DOCTOR OF PHILOSOPHY

Examining the relationship between paired associate learning and reading ability in adults and children

Lira Calabrich, Simone

Award date:
2023

Awarding institution:
Bangor University

Link to publication

Examining the relationship between paired associate learning and reading ability in adults and children

Simone Lira Calabrich

Thesis submitted to the School of Human and Behavioural Sciences, Bangor University, in partial fulfilment of the requirement for the degree of Doctor of Philosophy

Bangor, United Kingdom

May 2023

'Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.'

Rwy'n cadarnhau fy mod yn cyflwyno'r gwaith gyda chytundeb fy Ngrichwyliwr (Goruchwylwyr)'

'I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.'

I confirm that I am submitting the work with the agreement of my Supervisor(s)'

Acknowledgements

Pursuing a PhD degree is a challenging and demanding process that requires dedication and persistence. Pursuing a PhD degree during a pandemic has added a layer of difficulty and unpredictability that cannot be overstated. Despite the challenges and uncertainties brought on by the pandemic, I am grateful for the opportunity to have completed my PhD and for the support and guidance that I have received along the way. In this acknowledgments page, I would like to express my appreciation to the individuals who have made this achievement possible.

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof Manon Jones, for believing in me and for guiding me throughout my PhD journey. Her unwavering support, patience, and guidance throughout my PhD were invaluable. I would like to thank her for always pushing me to become a better researcher and academic writer. I feel incredibly lucky to have had such a supportive supervisor by my side, who helped me overcome the challenges I faced during this process.

I would also like to thank my second supervisor, Dr Gary Oppenheim, for his insights and guidance on statistics in the early stages of my PhD. Additionally, I would also like to thank Dr Cameron Downing for being a kind, supportive, and encouraging collaborator, and for his unique contribution to this thesis.

I would like to express my sincere appreciation to the Department of Psychology at Bangor University for funding my PhD studies, and for their support throughout my academic journey. I would also like to express my gratitude to all the module organisers I have worked with as a graduate instructor, as well as my teaching mentor, for their invaluable support and guidance. I am grateful to be able to call

i

I dedicate this thesis to my three Marias.

Summary

The ability to establish *visual-phonological* associations correlates strongly with reading ability. Yet, the cognitive processes underlying paired associate learning trajectories of individuals with typical and atypical reading profiles are still somewhat unclear. Here, this gap is addressed *via* three research questions: 1) How do task-irrelevant episodic details modulate visual-phonological binding performance in individuals with and without dyslexia?, 2) Is there a specific cross-modal binding deficit in adults with developmental dyslexia, even in the absence of spoken output demands?, 3) Can online measures elucidate paired associate learning mechanisms in beginning readers? And what is the relationship between paired associate learning response accuracy and reading outcomes in these children? To this end, I created novel visual-phonological associations involving Kanji characters, unfamiliar to the participants, and consonant-vowel-consonant pseudowords that follow English phonotactics. Over four empirical chapters, I establish that: 1) when retrieving newly learnt visual-phonological associations from memory, individuals with dyslexia appear to rely heavily on episodic cues, particularly on consistent contextual information, a behavioural pattern that suggests that this group may have more fragile memory representations than their typical reading counterparts; 2) the evidence for a persistent cross-modal binding deficit in adult developmental dyslexia in the absence of phonological output remains somewhat unclear. Different experimental paradigms employed in this thesis yielded different results, thus highlighting the need for further investigation to determine the circumstances under which PAL without phonological output elicits a deficit in dyslexia; 3) the relationship between paired associate learning and reading appears to be mediated by rapid automatised naming, an ability that involves cross-modal associative mechanisms that are similar to those employed in PAL, albeit with overlearned stimuli; 4) even when accurate, individuals at the lower end of the reading spectrum are less confident about their knowledge of newly learnt visual-phonological associations. This thesis has made significant contributions to the field of reading research by shedding new light on some of the cognitive processes involved in paired associate learning of visual-phonological associations.

Table of Contents

List of Figures

Chapter 5. How Specific is the Paired Associate Learning Deficit in Dyslexia to Cross-Modal Working Memory?

Chapter 6. What the Hand in Motion Reveals about Reading: Children's Decision-Making Processes in Paired Associate Learning and its Relationship to Reading Outcomes

Appendices

List of Tables

Chapter 6. What the Hand in Motion Reveals about Reading: Children's Decision-Making Processes in Paired Associate Learning and its Relationship to Reading Outcomes

Appendices

Chapter 1

Literature Review and Thesis Aims

1. Chapter Overview

In this chapter, I will outline the core theoretical foundations of my thesis and the ensuing research questions, which are addressed in the subsequent empirical chapters.

1.1. The Importance of Reading

Reading is a complex and multi-faceted skill that is essential for personal, social, and professional success. In an age where access to information is at our fingertips, the ability to read and comprehend written text is paramount.

Individuals with a typical reading profile often take the ability to read for granted, not fully appreciating the complexity and sophistication needed to acquire this uniquely human skill. Some may even view reading as an effortless and automatic process, when, in reality, it is a highly specialised skill that requires several cognitive, linguistic, and cultural abilities that are gradually learnt and developed over several years (Goswami, 2015; Hulme et al., 2007; Wolf, 2008). Indeed, for some individuals, learning to read requires a great deal of effort, practice, and support. In the next sections, I will briefly describe how reading develops, and how this ability can be so effortful for some individuals. I will then outline the research questions that will be addressed in this thesis.

1.2. How Typical Reading Develops

Learning to read involves two related but distinct skills: the ability to translate printed words into speech (*i.e.,* decoding) and the ability to extract meaning from text (*i.e.,* comprehension) (Snowling & Hulme, 2021). Before a child can comprehend what

they are reading, they must first be able to decode the individual words in the text (Gough & Tunmer, 1986). As children gradually develop their decoding skills, a process that can vary in speed depending on the opacity of the orthography (Goswami, 2015), they become more confident readers and better equipped at comprehending what they are reading (Snowling & Hulme, 2021). Over time, novice readers become able to shift their attention to the global meaning of the text rather than struggling to decipher individual words. Decoding is thus a crucial step in reading development as it lays the foundation for comprehension and overall reading success.

The key role played by decoding in literacy development is highlighted in the 'Simple View of Reading', an influential framework that postulates that decoding skills, coupled with linguistic comprehension ability, predict ultimate reading comprehension (Gough & Tunmer, 1986). Within this view, knowledge of letter-sound correspondence rules, crucial to decoding processes, is considered critical to the development of word recognition skills, particularly in alphabetic languages (Gough & Tunmer, 1986). According to this framework, the *product* of decoding ability and linguistic comprehension predicts reading performance far more significantly than the independent contribution of these two separate components (Gough & Tunmer, 1986). It is, therefore, essential that the relationships between print and sound, and print and meaning receive balanced attention in instructional approaches to reading development (Snowling & Hulme, 2021).

Proficient reading can thus be conceptualised as automatic and effortless comprehension that does not rely on strenuous decoding processes (Snowling & Hulme, 2021). To achieve this automatisation, novice readers must initially rely on their pre-existing language skills, which, if intact, lay the necessary foundation for the

development of orthographic representations of words (Goswami, 2015; Seidenberg, 2005; Snowling & Hulme, 2021). This view of reading development is captured by Seidenberg and McClelland's (1989) influential computational model of word recognition. According to Seidenberg and McClelland (1989), acquiring word recognition skills entails three core features: semantic, phonological, and orthographic knowledge. Within this connectionist view, a computational model also known as the 'triangle model' (See Figure 1), orthography is connected to phonology via semantics, and, as a result, acquiring the meaning of a written word can be accomplished in three ways: indirectly, via a print-to-sound-to-meaning pathway, directly, via a print-to-meaning pathway, or a via a combination of these two routes (Chang et al., 2020; Harm & Seidenberg, 2004). The sound-mediated pathway has been suggested to contribute heavily to written word comprehension in the early stages of reading development (Harm & Seidenberg, 2004). Implementing a print-to-sound-to-meaning instructional approach is thus likely to be highly beneficial to novice readers when learning to acquire arbitrary mappings between written forms and semantics (Chang et al., 2020), further highlighting the fundamental role played by decoding ability. Prior to engaging in reading acquisition, pre-literate children demonstrate well-developed phonological and semantic knowledge. According to this connectionist view, to succeed at reading, children must develop links that connect the previously learnt phonological and semantic codes to the newly learnt orthography (Seidenberg, 2005; Seidenberg & McClelland, 1989).

Figure 1

Depiction of the architecture of the connectionist computational model of reading

commonly known as 'The Triangle Model', by Seidenberg and McClelland (1989)
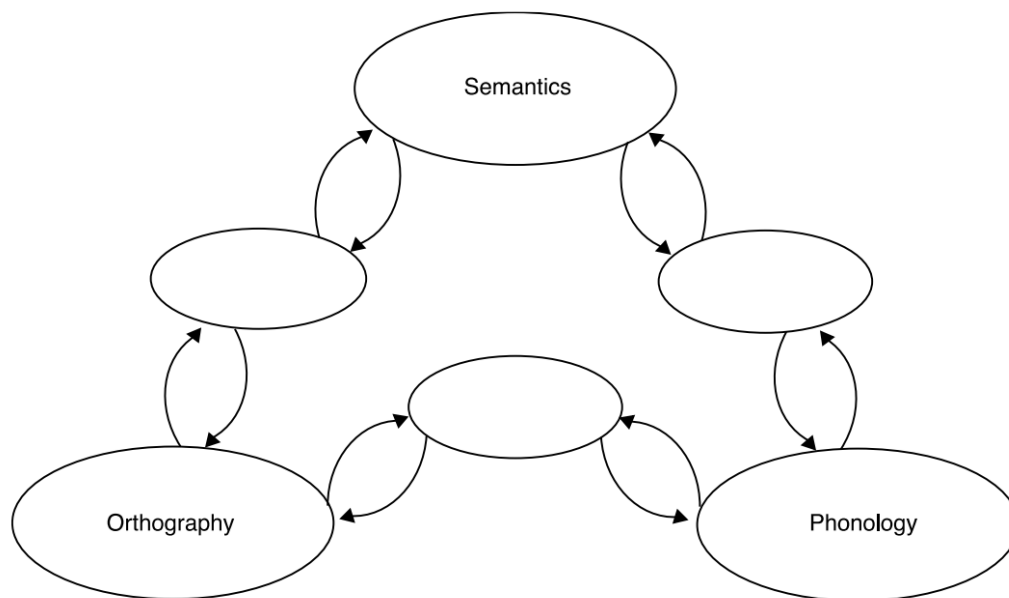
Note. This diagram, adapted from Seidenberg and McClelland (1989), shows a simulation of the language pathways involved in reading (*i.e.*, semantics, orthography, and phonology). The blank ovals represent the hidden units in the model, which enable learning of more complex mappings.

Seidenberg and McClelland's (1989) triangle model proposes that all three processing pathways (*i.e.*, orthographic, phonological, and semantic) operate in parallel during word recognition. There are, however, alternative models of word reading, such as the dual-route model (Coltheart et al., 2001), that propose a cascaded – or serial – processing approach for word reading. According to the dual-route model, there are two separate pathways for reading (See Figure 2), a lexical

and a non-lexical pathway. *Via* the lexical – and more direct – route, whole words are retrieved from memory, whereas word recognition *via* the non-lexical route involves reliance on spelling-to-sound correspondences (Coltheart et al., 2001).

Figure 2

Adapted representation of Coltheart and colleagues' (2001) dual-route model of reading aloud
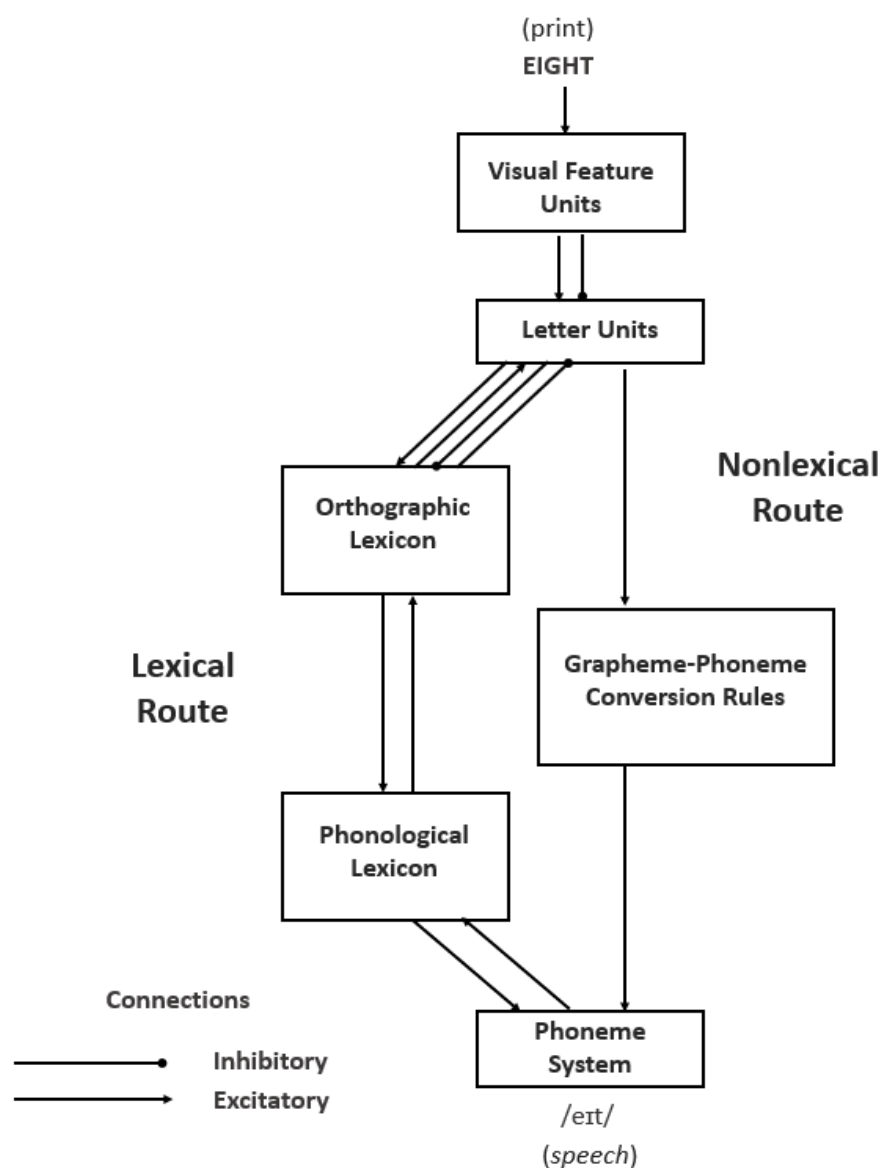
Novice readers, particularly in alphabetic orthographies, must learn systematic mappings between graphemes and phonemes, a mechanism known as the alphabetic principle, which is the knowledge that written letters represent the sounds of spoken language (Liberman et al., 1989; Rozin & Gleitman, 1977; Snowling et al., 2020). Associations between graphemes and phonemes are generally acquired within the first year of formal reading instruction, but full automatisation is expected to emerge after several additional years of instruction and practice (Blomert, 2011; Froyen et al., 2009). Acquisition of the alphabetic principle can be construed as a stepping stone for children to move from overt decoding to being able to read for meaning, a process that gradually automatises as a function of repeated exposure (Ehri, 2005; Ehri & Saltmarsh, 1995; M. Jones et al., 2018; Snowling et al., 2020).

Indeed, over time, as novice readers repeatedly encounter words in their reading development, they begin to recognise these words as whole units rather than focusing on their individual graphemes or phonemes, a process that is known as sight word recognition (Ehri, 1995, 2000, 2014). Sight words can be conceptualised as any words that have been practised sufficient times to be read from memory by sight without need for active decoding anymore (Ehri, 2014). For novice readers, learning sight words is essential for developing comprehension skills (Ehri, 2000, 2005, 2014). Sight word recognition is thought to consist of four somewhat overlapping learning stages: pre-alphabetic, partial-alphabetic, full-alphabetic, and consolidated alphabetic stages (Ehri, 2005).

In the pre-alphabetic stage of sight word learning, children rely on visual cues and contextual information to recognise words. They do not yet understand the relationship between graphemes and phonemes and are thus still unable to rely on decoding strategies (Ehri, 2005). To illustrate, British children in the pre-alphabetic

stage of sight word learning may recognise the word "Tesco" by the red font with the blue dotted underline logo. It is in the partial-alphabetic stage that children begin to learn and recognise individual letters and their corresponding letter names and/or sounds. In that stage, children may be able to decode words with simple syllable structures, but still rely heavily on visual and contextual cues to predict what the words are because they only have partial knowledge of the alphabetic system (Ehri, 2005). In the full-alphabetic stage, children begin to have a strong understanding of letter-sound relationships and can decode most words. They are able, for example, to break words down into their component phonemes and blend them together. When encountering words with irregular spelling, however, novice readers in the full-alphabetic stage will still likely rely on sight recognition for these items. Sight word recognition becomes an automatic and effortless process in the consolidated alphabetic stage, the stage during which novice readers become increasingly familiar with recurring letter patterns (Ehri, 2005). During the consolidated alphabetic stage, most typically developing children have a large repertoire of high-frequency words, and are able to quickly recognise and read irregularly spelled words without the need of sounding them out (Ehri, 2005).

Teaching decoding strategies, therefore, has the potential to provide novice readers with a strong alphabetic foundation, which, in turn, will help them acquire and expand their sight vocabulary (Ehri, 2005). Over time, whilst the process of decoding per se will no longer be needed to retrieve words from memory, memory retrieval can still be considered phonological in nature, as it entails the rapid activation of graphophonemic connections to retrieve the words' phonetic and semantic information from long-term memory (Ehri, 2005).

Because novice readers must learn essentially arbitrary associations between visual characters and their phonological and semantic representations (Litt et al., 2019), it follows that learning to read also requires intact associative learning mechanisms (Hulme et al., 2007; Warmington & Hulme, 2012). *Cross-modal* (*i.e.*, visual-verbal) associative learning skills, in particular, has been demonstrated to play a crucial role in literacy acquisition (Albano et al., 2016; Blau et al., 2009, 2010; Di Pietro et al., 2023; Garcia et al., 2019; Hulme et al., 2007; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Litt et al., 2019; Litt & Nation, 2014; Toffalini et al., 2019; Warmington & Hulme, 2012; Windfuhr & Snowling, 2001). In the working memory literature, the process of associating two seemingly arbitrary features is known as 'binding' (Baddeley et al., 2011), an operation that is assumed to take place in the episodic buffer, the latest addition to Baddeley's influential multi-component working memory model[1] (Baddeley, 2000; Baddeley et al., 2010, 2011). The episodic buffer not only integrates information that stems from different sources – or modalities – but also creates a single unified unit that is a coherent representation of the input being processed (Baddeley, 2000; Baddeley et al., 2010).

The ability to encode and subsequently retrieve cross-modal bindings (*e.g.*, associations between letters and their corresponding letter sounds) as a single audiovisual unit is, therefore, a fundamental skill in reading development (Blomert, 2011; Brockmole & Franconeri, 2009; Litt et al., 2019). This skill is often assessed through so-called paired associate learning tasks, a commonly used episodic memory paradigm that is used to measure an individual's ability to learn and accurately recognise and/or recall arbitrary pairings of unrelated stimuli (Scorpio et

---

[1] The four components in Baddeley's (2000) working memory model are the central executive, the phonological loop, the visual-spatial sketchpad, and the episodic-buffer

al., 2018). Acquisition of letter-sound knowledge and early development of a sight vocabulary are examples of processes that require intact paired associate learning mechanisms (Hulme et al., 2007). Children and adults with typically developed reading profiles consistently demonstrate intact associative learning skills of novel cross-modal bindings (*e.g.*, Clayton et al., 2018; Hulme et al., 2007; Jones et al., 2018; Kalashnikova et al., 2021; Litt et al., 2013, 2019; Litt & Nation, 2014; Messbauer & de Jong, 2003).

Relatedly, the associative binding processes that are characteristic of episodic memory, conceptualised as the ability to encode, store, and retrieve an event by binding it to its associated context (Tulving, 1972), appear to also be implicated in reading development (Ghetti & Bunge, 2012). To illustrate, a child might recall the first time they encountered the letter "T" in a storybook, or the shape of the letter "Z" when they first learned to write it. While these episodic memory cues are useful in helping to gradually strengthen newly learnt visual-phonological associations, continued reliance on these cues over time could be indicative of weaker, less confident memory representations (M. Jones et al., 2018; Scholz et al., 2011, 2016). Instead, with time, multiple exposures to a given letter or word and associated phonological sequences should abstract to a representation with minimal contextual or episodic features, through a process of statistical learning.

Statistical learning refers to the brain's ability to detect and learn regularities in the environment (Altmann, 2017). Through repeated exposure to an input over several occurrences (*i.e.*, episodes), we incrementally develop sensitivity to the statistical properties of the input encoded within each individual occurrence (Altmann, 2017). In initial exposures to a stimulus, we display a tendency to rely on episodic cues, such as the temporal and spatial features of the input, to retrieve an

item from memory (El-Kalliny et al., 2019; Tulving, 1972). Over time, however, memory representations become more abstract, and reliance on episodic details tend to subside (Altmann, 2017; Squire & Zola, 1998; Stark & Stark, 2016). In the context of literacy development, being able to track simple statistical patterns, such as stimulus repetition and sequences, is a robust indicator of reading proficiency (Ahissar, 2007), translating into more fluent and less effortful reading (M. Jones et al., 2018).

In summary, reading acquisition is a process of learning associations between visual characters and their phonological and semantic representations. The formation of visual-phonological associations – including the cognitive mechanisms that enable these associations, such as working memory, decision-making, episodic memory and statistical learning – are central to this thesis. Whilst most children eventually learn to read fluently and effortlessly, a proportion of individuals will struggle to read all of their lives, and it is important to understand the learning differences between typically developing readers and individuals who struggle to read, such as those with *developmental dyslexia*. In the next section, I briefly review developmental dyslexia, including its characteristics and purported causes.

1.3. Developmental Dyslexia

Developmental dyslexia (hereafter also referred to as dyslexia) is a neurodevelopmental disorder that affects an individual's ability to read, write, and spell, despite having access to adequate educational and socioeconomic opportunities and no severe neurological deficits (Goswami, 2015; Gough & Tunmer, 1986; Jaffe-Dax et al., 2017; Snowling et al., 2019, 2020; Vellutino et al., 2004; Windfuhr & Snowling, 2001). The condition emerges early in children's development and persists over time (Francks et al., 2002; Jaffe-Dax et al., 2017; Snowling et al.,

2020). The severity of each individual's deficit is considered to lie on a continuum, ranging from very mild to severe (Vogel, 2003). Previous research has suggested that developmental dyslexia has a genetic component (Francks et al., 2002). Children who have a parent or sibling with dyslexia are at a higher risk of developing the condition themselves (Snowling et al., 2012). However, despite the condition being often inherited, not all individuals with a family history of dyslexia will necessarily develop the condition, as there are likely to be multiple genetic and environmental factors at play (Francks et al., 2002).

Children with developmental dyslexia may display no positive response to literacy interventions and will generally struggle to catch up with their peers (Snowling et al., 2020). The condition can manifest in alphabetic and non-alphabetic languages of varying degrees of orthographic opacity (Goswami, 2015). Learning to read in languages with opaque orthographies (*i.e.*, languages with inconsistent mappings between graphemes and phonemes), such as English and French, is reportedly more challenging, particularly for individuals with dyslexia, relative to literacy acquisition in languages with transparent alphabetic systems, such as Spanish and Welsh (Landerl et al., 2013; Seymour et al., 2003). However, when simultaneously learning to read in languages that vary in their orthographic depth, bilinguals with dyslexia have been shown to benefit from cross-linguistic transfer: the severity of the deficits in the less consistent language appears to be reduced if the individual has also learnt to read in a language that has a more consistent orthography (Lallier et al., 2018)[2]. This is because phonological reading strategies are reportedly easier to master in languages with transparent writing systems but

---

[2] For an account on the impact of bilingualism on literacy development when learning to read in two alphabetic orthographies, see Grain Size Accommodation Hypothesis by Lallier and Carreiras (2018)

they can transfer cross-linguistically to modulate reading in the inconsistent language (Lallier et al., 2010, 2013, 2016).

