flannel | cough | delk |
| free | eye | doash |
| hand | friend | drick |
| life | gauge | farl |
| long | give | framp |
| luck | good | ganten |
| market | head | gop |
| marsh | iron | grenty |
| middle | island | gurve |
| mist | lose | hest |
| navy | meringue | norf |
| need | pint | peef |
| nerve | pretty | peng |
| peril | routine | pite |
| plant | shoe | pofe |
| pump | soul | rint |
| stench | sure | seldent |
| tail | tomb | spatch |
| take | wolf | stendle |
| weasel | work | tapple |
| wedding | yacht | trope |

**Appendix L.** Linguistic Properties Effect task stimuli organised according to length.

| Word | Raw Frequency | Freq group | Raw Imageability | Image group | AoA[a] | Bigram | Ortho_n |
|---|---|---|---|---|---|---|---|
| | | | 3 letters | | | | |
| ail | 0.76 | Low | 391 | Low | 9.71 | 2387.5 | 12 |
| aim | 1.18 | Low | 383 | Low | 6.72 | 1719 | 11 |
| arm | 65.41 | High | 593 | High | 3.26 | 3549 | 7 |
| axe | 4.88 | Low | 597 | High | 6.11 | 213 | 12 |
| bad | 545.18 | High | 388 | Low | 2.79 | 1876 | 23 |
| bin | 5.37 | Low | 562 | High | 4.68 | 7989.5 | 21 |
| buy | 192.43 | High | 397 | Low | 5.56 | 577.5 | 10 |
| ear | 32 | High | 597 | High | 3.63 | 4573 | 12 |
| fly | 85 | High | 582 | High | 3.05 | 2417.5 | 7 |
| gem | 2.47 | Low | 572 | High | 7.68 | 2130 | 9 |
| ice | 79.55 | High | 635 | High | 3.86 | 3574.5 | 6 |
| lie | 120.25 | High | 385 | Low | 3.75 | 4965 | 15 |
| lye | 0.61 | Low | 372 | Low | 12.67 | 2132.5 | 8 |
| nor | 35.98 | High | 243 | Low | 8.61 | 3418 | 13 |
| ode | 0.02 | Low | 344 | Low | 10.84 | 2988 | 6 |
| pea | 3.9 | Low | 568 | High | 3.95 | 3159 | 14 |
| pie | 28.75 | High | 604 | High | 3.67 | 2988 | 15 |
| pig | LF | Low | 635 | High | 3.84 | 1724.5 | 16 |
| wad | 1.75 | Low | 370 | Low | 8.78 | 1631.5 | 19 |
| yet | 341.73 | High | 237 | Low | 6.96 | 1684 | 15 |
| | | | 4 Letters | | | | |
| arch | 3.69 | Low | 604 | High | 8.26 | 3409 | 3 |
| atom | 2.75 | Low | 390 | Low | 9.68 | 4447.67 | 1 |
| beak | 2.1 | Low | 579 | High | 5.42 | 1904 | 12 |
| beef | 19.71 | High | 625 | High | 6.58 | 1540.67 | 6 |
| calf | 2.96 | Low | 559 | High | 6.63 | 3319.33 | 5 |
| data | 25.61 | High | 375 | Low | 10.42 | 4337.67 | 3 |
| desk | 43.9 | High | 614 | High | 5.56 | 6044 | 3 |
| ease | 19.1 | High | 302 | Low | 9.11 | 3399.67 | 9 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| foam | 3.51 | Low | 583 | High | 6.15 | 1301.33 | 4 |
| glad | 171.37 | High | 347 | Low | 5.79 | 2423.33 | 3 |
| grin | 2.71 | Low | 567 | High | 5.84 | 7415.33 | 9 |
| isle | 1.22 | Low | 390 | Low | 9.65 | 4430.67 | 3 |
| noon | 18.12 | High | 317 | Low | 5.89 | 3951.67 | 9 |
| norm | 3.33 | Low | 142 | Low | 9.84 | 2640.67 | 6 |
| pact | 3.76 | Low | 364 | Low | 9.37 | 2287.33 | 8 |
| pipe | 19.39 | High | 617 | High | 8.17 | 2016.67 | 7 |
| post | 32.43 | High | 607 | High | 8.11 | 3874.33 | 14 |
| tank | 25.61 | High | 572 | High | 7.17 | 3853.33 | 13 |
| toll | 3.35 | Low | 302 | Low | 8.37 | 3159 | 14 |
| tour | 30.8 | High | 354 | Low | 7.17 | 3101.67 | 9 |
| 5 Letters | | | | | | | |
| blame | 58.78 | High | 352 | Low | 6.48 | 2929.25 | 8 |
| blend | 4.27 | Low | 315 | Low | 10.17 | 4791.75 | 5 |
| brave | 31.71 | High | 329 | Low | 6 | 2881.5 | 7 |
| broom | 4.76 | Low | 595 | High | 5.5 | 2435.25 | 4 |
| cello | 1.86 | Low | 613 | High | 9.89 | 3319.25 | 4 |
| curry | 2.43 | Low | 587 | High | 13.28 | 1567.5 | 4 |
| flank | 4.63 | Low | 275 | Low | 12.15 | 3285.5 | 7 |
| folly | 1.45 | Low | 326 | Low | 10.67 | 2910 | 10 |
| pause | 5.39 | Low | 359 | Low | 6.75 | 2580.75 | 3 |
| plane | 95.53 | High | 556 | High | 4.95 | 4529 | 6 |
| prime | 18.29 | High | 209 | Low | 8.83 | 3425.75 | 7 |
| prove | 70.39 | High | 221 | Low | 7.84 | 2815 | 8 |
| purse | 19.76 | High | 640 | High | 5.53 | 2874.75 | 8 |
| relay | 2.98 | Low | 319 | Low | 9.15 | 4270.5 | 4 |
| river | 55.47 | High | 608 | High | 4.9 | 6078.25 | 13 |
| slave | 18.43 | High | 580 | High | 7.84 | 2410.5 | 7 |
| spear | 4.55 | Low | 557 | High | 7.22 | 3472.25 | 5 |
| trace | 19.39 | High | 239 | Low | 6.84 | 3724.75 | 7 |
| wheat | 5.75 | Low | 577 | High | 6.53 | 3837 | 1 |
| wound | 26.53 | High | 570 | High | 7.52 | 2684.75 | 8 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | 6 Letters | | | | |
| battle | LF | Low | 597 | High | 6.95 | 3811 | 6 |
| beagle | 1.25 | Low | 362 | Low | 7.39 | 2848.2 | 1 |
| bullet | 38.24 | High | 608 | High | 6.7 | 3372 | 6 |
| cannon | 1.65 | Low | 615 | High | 7.9 | 4405 | 4 |
| carbon | 5.24 | Low | 390 | Low | 10.55 | 4081.8 | 3 |
| decent | 28.1 | High | 246 | Low | 7.47 | 4507.4 | 4 |
| dinner | 202.67 | High | 590 | High | 3.99 | 7524.4 | 5 |
| former | 18.27 | High | 196 | Low | 8.45 | 4780 | 5 |
| funnel | 1.1 | Low | HI | High | 8.44 | 2758.4 | 3 |
| gravel | 1.43 | Low | 563 | High | 7.19 | 3088.2 | 5 |
| jacket | 33.41 | High | 632 | High | 3.95 | 1992.4 | 2 |
| manner | 11.53 | High | 342 | Low | 6.83 | 5999.4 | 5 |
| motion | 18.96 | High | 386 | Low | 6.78 | 5112.4 | 3 |
| parish | 3.02 | Low | 396 | Low | 10.79 | 4517 | 4 |
| pastry | 1.92 | Low | 593 | High | 7.21 | 3420.6 | 2 |
| picket | 2.22 | Low | 303 | Low | 10.79 | 2618.8 | 6 |
| simple | 89.31 | High | 195 | Low | 6.89 | 2895 | 6 |
| strife | 1.67 | Low | 234 | Low | 11.78 | 3819.2 | 5 |
| summer | 78.67 | High | 618 | High | 4.33 | 4049.4 | 7 |
| wallet | 22.8 | High | 605 | High | 5.89 | 4129.4 | 5 |
| | | | 7 Letters | | | | |
| alcohol | 16.57 | Low | 598 | High | 9 | 2670.5 | 0 |
| analogy | 1.14 | Low | 267 | Low | 11.33 | 3391.67 | 0 |
| concept | 10.84 | High | 258 | Low | 11.44 | 3339.83 | 2 |
| contest | 18.78 | High | 278 | Low | 7.44 | 7852.17 | 3 |
| curious | 26.22 | High | 254 | Low | 6.64 | 3523.5 | 2 |
| darling | 129.67 | High | 270 | Low | 5.79 | 6149 | 2 |
| feather | 6.63 | Low | HI | High | 4.67 | 5276.67 | 3 |
| Measles | 2.1 | Low | 559 | High | 9.28 | 5005.67 | 0 |
| miracle | LF | Low | 367 | Low | 6.8 | 3477.5 | 0 |
| picture | 138.45 | High | 581 | High | 4.05 | 3439 | 0 |
| premium | 1.73 | Low | 205 | Low | 9.95 | 2836.67 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| quality | 18.57 | High | 349 | Low | 8.78 | 3150.33 | 1 |
| theatre | 24.04 | High | 566 | High | 6.94 | 6519.33 | 1 |
| tobacco | 6.98 | Low | 601 | High | 7.39 | 2211.33 | 0 |
| tractor | 3.73 | Low | 585 | High | 5.5 | 3671.17 | 1 |
| traffic | 28.51 | High | 599 | High | 6.22 | 2793.5 | 0 |
| treason | 5.08 | Low | 363 | Low | 11.37 | 4609.17 | 0 |
| tribute | 5.24 | Low | 386 | Low | 10.44 | 3634 | 1 |
| vehicle | 22.61 | High | 619 | High | 6.58 | 3031.17 | 1 |
| village | 33.57 | High | 578 | High | 7.84 | 2770.83 | 2 |
| 8 Letters | | | | | | | |
| attitude | 26.08 | High | 321 | Low | 6.78 | 4313 | 2 |
| basement | 21.06 | High | 571 | High | 6.74 | 3810.57 | 2 |
| beverage | 2.33 | Low | 565 | High | 6.63 | 4189.71 | 1 |
| confused | 32.41 | High | 262 | Low | 8.02 | 4219.71 | 3 |
| darkness | 17.49 | High | 622 | High | 4.83 | 4232.29 | 1 |
| knitting | 2.25 | Low | 578 | High | 7.24 | 5951.29 | 1 |
| marriage | 77.06 | High | 556 | High | 7.17 | 3289.57 | 1 |
| mountain | 35.39 | High | 629 | High | 6.15 | 4869.14 | 1 |
| national | 38.16 | High | 400 | Low | 6.89 | 5925.71 | 2 |
| outbreak | 1.94 | Low | 359 | Low | 10.5 | 2698.43 | 0 |
| pleasant | 21.02 | High | 390 | Low | 7.67 | 4065.29 | 1 |
| possible | 114.04 | High | 249 | Low | 7.32 | 2947.71 | 1 |
| quantity | 1.86 | Low | 349 | Low | 11.11 | 4164.43 | 1 |
| rational | 5.02 | Low | 298 | Low | 10.06 | 6386.29 | 1 |
| sapphire | 1.2 | Low | 560 | High | 9.22 | 2527.71 | 0 |
| scorpion | 2.67 | Low | 596 | High | 7.84 | 3788.14 | 1 |
| shoulder | 26.2 | High | 577 | High | 4.5 | 4228.86 | 1 |
| swimming | 19.98 | High | 635 | High | 4.58 | 4281.29 | 3 |
| syllable | 0.82 | Low | 375 | Low | 8.1 | 2788.14 | 0 |
| wasteful | 0.59 | Low | 373 | Low | 7.11 | 3484.86 | 1 |
| woodland | 1.12 | Low | 608 | High | 7.21 | 2742.86 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 9 Letters | | | | | | | |
| avalanche | 3.39 | Low | 596 | High | 9.6 | 3578.63 | 0 |
| character | 38.16 | High | 372 | Low | 6.47 | 5538 | 0 |
| deduction | 1.71 | Low | 316 | Low | 10.89 | 4770.75 | 2 |
| destroyed | 30.84 | High | 360 | Low | | 5194.63 | 1 |
| expensive | 27.94 | High | 160 | Low | 6.33 | 2996.75 | 2 |
| footballs | 37.53 | High | 597 | High | 4.84 | 2364.71 | 1 |
| franchise | 2.37 | Low | 309 | Low | 12.58 | 3981.5 | 1 |
| gentleman | 41.86 | High | 559 | High | 6.89 | 4388.25 | 1 |
| hailstone | 0.04 | Low | 562 | High | 10.14 | 4034.75 | 0 |
| impressed | 19.06 | High | 299 | Low | | 5331.63 | 2 |
| landscape | 2.49 | Low | 608 | High | 9.89 | 3302.38 | 0 |
| plausible | 1.63 | Low | 278 | Low | 12.79 | 2811.13 | 1 |
| principle | 7.75 | High | 305 | Low | 9.68 | 4576.25 | 0 |
| professor | 69.57 | High | 587 | High | 10.89 | 3947.63 | 1 |
| projector | 1.22 | Low | 584 | High | 7.16 | 2417.38 | 1 |
| raspberry | 1.88 | Low | 636 | High | 5.33 | 3491.5 | 0 |
| receptive | 1.18 | Low | 302 | Low | 12.26 | 3692.38 | 1 |
| talkative | 0.9 | Low | 400 | Low | 8 | 4076.38 | 0 |
| telephone | 32.37 | High | 655 | High | 4.63 | 4643.88 | 1 |

Note: a) AoA= age of acquisition.