Dyslexia can manifest in individuals with varying levels of intelligence (Snowling et al., 2020), and its incidence is estimated to be around 5-10% of the general population (Francks et al., 2002; Snowling et al., 2019). Approximately 40% of the individuals diagnosed with developmental dyslexia will have a comorbidity, such as Attention-Deficit-Hyperactivity-Disorder and/or dyscalculia (Moll et al., 2020), a considerably high rate that not only weakens proposed single-deficit approaches to dyslexia (*e.g.*, the phonological deficit hypothesis; Frith, 1997; Snowling et al., 2020; Vellutino et al., 2004) but also highlights the complexity of this neurodevelopmental condition (Moll et al., 2020; Snowling et al., 2020; Stein, 2023). Growing evidence suggests that dyslexia may result from a combination of multiple risk factors (Moll et al., 2020; O'Brien & Yeatman, 2021; Pennington, 2006; Ramus et al., 2003; Snowling et al., 2019; Stein, 2023), including difficulties unrelated to literacy, such as cognitive and sensorimotor impairments (Carroll et al., 2016; Jaffe-Dax et al., 2017; Lallier et al., 2010; Szmalec, Loncke, et al., 2011).

There are several theories that attempt to explain the underlying causes of developmental dyslexia (cf. Stein, 2023 for an overview). One of such theories, the magnocellular deficit theory, proposes that individuals with dyslexia may have difficulty processing rapidly changing visual stimuli due to a deficit in the magnocellular pathway of the visual system (Stein, 2023; Stein & Talcott, 1999). Deficits in this system, which is responsible for processing visual motion and spatial information, have been proposed to lead to visual instability and, as a consequence, difficulties in controlling eye movements during reading (Stein & Talcott, 1999). However, the core difficulties associated with dyslexia have been suggested to be of

a predominantly phonological nature (Frith, 1997; Snowling et al., 2020; Vellutino et al., 2004). Indeed, one of the most commonly accepted hypotheses for the proximal cause of developmental dyslexia is the phonological deficit hypothesis, according to which individuals with dyslexia have an impaired ability to process and understand speech sounds (Snowling & Hulme, 1994; Stanovich, 1994). This deficit appears to be present even in some pre-literate children, who later receive a diagnosis of developmental dyslexia (Snowling et al., 2019).

The ability to manipulate and analyse the sounds of spoken language is commonly known as phonological awareness, and it is suggested to be a critical skill to the early stages of reading development, particularly in alphabetic orthographies (Wagner & Torgesen, 1987). Individuals with impaired reading, such as those with developmental dyslexia, tend to perform more poorly on tests that tap phonological processing, relative to typical readers (Wagner & Torgesen, 1987). However, it is important to note that impaired phonological processing skills have also been observed in children with reading disorders other than dyslexia (Stein, 2023), highlighting the fact that phonological difficulties alone cannot account for a developmental dyslexia diagnosis (Pennington, 2006).

As well as showing phonological deficits, readers with developmental dyslexia are also demonstrably impaired in acquiring the necessary skills for decoding and spelling (Ahissar, 2007; Goswami, 2015; Moll et al., 2020; Snowling et al., 2012, 2020). Children diagnosed with developmental dyslexia generally experience difficulty in grasping the alphabetic principle (Rozin & Gleitman, 1977), and thus struggle to learn systematic mappings between graphemes and phonemes efficiently (Snowling et al., 2020). Ultimately, difficulty in establishing mappings between graphemes and phonemes may have a cascade effect on individuals with dyslexia's

ultimate reading proficiency levels as such deficit can hinder their ability to access the meaning of words rapidly and effortlessly (Snowling et al., 2020).

Arguably, it has been proposed that it is the difficulty in processing and understanding the sound structure of speech that impairs the ability to establish grapheme-phoneme mappings (Snowling et al., 2020). However, difficulty in learning mappings between graphemes and phonemes indicates that individuals with developmental dyslexia may also have underspecified associative learning skills (Calabrich et al., 2021b; Hulme et al., 2007; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Kalashnikova et al., 2021; Litt et al., 2013, 2019; Litt & Nation, 2014; Messbauer & de Jong, 2003; Snowling & Hulme, 2021; H. C. Wang et al., 2017; Wimmer et al., 1998). Indeed, a growing body of evidence suggests that a *visual-phonological* binding deficit may be a key factor in developmental dyslexia in languages with alphabetic orthographies (Aravena et al., 2013; Blomert, 2011; Fraga González et al., 2015; Garcia et al., 2019; Hulme et al., 2007; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Litt et al., 2019; Toffalini et al., 2019, 2018; H. C. Wang et al., 2017; Žarić et al., 2014, 2015), but also in non-alphabetic languages (Georgiou et al., 2017). Notably, the ability to establish *visual-phonological* associations, rather than associations within-modally and/or with any two types of stimuli, appears to be a robust correlate of reading ability, discriminating readers with dyslexia from typical readers (Hulme et al., 2007; Warmington & Hulme, 2012). Indeed, this cross-modal associative learning ability has been shown to make a separate contribution to literacy development from measures of phonological processing skills (Hulme et al., 2007; Warmington & Hulme, 2012). However, this deficit has also been proposed to emerge only when a phonological output is required (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014). On the whole,

deficient visual-phonological associative ability appears to underlie some of the reading-related difficulties in developmental dyslexia to a significant extent, particularly affecting efficient letter learning (Hulme et al., 2007).

Letter learning can be construed as an early stage of visual-phonological associations formation process, also known as paired associate learning (Hulme et al., 2007). A related process that involves similar cross-modal associative mechanisms, albeit with overlearned stimuli and speeded retrieval from long-term memory, is the so-called rapid automatised naming (RAN) ability (Warmington & Hulme, 2012). Rapid automatised naming is a well-known predictor of reading fluency that reflects the ability to accurately name out loud lists of high-frequency overlearned items, such as letters and/or digits, as fast as possible (Denckla & Rudel, 1976; Kirby et al., 2010). In addition to demonstrating deficits in phonological processing, individuals with developmental dyslexia consistently perform more slowly on tasks measuring rapid automatised naming ability relative to their typical counterparts (Kirby et al., 2010; Wagner et al., 1999; Warmington & Hulme, 2012; Wolf & Bowers, 1999). Interestingly, this double deficit (*i.e.*, deficient phonological processing skills, and impaired rapid automatised naming ability) is not generally observed in individuals with other types of reading disorders, and is often considered a hallmark of developmental dyslexia (Stein, 2023).

Rapid automatised naming ranks amongst the strongest predictors of developmental dyslexia (Kirby et al., 2010; Norton & Wolf, 2012; Snowling & Hulme, 2021), accounting for unique variance in reading ability even after the effects of phonological processing have been controlled for (Warmington & Hulme, 2012). In developmental dyslexia, rapid automatised naming deficits have been shown to persist well into adulthood (M. Jones et al., 2009, 2016; Shaywitz & Shaywitz, 2005),

and to emerge across languages of varying degrees of orthographic opacity (Wimmer et al., 2000). In transparent languages, such as Welsh and Spanish, rapid automatised naming ability is a generally stronger predictor of reading skill than phoneme awareness (Georgiou et al., 2005; Kirby et al., 2010; Landerl & Wimmer, 2000; Wimmer, 1993; Wimmer et al., 1998, 2000; Zugarramurdi et al., 2022), a pattern of results that highlights the heterogeneity and multi-deficit nature of developmental dyslexia.

Despite similarities in the cross-modal associative mechanisms that underlie paired associate learning of novel visual-phonological mappings and rapid automatised naming, PAL and RAN have been suggested to play differential roles in literacy acquisition: in essence, PAL appears to tap the accuracy with which visual-phonological associations are initially acquired, whereas RAN taps the speed with which these associations are retrieved from long-term memory once they are fully mastered (Kirby et al., 2010; Warmington & Hulme, 2012). This distinction is evident in the generally slow and effortful reading process displayed by individuals with dyslexia who despite having fully mastered decoding skills might still struggle to retrieve well-established visual-phonological associations from long-term memory in a time-efficient manner (Warmington & Hulme, 2012).

Overall, the impaired visual-phonological associative ability typically observed in individuals with dyslexia, a deficit that has been shown to persist well into adulthood (M. Jones et al., 2018; M. Jones, Branigan, et al., 2013), provides support for a specific working memory deficit in developmental dyslexia (Beneventi et al., 2010a; M. Jones, Branigan, et al., 2013; Peng et al., 2018). Indeed, cortical areas commonly associated with working memory, such as the left superior parietal lobule and the right inferior prefrontal gyrus, appear to be underactive for those with

dyslexia, compared to their typical reader counterparts (Beneventi et al., 2010a). *Cross-modal* working memory ability, in particular, has been demonstrated to be impaired in children and adults with developmental dyslexia (Albano et al., 2016b; Calabrich et al., 2021a; Garcia et al., 2019b; M. Jones, Branigan, et al., 2013; M. Jones et al., 2018; Toffalini et al., 2019b).

Inefficiency in forming novel visual-phonological associations (*i.e.*, cross-modal bindings) in developmental dyslexia has also been linked to impaired statistical learning (M. Jones et al., 2018). Indeed, individuals with developmental dyslexia are prone to forgetting previously encoded stimuli (Jaffe-Dax et al., 2015, 2016, 2017), a deficiency that can potentially hinder processing of subsequent stimuli, thus restricting this group's ability to track statistical patterns. Crucially, sensitivity to statistical regularities in the input has been shown to be a fundamental skill for efficient acquisition of novel visual-phonological associations (M. Jones et al., 2018). Inability to track statistical patterns in the stimuli can lead to over-reliance on episodic memory cues during acquisition of novel visual-phonological associations, which can significantly slow down the process of learning (M. Jones et al., 2018).

1.4. Thesis Aims

The primary aim of this thesis is to examine and further delineate the cognitive processes underlying paired associate learning trajectories of individuals with typical and atypical reading profiles. Below are the primary research questions (RQ) which will be addressed here:

1.4.1. Research Question 1

RQ1 How do task-irrelevant episodic details modulate visual-phonological binding performance in individuals with and without dyslexia?

Research Question 1 (Chapters 3 and 4) is specifically concerned with the *statistical properties contributing to acquisition of novel visual-phonological associations in readers with dyslexia and typical readers' learning profiles*.

The studies reported in Chapters 3 and 4 emulate an eye-tracking paradigm, originally developed by Jones et al. (2018), on the role played by episodic memory cues and statistical learning in acquisition of novel visual-phonological associations in typical readers and readers with dyslexia. Jones and colleagues (2018) examined if, and the extent to which, individuals with and without dyslexia would detect statistical regularities across multiple exposures, and thus make use of statistical learning to facilitate memory recall. They also investigated whether directing one's gaze to a screen location which had been previously occupied by a given stimulus would help retrieve that stimulus (*i.e.*, the "looking-at-nothing" phenomenon; see Chapter 2 for a detailed description).

In their study, Jones and colleagues (2018) showed that individuals with dyslexia were sensitive to location information (*i.e.*, episodic cues) present only in the most recent trial, whereas individuals with a typical reading profile were sensitive to longer-range statistical regularities detected over *multiple* exposures.

To expand on Jones and colleagues' (2018) findings, building on their original looking-at-nothing paradigm, here, I designed a paired associate learning task in which the consistency of spatial (*i.e.*, item screen location) and contextual (*i.e.*, item co-occurrences) properties of the novel visual-phonological associations was

manipulated during the encoding phase. This manipulation was introduced to help us further understand the extent to which reliance on episodic memory contributes to the process of paired associate learning of novel cross-modal bindings in adults. Chapter 3 tackles this issue in a small-scale study that included only typical readers, whereas Chapter 4 compares learning differences between individuals with and without a developmental dyslexia diagnosis.

Furthermore, unlike Jones and colleagues (2018) who only looked at recall, the studies reported in Chapters 3 and 4 also include measures of recognition. Including measures of recognition here allows us to also examine a previous claim in the paired associate learning literature that individuals with dyslexia only show a cross-modal binding deficit in tasks that require a spoken output (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014).

1.4.2. Research Question 2

RQ2 Is there a specific cross-modal binding deficit in adults with developmental dyslexia, even in the absence of spoken output demands?

Research Question 2 (Chapter 5) examines the *specificity of the binding deficit in developmental dyslexia.*

Previous studies have shown that the ability to establish *visual-phonological* associations, rather than associations within-modally and/or with any two types of stimuli, appears to correlate strongly with reading ability, discriminating readers with dyslexia from typical readers (Hulme et al., 2007; Warmington & Hulme, 2012). However, this binding deficit has been proposed to emerge only in tasks that require a verbal output (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014), as outlined

above. Findings from the study reported in Chapter 4 of this thesis, however, suggest that a cross-modal binding deficit appears to persist in the adult population with dyslexia even in the absence of spoken output demands (Calabrich et al., 2021c). Here, to further explore this issue, I adapted the well-known working memory n-back paradigm to examine adults with and without dyslexia's ability to store visual-visual associations (within-modal binding), auditory-auditory associations (within-modal binding), and visuo-auditory associations (cross-modal binding). As will be further explained in Chapter 5, cognitive load was equated across the unimodal and cross-modal conditions of the n-back task to allow us to dissociate modality demands across the three different conditions.

1.4.3. Research Question 3

RQ3 Can online measures elucidate Paired Associate Learning mechanisms in beginning readers? And what is the relationship between paired associate learning response accuracy and reading outcomes in these children?

Research Question 3 (Chapter 6) examines *paired associate learning and its relationship to reading in school-aged children.*

As briefly outlined above, previous research shows that paired associate learning and rapid automatised naming, two related processes that appear to employ similar underlying cross-modal associative mechanisms, are significant predictors of reading skill (Hulme et al., 2007; Kalashnikova et al., 2021; Litt & Nation, 2014; Messbauer & de Jong, 2003; Toffalini et al., 2018, 2019; Warmington & Hulme, 2012; Wimmer et al., 2000; Windfuhr & Snowling, 2001). Despite their relatedness, to date, however, there is not a substantial amount of evidence on whether paired associate learning contributes to reading *indirectly via* rapid automatised naming.

Collecting such evidence is crucial to elucidate whether the skills underpinning paired associate learning ability may potentially represent a precursor to those underpinning rapid automatised naming, an ability that reflects rapid access to highly automatised visual-phonological associations (Lervåg & Hulme, 2009). For a fuller picture, I will also collect data on the potential indirect effects of paired associate learning accuracy on reading *via* another crucial skill, namely phoneme awareness.

Innovatively, in Chapter 6, I will also examine the decision-making process underlying paired associate learning of novel cross-modal bindings in school-aged children using mouse-tracking technology. By tracking participants' mouse cursor trajectories as they learn novel visual-phonological associations (see methodological considerations in Chapter 2), I aim to probe whether cognitive conflict (*i.e.*, indecision) will emerge during paired associate learning, and, if so, whether this conflict will vary as a function of reading ability. To the best of my knowledge, this approach has never been adopted in previous paired associate learning paradigms.

1.5. Chapter Summary

In this chapter, I provided a brief overview of how typically developing children learn to read, and outlined some of the potential causes of developmental dyslexia. As mentioned above, while there is no definitive consensus on the causes of dyslexia, previous research suggests that it is a complex and multifaceted condition with multiple underlying factors that contribute to its emergence and development. Therefore, due to the complexity of the condition, this introductory chapter was not an exhaustive review of the topic, but it highlighted some of the key factors that contribute to the development of reading skills and dyslexia. Similarly, whether

paired associate learning deficits are considered a cause of dyslexia or the consequence of more basic sensory/linguistic difficulties is a complex question that has not yet been definitively answered and one which would likely require a longitudinal design. This thesis is unable to address this question definitively due to its cross-sectional design.

In this chapter, I also introduced the research questions that will be addressed in the next four empirical chapters of this thesis. In summary, addressing research question 1 (*i.e.*, *How do task-irrelevant episodic details modulate visual-phonological binding performance in individuals with and without dyslexia?*) will enhance our knowledge of potential differences in statistical learning patterns among individuals with different reading abilities. The two studies addressing question 1 (*i.e.*, Chapters 3 and 4) build upon prior work by Jones and colleagues (2018), which revealed clear differences in statistical learning patterns in adults with and without developmental dyslexia during acquisition of novel visual phonological associations. If our results indicate distinct statistical learning patterns in individuals with and without dyslexia, this will not only expand on Jones and colleagues' (2018) findings, but it will also provide further evidence to strengthen Jaffe-Dax and colleagues' (2015; 2016; 2017) argument that individuals with dyslexia learn differently from environmental cues. Similarly, if our investigation reveals that task-irrelevant episodic details modulate visual-phonological binding performance in individuals with dyslexia, it will make a valuable contribution to research focusing on the anchoring deficit hypothesis (Ahissar, 2007), according to which individuals with dyslexia struggle to form a stable internal reference, a difficulty that can hinder their ability to perform tasks that require them to compare stimuli. In essence, if readers with dyslexia are unable to form a stable internal reference, as proposed by the anchoring deficit hypothesis (Ahissar,

2007), then participants with dyslexia in our study will likely rely more on episodic information to remember and compare stimuli, relative to their typical reader counterparts.

Addressing research question 2 (*i.e.*, *Is there a specific cross-modal binding deficit in adults with developmental dyslexia, even in the absence of spoken output demands?*) will help us determine whether paired associate learning deficits are specifically visual-phonological in nature, as observed in previous research (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014) or if they might be more broadly related to the ability to bind any two stimuli, a finding which would align with current multi-component models of working memory (Baddeley, 2000; Baddeley et al., 2010, 2011). If our investigation reveals that individuals with dyslexia are impaired in binding any two stimuli, this will challenge previous assumptions in the paired associate learning literature that emphasise that the primary driver in binding difficulty involves a verbal output (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014) or, at the very least, a phonological component (Frith, 1997; Snowling et al., 2020; Vellutino et al., 2004).

Finally, addressing research question 3 (*i.e.*, *Can online measures elucidate Paired Associate Learning mechanisms in beginning readers? And what is the relationship between paired associate learning response accuracy and reading outcomes in these children?*) will help us (a) test the feasibility of using online measures, such as continuously recorded mouse-trajectories, in paired associate learning research, as well as (b) examine the relationships between phonological awareness, rapid automatised naming and paired associate learning. Testing the feasibility of using online measures in paired associate learning research is important because online measures can provide more precise and nuanced data on cognitive

processes than traditional offline measures, such as accuracy and reaction time. Similarly, phonological awareness has long been recognised as a key factor in reading (Frith, 1997; Snowling et al., 2020; Vellutino et al., 2004), but the interplay between RAN, paired associate learning, and their combined influence on reading outcomes remains relatively under-researched. Addressing this gap is crucial since rapid automatised naming and paired associate learning appear to share similar cross-modal associative mechanisms but the former may represent a more advanced or later version of the foundational paired associate learning process as it involves retrieval of overlearned stimuli from long-term memory (Warmington & Hulme, 2012).

All in all, by investigating these questions, I hope to contribute to the field of reading research by adding another small but vital piece to the developmental dyslexia puzzle, thus moving a step closer towards a gradually more complete picture of reading development.

In the next chapter, I outline the methodologies used in this thesis. Taken together, I hope that the data yielded from this thesis will lend insight into the statistical learning processes underpinning PAL, the locus of the binding deficit in terms of its specificity to cross modal features, and the emergence of the learning profile in children.

Chapter 2

Methodological Considerations

2. Chapter Overview

In this chapter, I will outline and provide a justification for the methodologies used in this thesis. Due to the Covid-19 pandemic, which forced our research labs to close for several months between 2020 and 2022, all the data reported in this thesis were collected online using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), a cloud-based research platform that has been shown to provide reasonable accuracy and precision in web-based collection of behavioural data (Anwyl-Irvine et al., 2021).

In the first part of this chapter, I will provide a general overview on how the Covid-19 pandemic prompted experimental researchers to adapt their studies for online data collection, and will briefly outline the experimental controls that were put in place to maximise data quality. After that, I will provide a short description of webcam-based eye tracking and mouse-tracking, two methods I used in studies reported in this thesis (see Chapters 3, 4, and 5). I will then include a description of the stimuli design as well as the cognitive and literacy measures administered in the studies reported here. Finally, I will briefly outline the statistical analyses approaches undertaken in this thesis.

2.1. Shifting from Lab-Based to Online Research

The Covid-19 pandemic and the resulting closure of lab spaces prompted experimental researchers to swiftly adapt their studies for online administration. This shift presented both opportunities and challenges. Online research allows for a larger, more diverse, and potentially less biased participant pool, as individuals can be conveniently recruited from anywhere in the world in a cost-efficient and timely manner (Grootswagers, 2020; Sauter et al., 2020). One key challenge, however, is

the potential loss of experimental control as participants could be completing the study from a variety of different locations (*e.g.*, their homes, parks, vehicles), under diverse environmental conditions (*e.g.*, daylight, artificial light, in a noisy room), and on different devices (*e.g.*, phones, tablets, laptops). This heterogeneity makes it challenging to ensure that all participants are experiencing the same experimental conditions, which is a golden standard of research practice, upheld in highly controlled laboratory settings. Additionally, the lack of face-to-face interaction can make it harder to ensure that participants are fully engaged and motivated to complete the study adequately and/or in a timely manner. Here, I describe three strategies that I employed to enhance the quality of the data collected: 1) Providing clear instructions; 2) Having research assistants available to support child participants in remote testing; and 3) making use of attention checks. Further strategies (*e.g.*, data filtering) will be described in the subsequent eye-tracking and mouse-tracking sections.

2.1.1. Instructions in Online Research

Providing clear and detailed instructions for participants is crucial for the success of online behavioural studies (Grootswagers, 2020; Sauter et al., 2020). In traditional laboratory-based studies, experimenters are present to clarify any questions or concerns that participants may have. However, in online studies, participants are often completing tasks independently. Therefore, it is important to ensure that participants have a clear understanding of what is expected of them. Clear instructions help to ensure that participants are following the correct procedures and enhances the chances that the data collected is valid and reliable (Grootswagers, 2020; Sauter et al., 2020).

Similarly, clear instructions are also important for ensuring participants' comfort and the quality of the data collected remotely. Participants who are uncertain about what they are supposed to do may potentially become anxious or stressed, which can influence their performance and the data collected. Providing clear instructions can help to alleviate these concerns as well as increase participant comfort. Pictorial instructions, in particular, can be especially helpful for online studies that involve tasks that are difficult to describe in written form only (Sauter et al., 2020). Figure 3 shows some of the pictorial instructions provided to participants in the studies reported in Chapters 3 and 4 on how to position themselves for the eye-tracking calibration procedure.