**Appendix M.** The Partial Report task letter strings (target letter in each string shaded).

| Position 1 | Position 2 | Position 3 | Position 4 | Position 5 |
|---|---|---|---|---|
| T | **M** | P | F | B |
| P | D | S | T | **B** |
| S | D | **T** | P | B |
| B | P | **S** | T | D |
| **P** | B | D | S | T |
| B | P | S | **M** | H |
| **D** | S | R | H | P |
| F | R | **P** | B | H |
| **T** | B | S | P | F |
| F | S | H | B | **R** |
| B | L | M | F | **H** |
| T | **F** | B | R | S |
| D | **S** | F | M | H |
| D | F | **L** | M | R |
| R | L | F | **T** | D |
| **H** | R | B | M | P |
| R | M | P | **L** | S |
| M | T | **R** | F | H |
| S | **B** | T | L | P |
| **B** | T | M | L | R |
| **F** | M | D | R | L |
| R | P | B | S | **D** |
| **L** | T | B | D | M |
| L | M | T | **S** | F |
| M | **D** | R | B | T |
| S | D | R | M | **P** |
| P | **H** | L | R | M |
| H | R | L | **F** | D |
| T | **R** | P | D | S |
| R | D | **B** | L | F |
| D | B | M | P | **F** |

| | | | | |
|---|---|---|---|---|
| P | T | H | D | B |
| P | F | S | B | M |
| M | S | D | H | L |
| S | F | P | H | L |
| H | F | M | D | S |
| T | H | F | R | L |
| M | P | R | L | F |
| F | L | M | R | D |
| L | B | H | P | M |
| D | H | F | T | L |
| M | L | D | H | T |
| L | S | T | P | M |
| S | H | L | D | T |
| B | L | T | H | R |
| H | M | L | F | R |
| R | H | F | S | T |
| F | P | H | T | S |
| L | R | H | B | P |
| H | T | D | S | B |

**Appendix N.** Word length regression analysis results for each individual, including proportion of variance explained by the model – R, Adjusted $R^2$ (in brackets), upper and lower 95% confidence intervals (CI) and the ANOVA statistic.

| ID[a] | R ($AR^2$) | 95% CI | ANOVA |
|---|---|---|---|
| 22 | .000 (-.007) | -0.011, 0.012 | *F*(1,137)= .01, *p*=.939 |
| 23[#] | .036 (.028) | 0.001, 0.018 | *F*(1,123)= 4.58, *p*=.034* |
| 24[#] | .001 (-.007) | -0.008, 0.011 | *F*(1,127)= .01, *p*=.757 |
| 25[#] | .211 (.203) | 0.024, 0.055 | *F*(1,98)= 26.20, *p*< .001* |
| 28[#] | .003 (-.005) | -0.007, 0.013 | *F*(1,123)= .37, *p*=.547 |
| 29[#] | .133 (.122) | 0.014, 0.050 | *F*(1,80)= 12.24, *p*=.001* |
| 31[#] | .091 (.080) | 0.008, 0.046 | *F*(1,82)= 8.25, *p*=.005* |
| 32[#] | .034 (.026) | 0.000, 0.027 | *F*(1,115)= 4.08, *p*=.046 |
| 33[#] | .117 (.110) | 0.013, 0.041 | *F*(1,114)= 15.18, *p*< .001* |
| 34[#] | .028 (.018) | -0.002, 0.027 | *F*(1,103)= 2.95, *p*=.089 |
| 35[#] | .126 (.119) | 0.019, 0.053 | *F*(1,119)= 17.20, *p*< .001* |
| 37 | .008 (.001) | -0.005, 0.017 | *F*(1,135)= 1.08, *p*=.301 |
| 38[#] | .185 (.166) | 0.015, 0.068 | *F*(1,43)= 9.75, *p*=.003* |
| 39[#] | .143 (.136) | 0.023, 0.057 | *F*(1,125)= 20.80, *p*< .001* |
| 40[#] | .002 (-.006) | -0.007, 0.013 | *F*(1,128)= .29, *p*=.594 |
| 41[#] | .029 (.021) | -0.001, 0.0027 | *F*(1,120)= 3.63, *p*=.059 |
| 43[#] | .055 (.038) | -0.002, 0.038 | *F*(1,56)= 3.25, *p*=.077 |
| 44 | .042 (.035) | 0.003, 0.031 | *F*(1,134)= 5.86, *p*=.017* |
| 45[#] | .110 (.102) | 0.007, 0.023 | *F*(1,121)= 14.91, *p*< .001* |
| 46[#] | .059 (.051) | 0.005, 0.035 | *F*(1,114)= 7.19, *p*=.008* |
| 47[#] | .029 (.021) | 0.000, 0.029 | *F*(1,124)= 3.75, *p*=.055 |
| 48[#] | .000 (-.008) | -0.012, 0.011 | *F*(1,123)= .00, *p*=.956 |
| 49[#] | .101 (.093) | 0.010, 0.036 | *F*(1,114)= 12.76, *p*=.001* |
| 50[#] | .012 (.004) | -0.005, 0.023 | *F*(1,122)= 1.50, *p*=.224 |
| 51[#] | .003 (-.004) | -0.013, 0.006 | *F*(1,136)= .46, *p*=.497 |
| 52[#] | .002 (-.008) | -0.012, 0.019 | *F*(1,94)= .21, *p*=.649 |
| 54[#] | .163 (.146) | 0.011, 0.053 | *F*(1,50)= 9.75, *p*=.003* |
| 55[#] | .031 (.022) | -0.001, 0.031 | *F*(1,107)= 3.43, *p*=.067 |
| 57 | .114 (.107) | 0.012, 0.038 | *F*(1,119)= 15.38, *p*< .001* |

| | | | |
|---|---|---|---|
| 58[#] | .000 (-.008) | -0.009, 0.011 | $F(1,130)= .02$, $p=.885$ |
| 59[#] | .085 (.076) | 0.009, 0.042 | $F(1,104)= 9.63$, $p=.002$* |
| 60 | .039 (.032) | 0.001, 0.016 | $F(1,125)= 5.14$, $p=.025$* |
| 61[#] | .144 (.137) | 0.016, 0.040 | $F(1,122)= 20.55$, $p< .001$* |
| 62 | .021 (.013) | -0.001, 0.016 | $F(1,130)= 2.76$, $p=.099$ |
| 63[#] | .099 (.092) | 0.012, 0.040 | $F(1,127)= 13.91$, $p< .001$* |
| 64[#] | .071 (.051) | -0.002, 0.047 | $F(1,46)= 3.53$, $p=.067$ |
| 65[#] | .094 (.085) | 0.008, 0.033 | $F(1,97)= 10.06$, $p=.002$* |
| 66[#] | .068 (.061) | 0.006, 0.029 | $F(1,130)= 9.53$, $p=.002$* |
| 67 | .053 (.046) | 0.005, 0.030 | $F(1,132)= 7.45$, $p=.007$* |
| 68 | .003 (-.004) | -0.005, 0.010 | $F(1,135)= .43$, $p=.514$ |
| 69[#] | .004 (-.004) | -0.006, 0.013 | $F(1,114)= .49$, $p=.486$ |
| 71 | .004 (-.004) | -0.008, 0.004 | $F(1,134)= .50$, $p=.482$ |
| 72 | .002 (.015) | -0.001, 0.014 | $F(1,136)= 3.06$, $p=.082$ |
| 73[#] | .088 (.081) | 0.007, 0.024 | $F(1,132)= 12.71$, $p=.001$* |
| 74 | .000 (-.007) | -0.006, 0.007 | $F(1,135)= .01$, $p=.915$ |
| 75[#] | .036 (.029) | 0.002, 0.033 | $F(1,130)= 4.88$, $p=.029$* |
| 76 | .145 (.139) | 0.012, 0.030 | $F(1,132)= 22.46$, $p< .001$* |
| 77 | .016 (.008) | -0.002, 0.013 | $F(1,130)= 2.06$, $p=.153$ |
| 78 | .088 (.080) | 0.005, 0.018 | $F(1,130)= 12.47$, $p=.001$* |

**Appendix O.** Individual accuracy and response latency scores on the sub-lists of the Coltheart and Leahy task. For the accuracy-based scores: light grey indicates the participant fell within the impaired band, dark grey indicates the participant fell within the severely impaired band. # indicates reader is impaired according to reading age.

| Participant | Regular words | | Irregular words | | Non-words | |
|---|---|---|---|---|---|---|
| | ACC | RT | ACC | RT | ACC | RT |
| 22 | 0.97 | 0.60 | 0.83 | 0.75 | 0.93 | 1.09 |
| 23# | 0.93 | 0.66 | 0.67 | 0.74 | 0.67 | 0.84 |
| 24# | 0.93 | 0.74 | 0.7 | 0.65 | 0.77 | 1.11 |
| 25# | 0.87 | 0.90 | 0.63 | 0.74 | 0.53 | 1.57 |
| 28# | 0.83 | 0.55 | 0.67 | 0.57 | 0.7 | 1.16 |
| 29# | 0.77 | 1.20 | 0.57 | 0.82 | 0.37 | 1.92 |
| 31# | 0.93 | 1.70 | 0.63 | 1.58 | 0.5 | 3.38 |
| 32# | 0.93 | 0.74 | 0.67 | 0.83 | 0.53 | 1.40 |
| 33# | 0.87 | 1.05 | 0.77 | 1.11 | 0.57 | 1.69 |
| 34# | 0.90 | 0.59 | 0.67 | 0.59 | 0.3 | 1.30 |
| 35# | 0.93 | 0.93 | 0.73 | 0.84 | 0.57 | 2.21 |
| 37 | 1.00 | 0.65 | 0.83 | 0.73 | 0.83 | 0.94 |
| 38# | 0.73 | 1.44 | 0.6 | 1.01 | 0.33 | 4.29 |
| 39# | 1.00 | 0.89 | 0.7 | 1.15 | 0.73 | 1.78 |
| 40# | 1.00 | 0.85 | 0.8 | 1.17 | 0.77 | 1.66 |
| 41# | 0.93 | 1.33 | 0.73 | 0.91 | 0.6 | 2.76 |
| 43# | 0.80 | 2.05 | 0.6 | 1.44 | 0.45 | 3.74 |
| 44 | 1.00 | 0.66 | 0.77 | 0.98 | 0.83 | 1.33 |
| 45# | 0.97 | 0.65 | 0.6 | 0.67 | 0.87 | 0.89 |
| 46# | 0.93 | 0.95 | 0.73 | 1.07 | 0.8 | 1.52 |
| 47# | 0.93 | 0.97 | 0.77 | 1.05 | 0.57 | 2.50 |
| 48# | 0.90 | 1.21 | 0.77 | 1.03 | 0.6 | 1.34 |
| 49# | 0.83 | 0.80 | 0.57 | 0.98 | 0.6 | 1.53 |
| 50# | 0.90 | 1.14 | 0.77 | 1.01 | 0.77 | 1.35 |
| 51# | 0.97 | 0.68 | 0.77 | 0.69 | 0.7 | 0.85 |
| 52# | 0.87 | 1.60 | 0.7 | 2.12 | 0.43 | 2.27 |
| 54# | 0.77 | 2.16 | 0.47 | 1.17 | 0.23 | 3.64 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 55# | 0.93 | 1.30 | 0.63 | 1.46 | 0.63 | 3.58 |
| 57 | 0.93 | 0.89 | 0.8 | 1.12 | 0.83 | 1.59 |
| 58# | 0.90 | 0.79 | 0.73 | 0.87 | 0.7 | 1.02 |
| 59# | 0.87 | 1.79 | 0.73 | 1.00 | 0.33 | 4.28 |
| 60 | 1.00 | 0.77 | 0.77 | 0.81 | 0.83 | 1.02 |
| 61# | 0.93 | 0.89 | 0.73 | 0.83 | 0.63 | 1.40 |
| 62 | 1.00 | 0.75 | 0.8 | 0.81 | 0.7 | 1.12 |
| 63# | 0.97 | 0.65 | 0.87 | 0.94 | 0.63 | 1.13 |
| 64# | 0.67 | 1.39 | 0.57 | 1.12 | 0.43 | 2.74 |
| 65# | 0.90 | 0.54 | 0.77 | 0.60 | 0.37 | 0.64 |
| 66# | 1.00 | 0.71 | 0.8 | 0.76 | 0.8 | 1.41 |
| 67 | 1.00 | 0.58 | 0.7 | 0.62 | 0.87 | 0.78 |
| 68 | 1.00 | 0.54 | 0.77 | 0.56 | 0.97 | 0.77 |
| 69# | 0.87 | 0.85 | 0.7 | 0.84 | 0.57 | 1.31 |
| 71 | 1.00 | 0.58 | 0.93 | 0.62 | 0.93 | 0.81 |
| 72 | 0.97 | 0.73 | 0.87 | 0.78 | 0.83 | 0.95 |
| 73# | 1.00 | 0.65 | 0.8 | 0.78 | 0.93 | 0.80 |
| 74 | 1.00 | 0.53 | 0.97 | 0.57 | 0.83 | 0.67 |
| 75# | 0.87 | 0.69 | 0.73 | 0.60 | 0.83 | 1.41 |
| 76 | 1.00 | 0.63 | 0.7 | 0.73 | 0.87 | 1.09 |
| 77 | 1.00 | 0.55 | 0.73 | 0.58 | 0.8 | 0.77 |
| 78 | 1.00 | 0.57 | 0.8 | 0.60 | 0.93 | 0.72 |

**Appendix P.** Complete correlation matrix for the Heterogeneity study, p-values in brackets: ** *p* <.01; * *p* <.05.

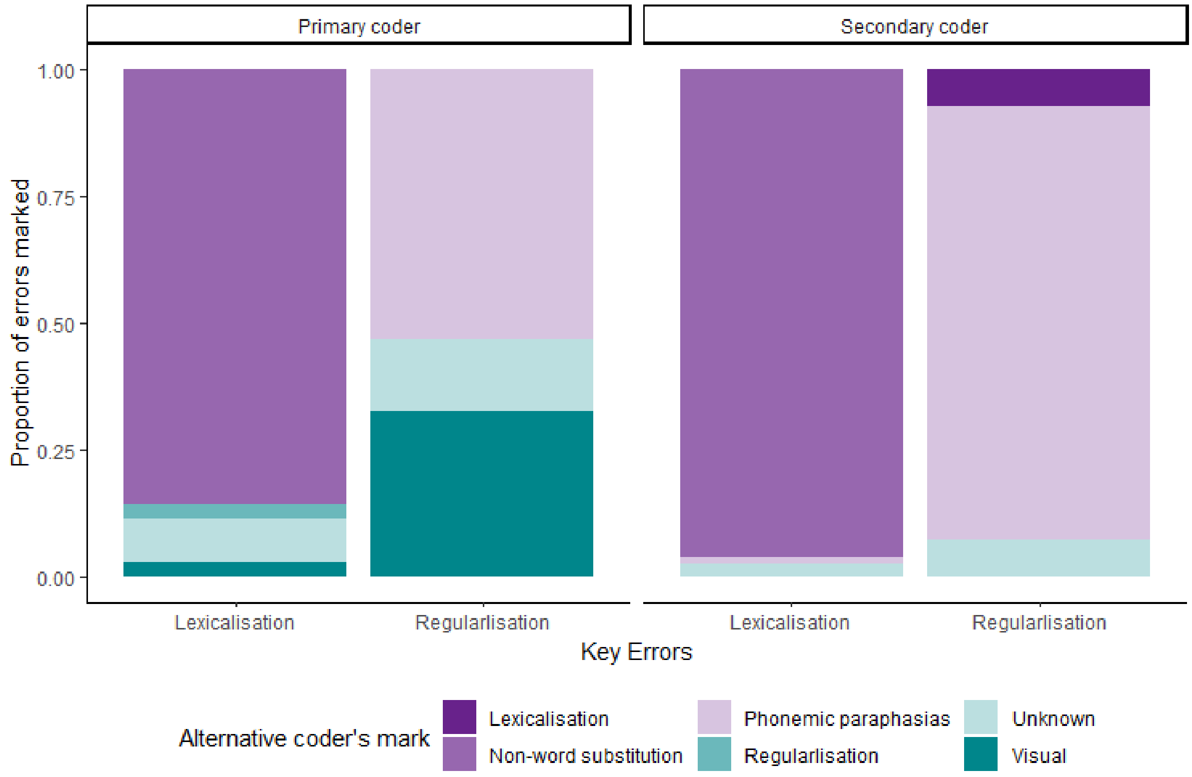
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. RA | -.560** | -.590** | -.206 | -.576** | -.412** | .537** | .716** | -.595** | -.145 | -.511** | -.661** |
| | (< .001) | (< .001) | (=.155) | (< .001) | (=.003) | (< .001) | (< .001) | (< .001) | (=.320) | (< .001) | (< .001) |
| 2. Image | | .621** | -.140 | .438** | .138 | -.252 | -.592** | .566** | -.084 | .457** | .596** |
| E_ACC | | (< .001) | (=.336) | (=.002) | (=.343) | (=.081) | (< .001) | (< .001) | (=.556) | (=.001) | (< .001) |
| 3. Freq | | | -.087 | .441** | .380** | -.320* | -.672** | .480** | -.010 | .494** | .595** |
| E_ACC | | | (=.554) | (=.001) | (=.007) | (=.025) | (< .001) | (<.001) | (=.943) | (< .001) | (< .001) |
| 4. Image | | | | .227 | .243 | -.026 | -.067 | .129 | -.099 | .348* | .064 |
| E_RT | | | | (=.117) | (=.093) | (=.859) | (=.647) | (=.376) | (=.500) | (=.014) | (=.647) |
| 5. Freq | | | | | .350* | -.430** | -.578** | .522** | .126 | .598** | .290* |
| E_RT | | | | | (=.014) | (=.002) | (< .001) | (< .001) | (=.388) | (< .001) | (=.040) |
| 6. WLE | | | | | | -.192 | -.429** | .242 | .318* | .388** | .227 |
| | | | | | | (=.187) | (=.002) | (=.093) | (=.026) | (=.006) | (=.117) |
| 7. VA | | | | | | | .258 | -.377** | .062 | -.259 | -.389** |
| span | | | | | | | (=.074) | (=.008) | (=.671) | (=.073) | (=.005) |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8. Real | | | | | | | | | -.565$^{**}$ | -.255 | -.572$^{**}$ | -.412$^{**}$ |
| ACC | | | | | | | | | (< .001) | (=.077) | (< .001) | (=.003) |
| 9. Real | | | | | | | | | | .061 | .690$^{**}$ | .513$^{**}$ |
| RT | | | | | | | | | | (=.678) | (< .001) | (< .001) |
| 10. Irreg | | | | | | | | | | | .062 | -.169 |
| E_ACC | | | | | | | | | | | (=.670) | (=.247) |
| 11. Non | | | | | | | | | | | | .514$^{**}$ |
| E_RT | | | | | | | | | | | | (< .001) |
| 12. Non | | | | | | | | | | | | |
| E_ACC | | | | | | | | | | | | |

**Appendix Q.** Potential errors and the descriptions given to each judge for coding reading errors on the Coltheart and Leahy task.

| Error | Definition |
|---|---|
| Lexicalisation | Sounding out a non-word word as a real word (i.e. vun> "van") |
| Regularisation | Irregular word read as non-word that follows the sight-sound rules ( i.e. blood> "blude", meringue> "mer-ing-goo") |
| Non-word substitution | Sounding out a non-word as another non-word (i.e. vun> "wun") |
| Phonemic paraphasias | Substitution of real word with non-word that maintains at least some features (i.e. gingerbread> "gingerjed") |
| Neologism | Substitution of a real word (regular or irregular) with non-word (complete gibberish, does not follow sight-sound rules; i.e. chance > "sanse") |
| Semantic | Synonyms (one real word for another with same meaning; i.e. negative> "minus") |
| Visual | Substitution with visually similar word (real word substitutions; i.e. carton > "carrot") |
| Derivational | Altered suffix/prefix (i.e. child> "children", science> "scientist") |
| Pass | Participant declines to answer ("pass" or "skip"), or no response |
| Unknown | Cannot place the error into any of the above categories |

**Appendix R.** Alternative coding patterns on trials where coders did not reach an agreement (only cases where one coder marked an error as either a Lexicalisation or a Regularisation error are shown). The X-axis shows what that coder marked the error, with the alternative marks given by the other coder shown within each bar.



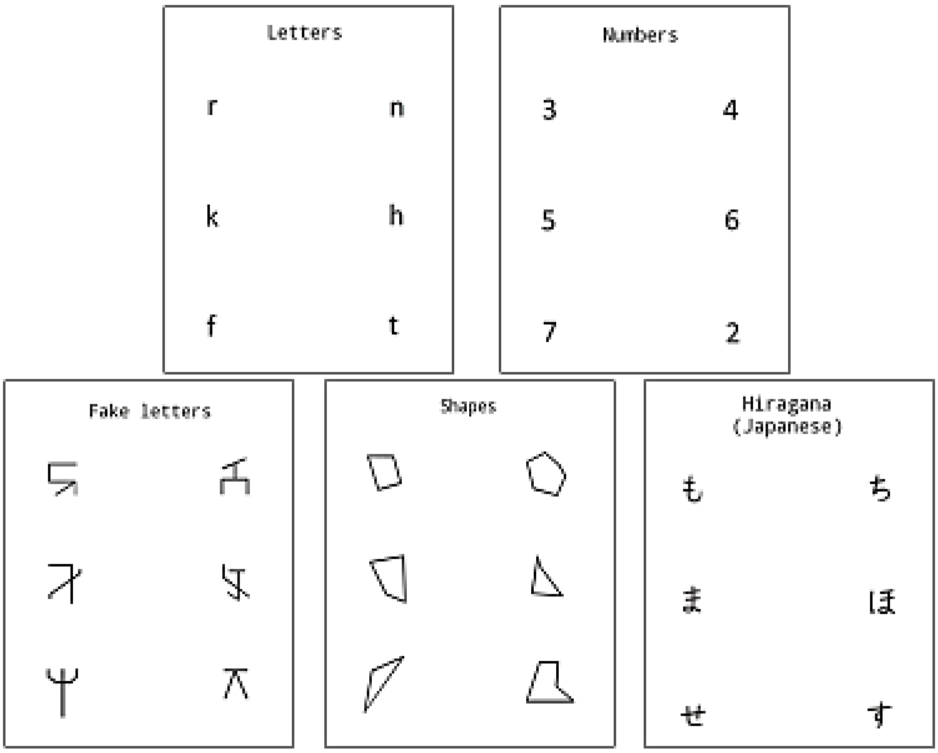
**Appendix S.** Supplementary analysis with the Verbal and Non-verbal categorisation task.

**Methods**

**Participants**. These analyses were completed with a subset (23) of participants from the Heterogeneity study. Of these participants, 13 were male and 10 were female, with a mean age of 13 years and 6 months (SD = 6 months).

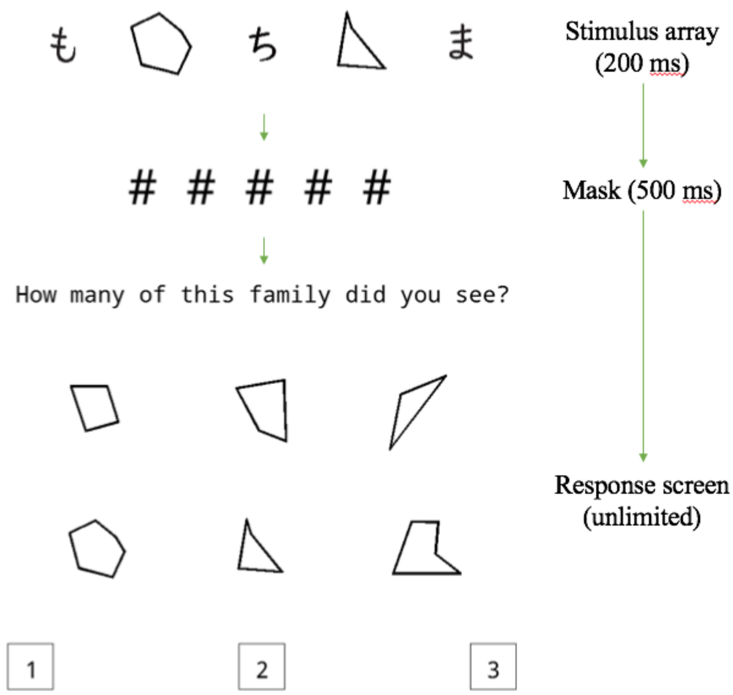
**Materials**. This task was replicated from Lobier et al (2012). Figure S.1 shows the five stimuli 'families' used in this task; two Verbal (letters and numbers) and three Non-verbal (Hiragana – Japanese symbols – fake letters and shapes). Each family consisted of six stimuli which participants did not need to remember. Throughout all stimuli were presented in black, on a white back ground. Prior to this task participants were asked whether they were familiar with Hiragana – one of the stimulus families – none were.

Figure S.1. The five stimulus families used in the in the Verbal/ Non-Verbal span task.




**Procedure**. Prior to the categorisation task, participants are introduced to each of the families and told that they need to be aware what the families were but did not need to remember the exact stimuli. To get them more familiar with the families, participants then completed an identification task where one of the stimuli was shown very briefly and they identified the family it belonged to. This was not included in the analyses and will not be discussed further. Figure S.2. shows the categorisation task. Participants were shown a five-element array for 200 ms. This array contained a combination of two families; with one, two or three stimuli from the target family, and the remainder from another family. Each family was the target for eight trials and each of the remaining families were the distractors for two of these eight trials. A mask was presented for 500ms before a response screen asked them: "How many of this family did you see?" Participants responded by selecting either '1', '2', or '3' on the screen. The response window was not limited, and the task proceed to the next trial once the participant responded. Each trial was separated by a 1000ms fixation dot. The task consisted of 40 trials.

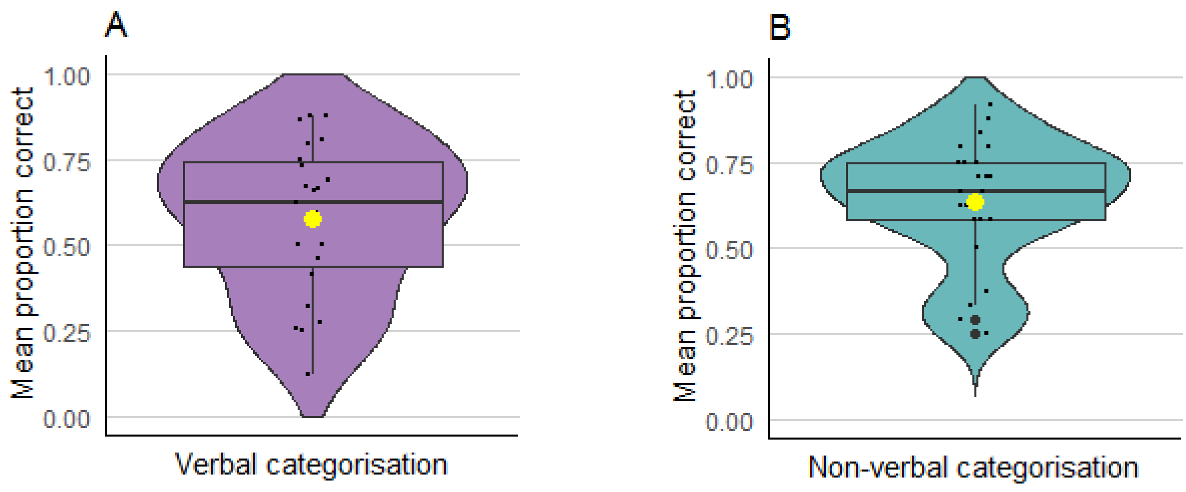
Figure S.2. The procedure of the Verbal/Non-verbal span task.




**Results**

**Spread.** Figure S.3 shows there was a good spread of scores on both the Verbal categorisation (Panel A) and Non-verbal categorisation (Panel B) scores. The Non-verbal categorisation score showed a marked tail end (3 participants scores well below the 95% confidence interval).

Figures S.3. Violin plots of Verbal categorisation (A) and Non-verbal categorisation (B). Black dots represent individual participants, yellow dots represent group mean, with 95% confidence intervals (black vertical line). Box plots show median (black central line), upper and lower quartiles.




**Correlation**. Table S.1 shows the results from the correlation analyses for the two categorisation scores. Verbal categorisation and Non-verbal categorisation were found to correlate with one another, however neither score correlated with any of the other key measures.

Table S.1. Correlation co-efficient and significance levels (in parentheses) for the Verbal categorisation and Non-verbal categorisation scores, with each of the previous key scores of reading proficiency and processes. *Significant relationships at the $p < .01$ level.

| | Verbal categorisation | Non-verbal categorisation |
|---|---|---|
| Non-verbal categorisation | .738 * (< .001) | *na* |
| VA span | .260 (.232) | .069 (.755) |
| Reading Age | .190 (.385) | .160 (.465) |
| Nonword Accuracy Diff | -.018 (.934) | .027 (.902) |
| WLE | .117 (.596) | .075 (.734) |

**Appendix T.** Screening study letter to Principal (2016 cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI
VICTORIA
UNIVERSITY OF WELLINGTON


| Emma Ashcroft | Dr. Carolyn Wilshire |
|---|---|
| **Masters Student** | **Senior Lecturer, School of Psychology** |
| Email: *[email]* | Email: *[email]* |


**A comparison of two different cognitive interventions for Secondary School students with reading difficulties.**
**Emma Ashcroft**
**Masters of Science in Cognitive and Behavioral Neuroscience**

Dear **Principal,**

My name is Emma Ashcroft and I am a Masters student working under the supervision of Dr Carolyn Wilshire in the School of Psychology at Victoria University. My thesis involves investigating dyslexia and developing suitable interventions to aid the development of reading skills. The aim of my research is to assess the efficacy of reading interventions specially designed to target different reading difficulties. I am particularly interested in trialling a new intervention for atypical "visual" forms of dyslexia, characterised by problems developing an efficient sight vocabulary, despite normal phonological skills. However, I will be trialling my methods on a range of individuals with different kinds of reading difficulties.

I am writing to enquire as to whether your school may be willing to participate in this research. I have chosen to write to you because [college] has a reputation within the community as a centre with specialised resources and a proactive approach to addressing reading difficulties. Our target group of participants is Year 9 to 10 students who are currently enrolled in remediation classes. If you were to take part, your role would be simply to disseminate the information packs to students in these classes, so they can share them with their families. This could be done in paper form (in a sealed envelope) or electronically, whichever works best for you. The information packs would provide our contact details, and parents who are interested could then respond to us directly. I have attached a copy of the information sheet to parents for your viewing.

The study is organised into two parts. In the first part, all students in special reading classes willing to participate will complete some simple reading and cognitive tasks. This will allow us to create a brief profile of their reading difficulties. This part takes around one hour.

In the second part of the study, we will be asking 2-4 participants from the first part if they wish to take part in a reading intervention trial, which is designed to target two very different cognitive skills crucial for reading: a) phonological awareness or the ability to convert text to sounds, and b) visual attention. We will then contrast the effectiveness of these two approaches for the different participants. This intervention would consist of a series of weekly sessions over a period of approximately four months.

These sessions would be available for completion on school premise, with your permission, or at students' homes outside school hours. A secondary supervisor will be present throughout all sessions, this can either be another researcher or a teacher. The secondary supervisor will not partake in the research.

Full consent will be obtained from all students and their families before they participate in any part of this research. Parents will also be asked for their permission to give students movie vouchers as appreciation for their participation. This research has been approved by the School of Psychology Ethics Committee (approval number 00000 22713).

After the research is complete, I will also be available to visit your school and discuss the overall findings with any interested staff. If students and their families wish, I will also provide teachers with feedback on individual children which may be useful for setting future goals.

Thank you for your time. I will be in touch with you within the next fortnight to hear your views on this proposal, or please feel free to email me on the address provided above.

Yours sincerely,
Emma Ashcroft

**Appendix U.** Screening parent information sheet (2016 cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA
UNIVERSITY OF WELLINGTON


| Emma Ashcroft | Dr. Carolyn Wilshire |
|---|---|
| **Masters Student** | **Senior Lecturer, School of Psychology** |
| Email: *[email]* | Email: *[email]* |


**A comparison of two different cognitive interventions for Secondary School students with reading difficulties.**
**Ethics approval number: 0000022713**

Dear **Parent,**

My name is Emma Ashcroft and I am a Masters student working under the supervision of Dr. Carolyn Wilshire in the School of Psychology at Victoria University. My Masters thesis involves investigating reading difficulties, with the aim of developing new interventions to improve students' reading skills.