Figure 3

Instructions on how to perform the calibration procedure in the webcam-based eye tracking tasks
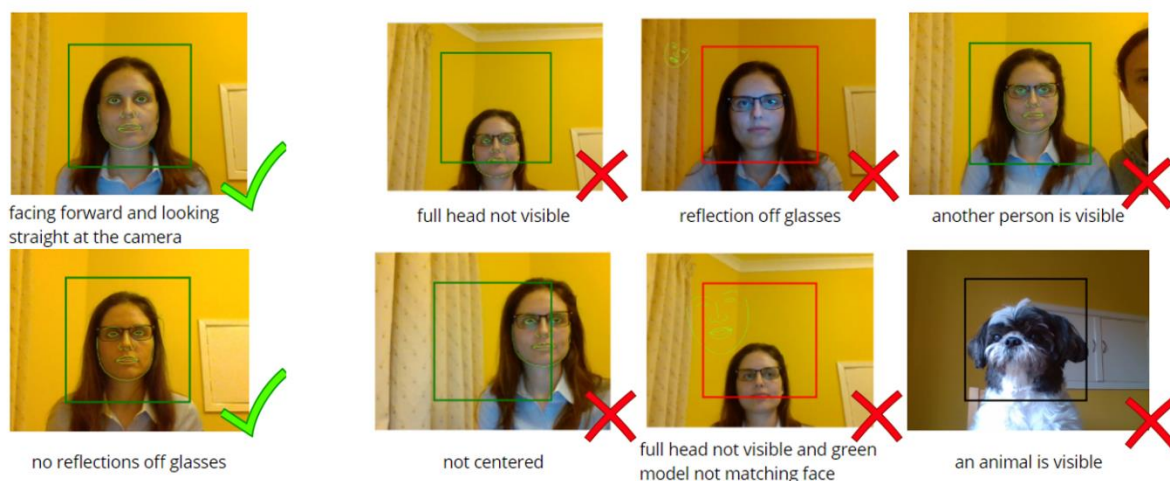
Likewise, video instructions can be helpful for providing a more interactive and engaging experience for participants. In the studies reported in Chapters 3 and 4,

participants were provided with pictorial instructions, whereas in the studies reported in Chapters 5 and 6, demonstration videos were also included. These were added with the goal of maximising the chances that participants understood what was expected of them during the online testing sessions.

2.1.2. Research Assistants Supporting Children During Online Testing

When conducting online research with children, such as the study reported in Chapter 6, I have found that having research assistants available to assist participants remotely can be an invaluable asset, and can help to ensure that the study is completed successfully and with higher-quality data. In essence, research assistants can monitor the child's engagement and their progress remotely, ensuring that they stay on task and do not lose focus. They can also help to troubleshoot any (technical) difficulties that may arise during the study. Additionally, research assistants can help to explain the instructions and procedures to the child in a way that is age-appropriate and easy for them to understand, and they can also answer questions or provide additional examples to help the child complete the experimental tasks correctly. This can be especially helpful for studies involving young children who may struggle with reading or understanding written instructions. In Chapter 6, I report findings from an online study conducted with school-aged children. In that study, trained research assistants were present via Microsoft Teams for the duration of the task providing support, reassurance, and encouragement to help the children feel more comfortable during the remote testing session.

2.1.3. Attention Checks in Online Research

Due to the nature of online research, which allows individuals to undertake a study from the comfort of their homes or anywhere of their choosing, participants may multitask, get interrupted, or become easily distracted (Grootswagers, 2020). Unsurprisingly, this can lead to inaccurate or incomplete datasets. To ensure that online participants are paying attention and following the instructions provided during an online study, we can use attention checks (Oppenheimer et al., 2009; Sauter et al., 2020).

There are several ways to implement attention checks in online research, such as including questions that require specific attention-related responses or by including unexpected events that require a reaction, such as a button press or a mouse click (Oppenheimer et al., 2009). Crucially, attention checks can be employed to confirm whether participants are actively engaged in the experimental task and are not just going through the motions or becoming distracted. In the three unassisted experiments reported in this thesis (see Chapters 3, 4, and 5), attention checks were implemented (*e.g.*, performing a mouse click at a specific location at regular intervals) to ensure our statistical analyses only included data from participants who were fully engaged with the experimental tasks. As a result, any indication that participants had left their computers unattended for a few minutes would prompt their immediate and automatic rejection from the study.

2.2. Methods Employed in the Studies Reported in this Thesis

Two examples of online methodologies increasingly being used in cognitive psychology research are webcam-based eye-tracking and mouse-tracking. Both of

these methodologies, which are low-cost, easy to implement, and non-invasive, were used in studies reported in this thesis (see Chapters 3, 4, and 5). In this subsection, I provide a brief overview of these two methodologies.

2.2.1. Webcam-Based Eye-Tracking

In cognitive science and reading research, eye tracking technology is used to investigate a wide range of issues, such as memory (*e.g.*, Herten et al., 2017), language (*e.g.*, Vaughan-Evans et al., 2020), and decision-making (*e.g.*, Fiedler & Glöckner, 2012). The use of eye-tracking technology has been a valuable tool in developmental dyslexia research due to its high-precision in measuring various aspects of visual processing, such as fixation durations, saccades lengths, and attentional control (Holmqvist et al., 2011). The use of eye-trackers has significantly contributed to advancing our understanding of developmental dyslexia, having been employed in investigations tackling a variety of relevant issues, such as rapid automatised naming (M. Jones, Ashby, et al., 2013; Kuperman et al., 2016), acquisition of novel visual-phonological associations (M. Jones et al., 2018), spoken language processing (Huettig & Brouwer, 2015), semantic judgements (Egan et al., 2023), phonological recoding (Blythe et al., 2020), to name a few.

Eye-tracking technology has also been extensively used to investigate the looking-at-nothing phenomenon, whereby people may sometimes move their eyes towards an empty location when attempting to recall an item that had been originally encoded in that now-empty location (De Groot et al., 2016; Ferreira et al., 2008; Richardson & Spivey, 2000). The premise of the looking-at-nothing phenomenon is that our brain may sometimes trigger reflexive eye movements to facilitate memory

retrieval, prompting us to direct our gaze towards a location previously occupied by a stimulus, even after the stimulus is no longer present and/or visible in that location (Altmann, 2004; Ferreira et al., 2008; Richardson & Spivey, 2000). This oculomotor behaviour indicates that the brain might generate an internal representation (*i.e.*, memory traces) of the stimulus that also includes its original spatial location (De Groot et al., 2016).

The looking-at-nothing phenomenon has been previously reported in the dyslexia literature by Jones et al. (2018), who examined the role played by episodic memory and statistical learning in acquisition of novel visual-phonological associations in typical readers and individuals with developmental dyslexia. In their study, Jones et al. (2018) expected a gradual reduction of looking-at-nothing behaviour as learning of cross-modal bindings unfolds and strengthens over time, which, if true, would be in line with previous findings on the phenomenon that showed a significant decrease in this oculomotor behaviour as a function of repetition (Scholz et al., 2011; Wantz et al., 2016). In general, participants with both typical and atypical reading profiles tended to gradually fixate less and less on target-relevant locations. In fact, the more strengthened the bindings became, the less the looking-at-nothing behaviour was observed to manifest. The looking-at-nothing behaviour displayed by participants in the study was interpreted by Jones and colleagues (2018) to reflect effortful retrieval. Their interpretation derived from the fact that their participants were, in general, predominantly less accurate whenever they fixated screen locations formerly occupied by the target as opposed to when they did not.

Two of the studies reported in this thesis (see Chapters 3 and 4) emulate the looking-at-nothing paradigm originally employed by Jones and colleagues (2018).

Innovatively, in these two studies, the looking-at-nothing phenomenon is probed *via* participants' personal webcams, an approach which, to the best of my knowledge, had never been done before. While adoption of this approach may have originally been driven by the lab closures during the pandemic, it is important to draw attention to its relevance to the field of experimental psychology research, particularly due to its cost-effective and accessible nature, whist also allowing data collection in real-world settings. Below is a description of the measures taken in the studies reported in this thesis to ensure collection of high-quality data whilst employing this approach.

Commercial eye trackers, commonly found in university research labs, use infrared light to track the position and movement of the eyes. Webcam-based eye-tracking, on the other hand, uses a webcam to remotely track the movement of a person's eyes in real-time (Papoutsaki et al., 2016). Essentially, through complex machine learning algorithms, pixels in the images of the participant's eyes and face are matched to specific features of the eyes, such as the pupils (Papoutsaki et al., 2016). Once the position of the eyes is identified, their movements are tracked in real-time, creating a record of where the individual likely looked on the computer screen and for how long. Open-source webcam-based eye-tracking libraries available online have been shown to reliably detect fixations and replicate findings of in-lab studies with relative accuracy as that shown in lab-based commercial eye-trackers (Papoutsaki et al., 2016; Semmelmann & Weigelt, 2018). In two of the studies reported in this thesis (Chapters 3 and 4), I used WebGazer.js (Papoutsaki et al., 2016), an open-source webcam-based eyetracking library hosted in Gorilla Experiment Builder (Anwyl-Irvine et al., 2020).

To account and control for the unavoidable variability inherent in webcam-based eye-tracking data collection (*e.g.*, different screen sizes and/or lighting

conditions), I employed a few strategies to enhance the quality of the data collected for this thesis. As will be described in Chapters 3 and 4, in order to account for differences in participants' screen sizes, I used normalised rather than raw coordinates in the statistical analyses. Specifically, *x* and *y* coordinates of participants' gazes were converted into a 'normalised' space, a necessary step that enables adequate comparison of eye movements detected across different screen sizes (See Figure 4).

Figure 4

Normalised predicted coordinates used for the eye-tracking and mouse-tracking data analyses

gorilla.sc

Gorilla Zone height: 789

1,1

907,441

Normalised Y = 441/789= -0.5589
Normalised X = (907-194)/1052 = -0.6778

0,0

Gorilla Zone x: 194

Screen width: 1440

Gorilla Zone width: 1052
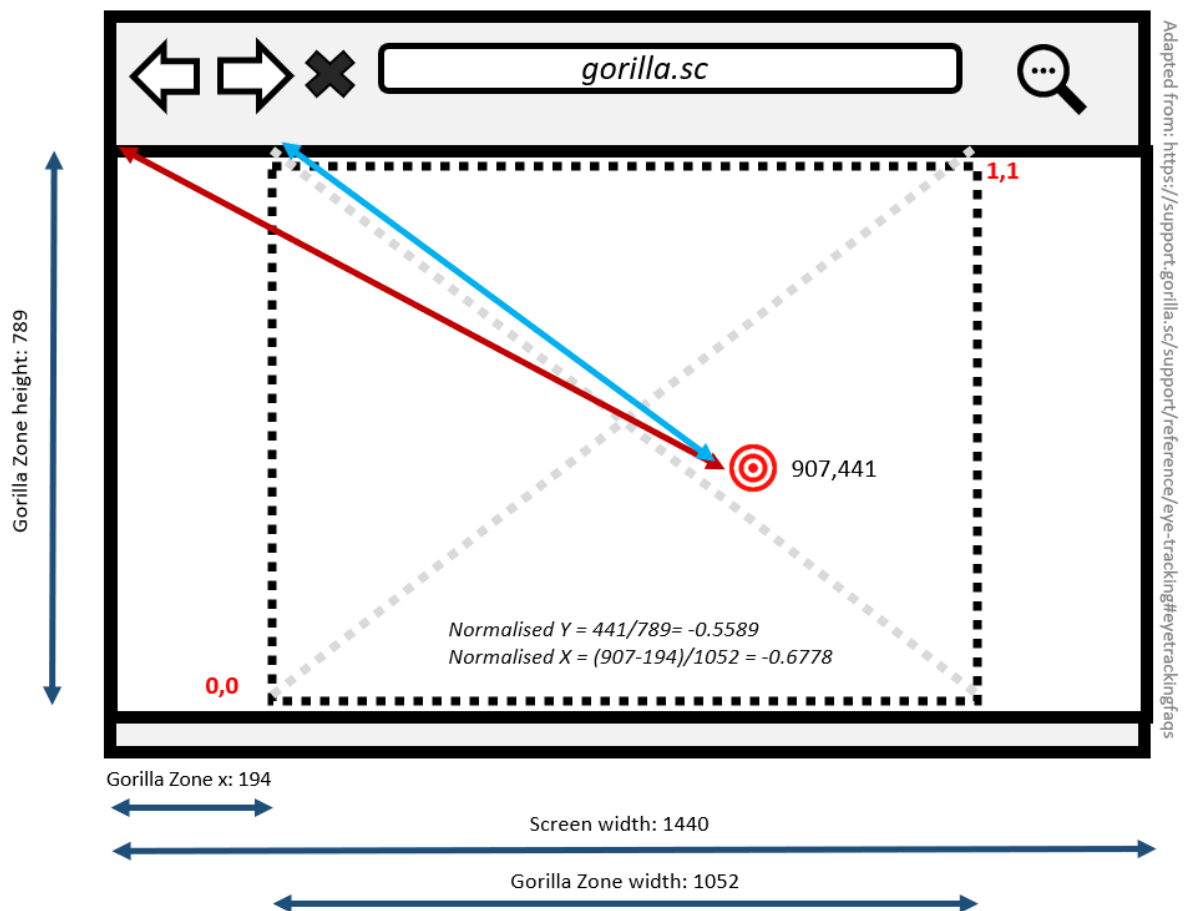
Note. This mock-up image, adapted from Gorilla Experiment Builder, is an illustration of how x and y coordinates are normalised into a unified space. By default, Gorilla Experiment Builder displays all experiments within a specific ratio frame (*i.e.*, 4:3), and normalised coordinates will always be relative to this ratio frame. This normalisation procedure makes eye-tracking and/or mouse-tracking coordinates comparable across different participants regardless of how big or small their screen is (*i.e.*, -0.5 and 0.5 will always refer to the centre of the screen).

Similarly, I used an innovative tool called Virtual Chinrest (Q. Li et al., 2020) in order to calculate participants' viewing distance. The Virtual Chinrest (Q. Li et al., 2020) performs this calculation by measuring participants' blind spot through their personal webcam, and uses the data gathered to ensure that visual stimuli are presented in a uniform way to all participants, irrespective of participants' screen sizes and/or their relative distance from the screen.

Finally, I made sure that the data cleaning process of the eye-tracking data was very rigorous and conservative. Webcam-based eyetracking uses machine learning classifier scores to determine how confident the method is in detecting a face through a webcam (See Figure 5). Whilst preparing to collect data for a pilot study using webcam-based eye tracking for the first time (see study reported in Chapter 3), I noticed that a range of different factors, such as whether participants were wearing glasses or not, the lighting conditions where they were, and where their laptops were placed, influenced how confident the algorithm would be at detecting faces. Therefore, to minimise the inclusion of data collected in suboptimal conditions, I excluded eye tracking estimates that showed low face detection values (*e.g.*, bottom picture in Figure 5).

Figure 5

Face detection procedure during webcam-based eye-tracking calibration phase
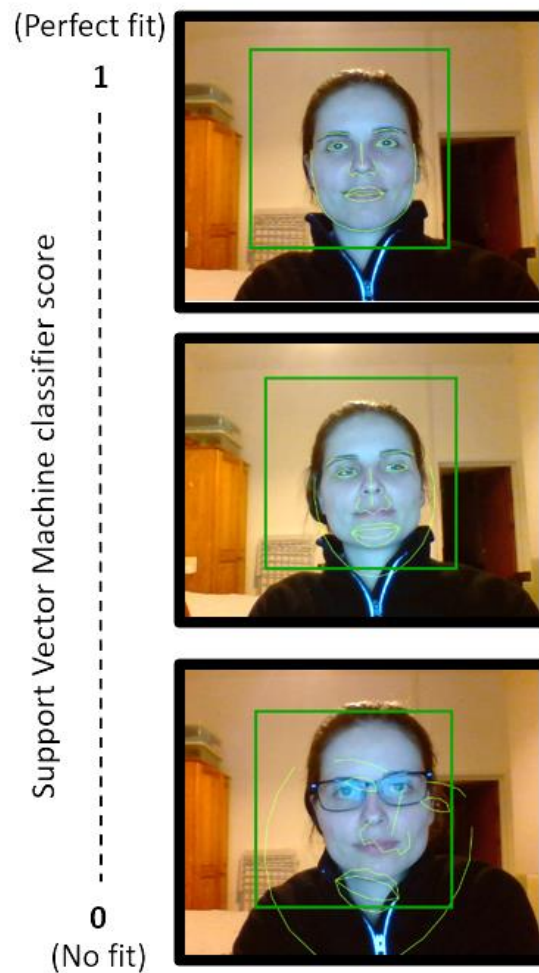
Note. The machine learning classifier score for the face detection procedure in webcam-based eyetracking ranges from 0 (no fit) to 1 (perfect fit). In the top picture, the green face detection 'mask' overlaps perfectly with the participant's facial features. Conversely, there is no overlap in the bottom picture, possibly due to the glare coming from the participant's glasses. Note that this image does not show an actual participant from any of the studies reported in this thesis, and has been included here for illustrative purposes only. It is also important to note that webcam-based eye-tracking does not record any videos and/or images during data collection.

2.2.2. Mouse-Tracking

In addition to using webcams to investigate a myriad of cognitive processes, the movement and clicks of an individual's computer mouse or touchpad can also be tracked in order to gain insight into the underlying cognitive processes involved in decision-making and other psychological phenomena (Kieslich et al., 2019, 2020; Maldonado et al., 2019). Some of the psychological constructs that have been investigated *via* mouse-tracking technology include cognitive control (*e.g.*, Ye & Damian, 2023), attention (*e.g.*, Egner et al., 2018), social perception (*e.g.*, Freeman & Johnson, 2016), emotion (*e.g.*, Neta et al., 2021), and decision-making (*e.g.*, Maldonado et al., 2019), to name a few.

To date, a few developmental dyslexia studies have employed mouse-tracking technology to investigate a range of topics, such as text readability (Rello & Bigham, 2017), cognitive control (Bundt et al., 2018), and word recognition (Lin & Lin, 2020). However, to the best of my knowledge, no previous paired associate learning study has included mouse movement data to investigate acquisition of novel cross-modal bindings in individuals with and without dyslexia.

In classic mouse-tracking paradigms, participants are presented with two response options, typically displayed in both top corners of the screen. Participants' hand trajectories are then recorded whilst they move their mouse cursor towards the response options. These mouse trajectories allow researchers to gather information on whether the competing – and unselected – response option was ever considered during the decision-making process, a highly sensitive measurement that can

complement accuracy and reaction time data (Freeman & Ambady, 2010; Freeman & Johnson, 2016; Kieslich et al., 2019, 2020).

Mouse tracking data can be analysed using a range of different indices. In Chapter 6, I report a novel paired associate learning task in which I look at three of the most commonly used mouse tracking indices, namely maximum absolute deviation, average deviation, and area under the curve (Kieslich et al., 2019, 2020; Wulff et al., 2021). Maximum absolute deviation (MAD) corresponds to the maximum distance measured between the participant's mouse trajectory and an idealised straight line between the start and end points of the response (See Figure 6) (Kieslich et al., 2019, 2020). Average deviation (AD) is the average distance between the participant's mouse trajectory and an idealised straight line between the start and end points of the response (See Figure 6) (Kieslich et al., 2019, 2020). Lastly, area under the curve (AUC) is the geometric area between the idealised trajectory and the observed trajectory, and reflects the overall strength of attraction towards each response option (Freeman et al., 2011; Freeman & Ambady, 2010; Kieslich et al., 2019, 2020). In sum, these three indices allow us to examine our participants' consideration of alternative choices and the level of hesitation or confidence in their final choice (Kieslich et al., 2019, 2020). MAD, AD and AUC values are highly correlated (Wulff et al., 2021), and larger values are indicative that the participant may have had more difficulty in choosing between the response options.

Figure 6

Depiction of the mouse movement curvature indices investigated in the current study
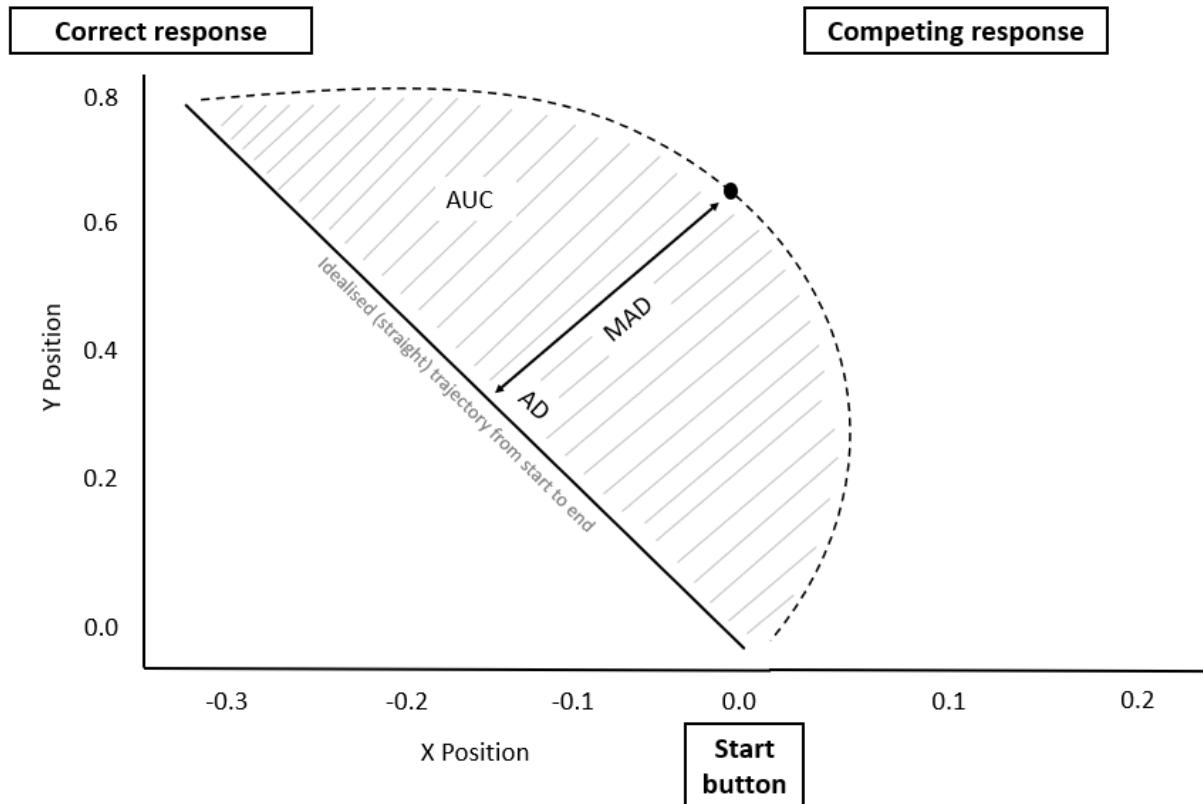
Note. The Maximum Absolute Deviation (MAD) represents the distance between the furthest point in the observed mouse trajectory (see dashed line) and the line that represents the shortest possible path (see solid line) connecting the start point and the correct response option. The Average Deviation (AD) is the mean deviation of the actual trajectory and the idealised direct path. Finally, the Area Under the Curve (AUC; see striped grey area) is the geometric area between the idealised and the observed mouse trajectories.

Due to the remote administration of the mouse-tracking task reported in this thesis (see Chapter 6), it was not possible to control for the type of hardware used by the participants. As a result, in order to account for differences in participants' screen

sizes, the statistical analyses of the mouse-tracking data also included normalised rather than raw coordinates. Similar to the webcam-based eye-tracking normalisation procedure described above, *x* and *y* coordinates of participants' mouse trajectories were converted into a 'normalised' space (See Figure 3). This step ensured adequate comparison of mouse trajectories detected across different screen sizes. It was also essential to only include mouse movement indices that would *not* require highly-specialised hardware for precise measurements of fine-grained hand movements. Our analyses, therefore, only focused on the three indices described above, namely Maximum Absolute Deviation, Average Deviation, and Area Under the Curve, as these are not highly complex mouse-tracking measurements that would require high levels of control in hardware use. Lastly, as will be described in detail in Chapter 6, the overall pre-processing of the mouse-tracking data was rigorous, including alignment of cursor start positions, thus accounting for potential cursor start variability across participants.

2.3. Background Literacy and Cognitive Measures

2.3.1. Literacy Measures

To assess whether the participants reported in this thesis demonstrate varying degrees of reading proficiency, with some exhibiting typical reading skills and others presenting with reading difficulties, a battery of background literacy measures was administered. Overall, we expected that individuals with a typical reading profile will score higher in all background literacy measures relative to those with atypical and poorer reading skills. However, it is important to note that some degree of variability in literacy abilities within each group should also be expected.