I am therefore looking for secondary school students who are currently in reading remediation classes and would be interested in participating in an assessment of their reading. If you agree, your child will be given several tasks to examine their reading and visual skills. We estimate the tasks will take no longer than one hour. Testing will take place at the school. Students are welcome to stop at any time. A digital audio recorder will be used during the reading task so that your child's responses can be transcribed later on.

Subsequent to this, we will be looking to recruit 2-4 students whose reading styles fit the profiles we are focusing on. This will consist of one-on-one sessions to target each student's reading problems. We will be inviting students to take part in the intervention phase of the research at a later date. You and your child do not need to decide whether or not you want to take part now.

Finally, the sessions are strictly confidential. Your child's name will never be released to anyone outside of the research team, and in written reports and publications of the research, they will be referred to only by a random combination of letters.

Attached is a more detailed information sheet and consent form. You are welcome to contact myself or my supervisor, Dr. Carolyn Wilshire, to find out more. Our details are given above. If you have any queries regarding ethics, please contact [Ethics administrator] ([email/ contact number]) who is chair of the Human Ethics Committee at Victoria University.

Yours Sincerely,
Emma Ashcroft

**Appendix V.** Screening parent consent form (2016 cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft | Dr. Carolyn Wilshire |
| **Masters Student** | **Senior Lecturer, School of Psychology** |
| Email: *[email]* | Email: *[email]* |


**A comparison of two different cognitive interventions for Secondary School students with reading difficulties.**
**Ethics approval number: 00000 22713**

**What is the purpose of this research?**
Reading difficulties are highly prevalent in primary and secondary school students, with some estimating up to 20% of children in New Zealand schools struggling to learn fluent reading skills. Given how important reading is for everyday communication and academic study, it is essential that intervention are developed for those who find reading particularly difficult. Importantly, these intervention need to target the specific needs of individuals'.

The goal of this study is to develop interventions which target different patterns of reading impairment.

**Who is conducting the research?**
Dr Carolyn Wilshire and Masters' student Emma Ashcroft from the School of Psychology at Victoria University will be conducting the research. Emma will subsequently write a Masters' thesis from the results of the research. This research has been approved under the delegated authority of the School of Psychology Ethics Committee (SoPHEC) at Victoria University.

**What is involved if you agree for your child to participate?**
If you and your child agree to participate in this study, Emma will visit your child's school to complete a session with them. Another researcher or the reading remediation teacher will be present in these sessions but they will not be actively involved. This session will last no longer than an hour and will comprise of three tasks.

The first task will ask students' to mimic a series of visual patterns using plastics blocks. The second task will ask students' to read words and sentences aloud. The third task will ask students to manipulate word sounds, according to instructions from the experimenter. The latter two tasks will be recorded using a digital tape recorded and later transcribed. Students reading accuracy will then be used to calculate an estimated reading age.

During the research you are free to withdraw your child (or any information they have provided) at any time up until the session has been completed, without having to give any reasons and without any penalty. Approximately 2-4 students will subsequently be invited to take part in the intervention phase of this study. In the event that we are able to tailor an intervention to suit your child, please indicate whether you wish to be contacted regarding this phase of the study below.

**What happens to the information you provide?**
Your child's information will be kept confidential. All data files will be coded using a random selection of letters. Any files which identify you will only be accessible by the investigators. Data *without identifying names* may appear in journal publications, conference presentations and will form part of a thesis submitted for assessment. Your child's name will never be used in any presentation or publication.

**Thank you for your participation.**
As a token of our appreciation, we will offer each child a small reward. This is in no way dependent on their performance.

If you would like to know the results of this study they can be emailed, or sent to you in another capacity, if you wish. Individual student's performance data will not be available for this phase of the study, as no formal diagnostic inferences may be made from this research.

If you have any further questions regarding this study, please contact one of us above.

**Statement of Consent**

I have been given a full explanation of this project, and have had an opportunity to ask questions and have had them answered to my satisfaction. I understand that my child will be given a number of tasks to examine his or her reading and other related cognitive skills. Following this, my child may be invited to take part in further sessions designed to improve their reading skills, which they may also take part in if we so wish.

I understand that the data obtained may appear in graduate student reports, poster presentations, conference presentations and peer-reviewed publications, but that in all cases, my child will be referred to only by a random stream of letters. All paper work and electronical files will be kept by Dr. Carolyn Wilshire in a secure place. Summaries of the data will be kept by Emma Ashcroft and Dr Wilshire and may be shared with other members of the research team.

I understand that I may withdraw my child (or any information we have provided) at any time up until the session has been completed, without having to give reasons and without penalty of any sort.

I agree that ___________________________, who is under my guardianship, may take part in this research.

Name: __________________________________________________

Signature: __________________________________________________

Date: __________________________________________________

If you would like to receive a summary of the results of this study once it is completed, please tick here: ☐ (Please note, individual data will not be available).

If we are able to tailor the intervention tasks to your child's needs, would you like to be contacted regarding this phase of the research? If so please tick here: ☐

Copy to:

(a) Parent
(b) Researcher (initial both copies below)

**Appendix W.** Screening student information and assent form (2016 cohort).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft | Dr. Carolyn Wilshire |
| **Masters Student** | **Senior Lecturer** |


**Reading study**

**Ethics approval number: 00000 22713**

Dear **Student,**

My name is Emma Ashcroft and I am a Masters student at Victoria University. For my project I am looking at reading difficulties, and how I can help students improve.

I would like to invite you to take part in some reading and Lego-style tasks. This would take one hour. Another adult would also be present for this session, but they will not be involved. If you do decide to take part, you can change your mind and stop at any time, even during the session.

I will tape what you say in the tasks on an audio recorder, and write them out later on. The details about what you say and do is private. After the study, I will work out how everyone performed as a group. I will send everyone a copy of the group results. No names will be mentioned in this report.

I will also write my Masters' thesis using the results added together. Your name will never mentioned.

After these sessions, I will be asking a small number of students (2-4) if they would like to take part in some new exercises designed to improve reading. I will be looking for people who "suit" the kind of exercises I'm planning. If I do ask you to take part, you can decide later on whether you want to do this.

If you wish to take part, please sign the form below. If you have any questions, your parents can email me on the address at the top of the previous page.

Yours Sincerely,

Emma Ashcroft

**Statement of Assent**

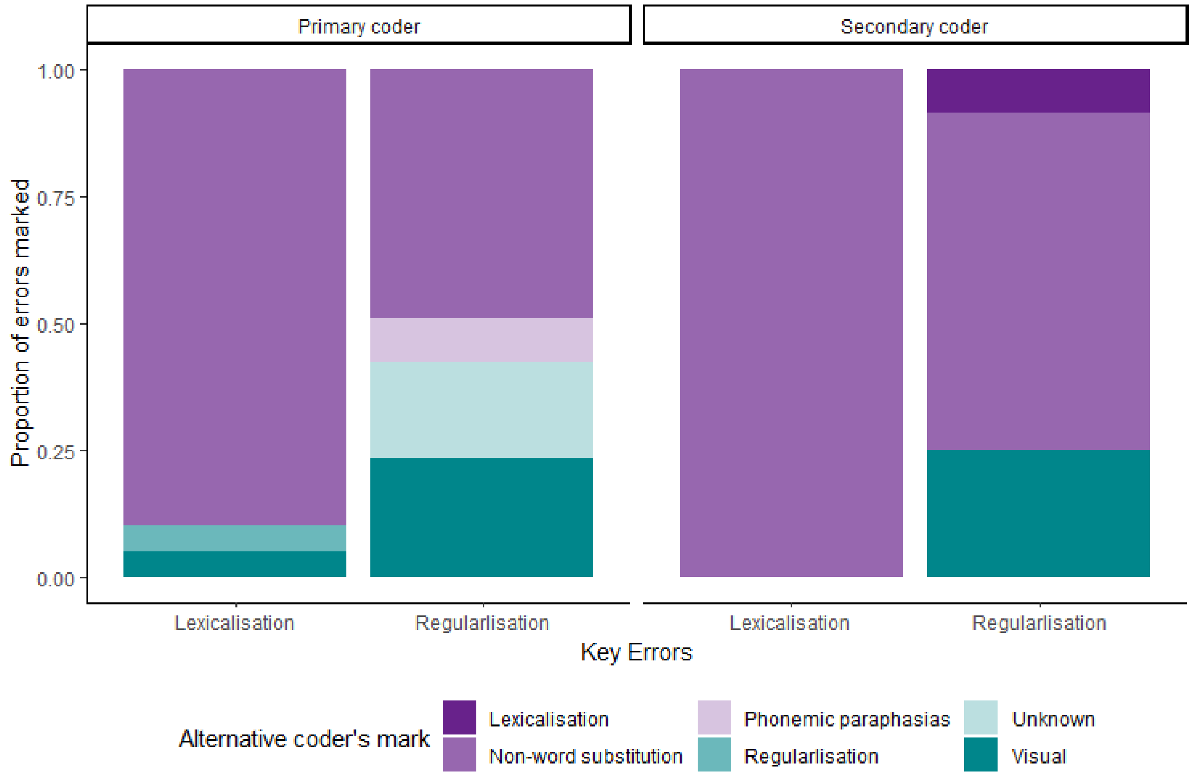
I, ____________________________, agree to take part in this research.

Signed: _____________________________________

Name: ______________________________________

Date: __________________________

**Appendix X.** Alternative coding patterns on trials where coders did not reach an agreement (only cases where one coder marked an error as either a Lexicalisation or a Regularisation error are shown). The X-axis shows what that coder marked the error, with the alternative marks given by the other coder shown within each bar.



**Appendix Y.** Intervention parent information sheet (2016 and 2017 cohorts).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA

UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft | Dr. Carolyn Wilshire |
| **[Masters/PhD] Student** | **Senior Lecturer, School of Psychology** |
| Email: *[email]* | Email: *[email]* |


**A comparison of two different cognitive interventions for Secondary School students with reading difficulties (Phase 2)**
**Ethics approval number: 00000 22713**

Dear **Parent,**

Your child recently took part in the first phase of our reading study. I am writing now to find out whether you and your child would be interested in participating in the second phase of the study, which will examine some new techniques for improving reading.

In this phase of the study, each student will complete a more thorough reading and cognitive skills assessment. They will then complete two different types of interventions that aim to improve different aspects of reading. More information is provided on the attached sheet.

The entire study will take approximately 12-18 sessions to complete (at roughly one session a week). You or your child are free to pull out at any time. These sessions will take place at the school during school time. Another researcher or teacher may be present throughout these sessions, but will not be actively involved.

Finally, these sessions are strictly confidential. Your child's name will never be released to anyone outside of the research team, and in written reports and publications of the research they will be referred to only by a random series of letters. However, we would be happy to discuss your child's results with you, and with their teachers should you both wish.

Attached is a more detailed information sheet and consent form. You are welcome to contact myself or my supervisor, Dr. Carolyn Wilshire, to find out more. Our details are given above. If you have any queries regarding ethics, please contact [Ethics administrator] ([email/ contact number]) who is chair of the Human Ethics Committee at Victoria University.

Yours Sincerely,
Emma Ashcroft

**Appendix Z.** Intervention parent consent form (2016 and 2017 cohorts).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON


| | |
|---|---|
| Emma Ashcroft | Dr. Carolyn Wilshire |
| **[Masters/PhD] Student** | **Senior Lecturer, School of Psychology** |
| Email: *[email]* | Email: *[email]* |


**A comparison of two different cognitive interventions for Secondary School students with reading difficulties (Phase 2): Information sheet**
**Ethics approval number: 00000 22713**

**What is the purpose of this research?**
Reading difficulties are highly prevalent in primary and secondary school students, with some estimating that up to 20% of children in New Zealand struggle to learn fluent reading skills. The overarching goal of this study is to develop interventions designed to target different patterns of reading impairment.

In this phase of the study, various reading exercises will be administered to try and improve reading skills. The effect of these tasks on different reading profiles will be analysed.

**Who is conducting the research?**
Dr Carolyn Wilshire and Masters' student Emma Ashcroft from the School of Psychology at Victoria University will be conducting the research. Emma will subsequently write a Masters' thesis from the results of the research. This research has been approved by the Human Ethics Committee of Victoria University.

**What is involved if you agree for your child to participate?**
If you and your child agree to take part, your child will participate in a series of sessions involving reading and other related activities. The sessions will be run by Emma Ashcroft and will take place at school during school time. Another researcher or the reading remediation teacher may be present during the session but they will not be actively involved. Each session will take approximately one hour.

There will be a total of 12-18 sessions in this phase of the research. In the first two sessions, Emma will conduct a detailed assessment of your child's reading and related cognitive skills. Subsequent session will be devoted to specific exercises aimed at improving various kinds of reading skills. Some exercises will involve the use of flashcards containing different words and word endings. Students will be asked to read these aloud, and practice some writing tasks. Other exercises will use a computer: your child may be asked to read a series of moving words, to make words up from smaller units, or break whole words into smaller units.

Within the sessions, further smaller reading tasks will be given to assess your child's ongoing

progress. Tasks which require students to respond verbally will be taped using a digital recorded and later transcribed.
During the research you are free to withdraw your child (or any information they have provided) at any time up until the completion of the research, without having to give any reasons.

**What happens to the information you provide?**
All data files will be coded and stored using a random selection of letters. Any files which identify your child will only be accessible by the investigators. Data *without identifying names,* and other identifying information, may appear in journal publications, conference presentations and will form part of a thesis submitted for assessment. Your child's name will never be used in any presentation or publication.

**Thank you for your participation.**
As a token of our appreciation, we would like to offer four-double movie vouchers to your child for their participation (to be given at regular intervals throughout the study). If you do not wish for your child to receive movie vouchers, alternatives can be discussed.

If you would like to know the results of this study, we would be happy to meet and discuss these with you. If you consent, we can also discuss with the reading remediation teachers, any techniques that were effective with your child. A written summary of the pooled data will also be available to all participants and the reading remediation teachers. Individuals will not be identified in this summary.

If you have any further questions regarding this study, please contact one of us above.

**Statement of Consent**

I have been given a full explanation of this project and have had an opportunity to ask questions and have had them answered to my satisfaction. I understand that my child will complete a number of language tasks and related activities across a series of hourly one-on-one sessions.

I understand that the data obtained may appear in graduate student reports, poster presentations, conference presentations and peer-reviewed publications, but that in all cases, my child will be referred to only by a random series of letters. The raw data will be kept by Dr. Carolyn Wilshire in a secure place. Summaries of the data will be kept by Emma Ashcroft and Dr Wilshire and may be shared with other members of the research team.

I understand that I may withdraw my child (or any information we have provided) at any time up until all sessions have been completed without having to give reasons and without penalty of any sort.

I agree that ___________________________, who is under my guardianship, may take part in this research.

Name: ______________________________________________

Signature: ______________________________________________

Date: ______________________________________________

*(please tick where appropriate)*

I would like to receive a summary of the results of this research ☐

I would like to discuss my child's performance ☐

I would like to share any relevant information with my child's reading teacher ☐ (my child will also be asked if they agree to this).

I agree to my child being given movie vouchers/I would like to discuss other options (*cross out whichever does not apply)*

Copy to:

(a) Parent
(b) Researcher (initial both copies below)

**Appendix AA.** Intervention student information and assent form (2016 and 2017 cohorts).

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA UNIVERSITY OF WELLINGTON

Emma Ashcroft
**[Masters/PhD] Student**

Dr. Carolyn Wilshire
**Senior Lecturer**

**Reading study**

**Ethics approval number: 00000 22713**

Dear **Student,**

My name is Emma Ashcroft. You recently took part in my study looking at reading skills. I would now like to invite you to the second part of the study.

In this second part, you would complete a series of sessions with different exercises designed to improve your reading. There will be 12-18 sessions in total.

In the first two sessions, you will complete some reading and other tasks, so that we can learn more about your thinking and reading style. I will then adapt exercises to fit your individual style.

There will be different kinds of exercises. Some involve reading words on a computer or on cards, others involve writing. Some exercises will ask you to separate words into different parts, or build words up from smaller parts.

On some tasks, I will tape what you say on an audio recorder, and write it out later on. Another adult will also be present for this session, but they will not be involved. If you do decide to take part, you can change your mind and stop at any time, even during the session.

After you have finished the sessions, I can talk to you and your parent(s)/ guardian(s) about what exercises helped you most. I will not discuss specific things you said in the sessions. If you wish, I can also talk to your reading teacher about what was most helpful.

I will be writing about the results for my PhD thesis, but in this, I'll refer to you only by random letters. No-one outside our research team will know your real name.
[I will also offer you some movie vouchers for taking part.]

If you wish to take part, please sign the form below. If you have any questions, your parents can email me on the address at the top of the previous page.

Yours Sincerely,
Emma Ashcroft

**Statement of Assent**

I, ____________________________, agree to take part in this research.

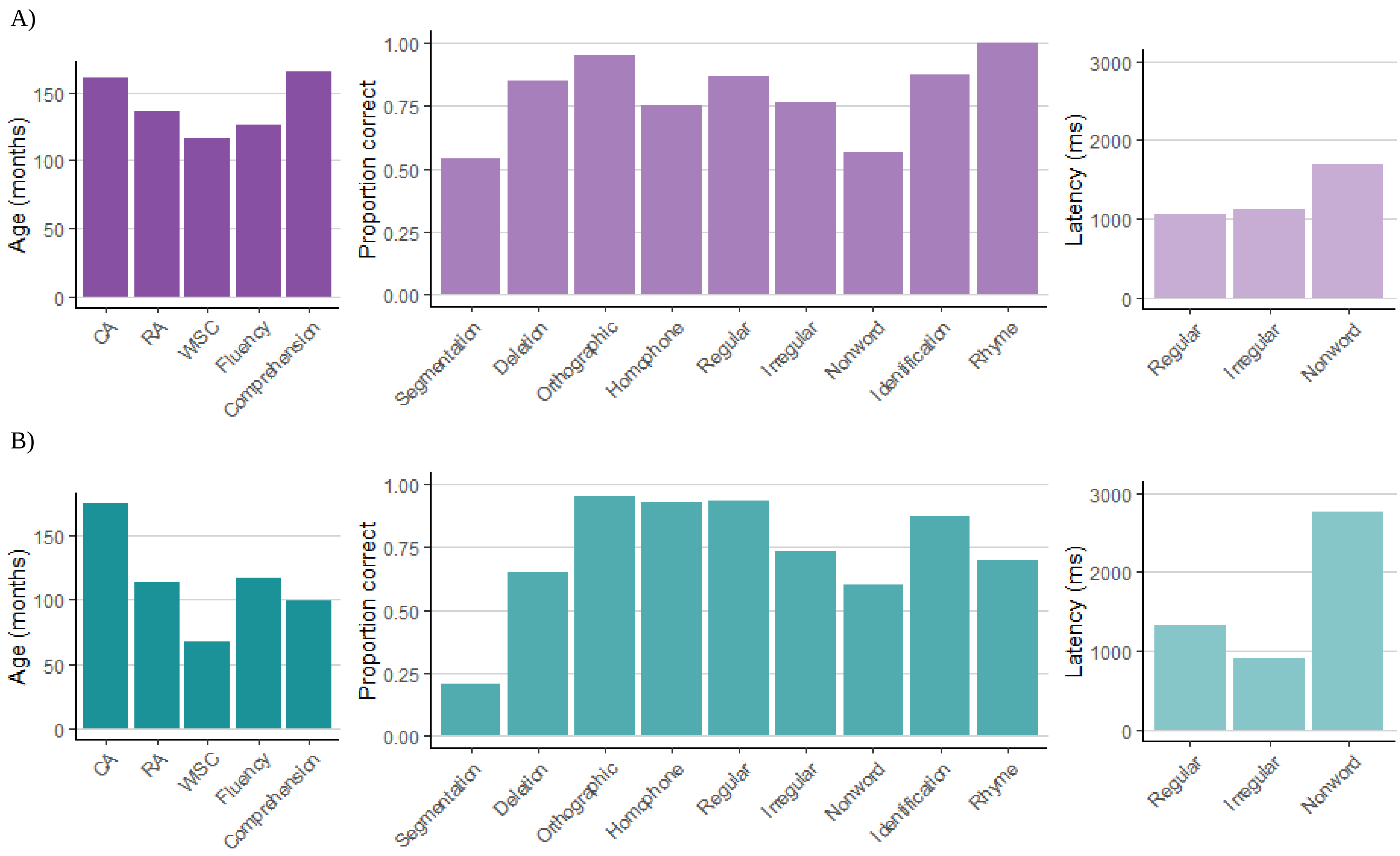
Signed: ____________________________________

Name: _____________________________________

Date: __________________________

I would be happy for Emma to talk to my reading teacher about what helped me most during the sessions: ☐

**Appendix AB.** Detail profiles of two additional participants that completed the Profile Building phase but not the Intervention study. Panel (A) = phonological case AR, Panel B = whole-word case BR.



**Appendix AC.** Phoneme segmentation task

| Word | Segmented word | Word | Segmented word |
|---|---|---|---|
| hib | /h/ /ɪ/ /b/ | plove | /p/ /l/ /ə/ /ʊ/ /v/ |
| fless | /f/ /l/ /ɛ/ /s/ | hif | /h/ /ɪ/ /b/ |
| zun | /z/ /ʌ/ /n/ | gless | /g/ /l/ /ɛ/ /s/ |
| glifs | /ˈg/ /l/ /ɪ/ /f/ /s/ | kun | /k/ /ʌ/ /n/ |
| wirb | /w/ /əː/ /b/ | glift | /g/ /l/ /ɪ/ /f/ /t/ |
| krilm | /k/ /r/ /ɪ/ /l/ /m/ | wirf | /w/ /əː/ /f/ |
| yek | /j/ /ɛ/ /k/ | frilm | /f/ /r/ ɪ/ /l/ /m/ |
| slee | /s/ /l/ /iː/ | yev | /j/ /ɛ/ /v/ |
| lekt | /l/ /ɛ/ /k/ /t/ | klee | /k/ /l/ /iː/ |
| zelts | /z/ /ɛ/ /l/ /t/ /s/ | leks | /l/ /ɛ/ /k/ /s/ |
| splog | /s/ /p/ /l/ /ɒ/ /g/ | delts | /d/ /ɛ/ /l/ /t/ /s/ |
| splove | /s/ /p/ /l/ /ə/ /ʊ/ /v/ | sploz | /s/ /p/ /l/ /ɒ/ /z/ |

**Appendix AD.** Phoneme deletion task

| Target Word | Deleted probe | Target Response | Target Word | Deleted probe | Target Response |
|---|---|---|---|---|---|
| jam | J | /am/ | soap | p | /səʊ/ |
| shout | sh | /aʊt/ | house | s | /haʊ/ |
| dear | d | /ɪə/ | shown | s | /həʊn/ |
| ran | r | /ran/ | pine | n | /pʌɪ/ |
| hold | h | /əʊld/ | cart | t | /kɑː/ |
| take | t | /eɪk/ | spoon | p | /suːn/ |
| make | k | /meɪ/ | sting | t | /sɪŋ/ |
| trade | d | /treɪ/ | frame | fr | /eɪm/ |
| splint | n | /splɪt/ | blame | b | /leɪm/ |
| smell | m | /smɛl/ | raft | f | /rat/ |

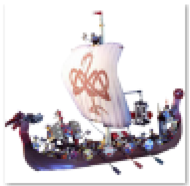
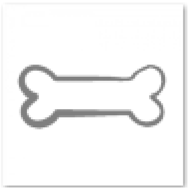
**Appendix AE.** Homophone selection task stimuli

| Correct homophone | Foil homophone | Description |
|---|---|---|
| four | for | A number |
| blue | blew | A colour |
| hear | here | A sense |
| piece | peace | A chunk |
| road | rode | A space for cars |
| week | weak | A period of time |
| wood | would | A tree |
| son | sun | A child |
| role | roll | A job position |
| groan | grown | A noise |
| break | brake | A holiday |
| mail | male | A letter |
| berry | bury | A fruit |
| plane | plain | A flying machine |
| sum | some | A math equation |
| weather | whether | A temperature pattern |
| war | wore | A fight |
| meat | meet | An animal |
| dye | die | A colouring process |
| sea | see | A body of water |

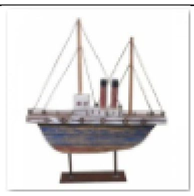
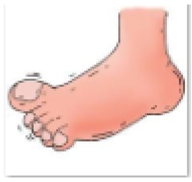
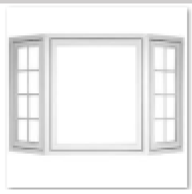
**Appendix AF.** Orthographic choice task stimuli

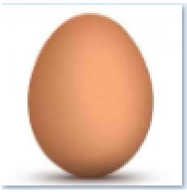
| Correct homophone | Foil homophone | Correct homophone | Foil homophone |
|---|---|---|---|
| bark | barc | cologne | calogne |
| bean | bene | column | collum |
| biscuit | biskuit | court | cort |
| blame | blaim | debt | det |
| blink | blinc | detour | detoor |
| bloom | blume | doubt | dout |
| brawl | brual | face | fais |
| built | bilt | feud | fude |
| by | bie | fly | fligh |
| chair | cheir | freight | frate |
| geyser | guyser | mine | mign |
| goat | gote | mischief | mischef |
| granite | grannit | monk | munk |
| guard | guard | more | mor |
| hoop | hupe | odd | od |
| journey | jurney | pageant | padgeant |
| lamb | lam | poultry | poltrey |
| leap | leep | pursuit | pursute |
| meant | ment | rich | ritch |
| menace | mennis | seize | seeze |
| sponge | spunge | watt | wot |
| tortoise | tortace | shriek | shreek |
| thumb | thum | ski | skee |
| vacuum | vacume | soap | sope |
| source | sorce | soon | sune |

**Appendix AG.** Phoneme identification task

| Word fragment probe | Image | Correct Response |
|---|---|---|
| “Ca-” |  | /t/ |
| “Hor-” |  | /s/ |
| “Fi-” |  | /ʃ/ |
| “kni-” |  | /f/ |
| “Shi-” |  | /p/ |
| “Bo-” |  | /n/ |
| “Car-” |  | /d/ |
| “Ga-” |  | /t/ |

**Appendix AH.** Rhyme Detection task (correct response in grey)

| Probe image | Image 1 | Image 2 | Image 3 |
|---|---|---|---|
|  |  |  |  |
| [Boat] | [Foot] | [Bike] | [Coat] (grey) |
|  |  |  |  |
| [Key] | [Cow] | [Tree] (grey) | [Door] |
|  |  |  |  |
| [Chair] | [Car] | [Table] | [Bear] (grey) |
|  |  |  |  |
| [House] | [Mouse] (grey) | [Horse] | [Window] |
|  |  |  |  |
| [Head] | [Hand] | [Bed] (grey) | [Eye] |

| | | | |
|---|---|---|---|
|  |  |  |  |
| [Sock] | [Clown] | [Clock] | [Shoe] |
|  |  |  |  |
| [Train] | [Rain] | [Tractor] | [Spoon] |
|  |  |  |  |
| [Egg] | [Bag] | [Spoon] | [Leg] |
|  |  |  |  |
| [Car] | [Star] | [Bike] | [Cake] |
|  |  |  |  |
| [Ball] | [Wall] | [Bell] | [Bag] |

**Appendix AI.** GPC selection task stimuli

| GPC | GPC (2) | Word | GPC | GPC (2) | Word |
|---|---|---|---|---|---|
| a-e | | flake | ay | | fray |
| a-e | | grape | ay | | hurray |
| a-e | | spade | ay | | sway |
| a-e | | slate | ay | | relay |
| a-e | | crane | ch | th | thatch |
| ai | th | faith | ch | | munch |
| ai | | stain | ch | ee | leech |
| ai | | wail | ch | o-e | choke |
| ai | | raisin | ck | | snack |
| ai | qu | quaint | ck | | rack |
| air | | dairy | ck | | bracket |
| air | | repair | ck | le | buckle |
| air | | fair | ck | | frock |
| air | | flair | dg/dge | | bridge |
| air | | chair | dg/dge | | ridge |
| ar | | collar | dg/dge | | hedgehog |
| ar | | alarm | dg/dge | | wedge |
| ar | sh | shark | ea | | scream |
| ar | | regular | ea | qu | squeal |
| ar | | smart | ea | ch | preach |
| are | | glare | ea | ch | cheat |
| are | | aware | ea | | yeast |
| are | | care | ear | | nearly |
| are | | rarely | ear | | year |
| are | sh | share | ear | | clear |
| au | | fraud | ear | | tears |
| au | ch | launch | ear | | beard |
| au | | sauce | ee | | deed |
| au | | haunt | ee | le | beetle |
| au | | haul | ee | | jeep |
| ay | | spray | ee | | keel |

| | | |
|---|---|---|
| er | | herd |
| er | | verse |
| er | | fern |
| er | | term |
| er | | nerve |
| ere | | whereby |
| ere | | Somewhere |
| ere | | coherence |
| ere | | adhere |
| ere | | here |
| ie | | shriek |
| ie | | yield |
| ie | | grief |
| ie | | pier |
| ie | | siege |
| i-e | | vine |
| i-e | | review |
| i-e | | slime |
| i-e | | stripe |
| i-e | | missile |
| ir | | thirst |
| ir | | skirt |
| ir | | first |
| ir | | circus |
| ir | | bird |
| kn | | knitter |
| kn | | knot |
| kn | dg/edge | knowledge |
| kn | | knoll |
| kn | | knight |
| le | | mumble |
| le | | fable |
| le | ng | jingle |

| | | |
|---|---|---|
| ng | | winger |
| ng | | cling |
| ng | | linger |
| ng | | hanger |
| oa | | croak |
| oa | | boast |
| oa | | foam |
| oa | | foal |
| oa | th | oath |
| o-e | qu | quote |
| o-e | | stole |
| o-e | | cope |
| o-e | | poke |
| oo | | spoon |
| oo | | spook |
| oo | | groom |
| oo | | brood |
| oo | | droop |
| ou | | generous |
| ou | | flavourful |
| ou | | accountant |
| ou | | ground |
| ou | | famous |
| ow | | willow |
| ow | | trow |
| ow | | bellow |
| ow | | crow |
| ow | | mow |
| oy | | coy |
| oy | | employ |
| oy | | destroy |
| oy | | royalty |
| oy | | boy |

| | | |
|---|---|---|
| ph | | phantom |
| ph | | aphid |
| ph | | photo |
| ph | a-e | phase |
| ph | | orphan |
| qu | | quilt |
| qu | | squash |
| sh | | shell |
| sh | | accomplish |
| sh | | hush |
| tch | | scratch |
| tch | | patch |
| tch | | stitch |
| tch | | ketchup |
| tch | | pitcher |
| th | | lather |
| th | | theft |
| u-e | | jute |
| u-e | | fuse |
| u-e | | vulture |
| u-e | | cure |
| u-e | | duke |
| ur | | nurse |
| ur | | burst |