It is also important to note that there are different perspectives on how reading ability is assessed. In clinical and educational/academic settings, for instance, a binary approach to categorising individuals as having dyslexia or not may be necessary to determine eligibility for specific interventions and/or accommodations. Similarly, in research, a binary approach can be relevant to inform practice. Reading ability, however, can also be treated as a continuous variable, on a scale ranging from poor to proficient, an approach that acknowledges the complexity and variability of this skill. It is important to note that both binary and continuous approaches to measuring reading ability have their advantages and disadvantages, but each can provide valuable insights to participants' reading skills. Nevertheless, a detailed description of the issues surrounding these approaches is beyond the scope of this thesis.

In this thesis, reading ability was treated as a binary variable in the adult studies (*i.e.*, Chapters 3, 4, and 5) but as a continuous variable in the child study (*i.e.*, Chapter 6). For the adult groups, the binary approach was used because it would be possible to have access to self-reported diagnoses, as assessed by the Adult Reading Questionnaire (ARQ; Snowling et al., 2012). For the child study, reading was treated as a continuous variable due to lack of access to formal and/or self-reported diagnoses of reading (dis)ability.

All literacy tests reported in this thesis are briefly described below, along with a justification for their inclusion.

2.3.1.1. Self-Reported Measure of Reading Ability

The Adult Reading Questionnaire (ARQ; Snowling et al., 2012) was administered to collect a self-reported measure of dyslexia and dyslexia-associated traits in adults.

This screening tool, reported in Chapters 3, 4, and 5, includes items that collect self-reported data on reading and spelling proficiency, as well as the frequency in which the respondents engage in reading and writing on an everyday basis (Snowling et al., 2012). From this, we can estimate the likelihood that the respondents may have dyslexia, with higher scores suggesting a higher risk of reading disability (Snowling et al., 2012). In the ARQ, respondents are also encouraged to disclose whether they have had a diagnosis of dyslexia in the past. One advantage of the ARQ is that it is relatively quick and easy to administer, making it a useful screening tool to measure risk of reading disability in adults. While it is important to acknowledge that self-reported measures may not provide an accurate picture of an individual's true abilities, the ARQ has been shown useful in identifying adults with dyslexia-associated traits and/or at risk for dyslexia in several studies (Asbjørnsen et al., 2021; Gooch et al., 2014; Leavett et al., 2014; Moll et al., 2015; Snowling et al., 2012, 2019).

2.3.1.2. Rapid Automatised Naming (RAN)

As previously mentioned, the Rapid Automatised Naming (RAN) test is a measure of an individual's ability to quickly name a series of visually presented items, such as letters, numbers, or pictures (Denckla & Rudel, 1976). In a RAN task, the time taken to name the series of items is recorded along with the accuracy rate. This information can be used as an indicator of an individual's automatised naming ability, which is thought to be an important component of reading fluency (Norton & Wolf, 2012). Research has shown that individuals with a reading impairment, such as those with developmental dyslexia, tend to perform more poorly on RAN tasks than their typical counterparts (Torgesen et al., 1999). Specifically, children and adults

with dyslexia generally display longer naming times and are more error-prone on RAN tasks relative to typical readers (Torgesen et al., 1999). This suggests that RAN performance may be a useful measure for identifying individuals who are at risk for reading difficulties. To complement the self-report measure of reading difficulty collected via the ARQ (Snowling et al., 2012), the rapid digit naming and the rapid letter naming subtests from the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999) were administered to participants reported in Chapters 4 and 5. For the child participants reported in Chapter 6, the rapid digit naming and the rapid letter naming subtests from the Multilanguage Assessment Battery of Early Literacy (MABEL; Caravolas et al., 2018) were administered.

2.3.1.3. Word Reading Fluency

Reading fluency involves the ability to recognise words automatically and with ease, an essential skill for proficient reading. Both children and adults with an atypical reading profile, such as those with developmental dyslexia, generally demonstrate poorer reading fluency skills relative to typical readers (Berninger et al., 2006). Lack of reading fluency is characterised by slow and hesitant reading, resulting in impaired reading comprehension (Samuels, 2002). To test for differences in reading fluency performance in the experimental groups reported in this thesis, the Sight Word Efficiency subtest from the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999) was administered in the studies reported in Chapters 4 and 5, and the One-Minute Word Reading Fluency subtest from the Multilanguage Assessment Battery of Early Literacy (Caravolas et al., 2018) was administered in the study reported in Chapter 6. Both tests measure an individual's ability to quickly

and accurately read a list of high-frequency words. These subtests are timed, and the scores are based on the number of words read correctly within the given time.

2.3.1.4. Phonological Awareness

Phonological awareness refers to the ability to recognise and manipulate the individual sounds (*i.e.*, phonemes) within words. A test that is commonly used to assess phonological awareness skills is the phoneme deletion test. The Phoneme Deletion subtest from the Multilanguage Assessment Battery of Early Literacy (Caravolas et al., 2018) was used in the study reported in Chapter 6. This test measures children's ability to delete specific phonemes from a word (*i.e.*, the onset or the coda) and then say the remaining word.

While we acknowledge that including other phonological tasks, such as phonological short-term memory and/or phoneme blending, could provide more comprehensive and/or alternative insights into our participants' phonological awareness skills, practical considerations weighed into our decision-making process. The phoneme deletion test from the Multilanguage Assessment Battery of Early Literacy (Caravolas et al., 2018) was the phonological awareness assessment of choice for the Remote Instruction of Language and Literacy programme, the larger language and literacy instruction project from which we pooled the participants in the study reported in Chapter 6. It is important to note that the phoneme deletion task is one of many measures of phonemic awareness, and we recognise the importance of including a broader range of phonological assessments in future studies to further explore the interplay between paired associate learning and phonological processing.

2.3.1.5. Word Reading Accuracy

Word reading accuracy can be construed as the ability to correctly identify and recognise written words while reading. Individuals with impaired reading ability generally perform less well in tasks requiring word reading accuracy skills, compared to those with a typical reading profile (Berninger et al., 2006). To test for differences in performance at word reading accuracy between the experimental groups reported in this thesis, the WRAT-4 Word Reading subtest (Wilkinson & Robertson, 2006), which is a standardised test that assesses an individual's ability to read words aloud accurately, was administered to participants in the studies reported in Chapters 4, 5 and 6.

2.3.1.6. Phonemic Decoding Efficiency

Phonemic decoding efficiency is a measure of an individual's ability to read pseudowords. Assessing phonemic decoding efficiency provides insights into an individual's ability to apply phonological awareness to decode unfamiliar words. The phonemic decoding efficiency subscale from the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999) was administered to participants in the studies reported in Chapters 4 and 5, and the One-Minute Pseudoword Reading Fluency subtest from the Multilanguage Assessment Battery of Early Literacy (MABEL; Caravolas et al., 2018) was administered to the children participating in the study reported in Chapter 6. These subtests are timed, and the scores are based on the number of words read correctly within the given time.

2.3.2. Cognitive Measures

To gain a comprehensive understanding of the cognitive abilities of the participants included in this thesis, a battery of background cognitive tests was also administered. For these cognitive measures, we did not expect to find any significant differences in performance between participants with typical and atypical reading profiles. However, it is important to note that, within each group, some degree of variability in cognitive abilities should also be expected. All cognitive tests are briefly described below, along with a justification for their inclusion.

2.3.2.1. Nonverbal and Verbal Intelligence

Nonverbal intelligence refers to our ability to understand and process information using predominantly visual and hands-on reasoning, whilst verbal intelligence refers to our use of language-based reasoning to analyse information and solve problems. Individuals with dyslexia and those with a typical reading profile do not generally differ in these two measures (Lyon et al., 2003). Despite the fact that intelligence quotient (IQ) has consistently been shown to be overall independent of reading ability (Cutting & Scarborough, 2006; Landi, 2010), it is important to administer tests of verbal and nonverbal IQ to the participants in the studies reported in this thesis to ensure that the experimental groups would not differ in these two measures to a great extent. By doing so, we can be more certain that any behavioural differences detected between our experimental groups will likely be reading specific.

In the studies reported in Chapters 4 and 5, the Matrix Reasoning and Similarities subtests of the Wechsler Abbreviated Scale of Intelligence (WASI) were administered as proxy measurements for participants' nonverbal and verbal IQ,

respectively. The Matrix Reasoning subtest measures an individual's ability to perceive and analyse visual-spatial relationships and to reason logically, whereas the Similarities subtest measures the ability to think abstractly and understand relationships between words. In the study reported in Chapter 6, the WRIT Matrices subtest (Glutting et al., 2000) was used to assess nonverbal intelligence. However, due to time constraints, no verbal intelligence data were collected from the children participating in the study reported in Chapter 6.

2.3.2.2. Digit Span

Digit span, a commonly used measure of working memory, refers to the longest list of digits that an individual can correctly remember and repeat back immediately after presentation. Digit span tasks can be administered in two forms: forwards and backwards. In the forward condition, participants hear a series of digits and are asked to repeat them back in the same order, whereas in the backwards condition, participants are asked to repeat them back in the reverse order. Digit span measures were included in the battery of background tests reported in Chapters 4 and 5 in this thesis to ensure that no participants with a significant deficit in working memory capacity would be included in our sample.

The digit span tasks reported in this thesis were adapted to online administration. Specifically, participants were required to type in their responses instead of repeating them back verbally. While this non-standard implementation of the digit span task could be deemed a limitation, previous studies using digit span tasks that required typed rather than verbal responses used this approach successfully (*e.g.*, Olsthoorn et al., 2014).

2.3.2.3. Spatial Span

The Spatial Span is another measure of working memory, specifically visuo-spatial working memory (Alloway et al., 2006). Spatial Span tasks can also be administered in forwards and backwards forms. However, instead of remembering lists of digits, participants are tested on their ability to remember and repeat back a series of spatial locations in the correct order, be it forwards or backwards. This measure was included to complement the digit span data as proxies of working memory capacity.

In in-person testing, when administering a spatial span task, physical boards are traditionally used, and participants are required to use hand gestures to provide their responses. The spatial span tasks reported in this thesis are an online adaptation that require participants to perform mouse-clicks on squares to recreate the randomly presented series of spatial locations.

2.3.3. Stimulus Design

As outlined in the introductory chapter, the central research questions in this thesis focus on paired associate learning of novel cross-modal bindings in individuals with typical and atypical reading profiles. In paired associate learning tasks, it is important to use novel and/or unfamiliar stimuli to control for any potential influence of prior knowledge or experience (Aravena et al., 2013).

The stimuli choice for the present thesis emulates experimental stimuli used in previous studies that looked at paired associate learning of novel cross-modal bindings in individuals with and without reading impairments (*e.g.*, Jones et al., 2018). Specifically, in all studies reported in this thesis, participants are required to

learn arbitrary associations between (a) pseudowords that obey the phonotactic constraints of the English language (*i.e.,* restrictions on permissible sound sequences and syllable structures), and (b) Kanji characters, logographic Chinese characters used in written Japanese, completely unfamiliar to the participants in the studies reported in this thesis (*e.g.*, /*pof*/ and 人).

All the pseudowords used here followed a monosyllabic consonant-vowel-consonant (CVC) structure, were recorded by native speakers of British English, had their overall quality optimised in Praat (Paul Boersma & van Heuven, 2001), and were generated *via* Wuggy, a pseudoword generator software (Keuleers & Brysbaert, 2010).

The Kanji characters used in this thesis had 3 strokes or less to keep visual complexity to a minimum. Rather than creating a completely novel script for this thesis, as previously observed in analogous paradigms (*e.g.*,Toffalini et al., 2019), I deliberately opted to use an existing script (*i.e.*, Kanji characters) as this would allow us to emulate the characteristics of graphemes as they naturally occur, an approach also taken in previous studies (*e.g.*, Aravena et al. 2013; Jones et al., 2018). This decision is particularly important because the way real world letters are shaped may not be a result of arbitrary cultural decisions, but, instead, their shapes might be a product of our own neural architectures (Dehaene, 2009).

2.3.4. Data Analyses

Data collected in online research is inherently noisier than that collected in controlled lab-based studies (Finley & Penningroth, 2015). Therefore, it is paramount that

researchers be more rigorous when cleaning and analysing datasets collected under these circumstances.

2.3.4.1. Mixed-Effects Models

In addition to the rigorous pre-processing steps briefly described above and further explained below in each empirical chapter, I used generalised and/or linear mixed-effects models to analyse the data in all four studies reported in this thesis. This statistical approach has become increasingly prominent in experimental psychology and cognitive neuroscience investigations due to its versatility and rigour (Barr et al., 2013; Bates et al., 2015). Generalised and linear mixed-effects models have several advantages over other statistical approaches, such as t-tests and ANOVAs. In particular, mixed-effects models excel at accommodating missing data and dealing with unbalanced data sets, generally provide superior fit to the data and increase generalisability of findings, and adequately take data dependence into account (Barr et al., 2013; Bates et al., 2015). The mixed-effects models reported in this thesis include maximal random effects structures (Barr et al., 2013), reverting to a 'parsimonious' model in case of convergence errors (Bates et al., 2015). This conservative approach was taken to minimise the occurrence of Type 1 error as well as to optimise generalisation of findings (Barr et al., 2013).

2.3.4.2. Signal Detection Theory

In Chapter 5, in addition to generalised and linear mixed-effects models, I also use signal detection theory to analyse participants' performance on an n-back task, an experimental paradigm commonly used in the working memory literature (Baddeley et al., 2010; He et al., 2022; Kirchner, 1958; W. Li et al., 2021; Pelegrina et al., 2015; Szmalec, Verbruggen, et al., 2011).

Signal detection theory is a quantitative theoretical framework used to assess an individual's ability to make judgements based on ambiguous and/or uncertain information (D. Green et al., 1966; Lynn & Barrett, 2014) . The overarching goal of signal detection theory is to quantify an individual's ability to distinguish between signal and noise in a given task. Essentially, the sensitivity of an individual to a given stimulus is measured by an index called d-prime (d'), also known as sensitivity index, which reflects the distance between the distributions of responses to the signal and noise. Higher d-prime values indicate that the individual is better able to discriminate the signal from the noise. Decision criterion (c'), on the other hand, reflects an individual's willingness to respond "yes" to a stimulus, and is influenced by factors such as perceived costs and benefits of making a correct or incorrect decision. Higher decision criterion values indicate that the individual is more conservative and/or cautious in their response, possibly requiring stronger evidence that the stimulus *is* indeed a signal before indicating that they have detected the signal (D. Green et al., 1966; Lynn & Barrett, 2014).

Signal detection theory provides a useful framework for analysing n-back data because this approach takes into account the different types of errors that participants can potentially make in this type of task (Meule, 2017). In sum, in an n-back task, participants are required to monitor a sequence of stimuli (*e.g.*, written letters) and indicate whether the current stimulus matches the one presented *n* trials earlier. Through a signal detection theory approach, participants' responses in a 2-back task can be classified into four different categories:

1. Hits: A hit would occur when the participant correctly identifies that the current stimulus matches the one presented two trials back;

2. Misses: A miss would occur when the participant fails to identify that the current stimulus matches the one presented two trials back;

3. False alarms: A false alarm would occur when the participant incorrectly identifies that the current stimulus matches the one presented two trials back, even though it does not;

4. Correct rejections: A correct rejection would occur when the participant correctly identifies that the current stimulus does not match the one presented two trials back, and refrains from responding.

Categorising responses in terms of hits, misses, false alarms, and correct rejections provides a more detailed and nuanced understanding of performance on the n-back task, allowing to differentiate potential lapses in attention (*i.e.*, a false alarm) from memory failures (*i.e.*, a miss), thus resulting in more sensitive data analysis (Meule, 2017).

2.3.4.3. Path Analyses

In Chapter 6, in addition to generalised and linear mixed-effects models, I also use path analyses, a subset of structural equation modelling.

Path analysis is a statistical approach that can be used to examine the relationships among multiple variables, allowing us to test causal hypotheses and determine the strength of the relationships between variables. Path analyses are commonly used in reading research to examine potential relationships between reading related skills, such as phonemic awareness and rapid naming, and reading outcomes, such as word reading accuracy and comprehension (*e.g.*, Warmington & Hulme, 2012b; Windfuhr & Snowling, 2001).

2.4. Chapter Summary

This chapter laid the methodological foundation for the subsequent empirical chapters, in which I report four different paired associate learning studies designed to investigate acquisition of novel cross-modal bindings in individuals with typical and atypical reading profiles (see Table 1 for an overview of the methods used across the four experimental studies reported in this thesis). The studies reported in this thesis incorporate a unique approach by transitioning from in-lab to remote data collection, which presented its own set of challenges. Despite the challenges faced, all the control measures that were put in place, as briefly outlined above, enabled successful data collection, potentially opening up new possibilities for future research in experimental psychology.

Table 1

Overview of the methods used across the four experimental studies reported in this thesis.

Research Question	Study Title	Group	Sample Size	Age	Methodology	Background Tests	Statistical Analyses
How do task-irrelevant episodic details modulate visual-phonological binding performance in individuals with and without dyslexia?	Episodic Memory Cues in Acquisition of Novel Visual-Phonological Associations: a Webcam-Based Eye-Tracking Study	Typical readers	$N = 14$	$M = 22.6$, $SD = 4.21$	Webcam-based eye tracking; Behavioural measures (i.e., accuracy data and reaction time)	Adult Reading Questionnaire (Snowling et al., 2012)	Logistic mixed effects regression
	Audiovisual Learning in Dyslexic and Typical Adults: Modulating Influences of Location and Context Consistency	Typical readers and readers with dyslexia	$N = 70$	Typical readers: $M = 23.55$, $SD = 6.14$; Readers with dyslexia: $M = 28.17$, $SD = 7$	Webcam-based eye tracking; Behavioural measures (*i.e.*, accuracy data and reaction time)	Adult Reading Questionnaire (ARQ, Snowling et al., 2012); Word reading efficiency and phonemic decoding efficiency subscales of the Test of Word Reading Efficiency (TOWRE, Torgesen et al., 1999); Letter and digit versions of the Rapid Automatized Naming (RAN) subtest from the Comprehensive test of Phonological Processing (CTOPP, Wagner et al., 1999);	Logistic mixed effects regression

Research Question	Study Title	Group	Sample Size	Age	Methodology	Background Tests	Statistical Analyses
						Similarities subtest from the Weschler Adult Intelligence Scale (WAIS, Wechsler, 1981);	
						Matrix Reasoning from the Wechsler Abbreviated Scale of Intelligence (WASI, Wechsler, 1999; Computerised forward and backward digit span.	
Is there a specific cross-modal binding deficit in adults with developmental dyslexia, even in the absence of spoken output demands?	How Specific is the Paired Associate Learning Deficit in Dyslexia to Cross-Modal Working Memory?	Typical readers and readers with dyslexia	*N* = 96	Typical readers: *M* = 21.09, *SD* = 3.30; Readers with dyslexia*:* *M* = 23.34, *SD* = 3.73	Behavioural measures (i.e., accuracy data and reaction time)	Adult Reading Questionnaire (Snowling et al., 2012); Word reading efficiency and phonemic decoding efficiency subscales of the Test of Word Reading Efficiency (Torgesen et al., 1999); Wide Range Achievement Test's (WRAT4) Word Reading subtest (Wilkinson & Robertson, 2006); Letter and digit versions of the Rapid Automatized Naming (RAN) subtest from the Comprehensive test of Phonological Processing (Wagner et al., 1999); Matrix Reasoning from the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999);	Logistic mixed effects regression, linear mixed effects regression analyses, and signal detection theory.

Research Question	Study Title	Group	Sample Size	Age	Methodology	Background Tests	Statistical Analyses
						Similarities subtest from the Wechsler Adult Intelligence Scale (Wechsler, 1981); Computerised forward and backward digit span tests; and computerised forward and backward spatial span tests.	
Can online measures elucidate Paired Associate Learning mechanisms in beginning readers? And what is the relationship between paired associate learning response accuracy and reading outcomes in these children?	What the Hand in Motion Reveals about Reading: Children's Decision-Making Processes in Paired Associate Learning and its Relationship to Reading Outcomes	Unselected group of school-aged children	$N = 80$	$M = 9:98$, $SD = 1:10$	Mouse-tracking; Behavioural measures (i.e., accuracy data and reaction time)	WRAT-4 Word Reading subtest (Wilkinson & Robertson, 2006); MABEL One-Minute Word Reading Fluency subtest (Caravolas et al., 2018); MABEL One-Minute Pseudoword Reading Fluency subtest (Caravolas et al., 2018); Phoneme Deletion test (Caravolas et al., 2018); Rapid Automatised Naming tests of letters and digits (Caravolas et al., 2018); WRIT Matrices subtest (Glutting et al., 2000)	Logistic mixed effects regression and path analyses.

Chapter 3
Episodic Memory Cues in Acquisition of Novel Visual-Phonological Associations: a Webcam-Based Eye-Tracking Study

Abstract

When learning to bind visual symbols to sounds, to what extent do beginning readers track seemingly irrelevant information such as a symbol's position within a visual display? In this study, we used adult typical readers' own webcams to track their eye movements during a paired associate learning task that arbitrarily bound unfamiliar characters with monosyllabic pseudowords. Overall, participants' error rate in recognition (Phase 1) decreased as a function of exposure, but was not modulated by the episodic memory-based effect of 'looking-at-nothing'. Moreover, participants' lowest error rate in both recognition and recall (Phases 1 and 2) was associated with item consistency across multiple exposures, in terms of spatial and contextual properties (*i.e.*, stimulus' screen location and co-occurrences with specific distractor items during encoding). Taken together, our findings suggest that normally developing readers extract statistical regularities in the input during visual-phonological associative learning, leading to rapid acquisition of these pre-orthographic representations.

Keywords: Episodic memory; looking-at-nothing; paired associate learning; cross-modal binding; reading.

Episodic Memory Cues in Acquisition of Novel Visual-Phonological Associations: a Webcam-Based Eye-Tracking Study.

Associative learning is a key skill underlying reading development. In initial stages of literacy acquisition, written or printed symbols (*i.e.*, graphemes), which at first appear meaningless, gradually begin to evoke specific linguistic representations (*i.e.*, phonemes). Repeatedly accessing such phonological associations in response to visual stimuli (*i.e.*, letters) progressively automatises that process (Ehri, 2005; Ehri & Saltmarsh, 1995; M. Jones et al., 2018) resulting in the apparent effortlessness of skilled reading. Performance in visual-verbal versions of the paired associate learning task – an episodic memory paradigm which assesses the ability to accurately bind two distinct items together in memory (Scorpio et al., 2018) and retrieve them later as a single entity (Brockmole & Franconeri, 2009) - has been shown to discriminate typical readers from those with dyslexia (*e.g.*, Jones et al., 2018; Toffalini et al., 2018; Wang, Wass, & Castles, 2017). Paired associate learning performance accounts for unique variance in reading ability, and impairments to the underlying skills appear to result in clinically significant reading difficulties (Litt & Nation, 2014; H. C. Wang et al., 2017), supporting the assumption that the task taps abilities that are crucial for skilled reading acquisition.

Reading acquisition thus appears to build on episodic memory. In episodic memory, contextual properties, such as temporal and spatial information, are encoded alongside salient task features (Tulving, 1972). These properties, which share patterns of neural activity, can be used as cues to aid memory retrieval (El-Kalliny et al., 2019). To illustrate, if Event A is encoded in temporal proximity to Event B, exploiting the temporal relationship between the two events may facilitate their subsequent retrieval from the episodic memory system when needed (Tulving,

1972; El-Kalliny et al., 2019). Episodic memory-based investigations focusing on learning of arbitrary visual-phonological associations demonstrated that typical readers, but not individuals with dyslexia, are sensitive to consistent spatial cues presented *across multiple trials* (Albano et al., 2016; M. Jones et al., 2018; Toffalini et al., 2018). Typical readers' sensitivity to spatial cues extends to their oculomotor behaviour: when given a visual cue, they fixate blank screen locations previously occupied by a target item, resulting in greater probability of accurate phonological recall (Jones et al., 2018).