**Appendix AJ.** Cluster selection task stimuli

| able | ation | aise | arge | ation |
|---|---|---|---|---|
| tablecloth | paint | fundraiser | garage | stationary |
| disabled | maintenance | appraise | surcharge | communication |
| fable | dainty | praise | charge | illustration |
| capable | saint | raise | enlarge | imagination |
| vegetable | taint | mayonnaise | barge | national |
| audi | aught | aunt | auto | cess |
| audition | slaughter | flaunt | autonomy | necessary |
| audible | distraught | taunt | automobile | process |
| auditorium | draught | daunting | autograph | recess |
| audit | daughter | auntie | automated | inaccessibility |
| marauding | naughty | gaunt | autobiography | concession |
| eeze | eigh | ence | ench | ense |
| squeeze | weigh | confidence | wench | defense |
| wheeze | sleigh | difference | french | offense |
| freeze | eight | sentence | clench | suspense |
| tweezers | neigh | science | stench | nonsense |
| breeze | freight | experience | drench | license |
| erge | fore | ight | inge | inter |
| submerge | before | frighten | ginger | winter |
| detergent | forecast | knight | fringe | painter |
| verge | therefore | flight | linger | splinter |
| emerge | foreign | right | syringe | interactive |
| emergency | foreboding | delight | cringe | interest |
| itch | ment | ogue | oist | otch |
| bewitch | garment | analogue | joist | scotch |
| pitch | ailment | prologue | moist | notch |
| pitcher | elementary | catalogue | hoist | blotch |
| snitch | moment | rogue | boisterous | hopscotch |
| hitch | instrument | dialogue | soloist | botch |

| ough | ounce | ound | port | spect |
|---|---|---|---|---|
| cough | bounce | wound | portrait | spectacular |
| although | pounce | compound | airport | inspect |
| dough | announce | around | import | suspect |
| thorough | ounce | surround | support | aspect |
| tough | pronounce | mound | portable | respect |

| tract | tude | udge | unch |
|---|---|---|---|
| subtract | multitude | nudge | munch |
| distraction | attitude | judge | scrunch |
| traction | altitude | begrudge | haunch |
| attractive | student | smudge | launch |
| extract | latitude | budget | brunch |

**Appendix AK.** Accuracy and reading latencies (ms) for each participant on the GPC selection test.

| | HE | | | MI | | | OK | | | PI | | | FR | | | SH | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Phonological | | | Phonological | | | Phonological | | | Phonological | | | Whole-word | | | Whole-word | | |
| | GPC | ACC | RT | GPC | ACC | RT | GPC | ACC | RT | GPC | ACC | RT | GPC | ACC | RT | GPC | ACC | RT |
| 1 | th | .60 | 1301 | ph | 0.40 | 862 | ie | 0.40 | 1962 | ie | 0.60 | 2227 | le | 0.60 | 1780 | oa | 0.60 | 2162 |
| 2 | air | .80 | 843 | ai | 0.40 | 788 | ai | 0.40 | 1503 | ph | 0.60 | 1906 | oo | 0.60 | 1430 | ie | 0.60 | 2093 |
| 3 | ay | .80 | 907 | ch | 0.57 | 854 | th | 0.40 | 1448 | ow | 0.60 | 1462 | qu | 0.60 | 1275 | th | 0.60 | 1500 |
| 4 | ee | .80 | 1039 | ere | 0.60 | 1295 | ng | 0.40 | 1367 | th | 0.80 | 6869 | ea | 0.60 | 1050 | ck | 0.60 | 856 |
| 5 | kn | .80 | 993 | th | 0.60 | 1171 | ph | 0.60 | 2387 | oa | 0.80 | 3972 | ai | 0.60 | 1023 | ee | 0.60 | 785 |
| 6 | ng | .80 | 1278 | oy | 0.60 | 1098 | oa | 0.60 | 1250 | ng | 0.80 | 3654 | ow | 0.60 | 934 | ai | 0.80 | 5154 |
| 7 | oo | .80 | 818 | oo | 0.60 | 907 | qu | 0.60 | 995 | ck | 0.80 | 1588 | kn | 0.80 | 2109 | ow | 0.80 | 3500 |
| 8 | ow | .80 | 1014 | qu | 0.60 | 855 | au | 0.60 | 987 | air | 0.80 | 1062 | ph | 0.80 | 1939 | oo | 0.80 | 1910 |
| 9 | ph | .80 | 805 | au | 0.60 | 842 | ch | 0.71 | 1390 | oy | 0.80 | 992 | ie | 0.80 | 1848 | ere | 0.80 | 1767 |
| 10 | u-e | .80 | 947 | u-e | 0.60 | 842 | i-e | 0.80 | 2291 | ay | 0.80 | 934 | ee | 0.80 | 1801 | tch | 0.80 | 1294 |
| 11 | ere | 1 | 1229 | ck | 0.80 | 1081 | ere | 0.80 | 1646 | are | 0.80 | 842 | o-e | 0.80 | 1520 | le | 0.80 | 1186 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | le | 1 | 1203 | ow | 0.80 | 985 | ea | 0.80 | 1335 | a-e | 0.83 | 1805 | tch | 0.80 | 1323 | ph | 0.80 | 1117 |
| 13 | oa | 1 | 1127 | o-e | 0.80 | 954 | ou | 0.80 | 1326 | ere | 1.00 | 4038 | ck | 0.80 | 1119 | qu | 0.80 | 1073 |
| 14 | ai | 1 | 1095 | ou | 0.80 | 913 | ee | 0.80 | 1252 | ai | 1.00 | 3549 | i-e | 0.80 | 1079 | ou | 0.80 | 1015 |
| 15 | ou | 1 | 1087 | ay | 0.80 | 888 | ow | 0.80 | 1140 | ch | 1.00 | 3374 | air | 0.80 | 944 | ay | 0.80 | 770 |
| 16 | ch | 1 | 1009 | ear | 0.80 | 885 | le | 0.80 | 966 | le | 1.00 | 2600 | ch | 0.86 | 1400 | sh | 0.80 | 542 |
| 17 | ea | 1 | 982 | i-e | 0.80 | 822 | oo | 0.80 | 915 | i-e | 1.00 | 2522 | oy | 1.00 | 3686 | air | 1.00 | 1659 |
| 18 | are | 1 | 979 | sh | 0.80 | 751 | ay | 0.80 | 833 | ear | 1.00 | 2454 | ere | 1.00 | 2914 | u-e | 1.00 | 1634 |
| 19 | oy | 1 | 948 | ea | 0.80 | 740 | u-e | 0.80 | 762 | oo | 1.00 | 2322 | th | 1.00 | 2739 | o-e | 1.00 | 1606 |
| 20 | ir | 1 | 931 | ie | 1.00 | 1311 | ck | 1.00 | 2375 | qu | 1.00 | 2220 | oa | 1.00 | 2535 | ng | 1.00 | 1257 |
| 21 | dg | 1 | 913 | ng | 1.00 | 1309 | ear | 1.00 | 1795 | u-e | 1.00 | 2087 | ou | 1.00 | 1781 | a-e | 1.00 | 1090 |
| 22 | tch | 1 | 891 | le | 1.00 | 1253 | oy | 1.00 | 1493 | ee | 1.00 | 2081 | ir | 1.00 | 1615 | i-e | 1.00 | 1036 |
| 23 | a-e | 1 | 863 | kn | 1.00 | 1161 | a-e | 1.00 | 1293 | ar | 1.00 | 1953 | ng | 1.00 | 1597 | au | 1.00 | 920 |
| 24 | qu | 1 | 855 | dg | 1.00 | 1000 | ir | 1.00 | 1240 | au | 1.00 | 1780 | u-e | 1.00 | 1529 | er | 1.00 | 906 |
| 25 | sh | 1 | 845 | oa | 1.00 | 979 | o-e | 1.00 | 1109 | ou | 1.00 | 1744 | ar | 1.00 | 1447 | ch | 1.00 | 855 |
| 26 | ur | 1 | 835 | ee | 1.00 | 912 | kn | 1.00 | 1098 | kn | 1.00 | 1683 | ear | 1.00 | 1381 | ear | 1.00 | 849 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | au | 1 | 833 | are | 1.00 | 903 | air | 1.00 | 1072 | ir | 1.00 | 1599 | ay | 1.00 | 1371 | oy | 1.00 | 832 |
| 28 | ear | 1 | 812 | er | 1.00 | 897 | are | 1.00 | 949 | sh | 1.00 | 1529 | a-e | 1.00 | 1359 | ir | 1.00 | 801 |
| 29 | ie | 1 | 791 | tch | 1.00 | 880 | dg | 1.00 | 931 | ea | 1.00 | 1508 | sh | 1.00 | 1345 | ea | 1.00 | 717 |
| 30 | er | 1 | 767 | air | 1.00 | 844 | tch | 1.00 | 922 | o-e | 1.00 | 1419 | au | 1.00 | 1304 | kn | 1.00 | 701 |
| 31 | ck | 1 | 731 | ar | 1.00 | 772 | ar | 1.00 | 871 | ur | 1.00 | 1345 | are | 1.00 | 1283 | are | 1.00 | 682 |
| 32 | i-e | 1 | 724 | ir | 1.00 | 772 | er | 1.00 | 851 | tch | 1.00 | 1150 | dg | 1.00 | 1216 | ar | 1.00 | 672 |
| 33 | ar | 1 | 721 | a-e | 1.00 | 765 | sh | 1.00 | 839 | dg | 1.00 | 1084 | er | 1.00 | 1210 | dg | 1.00 | 644 |
| 34 | o-e | 1 | 702 | ur | 1.00 | 710 | ur | 1.00 | 716 | er | 1.00 | 973 | ur | 1.00 | 1184 | ur | 1.00 | 639 |

**Appendix AL.** Accuracy and reading latencies (ms) for each participant on the Cluster selection test.

| | HE Phonological | | | MI Phonological | | | OK Phonological | | | PI Phonological | | | FR Whole-word | | | SH Whole-word | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | GPC | ACC | RT | GPC | ACC | RT | GPC | ACC | RT | GPC | ACC | RT | GPC | ACC | RT | GPC | ACC | RT |
| 1 | oist | 0.60 | 975 | ogue | 0.00 | na | audi | 0.00 | na | audi | 0.20 | 11553 | tude | 0.60 | 1705 | auto | 0.20 | 3521 |
| 2 | cess | 0.80 | 1636 | audi | 0.20 | 1710 | udge | 0.20 | 962 | eigh | 0.20 | 843 | eigh | 0.60 | 1541 | aise | 0.40 | 2968 |
| 3 | audi | 0.80 | 1422 | itch | 0.40 | 1613 | unch | 0.40 | 2191 | auto | 0.60 | 4522 | cess | 0.60 | 1509 | fore | 0.40 | 1150 |
| 4 | arge | 0.80 | 1078 | eigh | 0.40 | 1018 | aunt | 0.40 | 1659 | oist | 0.60 | 2589 | ench | 0.60 | 1197 | eigh | 0.60 | 6663 |
| 5 | aught | 0.80 | 991 | ment | 0.40 | 1007 | aise | 0.60 | 2112 | fore | 0.60 | 1282 | unch | 0.60 | 893 | audi | 0.60 | 2990 |
| 6 | tract | 0.80 | 936 | ounce | 0.40 | 953 | auto | 0.60 | 1726 | aught | 0.75 | 2620 | fore | 0.60 | 856 | arge | 0.60 | 2394 |
| 7 | ough | 0.80 | 882 | aint | 0.40 | 883 | inge | 0.60 | 1418 | tude | 0.80 | 3894 | aught | 0.60 | 770 | udge | 0.60 | 2362 |
| 8 | ence | 0.80 | 759 | aught | 0.60 | 1354 | ation | 0.60 | 1360 | ough | 0.80 | 3784 | audi | 0.80 | 2851 | cess | 0.60 | 1166 |
| 9 | udge | 1.00 | 1677 | auto | 0.60 | 1092 | aint | 0.60 | 1126 | ation | 0.80 | 3663 | inge | 0.80 | 1593 | oist | 0.80 | 6653 |
| 10 | auto | 1.00 | 1643 | cess | 0.60 | 1050 | aught | 0.60 | 917 | aint | 0.80 | 3037 | erge | 0.80 | 1587 | ment | 0.80 | 2836 |
| 11 | ogue | 1.00 | 1376 | fore | 0.60 | 906 | ough | 0.60 | 842 | inge | 0.80 | 2274 | auto | 0.80 | 1470 | ough | 0.80 | 2045 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | tude | 1.00 | 1250 | oist | 0.60 | 853 | cess | 0.60 | 782 | inter | 0.80 | 2128 | able | 0.80 | 1426 | aught | 0.80 | 1875 |
| 13 | eigh | 1.00 | 1132 | tract | 0.80 | 1328 | erge | 0.80 | 2317 | aunt | 0.80 | 1949 | aunt | 0.80 | 1364 | erge | 0.80 | 1790 |
| 14 | aise | 1.00 | 1016 | aise | 0.80 | 1277 | oist | 0.80 | 2006 | arge | 1.00 | 4795 | ough | 0.80 | 1352 | inge | 0.80 | 1629 |
| 15 | itch | 1.00 | 997 | tude | 0.80 | 1225 | ounce | 0.80 | 1989 | ense | 1.00 | 4123 | ment | 0.80 | 1320 | ogue | 0.80 | 1328 |
| 16 | ench | 1.00 | 995 | ense | 0.80 | 1219 | port | 0.80 | 1944 | ence | 1.00 | 3851 | arge | 0.80 | 1212 | able | 0.80 | 1273 |
| 17 | erge | 1.00 | 952 | able | 0.80 | 1141 | ogue | 0.80 | 1495 | ment | 1.00 | 3507 | ense | 0.80 | 1179 | ound | 0.80 | 1019 |
| 18 | aunt | 1.00 | 938 | erge | 0.80 | 1119 | tude | 0.80 | 1449 | udge | 1.00 | 3438 | ogue | 0.80 | 1131 | port | 0.80 | 953 |
| 19 | ation | 1.00 | 931 | inge | 0.80 | 1097 | ment | 0.80 | 1447 | able | 1.00 | 3365 | udge | 0.80 | 941 | unch | 0.80 | 829 |
| 20 | unch | 1.00 | 929 | ound | 0.80 | 1032 | eigh | 0.80 | 1420 | cess | 1.00 | 3302 | ation | 1.00 | 1895 | aint | 1.00 | 1681 |
| 21 | aint | 1.00 | 918 | ough | 0.80 | 1021 | able | 0.80 | 1180 | aise | 1.00 | 2840 | ounce | 1.00 | 1816 | ench | 1.00 | 1345 |
| 22 | inter | 1.00 | 908 | udge | 0.80 | 1008 | ense | 0.80 | 1128 | ench | 1.00 | 2638 | ence | 1.00 | 1712 | aunt | 1.00 | 1076 |
| 23 | ight | 1.00 | 902 | ation | 0.80 | 992 | tract | 0.80 | 1016 | tract | 1.00 | 2296 | inter | 1.00 | 1474 | ense | 1.00 | 1007 |
| 24 | able | 1.00 | 898 | ight | 0.80 | 978 | arge | 1.00 | 2367 | ogue | 1.00 | 2192 | spect | 1.00 | 1460 | ation | 1.00 | 984 |
| 25 | fore | 1.00 | 879 | ence | 0.80 | 909 | inter | 1.00 | 1885 | erge | 1.00 | 2022 | tract | 1.00 | 1458 | ence | 1.00 | 966 |
| 26 | ounce | 1.00 | 877 | arge | 1.00 | 1391 | fore | 1.00 | 1726 | ound | 1.00 | 1997 | aise | 1.00 | 1185 | tract | 1.00 | 963 |

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | eeze | 1.00 | 867 | aunt | 1.00 | 1194 | itch | 1.00 | 1684 | unch | 1.00 | 1979 | eeze | 1.00 | 1165 | ounce | 1.00 | 854 |
| 28 | ment | 1.00 | 866 | eeze | 1.00 | 1115 | ence | 1.00 | 1336 | ounce | 1.00 | 1921 | oist | 1.00 | 1123 | spect | 1.00 | 823 |
| 29 | inge | 1.00 | 855 | port | 1.00 | 1081 | ound | 1.00 | 1230 | spect | 1.00 | 1660 | itch | 1.00 | 1100 | tude | 1.00 | 817 |
| 30 | spect | 1.00 | 838 | inter | 1.00 | 1071 | ight | 1.00 | 1080 | eeze | 1.00 | 1493 | ight | 1.00 | 1099 | ight | 1.00 | 812 |
| 31 | ound | 1.00 | 786 | unch | 1.00 | 1028 | ench | 1.00 | 1048 | port | 1.00 | 1461 | aint | 1.00 | 1013 | itch | 1.00 | 796 |
| 32 | port | 1.00 | 763 | ench | 1.00 | 1013 | eeze | 1.00 | 947 | ight | 1.00 | 1311 | ound | 1.00 | 882 | eeze | 1.00 | 733 |
| 33 | ense | 1.00 | 732 | spect | 1.00 | 910 | spect | 1.00 | 889 | itch | 1.00 | 973 | port | 1.00 | 760 | inter | 1.00 | 732 |

**Appendix AM.** Intervention outline

| Block | Target block 1 | | | | | | Non-target block | | | | | Target block 2 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant | HE | MI | OK | PI | FR | SH | HE | MI | OK | FR | SH | HE | FR |
| 1 | th | ph | ai | ie | tude | auto | oist | ogue | audi | le | oa | le | able |
| 1 | / | / | ie | ph | / | / | / | / | / | / | ie | / | / |
| 2 | air | ai | ng | ow | eigh | aise | cess | audi | udge | oo | th | oa | aunt |
| 2 | ay | ch | th | th | cess | fore | audi | itch | unch | qu | ck | ai | ough |
| 3 | ee | ere | ph | oa | ench | eigh | arge | eigh | aunt | ea | ee | ou | ment |
| 3 | kn | th | qu | ng | unch | audi | aught | ment | aise | ai | ai | ch | arge |
| 4 | ng | oy | au | ck | fore | arge | tract | ounce | auto | ow | ow | ea | ense |
| 4 | oo | oo | oa | air | aught | udge | ough | aint | inge | kn | oo | are | ogue |
| 5 | ow | qu | ch | oy | audi | cess | ence | aught | ation | ph | ere | oy | udge |
| 5 | ph | au | i-e | ay | inge | oist | udge | auto | aint | ie | tch | ir | ation |
| 6 | u-e | u-e | ere | are | erge | ment | auto | cess | aught | ee | le | dg | ounce |
| 6 | ere | ck | ea | a-e | auto | / | ogue | fore | / | o-e | ph | tch | ence |

**Appendix AN.** Intervention participant's target GPC word list

| ai | air | are | au | ay |
|---|---|---|---|---|
| quaint | doubt | awareness | fraud | fray |
| bait | accountant | glare | saucer | array |
| braid | encounter | rarely | saunter | relay |
| frail | flourish | share | vault | spray |
| hail | lounge | dare | daunting | sway |
| maiden | mountain | flare | gaunt | crayfish |
| raid | compound | caregiver | laundry | betray |
| claim | council | declare | exhausted | delay |
| strain | discount | software | authorise | stray |
| tailor | voucher | square | taunt | mayonnaise |
| ch | ck | dg | ea | ee |
| leech | bracket | judge | appeal | beeper |
| poacher | backfire | nudge | preach | sleek |
| chemistry | cricketer | bridge | squeal | feeble |
| chant | flock | hedgehog | leash | sleeve |
| charmless | locker | knowledge | squeak | trainee |
| chatterbox | packet | dislodge | bleat | reef |
| chisel | pickled | budget | peat | agree |
| mechanism | dockside | ledge | gleam | beekeeper |
| hunch | flicker | begrudge | beaker | cheeky |
| perch | fickle | wedge | decrease | exceed |
| ere | ie | ir | kn | le |
| somewhere | nutrient | mirth | knight | supple |
| adhere | species | quirk | knead | battle |
| atmosphere | grieve | thirst | knife | castle |
| coherence | siege | affirm | knitter | adaptable |
| hemisphere | diesel | encircled | knack | suggestible |
| insincere | fiend | confirm | kneeling | bristle |
| interfere | niece | fir | slipknot | compile |
| sphere | priest | girl | knowhow | cubicle |
| there | brie | birch | penknife | wrangle |
| severe | thief | thirty | acknowledge | tremble |

| ng | oa | o-e | oo | ou |
|---|---|---|---|---|
| lingering | foal | cope | groovy | doubt |
| belong | bloated | awesome | monsoon | accountant |
| engage | moaned | quote | taboo | encounter |
| bangle | groaned | console | scooter | flourish |
| bungalow | coaxing | encore | voodoo | lounge |
| clang | cloak | dome | trooped | mountain |
| cliffhanger | abroad | grove | doom | compound |
| congratulate | gloat | tote | brood | council |
| dungarees | soar | restore | groom | discount |
| finger | oath | nitrogen | maroon | voucher |
| **ow** | **oy** | **ph** | **qu** | **tch** |
| bellow | boy | emphasis | quail | patch |
| burrow | coy | phase | quarrel | scratch |
| bestow | employ | autograph | quit | stitch |
| harrowing | boysenberry | dolphin | squawk | botch |
| marrow | loyal | nephew | equality | dispatch |
| narrow | ploy | euphoric | frequency | fetching |
| sorrow | soy | triumph | inadequate | tetchy |
| sparrow | flamboyant | pheasant | quad | catchphrase |
| stow | enjoyed | physics | tranquilizer | clutch |
| swallow | oysters | saxophone | squeeze | glitch |
| **th** | **u-e** | | | |
| thrill | cure | | | |
| seventh | fuse | | | |
| aftermath | immune | | | |
| thawed | amused | | | |
| withered | computer | | | |
| dither | tribute | | | |
| moth | flute | | | |
| thorn | fume | | | |
| empathy | mute | | | |
| loath | prune | | | |

**Appendix AO.** GPC training sentence stimuli

| GPC | Sentence |
|---|---|
| ai | The frail maiden who wore braids watched the tailor as he went to raid the bait in the hail. |
| air | The millionaire was in despair at the airport because she had lost her diamond solitaire ring at the fairground earlier in the day. |
| are | People in the square rarely dared to share the shade, however one day the sun's glare was too much. |
| au | The exhausted girl felt it was daunting when the gaunt man at the laundry started to taunt her. |
| ay | The crayfish got delayed as he swayed and strayed into the spray of the water. |
| ch | The charmless man was a chatterbox with the lady who had a hunch while she chanted to the leech in the hutch. |
| ck | The cricketer took the packets of pickled onions out of his locker and gave some to flock of sheep in the dockside. |
| dg | The hedgehog had knowledge about who had used the wedge to keep the bridge gate open while they stole the cartridge. |
| ea | The bleating sheep pulled off his leash with a squeak as he jumped over the peat with a gleam in his eye. |
| ee | The sleek but feeble mouse was cheeky when he went down the beekeepers' sleeve at the reef. |
| ere | Somewhere there on the atmosphere of the southern hemisphere there was a severe hole that meant unhealthy gases were able to interfere. |
| ie | The priest grieved as his niece was a fiend and a thief and stole diesel. |
| ir | The thirsty girl encircled the birch and fir trees thirty times. |
| kn | The knight got the knack of kneading the dough and used a knife to cut it while kneeling. |

| | |
|---|---|
| le | There was a tremble in the castle during the battle and the adaptable woman was trying to wrangle the animals. |
| ng | The bungalow belonged to the man in dungarees who wore a bangle as he went to congratulate engaged woman. |
| oa | The bloated foal moaned and groaned but didn't need coaxing to take off his cloak. |
| o-e | The awesome child carried a tote while in the dome at the grove watching the encore. |
| oo | The groovy groom got caught in the monsoon on his scooter while he trooped to the church. |
| ou | The council gave out a discount voucher to the accountant for her work with the compound on the mountain. |
| ow | The sparrow was full of sorrow as he watched the harrowing swallow stow a marrow in the burrow. |
| oy | The flamboyant boy enjoyed his boysenberry milkshake and oysters so much he ate them all quickly. |
| ph | The woman's nephew is good at physics, plays the saxophone, likes pheasants and dolphins and collects autographs. |
| qu | The squawking quail squeezed out of the cage and quit after she had a quarrel with the other birds. |
| tch | The seamstress tried to stitch the patch onto the fetching jeans but she botched it up and scratched herself. |
| th | He got a thrill on his seventh birthday when the snow thawed and saw a withered moth on a thorn. |
| u-e | The girl was amused as she ate a prune when the mute computer fumed and sounded like a flute. |

**Appendix AP.** Intervention participant's cluster word list

| able | aint | arge | ation | audi |
|---|---|---|---|---|
| capable | acquaintance | argent | abbreviation | applauding |
| fable | certainty | barge | acceleration | audible |
| imaginable | constraint | discharged | articulation | audience |
| liable | maintenance | enlarged | circulation | audio |
| personable | mountaintop | large | combination | audiologist |
| reasonable | paint | overcharge | illustration | audit |
| sable | plaintiff | recharge | imagination | audition |
| syllable | plaintive | surcharge | national | auditorium |
| tableau | restraint | target | nationality | defrauding |
| vegetable | taint | undercharged | stationary | marauding |
| **aught** | **aunt** | **auto** | **cess** | **eigh** |
| caught | auntie | autobiography | access | bobsleigh |
| daughter | daunting | autograph | cessation | counterweight |
| distraught | flaunt | autoimmune | concession | eight |
| draught | jaunty | automated | inaccessibility | freight |
| fraught | gaunt | automobile | incessant | height |
| haughty | gauntlet | autonomic | necessary | heightened |
| laughter | haunted | autonomous | princess | lightweight |
| naught | jaunt | autopilot | process | neigh |
| naughty | saunter | autopsy | recession | neighbourhood |
| onslaught | taunt | semiautomatic | succession | weight |

| ence | ench | ense | erge | fore |
|---|---|---|---|---|
| adolescence | benchmark | condense | allergen | aforementioned |
| circumference | clench | dispense | concierge | before |
| confidence | disenchanting | ensemble | detergent | foreboding |
| conscience | drench | expense | diverge | forecast |
| evidence | enchanting | immense | emerge | forehead |
| experience | penchant | intense | emergency | foreign |
| inconvenience | quench | nonsense | energetic | foresee |
| licence | stench | sense | poltergeist | forest |
| patience | trench | suspense | submerge | pinafore |
| science | wench | tense | verge | therefore |
| **inge** | **itch** | **ment** | **ogue** | **oist** |
| contingency | bewitch | ailment | brogues | boisterous |
| cringe | chitchat | argumentative | catalogue | cloistered |
| ginger | ditched | commentary | collogue | egoistic |
| harbinger | glitch | developmental | dialogue | foist |
| ingest | hitch | equipment | drogue | hoist |
| linger | pitch | excitement | monologue | joist |
| singed | pitcher | government | prologue | moist |
| stringent | snitch | instruments | rogue | moisture |
| syringe | switch | mentor | synagogue | oboist |
| tinged | twitch | moment | travelogue | soloist |

| ough | ounce | tract | tude | udge |
|---|---|---|---|---|
| although | bounced | abstract | amplitude | begrudged |
| cough | denounce | attractive | aptitude | budgerigar |
| dough | flounce | contract | fortitude | budget |
| enough | mispronounce | distracted | gratitude | drudge |
| hiccough | ounces | extract | latitude | dudgeon |
| thorough | pounced | extraction | longitude | fudge |
| though | pronounce | protracted | magnitude | judge |
| through | renounce | retract | multitude | misjudge |
| tough | trounced | traction | platitude | smudged |
| trough | announced | tractor | student | trudge |

| unch |
|---|
| brunch |
| crunch |
| haunch |
| launched |
| lunch |
| munch |
| scrunched |
| staunch |
| truncheon |
| uncharitable |

**Appendix AQ.** Visual training sentence stimuli

| Cluster | Sentence |
|---|---|
| able | The fable told the story of a man who was very reasonable and personable but not capable of eating vegetables. |
| aint | The maintenance woman went up to the mountaintop with the paint tins and had certainty it would impress her acquaintance. |
| arge | The large man paid a surcharge after he drove into the enlarged barge like a target. |
| ation | The boy used the stationary to draw an illustration of the combination of old and new national flags from his imagination. |
| audi | The audience were applauding in the auditorium because the audio from the music audition was very good. |
| aught | The naughty daughter was distraught because the draught blew away the leaves which she had caught. |
| aunt | The haunted house was daunting to the gaunt girl but her auntie was busy flaunting her jewels. |
| auto | The autopsy doctor got in her automobile after signing her autograph on her autobiography about autoimmune health. |
| cess | The process was necessary to ensure the princess had access to the concession stand. |
| eigh | The height and weight of the freight train meant that it could not get into the eight neighbourhoods to deliver the goods. |
| ence | The science teacher had confidence and patience because of all the experience she had since she got her licence. |
| ench | The stench of cheese from the trench was disenchanting and did not quench the thirst of the girl who was drenched. |

| | |
|---|---|
| ense | The tense man was in immense suspense about the expense of his holiday while trying to condense his suitcase. |
| erge | The emergency kit was on the verge of being submerged in the detergent by the energetic child. |
| fore | The foreign man took a hat to the forest to keep his forehead dry as the forecast said it would rain before he left the house. |
| inge | The ginger man cringed and felt a tinge when the syringe went in his arm and didn’t want it to linger. |
| itch | The man ditched the pitcher of water when his arm twitched making him scream in a high pitch and switch to a water bottle. |
| ment | There was excitement in the government the moment the new equipment arrived for cleaning the instruments. |
| ogue | The rogue woman bought a catalogue and a book, she liked reading the dialogue and the monologues in the prologue. |
| oist | The egoistic girl was very boisterous as she hoisted the piano for the soloist through the moist grass. |
| ough | Although the man had a cough, he was still thorough enough when he made the dough mixture. |
| ounce | The doctor announced the dog’s weight in ounces after it trounced the car when it pounced and bounced on it. |
| tract | The attractive chef was distracted by the abstract painting of a tractor when he was trying to sign the contract. |
| tude | There were a multitude of reasons the student had a lot of gratitude for the magnitude of work the teacher did to help them understand amplitude. |
| udge | The judge checked the budget because he begrudged the fudge the man had eaten and smudged on the paper. |
| unch | After lunch the staunch woman scrunched up the bunch of flowers and launched them at the bin. |