Returning to a spatial location in which salient information was originally presented is an unconscious oculomotor behaviour that is triggered by the reactivation of internal memory representations (Ferreira et al., 2008; Richardson & Spivey, 2000). This behaviour is believed to play a functional role in memory retrieval (Richardson & Spivey, 2000; Scholz et al., 2018), modulating retrieval of both visual and auditory information (Scholz et al., 2016). The phenomenon seems to occur even when encoding of spatial information is task-irrelevant (Richardson & Spivey, 2000) and thus encoded incidentally. The episodic memory-based effect of 'looking at nothing' when trying to remember something gradually diminishes as learning unfolds and representations strengthen over time (Scholz et al., 2016; Wantz et al., 2016).

To date, however, the effect of presentation consistency in the episodic trace on visual-phonological binding accuracy in typical readers is relatively underexplored. Here, we begin to elucidate the cognitive underpinnings of efficient orthographic-phonological representations in typical readers.

The Current Study

We examine whether typical readers efficiently use a combination of spatial and contextual cues to aid learning of novel cross-modal bindings, taking a full and accurate snapshot of the episodes to facilitate the visual-phonological binding. To test this, we designed a paired associate learning task in which we manipulated consistency of stimuli's spatial locations and their co-occurrences across multiple exposures. Our goal is to probe whether these episodic cues, when combined, modulate recognition and recall of novel visual-phonological associations in typical readers. We also examined whether 'looking-at-nothing' behaviour would emerge in the current study at the trial level, and if so, whether directing one's gaze towards relevant empty screen locations would aid recognition of the novel associations.

We tracked participants' eye movements remotely with their webcams during a paired associate learning task in which Kanji characters – which were unfamiliar to these native British English speakers – were arbitrarily but consistently bound to monosyllabic pseudowords adhering to phonotactic constraints in English. On each trial, as in Jones et al. (2018), participants were prompted to encode three characters, one at a time, along with their corresponding pseudowords. An auditory cue with the target pseudoword followed the encoding phase. After a blank screen, during which we tracked participants' eye movements, participants were then tested on their ability to recognise the corresponding character associated with the auditory cue. Our manipulation of consistency of stimuli's locations and intra-trial co-occurrences ('context', henceforth) resulted in four different trial types. Consistent location involved Kanji characters appearing in the same screen location across trials, whereas consistent context involved characters appearing with the same

distractor items across trials. A separate cued-recall task was administered to assess lasting retention of the visual-phonological associations.

Based on previous empirical findings that typical readers gradually automatise retrieval of visual-phonological associations over time (M. Jones et al., 2018), performance in later blocks should be superior as a function of repetition, which, in turn, will be an indication of incremental learning.

If typical readers are able to efficiently use *multiple* episodic cues present during encoding in order to aid recognition of the novel visual-phonological associations, then they should err less when both location and context are kept consistent across trials, as compared to when they are not. Furthermore, if encoding under the consistent location/ consistent context condition is indeed more robust than in the other conditions as a consequence of the regular episodic cues, then we will also observe longer-lasting retention of the bindings encoded under this condition (as assessed by a separate cued-recall task following the main recognition task) showing that typical readers not only efficiently detect regularities in the stimuli but also use them to their advantage.

Considering that visually revisiting empty screen locations previously occupied by targets has been shown to aid memory retrieval, we expected looking-at-nothing behaviour to also emerge in our study.

Finally, one unique methodological aspect of this study is its use of a webcam-based method for remote eye-tracking. Previous research on the role of looking-at-nothing behaviour in paired associate learning has been conducted in-lab with specialised hardware. Here, we set out to investigate the phenomenon remotely using WebGazer.js, an open-source webcam-based eye-tracking JavaScript library

(Papoutsaki et al., 2016) which has been shown to reliably detect fixations and replicate findings of in-lab cognitive science studies with reasonably comparable accuracy (Semmelmann & Weigelt, 2018). Without transmitting videos or pictures, WebGazer.js uses participants' webcams to infer on-screen gaze locations with an average error of approximately 100 pixels. Thus, this study provides a test of the method's suitability as a flexible, low-cost alternative for 'looking-at-nothing' research.

Method

Participants

Fourteen university students (age: $M = 22.6$, $SD = 4.21$, 13 females) participated remotely in this experiment. One additional participant was excluded due to an error rate more than three standard deviations above the group mean. All were native speakers of British English, recruited through Bangor University, and none reported any history of psychiatric and/or neurological diseases, visual acuity, hearing, or any other risk factors. Crucially, all participants self-reported normal or skilled reading ability in the Adult Reading Questionnaire (Snowling et al., 2012). All participants were naïve to the purpose of the experiment, and had never seen nor heard any of the stimuli before. The experiment was approved by the Bangor University Ethics Committee, and participants provided informed consent and received payment for participation.

Stimuli, Design and Procedure

Phase 1: Recognition Task The task was programmed and hosted on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants were not allowed to do the task on mobile phones or tablets. Participants' physical distance from the screen was calculated with the Virtual Chinrest task (Q. Li et al., 2020), which indicated an

average sitting distance of 50.88 cm from their monitors ($SD = 8.59$). Participants were instructed to sit still, and to avoid head movements and/or to look away from the screen during the task. Pictorial instructions were included in an attempt to collect higher data quality. A 5- point calibration was performed at the beginning of the main task and every 18 trials (*i.e.*, mid-block and at the onset of a new block). Calibration was re-attempted whenever the calibration prediction for at least one of the five calibration points approximated an incorrect one.

Thirty-six Kanji characters were arbitrarily matched to 36 monosyllabic pseudowords (*e.g.*, 'kig'), generated with Wuggy (Keuleers & Brysbaert, 2010) according to English-like phonotactic constraints. The auditory stimuli were recorded by a female native speaker of British English. Character-sound pairings were kept constant across the experiment such that each character was always bound to the same pseudoword.

Each trial began with a 1000-ms fixation cross. Then, three Kanji characters appeared in black on white background in triangle formation (See Figure 7). Each character occupied 20x20 units within a 4:3 window in Gorilla Experiment Builder's screen space. One at a time, each character would pseudo-randomly highlight in red while its corresponding pseudoword played in the background (participants were encouraged to use earphones or headphones to listen to the stimuli). A 1000-ms blank screen followed, and then a visual backward masking phase, during which hash symbols and numbers replaced the target stimuli on the screen to minimise iconic memory. Then the 'testing phase' began. A black dot appeared in the centre of the screen; participants were instructed to click on it to hear one of the three pseudowords: the 'target' for the testing phase. This clicking instruction also provided a crucial attention check: participants were automatically excluded from the

experiment if they failed to click on the dot within 10 seconds in three consecutive trials. A 1000-ms blank screen followed, during which participants' eye movements were recorded via WebGazer (Papoutsaki et al., 2016) with a sampling rate of 60 Hz. The three Kanji characters re-appeared; to encourage participants to encode character-sound associations, characters' spatial positions changed between the encoding and testing phases in two thirds of the trials[3]. Participants were prompted to click on the character that corresponded to the target audio. The characters remained on the screen for up to 5000 ms or until a mouse-click was detected, after which a 250-ms blank screen terminated the trial.

Figure 7

Timeline of a single trial in the recognition task
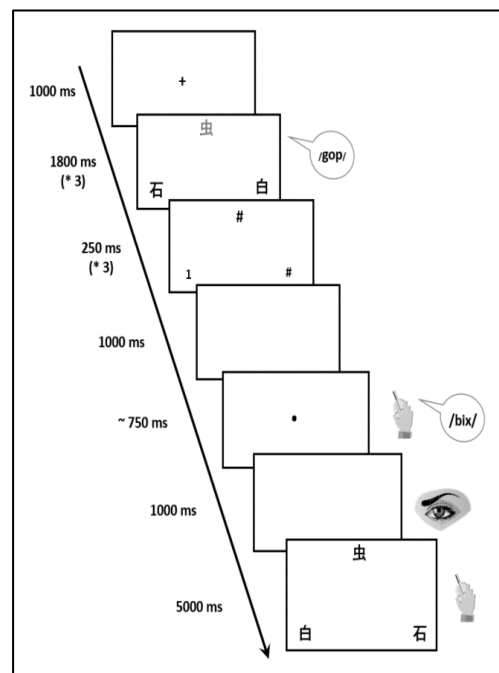
Note. The eye denotes the screen in which we expected to detect 'looks-at-nothing' behaviour*.*

---

[3] Due to the automatic and unconscious nature of the 'looking-at-nothing' behaviour (Ferreira et al., 2008); Richardson & Spivey, 2000), we did not expect this manipulation to prevent participants from re-fixating relevant screen locations.

We orthogonally manipulated two aspects of the encoding phase: 1. *Location consistency*: whether a target character consistently appeared in the same spatial location throughout the experiment, and 2. *Context consistency*: whether a target character consistently appeared with the same two other characters throughout the experiment. Thus, of the 36 Kanji characters, 18 always appeared in the same screen position, whereas 18 characters varied in position. Similarly, half of the stimuli consistently co-occurred with the same two other characters, whilst the remaining 18 did not have any fixed co-occurrences.

To ensure attention to the phonological component of the bindings, we interspersed cued-recall trials within each block at regular intervals (*i.e.*, every six trials). In each trial, a Kanji character was shown in the middle of the screen (see Figure 8), after which participants were prompted to articulate the corresponding pseudoword. The target for each interspersed recall trial (*N*= 36) was a character randomly selected from one of the six preceding recognition trials.

Figure 8

Timeline of a single trial in the cued-recall task
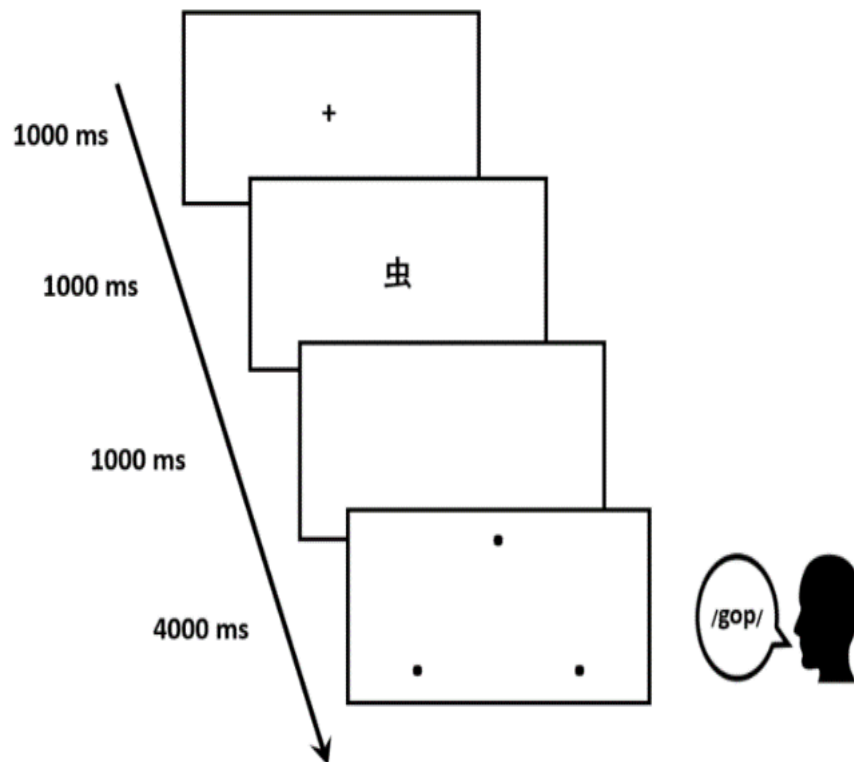
The 252 trials (216 recognition trials plus 36 interspersed cued-recall trials) were presented over 6 blocks, between which participants were encouraged to take short breaks. Trials' assignment to blocks was pseudo-randomised to ensure that all conditions were equally frequent within a block. Presentation of blocks and of trials within each block was randomised across participants to avoid order effects.

Five practice trials (*i.e.*, four recognition trials and one recall trial) representative of those used in the actual experiment were presented in order to familiarise the participants with the procedure. None of the practice items were used during the experiment. Feedback was provided to participants during the practice block, but not in the experimental trials.

Phase 2: Cued Recall A separate cued-recall task comprising the same visual-auditory stimuli from the previous task was administered on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020) immediately after Phase 1. The task consisted of a single block with 36 trials. Each trial, methodologically identical to the above mentioned interspersed cued-recall trials, started with a 1000-ms fixation cross, followed by a Kanji character presented in black on a white background (See Fig. 8). The character was presented in the center of the screen for 1000-ms, and occupied 20x20 units of screen space within a 4:3 window. Three black dots, presented in the same triangle formation as Phase 1, indicated that a voice response was required. Participants were allowed 4 seconds to provide a verbal response. A 250-ms blank screen terminated the trial. Trial presentation was randomised across participants to avoid order effects. Eye-tracking metrics were not recorded in this task.

Total experiment duration averaged 105 minutes. An automatic time limit of 150 minutes ensured that participants would complete the experiment in one sitting.

Data Analysis

Eye tracking. Eye-tracking metrics recorded by Gorilla Experiment Builder (Anwyl-Irvine et al., 2020) include a face convergence value column, which comprises a score ranging from 0 to 1 for the face model fit. The face convergence value indicates how strongly the image detected resembles a face: 0 means no fit and 1 means perfect fit. Gorilla's recommendation is to trust face convergence values over 0.5. We excluded eyetracking estimates below that threshold in our analyses.

Under ideal conditions, WebGazer.js (Papoutsaki et al., 2016) is able to generate up to 60 eyetracking estimates (*i.e.*, predictions) per second with x and y coordinates of where on the screen the subject is predicted to be looking. However,

the number of predictions largely varies depending on participants' hardware, lighting conditions, among other things. In addition to these predictions, Gorilla Experiment Builder (Anwyl-Irvine et al., 2020) translates the coordinate data into a 'normalised' space, in which -0.5 and 0.5 will always be the center of the screen regardless of its size. This normalisation allows eye movements detected across different screen sizes to be compared. We used the normalised coordinates in our analyses.

Regression analyses. Analyses used confirmatory logistic mixed effects regression, via the glmer::binomial function in the lme4 v1.1-23 library (Bates et al., 2015) in R v4.0.0 (R Core Development Team, 2020), including maximal random effects structures (Barr et al., 2013) reverting to a 'parsimonious' approach in the case of convergence errors (Bates et al., 2015). For the recognition task in Phase 1, error rate was modelled as a function of *Location consistency* ("LocationC", *i.e.*, whether a target character consistently appeared in the same spatial location throughout the experiment; consistent = -0.5, inconsistent = 0.5), *Context consistency* ("ContextC", *i.e.*, whether a target character consistently appeared with the same two other characters throughout the experiment; consistent = -0.5, inconsistent = 0.5), and *Block*, a predictor tracking target repetition, log-transformed to account for the fact that repetition effects follow a logarithmic function. Following Jones et al. (2018), to probe whether participants' looks back at blank screen locations previously occupied by targets would facilitate recognition of those items, we also included two eyetracking-related binomial predictors: (1. a binomial predictor indicating whether we identified fixations on any region of interest during the blank screen immediately preceding the testing phase ("FixatedAnyROI", no = −0.5, yes = 0.5), and 2. a nested binomial predictor indicating whether the participant fixated the former location of the target more than the former locations of the distractors ("PrimaryFixation", target = -

0.5, distractor = 0.5, no fixations = 0.0). All predictors were contrast-coded and centered. For Phase 2's cued-recall task, accuracy was modelled as a function of Location consistency and Context consistency, as described above.

Results

Phase 1 (Recognition Task)

We excluded 30 (out of 3024) trials without behavioural responses (*i.e.*, mouse clicks), leaving 2994 trials. The eye tracking procedure generated 52,204 fixation estimates across these 2994 behaviourally valid trials. We excluded 1.39% of those estimates (N=726), due to face convergence values below 0.5, indicating low-confidence eyetracking estimates. Finally, to address questions about looking-at-nothing behaviour, in this paper, we focus our analyses on just the 2093 behaviourally valid trials with at least one valid eyetracking estimate; as illustrated in Figure 9, this restricted dataset is behaviourally very similar to the larger dataset. The mean face convergence value for these remaining trials was 0.77 ($SD = 0.12$), suggesting a sufficient basis for estimating eye movements. Participants primarily fixated the former locations of the target in 17% (N =366) of these trials, former locations of distractors in 18% (N=386), the center of the screen in 41% (N=874), and elsewhere in 22% (N=467).

Error rate data are illustrated in Figure 9, and do not suggest floor or ceiling effects in the recognition task. As described in the Method section, we used logistic mixed effects regression to model error rates as a function of location consistency, context consistency, target repetition, and eye fixation patterns (Table 2). Participants benefitted from stimulus repetition, erring less in later blocks (OR:

0.36:1, $\beta_{log(Block)} = -1.02$, *SE* = .22 *p* < .001), and this benefit was stronger for targets that repeatedly appeared with the same distractors than those appearing with different distractors (OR: 1.93:1, $\beta_{log(Block)\ x\ Context} = 0.66$, *SE* = .22, *p* = .003). Finally, as illustrated in Figure 9, participants particularly benefitted from the combination of a consistent context with a consistent location (OR: 0.42:1, $\beta_{Location\ x\ Context} = -0.87$, *SE* = .44, p = .046).[4]

On average, participants correctly articulated 19.3 out of 36 pseudowords in the interspersed cued-recall trials (*SD* = 6.95). Since these trials were only included to ensure participant engagement with the task, they were not further analysed.

---

[4] In a post-hoc analysis, we examined the effect of varying stimuli positions between encoding and testing phases. Although participants erred significantly more when stimuli positions were mismatched across the two phases (OR: 2.23:1, $\beta_{EncodingVersusTestingPositions} = -1.02$, SE = .24 p < .001), the overall pattern of results indicated in the main analysis stayed largely the same ($\beta_{log(Block)} = -1.07$; $\beta_{Context} = 0.58$; $\beta_{log(Block)\ x\ Context} = 0.73$; $\beta_{Location\ x\ Context} = -0.90$; all *p*s < .05).

Figure 9

Error rate by condition in the Phase 1 recognition task
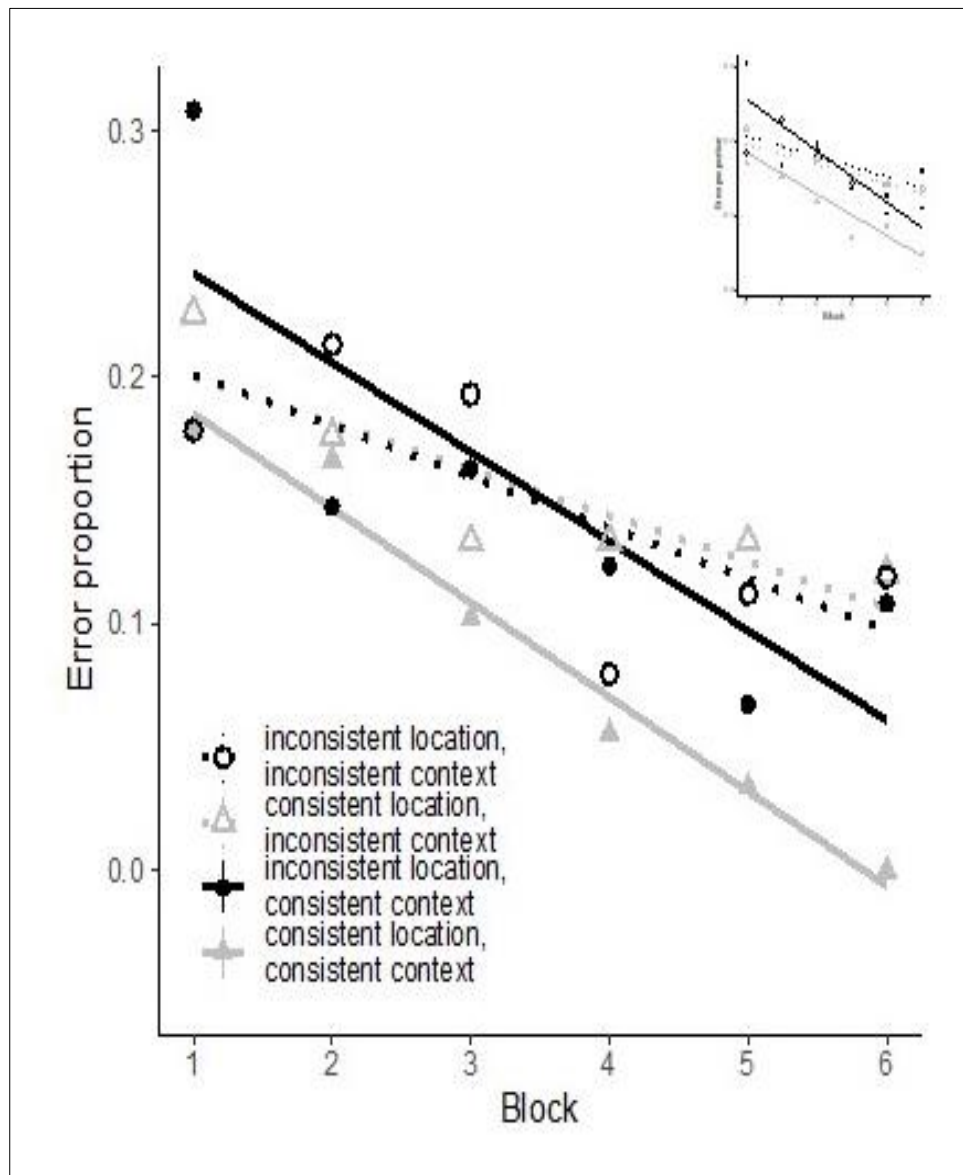
Note. The main figure depicts the pattern in the restricted dataset; the inset shows

the same pattern when including trials without valid eyetracking data.

Table 2

Summary of a logistic mixed effects regression analysis of recognition error frequency

	Coef (ß)	SE (ß)	*p*	OR (exp(ß))
(Intercept)	-2.49	0.37	**<.001**	0.08
log(Block)	-1.02	0.22	**<.001**	0.36
LocationC	0.23	0.33	0.489	1.26
ContextC	0.50	0.20	**0.011**	1.65
PrimaryFixation	0.39	0.29	0.175	1.48
FixatedAnyROI	-0.21	0.16	0.195	0.81
Block x LocationC	0.19	0.29	0.509	1.21
Block x ContextC	0.66	0.22	**0.003**	1.93
LocationC x ContextC	-0.87	0.44	**0.046**	0.42
Block x LocationC x ContextC	-0.32	0.45	0.477	0.73

Phase 2 (Cued-Recall Task)

Due to slow Internet connections, two participants' audio recordings from the cued-recall task failed to properly upload to Gorilla Experiment Builder's server, leaving a total of 12 participants for these analyses. On average, participants correctly articulated 20 out of 36 pseudowords in the cued-recall task (*SD* = 10.91). Participants' mean error proportions per trial type (*i.e.*, whether location and/or context were consistent) can be found on Table 3.

Table 3

Summary of subject-weighted mean error proportions in the Phase 2 cued-recall task

| | | Context | |
		Consistent	Inconsistent
Location	Consistent	.454 (.274)	.491 (.340)
	Inconsistent	.500 (.320)	.493 (.216)

We used logistic mixed effects regression to model error rates as a function of location consistency and context consistency (Table 4). As in the recognition task, these factors significantly interacted to affect cued recall performance (OR: 0.30:1, $\beta_{\text{Location x Context}}$ = -1.21, *SE* = .61, *p* = .049): as in the Phase 1 recognition task, target location consistency only appeared to affect error rates when the target had been consistently presented with the same pair of distractors.[5]

---

[5] Observed power for the significant results: Recognition task: $1-\beta_{\text{log(Block)}}$ = .99; $1-\beta_{\text{Context}}$ = .83; $1-\beta_{\text{log(Block) x Context}}$ = .84; $1-\beta_{\text{Location x Context}}$ = .58. Separate recall task: $1-\beta_{\text{Location x Context}}$ = .62. Due to the noisier nature of webcam-based eyetracking, we did not have a good basis for a pre-hoc power calculation for the current study. We intend to use the current findings to estimate sample and effect sizes that are suitable for the context of webcam-based eyetracking in future paired-associate learning/ looking-at-nothing experiments.

Table 4

Summary of a logistic mixed effects regression analysis of cued-recall error

frequency

	Coef (ß)	SE (ß)	*p*	OR (exp(ß))
(Intercept)	-0.32	0.46	0.481	0.73
LocationC	-0.12	0.31	0.690	0.88
ContextC	-0.17	0.29	0.562	0.84
LocationC x ContextC	-1.21	0.61	**0.049**	0.30

Discussion

In this study, we examined the conditions under which typical readers optimally learn to associate visual-phonological information, simulating the process of acquiring orthographic-phonological representations. Specifically, we investigated the extent to which ostensibly task-irrelevant episodic details modulate visual-phonological binding performance in typical readers. To this end, we tested whether encoding new visual-phonological associations over multiple exposures was modulated by whether targets consistently appeared in the same screen locations or with the same pair of non-target distractors. To assess whether visual attention, in the form of 'looking-at-nothing' behaviour, modulated these episodic effects, we also used participants' webcams to remotely track their eye movements.

Recognition accuracy for novel orthographic-phonological bindings improved with repetition (see Figure 9), in line with previous evidence in the paired-associate

learning literature (*e.g.*, Jones et al., 2018), and suggesting an incremental development of stable visual-phonological associations with repetition.

Recognition, as well as later recall, was also modulated by the consistency of extraneous cues that were present during encoding. Participants more accurately recognised visual symbols from associated nonword cues for targets that were consistently presented in *both* the same screen location and with the same distractor symbol/nonword pairs. This finding suggests that, during the process of building an episodic representation of a novel visual-phonological binding, typical readers not only incorporate all the features available at the time of encoding, a typical occurrence in episodic memory formation (Tulving, 1972), but they also appeared to use the consistent features as an aid to help them retrieve these representations from memory. This pattern also emerged in the subsequent cued-recall task, which demonstrated superior accuracy for the bindings that participants had encoded in the consistent location and consistent context condition, suggesting that multiple co-occurring statistical frequencies in the input enable typical readers to quickly acquire accurate visual-phonological bindings, even after relatively few exposures.

In our experiment, participants were prompted to encode three bindings in each trial. In the consistent context condition, all three bindings repeatedly co-occurred over the course of the experiment. We might speculate that participants encoded all three bindings and stored them together, such that when the locations of these items were *in*consistent across trials, separating one item representation from the others for recall became problematic.

It is worth noting that our superadditive interaction of location consistency and context consistency for novel orthographic/phonological bindings resembles on its surface, at least, a very well-known superadditive effect in which relative location

consistency interacts with context consistency to support perception and recall of overlearned orthographic-phonological bindings: 'the word superiority effect' (J. Baron & Thurston, 1973). This resemblance is intriguing because models of that effect often attribute it to robust connections between well-established representations (*e.g.* Rumelhart & McClelland, 1982). If a shared mechanism underpins both effects, our results would further demonstrate continuity between the earliest stages of binding acquisition and the distant goalpost of seemingly automatic skilled reading.

Although this study was partly motivated by previous reports that 'looking-at-nothing' modulates paired associate learning, we did not detect any such significant effects in this dataset. Contributing factors may simply be power and webcam-based eyetracking data quality: though the regression analysis identified trends in the expected directions, webcam-based eyetracking is still in its infancy, and thus, due to the inevitable increase in noise engendered by remote webcam-based eyetracking, the method used in our study may potentially not have detected fixations as consistently as specialised laboratory hardware.

Chapter 4

Audiovisual Learning in Dyslexic and Typical adults: Modulating Influences of Location and Context Consistency

Abstract

Learning to read involves efficient binding of visual to auditory information. Aberrant cross-modal binding skill has been observed in both children and adults with developmental dyslexia. Here, we examine the contribution of episodic memory to acquisition of novel cross-modal bindings in typical and dyslexic adult readers. Participants gradually learned arbitrary associations between unfamiliar Mandarin Chinese characters and English-like pseudowords over multiple exposures, simulating the early stages of letter-to-letter sound mapping. The novel cross-modal bindings were presented in consistent or varied locations (*i.e.*, screen positions), and within consistent or varied contexts (*i.e.*, co-occurring distractor items). Our goal was to examine the contribution, if any, of these episodic memory cues (*i.e.*, the contextual and spatial properties of the stimuli) to binding acquisition, and investigate the extent to which readers with and without dyslexia would differ in their reliance on episodic memory during the learning process. Participants were tested on their ability to recognise and recall the bindings both during training and then in post-training tasks. We tracked participants' eye movements remotely with their personal webcams to assess whether they would re-fixate relevant empty screen locations upon hearing an auditory cue - indicative of episodic memory retrieval - and the extent to which the so-called 'looking-at-nothing behaviour' would modulate recognition of the novel bindings. Readers with dyslexia both recognised and recalled significantly fewer bindings than typical readers, providing further evidence of their persistent difficulties with cross-modal binding. Looking-at-nothing behaviour was generally associated with higher recognition error rates for both groups, a pattern that was particularly more evident in later blocks for bindings encoded in the inconsistent location condition. Our findings also show that whilst readers with and

without dyslexia are capable of using stimulus consistencies in the input – both location and context – to assist in audiovisual learning, readers with dyslexia appear particularly reliant on consistent contextual information. Taken together, our results suggest that whilst readers with dyslexia fail to efficiently learn audiovisual binding as a function of stimulus frequency, they are able to use stimulus consistency – aided by episodic recall – to assist in the learning process.

Keywords: Episodic memory; looking-at-nothing; paired associate learning; cross-modal binding; reading.

Audiovisual Learning in Dyslexic and Typical Adults: Modulating Influences of Location and Context Consistency

Quickly binding visual forms to phonological forms is a fundamental skill in the initial stages of grapheme-phoneme learning, providing a foundation for the later development of integrated visual-phonological representations that are crucial for skilled reading. Most children are able to convert written letters and words into sounds effortlessly, and later retrieve them as a single audiovisual unit, eventually becoming proficient readers. However, some struggle to form novel audiovisual mappings, a difficulty that can persist well into adulthood (Blau et al., 2009; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013). Readers with developmental dyslexia exhibit indications of less-integrated grapheme-phoneme representations (Aravena et al., 2013, 2018; Blau et al., 2009, 2010; Blomert, 2011; Warmington & Hulme, 2012; Žarić et al., 2015), a deficit owing in part to their comparatively poorer cross-modal binding skills (Albano et al., 2016; Aravena et al., 2013; Garcia et al., 2019; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Toffalini et al., 2018, 2019; Žarić et al., 2015). Despite the well-known link between audiovisual integration and ultimate reading attainment, the cognitive mechanisms underlying typical and atypical cross-modal binding ability are not yet fully understood. Here, we examine how adults with dyslexia and typical readers may differ in their reliance on episodic memory cues as they acquire novel cross-modal bindings that vary in location-related and contextual consistency over the course of the learning process.

Learning to read requires establishing new representations in memory: not only separate representations for novel visual/orthographic and phonological forms, but also correspondences between them. A commonly used task to tap the acquisition of novel visual-phonological mappings is cross-modal *paired associate*

learning (PAL; *e.g.*, Calabrich et al., 2021; Jones et al., 2018; Wang et al., 2017;

Warmington & Hulme, 2012), in which participants must learn that a given visual

symbol is associated with a particular phonological sequence (typically a

pseudoword). This learning process is thought to emulate the associative

mechanisms underpinning grapho-phonological mappings in the early stages of

literacy development (Hulme et al., 2007; Warmington & Hulme, 2012). An extensive

body of research demonstrates that readers with dyslexia are generally more error

prone on such cross-modal PAL tasks, relative to typical readers (M. Jones et al.,

2018; Litt & Nation, 2014; Messbauer & de Jong, 2003; Toffalini et al., 2018; H. C.

Wang et al., 2017; Warmington & Hulme, 2012), and, crucially, performance on PAL

tasks correlates with individual differences in reading skill (Hulme et al., 2007;

Warmington & Hulme, 2012). In particular, visual-verbal PAL ability is a unique

predictor of both word recognition and nonword reading (Warmington & Hulme,

2012).

Whilst PAL tasks are useful in showing the relationship between visual-verbal

learning and reading ability, such paradigms do not typically elucidate the learning

mechanisms that distinguish good and poorer performance in PAL and reading.

However, in other learning contexts, the ability to track simple statistics, such as

stimulus repetition and sequences is a strong predictor of reading ability (Ahissar,

2007), and poorer readers are liable to forget previous exposures to perceptual

stimuli (Jaffe-Dax et al., 2015, 2016, 2017), potentially leading to 'noisier' processing

of a current stimulus. We can therefore reasonably extrapolate that statistical

tracking, implicating episodic memory and associated decay, may play an important

role in determining the effectiveness with which audiovisual associations can be

created and established over repeated exposures. Indeed, learning audiovisual

stimuli requires accurate encoding of temporal and spatial characteristics in order to appropriately bind visual and phonological features and to create a composite representation. Temporal and spatial properties, commonly encoded in episodic memory, share patterns of neural activity, and can be used as cues to aid memory retrieval when required (El-Kalliny et al., 2019; Tulving, 1972). In the context of language, episodic memory of the context in which a word is encountered plays an important role in acquisition (Stark & Stark, 2016). Through repetition and rehearsal, representations become gradually less episodic and more abstract, representative of an amalgam of consistent stimulus properties, with the result that specific episodic details, such as spatial and temporal properties, become less and less relevant (Squire & Zola, 1998; Stark & Stark, 2016). In literacy acquisition, this process also entails a gradual increase in automatisation of print reading, such that phonology is eventually accessed automatically and without recourse to an effortful retrieval process, implicating episodic memory resources (Ehri, 2005; Ehri & Saltmarsh, 1995; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; LaBerge & Samuels, 1974).

Recently, Jones and colleagues (2018) examined the role of statistical learning mechanisms and episodic memory in the context of a paired associate learning (PAL) task, in which groups of readers with dyslexia and typically reading adults learned to associate a sequence of unfamiliar characters (*i.e.*, Mandarin Chinese characters) with consistently paired pseudowords. As participants attempted to retrieve each target's corresponding pseudoword, their eye movements were tracked on the now-blank screen to examine whether they consulted the spatial location of the target item in order to support retrieval. Such 'looking-at-nothing' behaviour would imply re-activation of integrated memory representations: re-activating one of the target features, such as its phonological representation,

may automatically drive the memory system to re-activate other features as well, including its visuospatial location, mechanistically or epiphenomenally producing eye movements toward that location (Altmann, 2004; Ferreira et al., 2008; Johansson & Johansson, 2020; Kumcu & Thompson, 2020; Scholz et al., 2011) when rebinding the multiple features again. Although such looking-at-nothing movements can suggest both successful memory encoding and reactivation in the earliest stages of learning, they also become less frequent as learners consolidate a memory representation, presumably abstracting away such details (Scholz et al., 2011; Spivey, 2007). For readers with dyslexia in Jones and colleagues' (2018) study, fixating a target's former location within the current trial was associated with greater recall accuracy (which nevertheless stayed well below par, compared with the typical reader group), and fixating a distractor's former location was associated with lesser recall accuracy, both compared to a no-fixation baseline. For typical readers, in contrast, fixating a target's former location within the current trial was only associated with greater recall accuracy when it had also appeared in the same location in a previous trial. Thus, whilst readers with dyslexia showed sensitivity to location information for only the current trial, typical readers showed a sensitivity to longer-range statistical regularities over multiple exposures. This pattern suggests that typical readers use spatial location as a cue to retrieve such bindings, even when location information is task irrelevant, and this ability may specifically be impaired in readers with dyslexia (Albano et al., 2016; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Toffalini et al., 2018). Audiovisual learning is therefore modulated by the statistical sensitivity and associated episodic memory usage that individuals of different reading abilities bring to bear on the task. However, we are yet to discover how specific stimulus configurations during learning affect learning efficiency in

dyslexic and typical readers. In general, presenting multiple items in a temporally adjacent format increases the association between these items (El-Kalliny et al., 2019). However, isolating and retrieving *individual* memories encoded in temporal proximity can only succeed if the distinct memories were separated in neural space during encoding (El-Kalliny et al., 2019; Sheehan et al., 2018). In other words, our ability to discriminate between different past experiences that share similar features largely depends on the brain's capacity to store *distinct* activity patterns to represent *each* of these experiences (Madar et al., 2019). Readers with dyslexia have been shown to benefit from having novel cross-modal bindings presented in a fixed temporal order (Toffalini et al., 2018), but, to the best of our knowledge, there is no comprehensive study of how stimulus configurations during learning affect typical and dyslexic readers' capacity to learn reading-related items. This is an important next step, since dyslexic readers' reduced ability to create stable representations over multiple exposures is plausibly related to their inability to identify an item as distinct from other items presented in temporal and spatial proximity.

In the present study, we examine whether specific statistical properties of stimulus exposures differentially affect learning in adults with and without developmental dyslexia. To this end, we designed a paired associate learning task (adapted from Jones et al., 2018), in which we manipulated the consistency of the spatial and contextual stimulus properties during encoding. We created arbitrary associations between monosyllabic pseudowords – following English phonotactics (*e.g.*, /gɔp/) – and Mandarin Chinese characters (*e.g.*, 日). Our participants were unfamiliar with both the visual and auditory stimuli, thus ensuring an arbitrary relationship between these visual-verbal bindings, and simulating the early stages of orthography-to-phonology mapping.

In terms of accuracy, we predicted that, compared with typical readers, readers with dyslexia would show generally higher error rates, and a shallower function of learning (Albano et al., 2016; Aravena et al., 2013; Garcia et al., 2019; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Messbauer & de Jong, 2003; Toffalini et al., 2018, 2019). Further, whilst we predicted that consistently presenting targets in the same spatial location and/or in the context of the same alternatives would generally decrease error rates, we suspected that these consistency effects would disproportionately benefit readers with dyslexia: though previous work indicates that readers with dyslexia are less likely to track single-feature statistics (*e.g.*, location) over multiple exposures (M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Toffalini et al., 2018), providing *both* spatial (*i.e.*, item screen location) *and* contextual consistencies (*i.e.*, item co-occurrences) might prove particularly advantageous to help impaired readers bootstrap degraded representations/poorer retrieval of individual items. Indeed, readers with dyslexia are known to engage in chunking strategies such as whole word memorisation in order to avoid phonological sequencing, which is problematic in dyslexia (Ullman & Pullman, 2015).

To consider the possible role of implicit memory retrieval, we estimated participants' reference to episodic detail via a looking-at-nothing paradigm. During the main training and recognition task, we made novel use of webcam-based technology (WebGazer.js: Papoutsaki et al., 2016) to remotely track participants' eye movements as they viewed a blank screen immediately after hearing an auditory cue. Even though the use of webcam-based eye tracking in behavioural science is still in its infancy, previous investigations have demonstrated the method's suitability to detect fixations reliably and to replicate in-lab findings with minimal reduction in data quality (Bott et al., 2017; Semmelmann & Weigelt, 2018). With this approach,

we sought to ascertain whether looks to relevant blank screen locations would modulate recognition accuracy. Following previous work (M. Jones et al., 2018), we predicted that readers with dyslexia would have a stronger tendency to make errors following fixations to blank screen locations previously occupied by distractor items. We also expected repetition to diminish the link between accuracy and looking-at-nothing behaviours for all participants, reflecting direct access to increasingly abstracted memory representations (Ferreira et al., 2008; Richardson & Spivey, 2000; Scholz et al., 2011; Wantz et al., 2016). Finally, our factorial manipulation allows us to consider higher-order interactions, but it is challenging to derive and evaluate specific predictions for such interactions, and robustly assessing such interactions would require more power than our study provides (Button et al., 2013); as a compromise, we note such interactions but consider them primarily as invitations for future research.

In addition to the main training and recognition task, we collected three additional measures of item learning. We added (1) cued-recall trials at regular intervals in the main training task to test participants' ability to recall and verbalise the specific pseudoword associated with a given character. Moreover, to probe participants' longer-term memory, we tested participants' ability to (2) recall, and (3) recognise the bindings in two separate tasks administered approximately ten minutes after the main task. This approach allowed us to assess whether the episodic memory effects of spatial and contextual cues carried over and differentially modulated longer term retention of the bindings for the two reading groups. Due to the gradual consolidation process engendered by repeated exposures, we predict that performance in the subsequent post-training cued-recall and recognition tests would be less strongly modulated by episodic memory cues. We also predicted

overall higher error rates in recall than in recognition, given that recognition is wont to succeed even when recall fails (Tulving, 1982).

Methods

Participants

Thirty-five readers with dyslexia (age: $M = 28.17$, $SD = 7$; 23 females) and thirty-five typical readers (age: $M = 23.55$, $SD = 6.14$; 19 females) were tested remotely. All participants were native speakers of British English, recruited through Bangor University and Prolific[6]. A similar level of education was reported in both groups (dyslexia: $M = 15.8$ years, $SD = 2.37$; typical: $M = 14.8$ years, $SD = 2.11$; $p = .09$), and none of the participants reported any history of psychiatric and/or neurological diseases, visual and/or hearing impairments, or any other risk factors. Group membership (*i.e.*, typical reader or individual with dyslexia) was confirmed via a battery of literacy tests. All participants provided informed consent, were naïve to the purpose of the experiment, and were unfamiliar with the experimental stimuli. Participants received SONA credits or payment for their time. The experiment was approved by Bangor University's Ethics Committee.

Materials

Literacy and general cognitive ability measures

Participants' group membership was validated via a battery of eight short tests: (1) *Adult Reading Questionnaire* (ARQ, Snowling et al., 2012); (2) word reading efficiency and (3) phonemic decoding efficiency subscales of the *Test of Word Reading Efficiency* (TOWRE, Torgesen et al., 1999); (4) letter and (5) digit versions of the *Rapid Automatized Naming* (RAN) subtest from the *Comprehensive test of*

---

[6] *www.prolific.co*

Phonological Processing (CTOPP, Wagner et al., 1999); (6) *Similarities* subtest from the *Weschler Adult Intelligence Scale* (WAIS, Wechsler, 1981) as an index of verbal intelligence quotient (IQ); (7) *Matrix Reasoning* from the *Wechsler Abbreviated Scale of Intelligence* (WASI, Wechsler, 1999) as an index of nonverbal IQ; and (8) computerised forward and backward digit span tests in which participants first saw sequences of digits and were then prompted to type the digits in the same or reverse order. Tests 1 through 5 were administered shortly before the main training and recognition task, whereas the remaining were administered immediately after the main task.

Stimuli

Thirty-six consonant-vowel-consonant (CVC) pseudowords (*e.g.*, /gɔp/) were arbitrarily matched to thirty-six Mandarin Chinese characters (*e.g.*, 日), as in Jones and colleagues (2018). The pseudowords followed English phonotactic rules and were generated with Wuggy (Keuleers & Brysbaert, 2010), a multilingual pseudoword generator. The auditory stimuli were recorded by a female native speaker of British English and digitised at 44.1 kHz on Praat (P Boersma & Weenink, 2021). Each Mandarin Chinese character was consistently presented with the same CVC pseudoword over the course of the experiment.

Procedure

The experiment was programmed and deployed online on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). It included three tasks, presented in the same order to all participants: (1) training, *via* a six-block recognition task with interspersed cued-recall trials; (2) a single-block cued-recall test; and (3) a single-block recognition test.

Access to the experiment was restricted to desktop and laptop users only;

mobile phones and tablets were disallowed. Participants were instructed to wear

earphones or headphones, to place their computers on a desk, and to do the tasks

individually in a quiet and well-lit room. To minimise distraction and correct for

varying screen sizes and resolutions, participants were prompted to activate the full

screen mode on their computers before proceeding to the experimental tasks. On

average, participants sat 546.03 mm ($SD = 101.02$) from their computer screens as

estimated by the Virtual Chinrest task (Q. Li et al., 2020). The entire testing session

lasted approximately 130 minutes, including background tests, experimental tasks,

and calibrations. A time limit of 180 minutes automatically rejected any participants

exceeding this threshold.

Eye-tracking measures were assessed via WebGazer.js (Papoutsaki et al.,

2016) with an ideal sampling rate of approximately 60 Hz, dependent on each

participant's monitor's refresh rate (Anwyl-Irvine et al., 2020). Before each task,

participants completed a 5-point calibration procedure. A series of pictorial

instructions demonstrated appropriate head position during calibration and

experimental tasks. Failure to calibrate at least one of the points (*i.e.*, if the estimate

for a point was too close to another) resulted in an automatic repetition of the

calibration procedure. To account for participants' potential head drift and body

repositioning, re-calibration was performed in the middle of each experimental block

(*i.e.*, after 18 trials), and before the onset of each new block in training. Eye-tracking

estimates with face confidence values (*i.e.*, a score ranging from 0 to 1 estimating

the webcam-based eye-tracking machine learning model's confidence level in

detecting a human face) lower than .5 were excluded from the analyses. In the two

post-training tests, eye-tracking measures were recorded for exploratory purposes only and are not reported here.

Training: Recognition (with interspersed cued-recall trials)

Training emulates Calabrich and colleagues' (2021) main paradigm, originally based on Jones and colleagues' (2018) cued-recall paradigm. Each training trial consisted of an encoding phase and a testing phase. Each trial began with a 1000-ms fixation cross, followed by three Mandarin Chinese characters presented in black on a white background. The three characters were displayed in triangle formation (see Figure 10a), each occupying 20x20 units of Gorilla Experiment Builder's (Anwyl-Irvine et al., 2020) screen space. Each character's colour changed from black to red synchronously with auditory presentation of its corresponding pseudoword. The order in which character/pseudowords were highlighted/presented was fully counterbalanced across trials. At the end of this encoding phase, a 1000-ms blank screen was followed by a visual backward masking phase: hash symbols and numbers, presented in pseudorandomised order, momentarily replaced the characters to minimise visible persistence (see Figure 10b). The onset of the testing phase was signaled by the appearance of a small black dot presented in the centre of the screen. A click on the black dot would play the auditory cue that corresponded to the target (*i.e.*, one of the three pseudowords from the encoding phase). If no clicks were detected within 10 seconds, the trial would terminate. The requirement to click the black dot had the secondary purpose of introducing an inconspicuous attention check: if, in three consecutive trials, no clicks had been detected, the participant would be automatically excluded from the experiment as this would constitute a strong indication that their computer had been left unattended mid-task. A 1000-ms blank screen followed the black dot, during which participants' eye

movements were recorded. The three Mandarin Chinese characters then reappeared, and a mouse-click was expected: participants were instructed to select the character that corresponded to the auditory cue. In order to minimise auditory localisation bias and encourage our participants to attend to both visual *and* auditory features of the stimuli, the characters' screen position changed in two thirds of the trials once they reappeared in the testing phase. The characters remained on the screen for 5000-ms, or until a mouse-click was detected, whichever occurred first. A 250-ms blank screen was presented, at which point the trial ended. A total of 216 trials were evenly distributed over 6 blocks. Block and trial presentation were randomised across participants to avoid order effects.

As in Calabrich and colleagues (2021), we orthogonally manipulated two binomial factors in the encoding phase: 1. *Location consistency*: whether a visual-phonological association was consistently presented in the same spatial location throughout the experiment, and 2. *Context Consistency*: whether a visual-phonological association consistently co-occurred with the same items throughout the experiment. As a result, half of the 36 Mandarin Chinese characters were always presented in the same screen position across different trials (*i.e.*, six items would only appear in the top middle, six in the bottom left, and six in the bottom right), whilst the other half could appear in any of three possible screen locations with equal probability. Similarly, half of the stimuli would always appear within a specific triplet (*i.e.*, a target item along with the same two distractors, *e.g.*, items A, B, and C would always be presented together in each occurrence -- taking turns as targets and distractors across different trials -- and would never co-appear with any other items over the course of the experiment), whilst the remaining would not have any fixed co-occurrences. For each participant, each binding was therefore assigned to one of

four trial types: 1. *Inconsistent Location/Inconsistent Context*, 2. *Inconsistent Location/Consistent Context*, 3. *Consistent Location/Inconsistent Context*, and 4. *Consistent Location/Consistent Context*. Each binding appeared three times in each block: once as a target, and twice as a distractor. Each 36-trial block thus contained nine pseudorandomly ordered trials of each type.