**Appendix AR.** Pre-post test

| Regular words | Irregular words | Nonwords |
| --- | --- | --- |
| | Block 1 pre-post test | |
| adopt | boulder | baft |
| blares | bowl | bick |
| call | clue | bield |
| clinic | colonel | bine |
| despair | dough | bleaner |
| feet | gauge | buke |
| flannel | give | cloam |
| lack | gone | doash |
| land | good | drick |
| least | high | hain |
| long | know | maunch |
| marsh | liquid | peef |
| plant | meringue | peng |
| reason | minute | prane |
| sixteen | rich | reep |
| slide | soul | reetle |
| soil | sure | scarrow |
| stench | war | snay |
| tail | wool | trang |
| take | yacht | wist |
| | Block 2 pre-post test | |

| | | |
|---|---|---|
| cane | bouquet | boril |
| chance | bruise | borp |
| check | ceiling | brennet |
| display | choir | clemty |
| hand | circuit | ganten |
| hook | come | gop |
| luck | cough | greal |
| market | engine | hest |
| mist | eye | jart |
| name | head | moof |
| nerve | iron | norf |
| north | island | pite |
| nurse | live | prite |
| peril | mayor | seldent |
| plug | music | spatch |
| sat | ocean | stendle |
| scarf | shoe | sunten |
| sister | tomb | tribble |
| text | villain | troat |
| weasel | work | virth |

Block 3 pre-post test

| | | |
|---|---|---|
| bed | acid | aspy |
| brandy | blood | brinth |
| chicken | bomb | chike |

| | | |
|---|---|---|
| context | break | crat |
| cord | brooch | delk |
| curb | debt | farl |
| drop | diesel | framp |
| free | foreign | grenty |
| life | friend | gurve |
| middle | geyser | phurp |
| navy | limb | pofe |
| need | lose | quist |
| nine | much | rint |
| part | none | slont |
| please | once | stome |
| possum | pint | swull |
| prevent | pretty | tapple |
| pump | routine | trool |
| wait | shove | trope |
| wedding | wolf | vock |

**Appendix AS.** GPC training review (untrained) stimuli

| ai | air | are | au |
|---|---|---|---|
| wail | mohair | compare | haul |
| painstaking | repair | barely | thesaurus |
| reacquainted | staircase | farewell | laughable |
| ay | ch | ck | dg |
| essay | orchard | frock | porridge |
| decay | changeable | checkmate | ridge |
| layering | purchase | feedback | fidget |
| ea | ee | ere | ie |
| yeast | deed | here | yield |
| beachfront | creep | mere | shriek |
| colleague | keel | persevere | worried |
| ir | kn | le | ng |
| smirk | knot | mumble | winger |
| squirm | knoll | bauble | rung |
| swirl | knuckles | braille | amongst |
| oa | o-e | oo | ou |
| boast | pope | zoom | abound |
| upload | ozone | proof | scour |
| roam | corrode | loom | bountiful |
| ow | oy | ph | qu |
| mow | destroy | phantom | squash |
| hallow | royalty | pharmacy | quilt |
| owner | annoy | geography | request |
| tch | th | u-e | |
| latch | lather | duke | |
| sketch | wrath | vulture | |
| wretched | frothy | exclude | |

**Appendix AT.** Visual training review (untrained) stimuli

| | | | |
|---|---|---|---|
| able | aint | arge | ation |
| inescapable | dainty | barged | communication |
| stable | faint | charge | operational |
| tablecloth | saint | enlargement | population |
| audi | aught | aunt | auto |
| audiotape | granddaughter | flaunting | autobiographical |
| auditor | slaughter | taunting | automatic |
| inaudible | taught | undaunted | autonomy |
| cess | eigh | ence | ench |
| accessorise | airfreight | difference | enchase |
| excess | sleigh | fluorescence | happenchance |
| successful | weigh | offence | hench |
| ense | erge | fore | inge |
| incense | intergeneration | forego | fringe |
| recompense | overgenerous | forewarning | hinge |
| senseless | serge | unforeseen | ringer |
| itch | ment | ogue | oist |
| itchy | announcement | analogue | foisted |
| kitchen | compliment | epilogue | moisturize |
| witch | mention | vogue | tattooist |
| ough | ounce | tract | tude |
| borough | announce | contractual | altitude |
| breakthrough | bouncer | detract | attitude |
| ought | unannounced | subtract | solitude |
| udge | unch | | |
| budge | bunch | | |
| nudge | punch | | |
| sludge | unchanged | | |

**Appendix AU.** Full model outputs from the individual analyses of primary outcome measures.

| Participant | Profile | Intervention | Testing time | Intervention X Testing time |
|---|---|---|---|---|
| | | Accuracy | | |
| HE | P | $z$= 1.04, $p$ = .298 | $z$= 0, $p$ = 1 | $z$= 1.20, $p$ = .230 |
| MI | P | $z$= 0.41, $p$ = .682 | $z$= 0, $p$ = 1 | $z$= 0.80, $p$ = .427 |
| OK | P | $z$= 2.09, $p$ = .037* | $z$= 4.31, $p$ < .001** | $z$= 2.90, $p$ = .004** |
| PI | P | $z$= 5.51, $p$ < .001 | *na* | *na* |
| FR | WW | $z$= 0.54, $p$ = .588 | $z$= 0.63, $p$ = .526 | $z$= 1.34, $p$ = .181 |
| SH | WW | $z$= 0.04, $p$ = .965 | $z$= 0.48, $p$ = .632 | $z$= 0.34, $p$ = .731 |
| | | Latency | | |
| HE | P | $F$(1,153)= 0.10, $p$ = .755 | $F$(1,153)= 27.23, $p$ < .001 | $F$(1,153)= 20.65, $p$ < .001** |
| MI | P | $F$(1,91)= 1.18, $p$ = .280 | $F$(1,91)= 0.44, $p$ = .509 | $F$(1,91)= 18.56, $p$ < .001** |
| OK | P | $F$(1,83)= 2.32, $p$ = .131 | $F$(1,83)= 0.01, $p$ = .905 | $F$(1,83)= 2.62, $p$ = .109 |
| PI | P | $F$(1,46)= 35.0, $p$ < .001** | *na* | *na* |
| FR | WW | $F$(1,145)= 0.27, $p$ = .606 | $F$(1,145)= 21.89, $p$ < .001** | $F$(1,153)= 3.92, $p$ = .050* |
| SH | WW | $F$(1,86)= 0.22, $p$ = .644 | $F$(1,86)= 6.52, $p$ = .012** | $F$(1,86)= 11.14, $p$ = .001** |

**Appendix AV.** Full (simple) model outputs from the group analyses of primary outcome measures.

| Group | Intervention | Testing time | Intervention X Testing time |
|---|---|---|---|
| | | Accuracy | |
| Phonological | *z*= 0.13, *p* = .900 | *z*= 2.17, *p* = .030* | *z*= 0.51, *p* = .612 |
| Whole-word | *z*= 0.41, *p* = .680 | *z*= 0.47, *p* = 0.64 | *z*= 0.86, *p* = .390 |
| | | Latency | |
| Phonological | *F*(1,166)= 1.95, *p* = .164 | *F*(1,611)= 18.74, *p* < .001** | *F*(1,611)= 25.03, *p* < .001** |
| Whole-word | *F*(1,139)= 0.09, *p* = .765 | *F*(1,297)= 24.11, *p* < .001** | *F*(1,297)= 13.84, *p* < .001** |

**Appendix AW.** Full (complex) model outputs from the group analyses of primary outcome measures.

| Group | Intervention | Testing time | Word type | Intervention X Testing time | Intervention X Word type | Testing time X Word type | Intervention X Testing time X Word type |
|---|---|---|---|---|---|---|---|
| | | | | Accuracy | | | |
| Phonological | *z*= 0.03, *p* = .974 | *z*= 1.50, *p* = .134 | *z*= 2.32, *p* = .021* | *z*= 0.93, *p* = .352 | *z*= 0.02, *p* = .982 | *z*= 0.86, *p* = .388 | *z*= 0.86, *p* = .388 |
| Whole-word | *z*= 0.99, *p* = .324 | *z*= 0.32, *p* = .746 | *z*= 0.58, *p* = .565 | *z*= 0.69, *p* = .488 | *z*= 1.04, *p* = .298 | *z*= 0.20, *p* = .845 | *z*= 0.49, *p* = .623 |
| | | | | Latency | | | |
| Phonological | *F*(1,748)= 2.62, *p* = .106 | *F*(1,745)= 17.56, *p* < .001** | *F*(1,745)= 38.28, *p* < .001** | *F*(1,745)= 2.48, *p* = .084 | *F*(1,745)= 1.05, *p* = .351 | *F*(1,745)= 22.06, *p* < .001** | *F*(1,745)= 0.19, *p* = .824 |
| Whole-word | *F*(1,457)= 1.34, *p* = .248 | *F*(1,457)= 18.74, *p* < .001** | *F*(1,457)= 75.06, *p* < .001** | *F*(1,457)= 4.17, *p* = .016* | *F*(1,457)= 0.93, *p* = .396 | *F*(1,457)= 11.27, *p* < .001** | *F*(1,457)= 0.95, *p* = .386 |

**Appendix AX.** Simple group analyses of primary outcome measures.

| Dependent variable | Intervention | Testing time | Impairment profile | Intervention X Testing time | Intervention X Impairment profile | Testing time X Impairment profile | Intervention X Testing time X Impairment profile |
|---|---|---|---|---|---|---|---|
| Accuracy | *z*= 0.50, *p* = .618 | *z*= 2.21, *p* = .027* | *z*= 0.71, *p* = .481 | *z*= 0.61, *p* = .542 | *z*= 0.35, *p* = .726 | *z*= 1.58, *p* = .113 | *z*= 0.94, *p* = .345 |
| Latency | *F*(1,1070)= 4.53, *p* = .034* | *F*(1,1041)= 40.45, *p* < .001** | *F*(1,4)= 0.34, *p* = .589 | *F*(1,1041)= 0.00, *p* = .948 | *F*(1,152)= 0.07, *p* = .787 | *F*(1,1041)= 1.57, *p* = .210 | *F*(1,1041)= 34.05, *p* < .001** |

**Appendix AY.** Complex group analyses of primary outcome measures.

| DV | I | TT | IP | WT | I x TT | I x IP | I x WT | TT x IP | TT x WT | IP x WT | I x TT x IP | I x TT x WT | I x IP x WT | TT x IP x WT | I x TT x IP x WT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ACC | $z = 0.04$, $p = .972$ | $z = 1.50$, $p = .135$ | $z = 0.58$, $p = .561$ | $z = 2.32$, $p = .021$* | $z = 0.93$, $p = .353$ | $z = 0.75$, $p = .451$ | $z = 0.02$, $p = .983$ | $z = 1.05$, $p = .296$ | $z = 0.86$, $p = .390$ | $z = 0.87$, $p = .384$ | $z = 1.12$, $p = .263$ | $z = 0.86$, $p = .389$ | $z = 0.78$, $p = .437$ | $z = 0.64$, $p = .524$ | $z = 0.93$, $p = .355$ |
| RT | $F(1,1205) = 3.81$, $p = .051$ | $F(1,1202) = 37.84$, $p < .001$** | $F(1,4) = 0.25$, $p = .643$ | $F(2,1202) = 121.15$, $p < .001$** | $F(1,1202) = 0.01$, $p = .909$ | $F(1,1205) = 0.00$, $p = .999$ | $F(2,1202) = 0.71$, $p = .492$ | $F(1,1202) = 1.54$, $p = .215$ | $F(2,1202) = 1.98$, $p = .139$ | $F(2,1202) = 13.91$, $p < .001$** | $F(1,1202) = 31.57$, $p < .001$** | $F(2,1202) = 0.61$, $p = .543$ | $F(2,1202) = 7.12$, $p < .001$** | $F(2,1202) = 0.09$, $p = .916$ | $F(2,1202) = 0.92$, $p = .400$ |

Note: a) DV= Dependent variable, b) ACC= Accuracy, c) RT= Latency d) I= Intervention, e) TT= Testing time, f) IP=Impairment profile, g) WT= Word type

**Appendix AZ.** Full model outputs from the individual analyses of secondary outcome measures.

| Participant | Profile | Intervention | Testing time | Intervention X Testing time |
|---|---|---|---|---|
| | | Trained word Accuracy | | |
| HE | P | *z*= 1.15, *p* = .248 | *z*= 3.87, *p* < .001** | *z*= 0.16, *p* = .538 |
| MI | P | *z*= 0.43, *p* = .671 | *z*= 3.29, *p* < .001** | *z*= 2.17, *p* = .030* |
| OK | P | *z*= 0.06, *p* = .952 | *z*= 2.47, *p* = .013* | *z*= 1.38, *p* = .168 |
| PI | P | *z*= 3.11, *p* = .002** | *na* | *na* |
| FR | WW | *z*= 2.00, *p* = .046* | *z*= 2.48, *p* = .013* | *z*= 1.98, *p* = .048 |
| SH | WW | *z*= 1.06, *p* = .291 | *z*= 6.63, *p* < .001** | *z*= 0.00, *p* = .997 |
| | | Redundant word Accuracy | | |
| OK | P | *z*= 2.09, *p* = .834 | *z*= 1.39, *p* = .165 | *z*= 0.95, *p* = .341 |
| PI | P | *z*= 0.46, *p* = .645 | *na* | *na* |
| SH | WW | *z*= 0.68, *p* = .500 | *z*= 1.46, *p* = .145 | *z*= 0.05, *p* = .957 |
| | | Redundant word Latency | | |
| OK | P | *F*(1,58)= 1.94, *p* = .058 | *F*(1,30)= 3.56, *p* < .001** | *F*(1,30)= 0.80, *p* = .431 |
| PI | P | *F*(1,32)= 0.31, *p* = .759 | *na* | *na* |
| SH | WW | *F*(1,58)= 0.47, *p* = .643 | *F*(1,30)= 1.87, *p* = .070 | *F*(1,30)= 0.60, *p* = .554*F*(1,30)= 11.14, *p* = .001** |

**Appendix BA.** Full model output for the group analyses of secondary measure accuracy (trained words).

| Group | Intervention | Testing time | Intervention X Testing time |
|---|---|---|---|
| Trained word Accuracy | | | |
| Phonological | *z*= 3.12, *p* = .002** | *z*= 4.12, *p* < .001** | *z*= 0.43, *p* = .665 |
| Whole-word | *z*= 3.37, *p* = .001** | *z*= 3.86, *p* < .001** | *z*= 1.51, *p* = .132 |

**Appendix BB.** Full model outputs from the group analyses of secondary outcome measures.

| | Intervention | Testing time | Impairment profile | Intervention X Testing time | Testing time X Impairment profile | Intervention X Impairment profile | Intervention X Testing time X Impairment profile |
|---|---|---|---|---|---|---|---|
| | | | | Accuracy | | | |
| Trained word | $z= 3.02$, $p = .003$** | $z= 3.65$, $p < .001$** | $z= 0.86$, $p = .392$ | $z= 0.56$, $p = .577$ | $z= 0.91$, $p = .362$ | $z= 1.35$, $p = .178$ | $z= 1.08$, $p = 280$ |
| Redundant word | $z= 0.62$, $p = .534$ | $z= 0.82$, $p = 414$ | $z= 2.03$, $p = .042$* | $z= 0.43$, $p = .670$ | $z= 0.09$, $p = .925$ | $z= 0.38$, $p = .704$ | $z= 0.26$, $p = .798$ |
| | | | | Latency | | | |
| Redundant word | $F(1,85)= 2.80$, $p = .006$** | $F(1,109)= 2.29$, $p = .024$* | $F(1,4)= 1.61$, $p = .179$ | $F(1,109)= 0.64$, $p = .527$ | $F(1,136)= 2.07$, $p = .040$* | $F(1,109)= 0.59$, $p = .559$ | $F(1,109)= 0.89$, $p = .377$ |