We added cued-recall trials at regular intervals (*i.e.*, every six recognition trials) within each block. In each cued-recall trial, one of the 36 Mandarin Chinese characters appeared in the centre of the screen (see Figure 10c). Upon seeing this visual cue, participants were required to articulate the corresponding pseudoword. The target item for each of the interspersed cued-recall trials ($N = 36$) was randomly selected from one of the six preceding recognition trials. The purpose for the interspersed cued-recall trials was twofold: 1. to ensure participants were actively attempting to store the items in their memory beyond the temporal boundaries of each recognition trial, and 2. to afford participants the opportunity to practice saying the pseudowords aloud, since they would later be tested on their ability to recall the cross-modal bindings in the post-training cued-recall test. Participants were prompted to recall each binding once over the course of the task.

Figure 10

Training and post-training encoding and testing phases
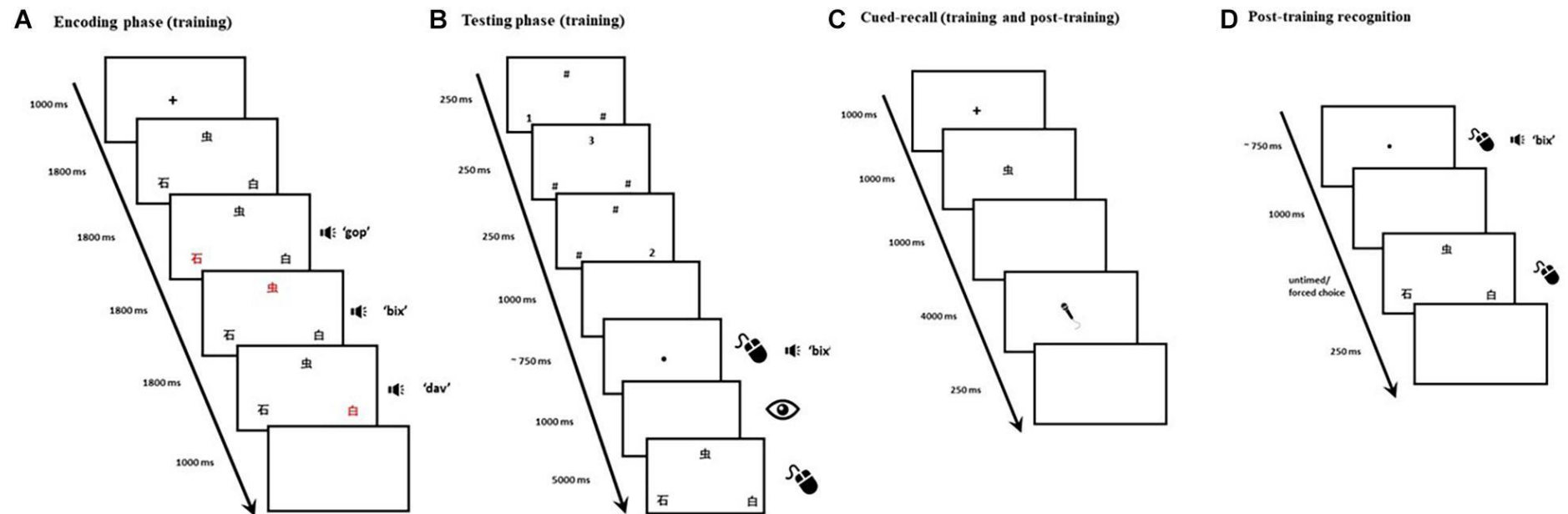
Note. Panels (A) and (B) depict the timeline of a single trial in the main training and recognition task. The encoding phase (A) was immediately followed by backward masking and then by the testing phase (B). Panel (C) depicts the timeline of a single cued-recall trial (both training and post-training). Panel (D) depicts the timeline of a single trial in the post-training recognition test. The eye depicts recording of onscreen fixations, the microphone depicts when a verbal response was expected, and the mouse illustrates when a click was expected.

To familiarise the participants with the experimental procedure, training was preceded by a practice block with four recognition trials and one cued-recall trial, using additional filler stimuli. Participants were provided with feedback after each practice trial, and were given the option of repeating the practice block if needed. Participants were encouraged to take short breaks between blocks, and were instructed to resume to the same position upon their return. Re-calibration ensured that accurate eye-movements were detected following these breaks.

Post-training cued-recall test

As in Calabrich and colleagues' (2021), a cued-recall test followed training. The post-training cued-recall test consisted of a single block with 36 randomly ordered trials (See Figure 10c), testing each of the previously trained bindings. A 1000-ms fixation cross started each trial, which was then followed by a Mandarin Chinese character presented centrally for 1000-ms in black on a white background. As in training, each character occupied 20x20 units of Gorilla Experiment Builder's screen space. A 1000-ms blank screen followed, and then a drawing of a grayscale microphone, presented in the centre of the screen, indicated that the voice recording had started and a verbal response was required. Participants were instructed they would have three seconds to provide a response. However, to ensure that the onsets of participants' responses were not inadvertently trimmed due to potential delays in the activation of the audio recording, voice recording effectively started 1000-ms before the microphone was shown. A 250-ms blank screen then appeared, ending the trial.

Post-training recognition test

A single-block recognition test, comprising the same visual-auditory stimuli from the previous tasks, was administered immediately after the post-training cued-recall test.

It consisted of 36 randomly ordered three-alternative forced-choice trials. These were similar to the recognition trials in the training task but lacked the encoding phase. Each trial began with a black dot presented on a white background in the centre of the screen (See Figure 10d). Upon clicking on the dot, participants would hear one of the 36 target pseudowords. A 1000-ms blank screen would follow, and three equidistant Mandarin Chinese characters would be presented in the same triangle formation as training. Participants were instructed to select the character which corresponded to the auditory cue they had just heard. A 250-ms blank screen was presented, at which point the trial ended.

Data analysis

To enable comparisons of eye movements across different screen sizes, we used normalised coordinates in our eye-tracking analyses wherein -0.5 and 0.5 always refer to the centre of the screen regardless of their size (Gorilla Experiment Builder; Anwyl-Irvine et al., 2020). We performed fixation detection on the normalised data for each individual participant via the 'detect.fixations' function in the 'saccades' v0.2.1 library (von der Malsburg, 2019) in R v4.0.0 (R. C. Team, 2020). Due to the noisier and low-frequency nature of webcam-based eye-tracking data, we set the 'smooth.coordinates' parameter to 'TRUE' to suppress noise, and set the 'smooth.saccades' to 'FALSE' to detect short saccades more reliably (von der Malsburg, 2019).

We used confirmatory logistic mixed effects regression, *via* the glmer::binomial function in the lme4 v1.1-23 library (Bates et al., 2015) in all analyses. All models included maximal random effects structures (Barr et al., 2013) reverting to a 'parsimonious' approach in the case of convergence errors (Bates et al., 2015). In all models, *subject* and *item* were included as random effects. For the

recognition trials from the training task—our richest source of data—we modelled

error rate as a function of six fixed effects and their interactions: 1. Group

membership (*Group*, *i.e.*, typical reader = -0.5, individual with dyslexia = 0.5); 2.

Context consistency (*Context*, *i.e.*, whether a target consistently co-occurred with the

same distractors over the course of the task; consistent = -0.5, inconsistent = 0.5); 3.

Location consistency (*Location*, *i.e.*, whether a target consistently appeared in the

same screen location over the course of the task; consistent = -0.5, inconsistent =

0.5); 4. Repetition effects (*log(Block)*, *i.e.*, Blocks 1 through 6; log-transformed); 5.

The presence of looking-at-nothing behaviour (*FixatedAnyROI*, *i.e.*, whether

participants re-fixated any of the regions of interest upon hearing the auditory cue;

no = -0.5, yes = 0.5); and 6. Primary fixation (*PrimaryFixation*, *i.e.*, the dominant

region of interest fixated upon hearing the auditory cue; target = -0.5, distractor =

0.5, none = 0.0), conceptually nested within *FixatedAnyROI*. All predictors were

contrast-coded and centred. In our pre-registration of this study, we conducted a

power analysis (see Appendix A) using the simR library (P. Green & Macleod, 2016)

to estimate a sample size with sufficient power for the interaction of primary

theoretical interest [Group x Context x Location]. Thus, when reporting the findings

below, we signpost significant higher order interactions that should be interpreted

with caution.

In the cued-recall trials embedded in the training task, and in the subsequent

post-training tests of cued-recall and recognition, we modelled error rate as a

function of the following three factors and their interactions, as described above: 1.

Group membership, 2. Context consistency, and 3. Location consistency. Cued-

recall errors were defined as any mis-articulations that deviated from the correct

pseudoword in at least one phoneme. Recognition errors were defined as any trial in which a participant clicked on a non-target character.

Results

Literacy and general cognitive ability measures

Background measures for both groups are summarised in Table 5. Participants with self-reported dyslexia diagnoses scored significantly higher on the Adult Reading Questionnaire (Snowling et al., 2012) than those without such diagnoses. As a group, readers with dyslexia correctly read significantly fewer words and pseudowords than did the typical readers. Similarly, typical readers were significantly faster at naming digits and letters than readers with dyslexia. There were no significant group differences on verbal and non-verbal IQ measures, nor on forward and backward digit span measures.

Table 5

Group scores on literacy and general cognitive ability measures

Test	Measure	Dyslexic N = 35			Typical N = 35			t	p	Cohen's d
		Range	M	SD	Range	M	SD			
TOWRE	Word reading rate[a]	35-102	74.60	19.26	70-104	90.63	9.25	4.42	<.001	-1.05
	Pseudoword reading rate[a]	17-60	41.11	11.24	34-63	53.97	7.27	5.68	<.001	-1.35
CTOPP	RAN digits[b]	11-28	16.46	4.1	9-21	13.31	2.61	3.82	<.001	0.91
	RAN letters[b]	12-30	17.23	4.09	10-18	13.51	2.34	34.11	<.001	1.11
WAIS	Verbal IQ[c]	11-32	22.66	4.14	15-30	23.31	3.74	0.69	.488	-0.16
WASI	Non-verbal IQ[c]	9-28	18.50	6.7	11-27	20.69	3.92	1.64	.105	-0.39
ARQ	Risk of reading impairment[d]	12-31	23.09	5.17	3-29	13.30	5.57	7.57	<.001	-1.82
	Forward digit span[e]	2-8	5.27	1.7	2-9	6.03	1.76	1.80	.076	-0.43
	Backward digit span[e]	2-8	4.26	1.7	2-8	5.06	1.76	1.92	.059	-0.46

[a] Number of correctly read items within 45 seconds.

[b] Raw scores in seconds.

[c] Raw scores.

[d] Higher scores represent greater likelihood of reading disability.

Training

Recognition task

A total of 491 (3.24%) recognition trials timed out (*i.e.*, no mouse click was detected) and were thus excluded, leaving the 14,629 trials for the behavioural analyses summarised in Table 6. Distributed across these behaviourally valid trials, the webcam-based eye tracking technique provided a total of 900,837 eye-tracking estimates in our screen of interest. We excluded approximately 3% of these estimates (*N* = 28,080) due to suboptimal face detection values (*i.e.*, face_conf < .5). The noise suppression and short saccade detection filtering excluded about 16% of the data, leaving a total of 12,145 trials (6,130 dyslexic, 6,015 typical) containing both the behavioural and eye tracking measures required for our planned analyses. In these trials, readers with and without dyslexia fixated regions of interest (ROIs) for targets and distractors in similar proportions (χ^2 (1) = 0.02, *p* = .88).

Error patterns common to both groups: As illustrated in Figure 11, both typical readers and readers with dyslexia benefitted from stimulus repetition, making fewer errors in each successive block (odds ratio: 0.32:1, $\beta_{log(Block)}$ = -1.13, *SE* = 0.08, *p* < .001). Participants made fewer recognition errors in context-consistent conditions, when a target consistently appeared with the same distractors (odds ratio: 1.35:1, $\beta_{Context}$ = 0.30, *SE* = 0.13, *p* = .018). As illustrated in Figure 12a, participants also showed some tendency to make fewer errors in location-consistent conditions, when a target consistently appeared in the same screen location (odds ratio: 1.20:1, $\beta_{Location}$ = 0.19, *SE* = 0.13, *p* = .153), but this effect was diminished for trials in which they fixated the former location of either a target or distractor (odds ratio: 0.39:1, $\beta_{Location\ x\ Context\ x\ FixatedAnyROI}$ = -0.94, *SE* = 0.48, *p* = .049). Repetition also interacted with location consistency to modulate the general looking-at-nothing effect, as

illustrated in Figure 12b: when a target appeared in varied screen positions, looking at any of the three blank regions of interest was associated with lower recognition error rates in the early blocks, but this pattern reversed in later blocks (odds ratio: 2.33:1, $\beta_{log(Block) \, x \, Location \, x \, FixatedAnyROI}$ = 0.85, SE = 0.36, p = .018).

Group effects: As illustrated in Figure 11, typical readers made significantly fewer errors than readers with dyslexia (odds ratio: 2.72:1, β_{Group} = 1.00, SE = 0.22, p < .001), but there was no significant difference in how the two groups performed as a function of repetition (odds ratio: 1.30:1, $\beta_{log(Block) \, x \, Group}$ = 0.26, SE = 0.15, p = .069). We predicted a stronger tendency for readers with dyslexia to err more when fixating screen locations previously occupied by distractors, as previously observed by Jones et al. (2018). However, this interaction did not come out significant in our study (odds ratio: 1.08, $\beta_{Group \, x \, PrimaryFixation}$ = 0.08, SE = 0.47, p = .864).

Table 6

Summary of subject-weighted mean error proportions in the training recognition task and interspersed cued-recall trials, post-training recognition and cued-recall tasks

		Context														
		Consistent								Inconsistent						
		M				*SD*				*M*				*SD*			
		TR[a]	TCR[b]	PTR[c]	PTCR[d]	TR	TCR	PTR	PTCR	TR	TCR	PTR	PTCR	TR	TCR	PTR	PTCR
	Consistent (dyslexic)	.173	.567	.225	.679	.132	.209	.213	.203	.227	.744	.241	.753	.112	.226	.173	.232
	Consistent (typical)	.091	.435	.082	.489	.072	.229	.136	.294	.110	.542	.140	.493	.089	.222	.167	.294
Location	Inconsistent (dyslexic)	.191	.676	.171	.673	.128	.196	.200	.222	.244	.621	.216	.716	.138	.208	.190	.245
	Inconsistent (typical)	.107	.520	.104	.466	.085	.212	.155	.223	.123	.430	.098	.428	.098	.230	.126	.273

[a] Training recognition

[b] Training cued-recall

[c] Post-training recognition

[d] Post-training cued-recall

Figure 11

Subject-weighted mean recognition error rates as a function of reading ability and repetition in the training and recognition task
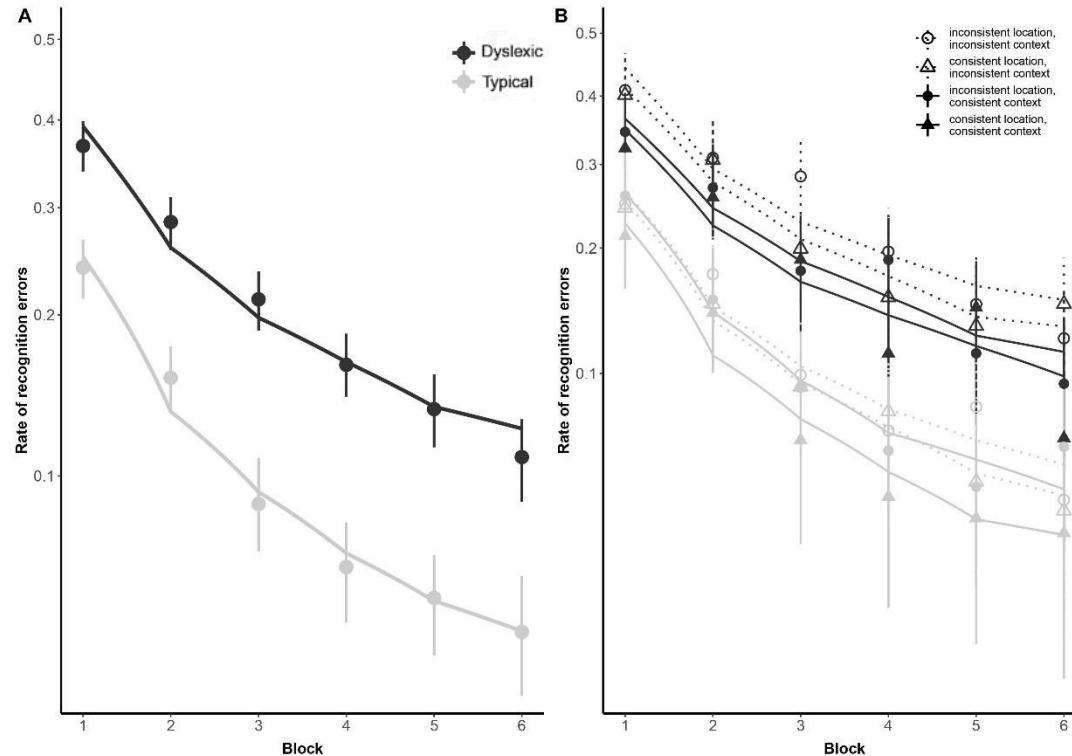
Note. Panel **(A)** depicts overall recognition error rates for readers with dyslexia and typical readers, whereas panel **(B)** outlines the same data broken down by trial type (*i.e.*, whether context and/or location was kept consistent during encoding). The *y*-axis is logit-scaled in both plots to match logistic regression error analyses. Point ranges/error bars represent bootstrapped confidence intervals, and lines represent logistic regression model fits.

Figure 12

Subject-weighted mean recognition error rates during training
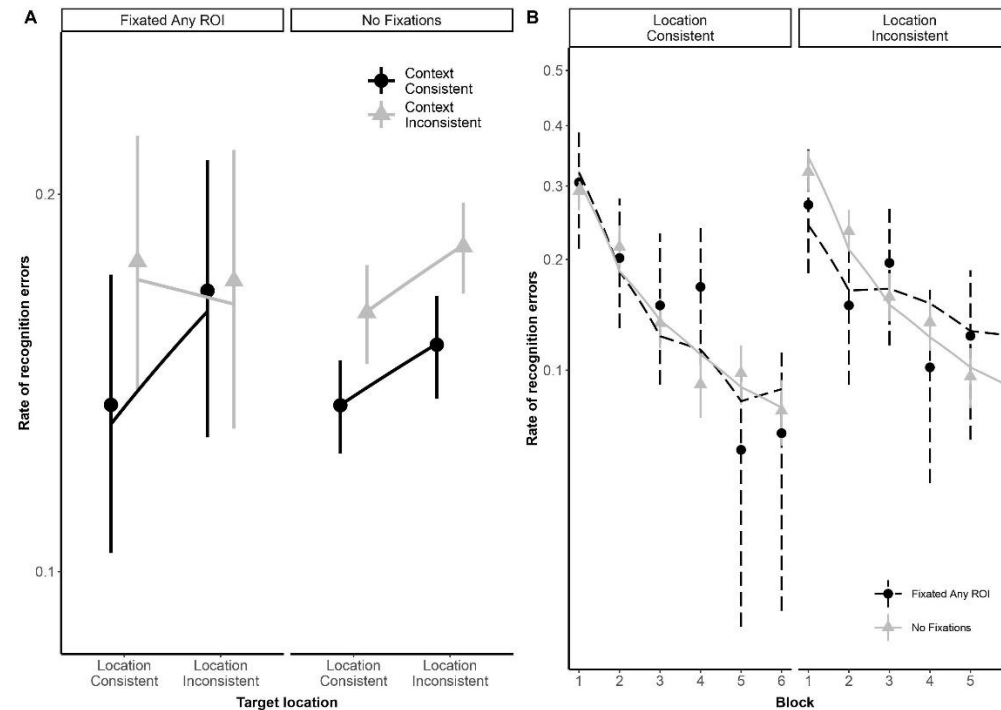
Note. Panel **(A)** shows subject-weighted mean recognition error rate as a function of context and location consistency in trials where participants looked at any of the three regions of interest (ROI), depicted by the "FixatedAnyROI" facet, compared to trials in which looking-at-nothing behaviour did not emerge. Panel **(B)** shows subject-weighted mean recognition error rate as a function of repetition (*i.e.*, blocks) and location consistency. In both panels, the *y*-axis is logit-scaled to match logistic regression error analyses. Point ranges/error bars represent bootstrapped confidence intervals, and lines represent logistic regression model fits.

Similarly, contrary to our prediction that spatial *and* contextual consistency would jointly decrease recognition error rates in general, albeit with a disproportionately stronger effect for readers with dyslexia, these two-way and three-way interactions also did not reach significance in the present study (odds ratio = 0.89:1, $\beta_{Context \times Location}$ = -0.12, *SE* = 0.26, *p* = .650; odds ratio = 1.18:1, $\beta_{Group \times Context \times Location}$ = 0.17, *SE* = 0.29, *p* = .563).

Our analysis yielded a higher-order interaction involving reading ability and eye movements. Specifically, a five-way interaction between block, group, context consistency, location consistency, and ROI fixation (odds ratio: 44.78:1, $\beta_{log(Block) \times Group \times Location \times Context \times FixatedAnyROI}$ = 3.80, *SE* = 1.38, *p* = .006; see Figure 13).

This interaction suggests differential sensitivity to presentation details, but we report it with caution because we did not anticipate the precise form of this interaction and, as noted earlier, the analysis lacks the necessary power to properly assess it (Button et al., 2013).

Figure 13

Higher-order interaction involving reading ability and eye movements during training
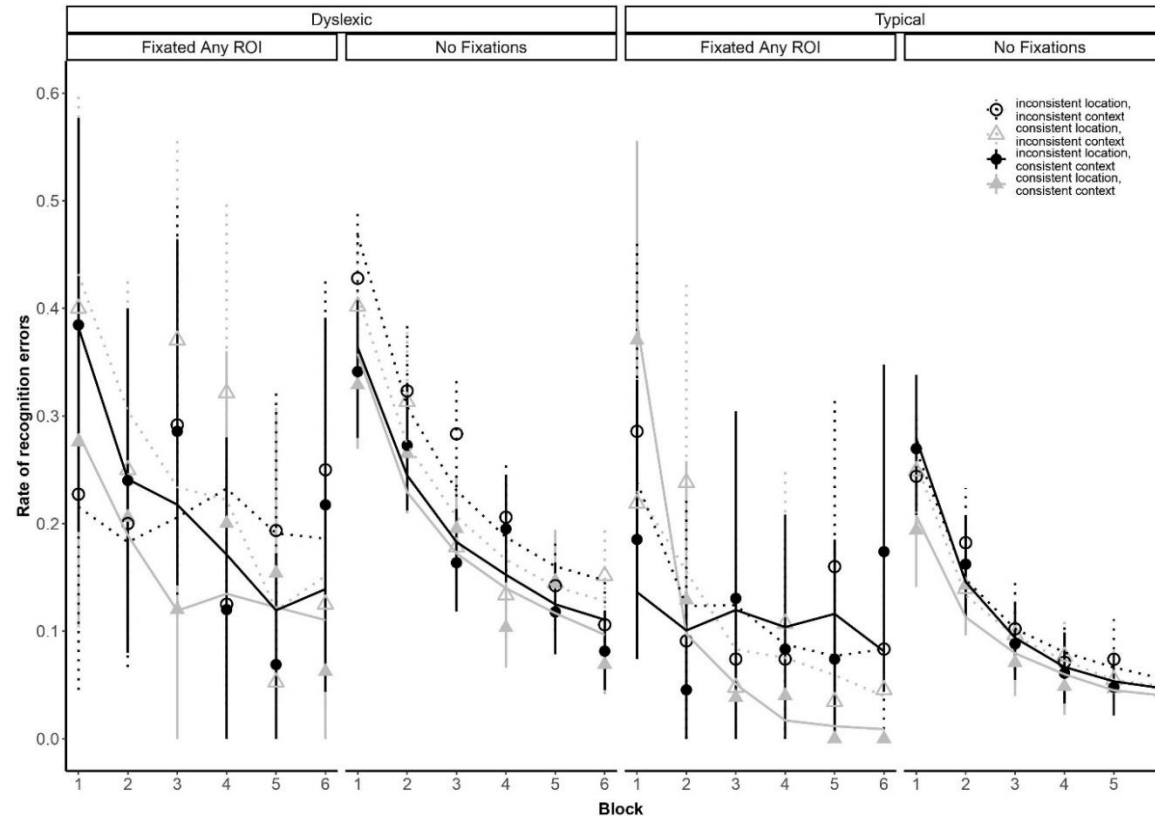
Note. Subject-weighted mean recognition error rate as a function of repetition (*i.e.*, blocks), group membership, context and location consistency, and whether participants looked at one of the three regions of interest (ROI). Point ranges/error bars represent bootstrapped confidence intervals, and lines represent logistic regression model fits.

Cued-recall trials

Due to a playback error which rendered some of the audio files unintelligible, we excluded 82 (3.25%) of the cued-recall trials that were interspersed in the training task, leaving the 2,438 analysable trials (1,210 dyslexic; 1,288 typical) summarised in Table 6. Overall, readers with dyslexia incorrectly recalled bindings more frequently than typical readers (odds ratio: 2.28:1, β_{Group} = 0.82, SE = 0.19, $p <$ .001). As illustrated in Figure 14a, location-consistency and context-consistency significantly interacted (odds ratio = 0.35:1, $\beta_{Context\,x\,Location}$ = -1.04, SE = 0.39, $p =$ .007), such that location-consistency only benefitted recall during training when context was also consistent, but the strength of this interaction did not significantly differ between groups (odds ratio = 0.80:1, $\beta_{Group\,x\,Context\,x\,Location}$ = -0.22, SE = 0.37, p = .549).

Post-training cued-recall test

We excluded 224 (8.88%) trials from the post-training cued-recall test, due to the playback error noted above, leaving the 2296 trials (1,113 dyslexic; 1,183 typical) summarised in Table 6. Overall, readers with dyslexia incorrectly recalled bindings more frequently than typical readers (odds ratio: 3.50:1, β_{Group} = 1.25, SE = .28, $p <$ .001), but as illustrated in Figure 14b they benefitted more from having consistently appeared with the same distractors during the training phase (odds ratio = 1.48:1, $\beta_{Group\,x\,Context}$ = .39, SE = .2, p = .047).

Figure 14

Rate of recall errors during training and post-training
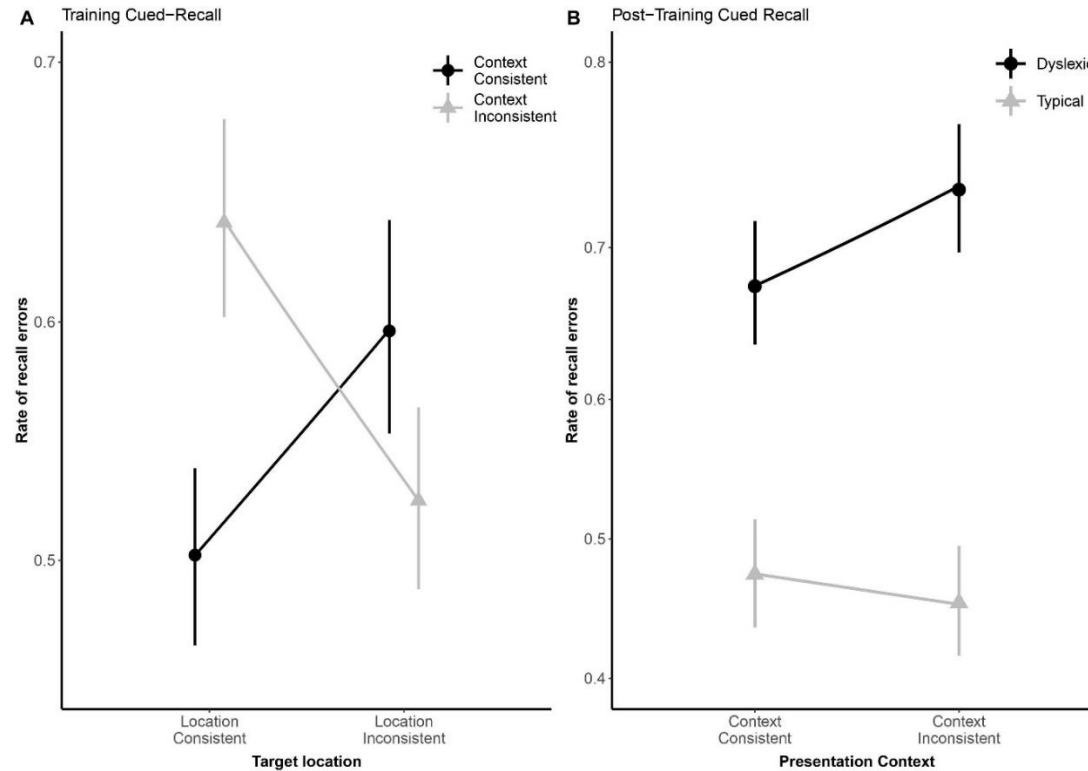
Note. Panel **(A)** shows subject-weighted mean recall error rate as a function of context and location consistency in the cued-recall trials interspersed in training. Panel **(B)** shows subject-weighted mean recall error rate as a function of reading ability and context consistency in the post-training cued-recall test. In both panels, the *y*-axis is logit-scaled to match logistic regression error analyses. Point ranges/error bars represent bootstrapped confidence intervals, and lines represent logistic regression model fits.

Post-training recognition test

Accuracy in the post-training recognition test is summarised in Table 7. Readers with dyslexia incorrectly recognised bindings more frequently than typical readers (odds ratio: 2.71:1, β_{Group} = .99, SE = 0.33, p =.003). No other effects or interactions approached significance.

A summary with the significant effects and interactions observed in all models can be found in Table 7. A complete list with all the effects and interactions can be found in the Supplementary Materials section (Appendix B).

Table 7

Summaries of logistic mixed effects regression analyses of error frequency

Recognition error frequency (training)				
	Coef (β)	SE (β)	p	OR (exp (β))
(Intercept)	-2.18	0.12	<.001	0.11
log(Block)	-1.13	0.08	<.001	0.32
Group (typical, dyslexic)	1.00	0.22	<.001	2.72
Context (consistent, inconsistent)	0.30	0.13	.018	1.35
log(Block) x Location x FixatedAnyROI	0.85	0.36	.018	2.33
Location x Context x FixatedAnyROI	-0.94	0.48	.049	0.39
log(Block) x Group x Location x Context x FixatedAnyROI	3.80	1.38	.006	44.78

Cued-recall error frequency (training)				
	Coef (β)	SE (β)	p	OR (exp (β))
(Intercept)	0.33	0.13	.011	1.39
Group (typical, dyslexic)	0.82	0.19	<.001	2.28
Location x Context	-1.04	0.39	.007	0.35

Cued-recall error frequency (post-training)				
	Coef (β)	SE (β)	*p*	OR (exp (β))
(Intercept)	0.44	0.16	.007	1.56
Group (typical, dyslexic)	1.25	0.28	<.001	3.50
Group x Context	0.39	0.19	.047	1.48

Recognition error frequency (post-training)				
	Coef (β)	SE (β)	*p*	OR (exp (β))
(Intercept)	-2.23	0.18	<.001	0.11
Group (typical, dyslexic)	0.99	0.33	.003	2.71

Response times

Although our predictions and power analyses concerned only accuracy data, for completeness, we also ran an analogous analysis of the response time data, reported in the Supplementary Materials section (See Appendix B). In sum, although readers with dyslexia were generally slower at recognising the bindings during training, response times for the accurate responses did not significantly differ between the two groups. In the post-training recognition test, however, typical readers accurately recognised the bindings significantly faster than readers with dyslexia.

Discussion

Efficient cross-modal binding (*e.g.*, mapping letters to letter sounds) is fundamental in the initial stages of literacy acquisition (Harm & Seidenberg, 1999; Seidenberg & McClelland, 1989), and this skill appears to be impaired in children and adults with

developmental dyslexia (Blau et al., 2009; M. Jones et al., 2009, 2018). Here, we examined whether dyslexic readers' ability to track stimulus consistencies across multiple exposures might contribute to their impaired audiovisual learning (relative to typical readers), more generally considering the contributions of statistical learning and associated episodic memory processes to the acquisition of novel cross-modal bindings. Our experiment simulated the incremental process of letter-sound acquisition by repeatedly presenting participants with arbitrary visual-phonological associations. We were primarily motivated by (1) the specific question of how episodic memory cues, such as consistent spatial and contextual properties, might modulate readers' acquisition of these novel bindings, and (2) more generally identifying differences in the learning characteristics of typical and dyslexic readers. This section is structured according to these objectives. To briefly summarise our main findings, we show that whilst all participants used stimulus consistencies in order to improve learning, readers with dyslexia may show a particular reliance on stimulus co-occurrence.

How statistical consistencies impact cross-modal binding for all participants

We examined the extent to which reliance on the consistency (or inconsistency) of spatial and contextual stimulus properties – presented across multiple exposures and trials – modulated binding performance. These effects were examined in the context of the main training task, but also in the recognition and recall post-tests. We also examined the extent to which participants would execute looks towards relevant blank screen locations previously occupied by targets, and their effect, if any, on recognition accuracy during the training task.

During training, all participants benefitted from a target's repeated presentation as part of the same three-stimulus set (*i.e.*, context consistency; see El-

Kalliny, 2019). Moreover, context interacted with location and screen fixations to modulate error rates: whilst *in*consistent contexts were overall detrimental to recognition (see above), recognition accuracy in these trials nevertheless improved in location-consistent trials, in which items were consistently presented in the same screen location. However, this pattern was predominantly observed in trials where participants *did not* fixate any of the relevant regions of interest. We suggest that since relevant spatial information had presumably already been encoded along with the bindings, re-fixating the empty screen locations in search of spatial retrieval cues may have been redundant, or even deleterious to performance[7]. This relationship is further modulated by stimulus repetition: recognition for stimuli presented in *in*consistent screen locations was found to be more accurate when participants did fixate relevant screen locations, but only during the initial exposures to these stimuli (reflected in performance on the earlier blocks). However, this pattern reversed as a function of block: participants eventually became less accurate following a fixation to a relevant screen location, following multiple exposures to the stimuli. For stimuli with inconsistent locations, therefore, attempts to use spatial location as a retrieval cue became increasingly – and perhaps unsurprisingly - error prone.

In the cued-recall trials interspersed in the training task, participants from both groups also exhibited lower error rates for items consistently encoded in fixed

---

[7] Note that characters' position was changed between encoding and testing phases in 2/3 of all trials during training, to avoid strategic responses. Pilot data suggested that when the positions were congruous across the two phases, participants appeared to rely more on their ability to localise the sound source (*i.e.*, top, bottom left, bottom right), rather than on their ability to bind the sound to its corresponding visual stimulus, pushing performance to ceiling during the main task. Also note that no interactions with 'location' approached significance in our response time analyses (See Supplementary Materials section, Appendix B), suggesting that participants were not particularly slower or faster as a function of location consistency.

locations *and* with fixed contexts. We speculate here that, while participants were still

being trained on the novel bindings, availability of *multiple* episodic memory cues

supported recall. In the absence of cues, however, or when only one consistent cue

was present, recall became more effortful, and thus less accurate.

Taken together, these findings show that *all participants, both typical and*

impaired, readily leveraged temporal and spatial consistencies to bootstrap

audiovisual learning over multiple exposures. Our findings are in line with the

regularity principle of statistical learning (Perry et al., 2010; Twomey et al., 2014;

Vlach & Sandhofer, 2011), in which the cognitive system structures inherent

environmental variability by integrating frequently occurring items by their co-

occurrence, or consistency. This enables us to build supraordinate categories for

words, and parts of words in the lexicon, and associated semantic webs. In real-

world learning contexts, both spatial location and context would presumably be

considerably more varied (though perhaps context less so), so the regularity

principle would lead beginning readers to average them out as noise. When we

increased the consistency of these features, however, readers appear to have

incorporated these co-occurrences into their proto-orthographic representations, thus

reinforcing our previous claim that even experienced readers track such information

as potentially meaningful (Jones et al., 2018).

Differential stimulus consistency effects on typical and dyslexic readers

Typical readers were more accurate than readers with dyslexia in all tasks, as in

Jones and colleagues' (2018) cued-recall study. The main recognition task also

suggested differences in the effect of stimulus consistencies on typical and dyslexic

readers' performance, in the form of a significant 5-way interaction. Such high-order

interaction is challenging to interpret, and based on pre-experiment simulations, we

did not expect to have power to accurately assess them. As others have noted (*e.g.* Button et al., 2013), low power increases the likelihood of false positives as well as false negatives in null hypothesis statistical testing. At present, we tentatively suggest this interaction may be understood as suggesting global differences emerging for errors that implicate re-fixations versus errors that proceed *via* direct access.

In the post-training recognition and cued-recall tests – the two tasks we administered to examine longer-term retention of the bindings – participants from the two reading groups recognised more bindings than they recalled, consistent with the general trend whereby recognition of previously studied items is often successful even when the items cannot be accurately recalled (Tulving, 1982). Overall, typical readers recognised *and* recalled twice as many bindings as did readers with dyslexia. We suggest that, given dyslexic readers' propensity to benefit less from multiple exposures during training (Ahissar, 2007), there are knock-on effects for later retrieval. Their comparatively worse performance in the two post-training tests is consistent with previous studies showing reduced long-term memory capacity in readers with dyslexia (Huestegge et al., 2014; Menghini et al., 2010).

In the post-training tests, one might reasonably predict that if repeated exposure to bindings is sufficient for participants to build strong representations to support recognition and recall, they may no longer rely on episodic cues to aid memory retrieval. Behavioural data showed that whilst this was indeed the case for the typical reader group, it was not the case for readers with dyslexia: compared to typical readers, they more frequently correctly recalled bindings which had *consistently* been trained with the same distractors. We suggest that dyslexic readers' reliance on episodic cues may be indicative of a more fragile memory

116

representation: bindings that are robustly represented in memory are accessed and retrieved via a direct visual-to-auditory route rather than via an indirect route that is dependent on seemingly irrelevant episodic cues (M. Jones et al., 2018). Our findings suggest that readers with dyslexia use context in order to support retrieval, consistent with previous findings, in which dyslexic readers benefited from item presentation in a fixed temporal order (Toffalini et al., 2018).

Taken together, our findings with respect to group differences show a deficit for readers with dyslexia in both recognising and recalling audiovisual bindings of novel items, in all tasks. This finding is in line with previous paired associate learning studies (M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Litt & Nation, 2014; Messbauer & de Jong, 2003; Toffalini et al., 2018, 2019; H. C. Wang et al., 2017; Warmington & Hulme, 2012). Even at the behavioural level, then, adult readers with dyslexia required substantially more repetition in order to achieve accuracy comparable to typical readers (see Figure 11), a pattern that is remarkably consistent with Saffran and colleagues' (1996) predictions that word learning in individuals with language disorders requires at least twice the exposure. Even these highly compensated adults with dyslexia were therefore relatively impervious to the effects of frequency on learning. Did this mean that they were insensitive to stimulus consistencies, which should, under normal circumstances, help in the statistical learning process? Our findings suggest not. Readers with dyslexia seemed perfectly able to use consistency in spatial location information to improve recall, which was on a par with the effect of location-consistency on their typically reading peers. This finding is at odds with the hypothesis that readers with dyslexia fail to use location information as a cue for cross-modal binding (cf. Jones et al., 2013; Toffalini et al., 2018), as typical readers are shown to do (Treisman & Gelade, 1980; Treisman &

Zhang, 2006). And it shows, moreover, that readers with dyslexia are in fact able to track longer-range statistical probabilities when the cues afforded across trials are highly salient and beneficial for item recognition. However, our findings showed a reader-type discrepancy in the use of context-consistency cues for item recognition: dyslexic readers' error rates decreased disproportionately compared with typical readers' when items were shown in a consistent context, (*i.e.*, item A appearing on each exposure with items B and C). Thus, readers with dyslexia showed an increased reliance on context consistency, suggesting that the entire episode (trial) was encoded as a whole. Previous studies have also noted a proclivity for chunking in dyslexia (Ullman & Pullman, 2015), in which memorisation of whole word forms is favoured over phonological decoding, leading to a disproportionate reliance on declarative memory for reading. We tentatively suggest that readers with dyslexia may use co-occurrences or consistencies to bootstrap their relative insensitivity to frequency: in a cognitive system that fails to efficiently integrate a current instance with previous exposures to that same item (Ahissar, 2007; Altmann, 2017), there may be a tendency to over-rely on episodic traces from within a single trial (as shown in the looks-at-nothing data), but also across trials (shown in an increased dependency on co-occurrences).

An important feature of this study is that testing was conducted *via* remote access to participants' personal webcams to collect eye-tracking data. Despite the rigorous controls and procedures documented in the methods and results sections, such convenience does not come without its possible limitations and challenges. Online data collection generally raises a number of questions, such as the participant's full capacity to understand and follow the instructions, length of task completion relative to similar in-lab studies, and the element of trust in participants'

self-reported data (such as dyslexia status, which we nevertheless mitigated to the extent that it is possible *via* objective literacy and cognitive measures). Collection of eye-tracking data *via* webcam-based eye tracking is a new and exciting method that requires highly stringent procedures in order to ensure the best possible data quality (see Bott et al., 2017 and Semmelmann & Weigelt, 2018 for empirical validation of web-based eye-tracking as a suitable experimental method). Here, we took careful design considerations such as providing pictorial as well as written instructions, adding frequent attention checks to ensure participants' computers were not left unattended mid-experiment, and enforcing an overall time limit to prevent excessively long breaks between tasks. We also employed a conservative filtering approach to exclude eye tracking estimates with low face detection values to avoid as much as possible fluctuation depending on variables such as lighting conditions and/or participants' sitting conditions. We also calculated participants' viewing distance, and avoided relying on fine-grained eye tracking analyses that would require sophisticated infrared technology.

Conclusion

This study aimed to shed further light on audiovisual learning differences in typical and dyslexic readers. Our findings show that all of our participants used consistencies in the input during stimulus exposure in order to improve recognition and recall of items. However, dyslexic readers showed a persistent difficulty in integrating items in memory, and an overreliance on episodic detail in order to assist in the retrieval process. These findings may be of clinical relevance in understanding the challenges facing apparently high functioning adults. Overall, our findings provide novel evidence on dyslexic readers' reduced ability to create abstracted representations in memory, relying instead on instance-based memory.

Data Availability Statement

The datasets generated for this study can be found on GitHub

[https://github.com/simOne3107/BindingExperimentLocationContextWebcamEyetrac

king].

Chapter 5

How Specific is the Paired Associate Learning Deficit in Dyslexia to Cross-Modal Working Memory?

Abstract

Proficient reading requires the ability to create, store, and retrieve cross-modal bindings (*e.g.*, letter and letter-sound mappings) from memory. This binding ability has been shown to be deficient in individuals with atypical reading skills, such as those with developmental dyslexia. However, research to date does not provide a well-controlled comparison of cross-modal and within-modal binding ability. Here, we sought to ascertain the specificity of a cross-modal binding deficit in dyslexia, in the absence of verbal output demands, *via* a novel manipulation of the well-known n-back paradigm. Our findings showed that adults with and without dyslexia were more accurate but slower to respond in the auditory-auditory condition of the n-back task, relative to the visual-visual and audio-visual conditions. This pattern of results is indicative that distinct stimulus modality presentations appear to influence encoding and processing mechanisms differently. Overall, our experimental paradigm was insufficiently sensitive to detect differences in working memory binding processes as a function of reading ability. The lack of evidence of a specific cross-modal binding deficit in adults with developmental dyslexia in the present study is inconclusive and needs further investigation.

Keywords: paired associate learning; cross-modal binding; reading; n-back; working memory

How Specific is the Paired Associate Learning Deficit in Dyslexia to Cross-Modal Working Memory?

Skilled reading requires the ability to process and store co-occurring visual-phonological information (*i.e.*, letter and letter sounds in alphabetic languages). As such, working memory – the limited-capacity cognitive system involved in the simultaneous storage and processing of information (Baddeley, 1986; Baddeley & Della Sala, 1996) – is assumed to play a fundamental role in reading development. Indeed, previous research has demonstrated that efficient working memory may be critical to the development of literacy skills such as phonological coding (Oakhill & Kyle, 2000), expressive vocabulary (Peng et al., 2018), and reading comprehension (Peng et al., 2018). Individuals with aberrant reading ability, such as those with developmental dyslexia, generally demonstrate less automated reading skills (Shaywitz & Shaywitz, 2005), which may stem from difficulty in encoding, consolidating and retrieving visual-phonological *associations* in memory (Hulme et al., 2007; Warmington & Hulme, 2012). Children with reading deficits perform poorly in tasks assessing such cross-modal working memory (Albano et al., 2016; Garcia et al., 2019; Toffalini et al., 2019). However, it remains unclear whether poorer readers experience greater difficulty in storing cross-modal representations *per se*, or whether similar difficulties can be elicited within the same modality (*i.e.*, either visual or phonology), signalling a more general impairment in cognitive load, for example. Here, we employ the n-back paradigm to investigate within-modal and cross-modal working memory in adults with and without developmental dyslexia. Our aim is to examine whether establishing cross-modal associations are specifically onerous for individuals with impaired reading skills.

Reading ability is assumed to correlate with working memory capacity (Baddeley et al., 1985; Vellutino et al., 2004). A review study (Jacob & Parkinson, 2015) and a meta-analysis (Peng et al., 2018) that looked at a potential link between working memory and reading ability found a moderate relationship between the two, with reading ability correlating most strongly with *verbal* working memory (Peng et al., 2018). The phonological loop and the central executive, two key components in the domain-general working memory model (Baddeley, 1986; Baddeley & Della Sala, 1996), are assumed to play a fundamental role in reading. Specifically, the phonological loop appears to support novice readers in basic reading skills, such as phonological awareness (Kibby et al., 2014; McDougall et al., 1994), whereas the central executive reportedly contributes to reading fluency and comprehension (Kibby et al., 2014).

In the most recent version of the multicomponent working memory model, Baddeley (2000) added the episodic buffer, a limited-capacity component that is intended to capture the ability to integrate and store bound features that stem from different sources, within- and cross-modally (Baddeley, 2000; Baddeley et al., 2010). In essence, the ability to efficiently create, store, and retrieve cross-modal bindings appears to be particularly relevant for reading (Blomert, 2011), especially in the early stages of literacy development when novice readers are expected to learn how to make seemingly arbitrary associations between visual and phonological information (*i.e.,* grapheme-phoneme connections, or letter and letter sound pairings). Previous research has provided evidence for impaired cross-modal binding ability in children and adults with atypical reading skills, such as those with developmental dyslexia (Albano et al., 2016; Calabrich et al., 2021b; Garcia et al., 2019; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Toffalini et al., 2019), further supporting the

existence of a potential link between (cross-modal) working memory and reading performance. Such a deficit has been previously suggested to only emerge in tasks that require a spoken output (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014; but see Calabrich et al., 2021; Poulsen, 2011). It is possible, however, that previous studies failed to capture differences in within-modal and cross-modal associative learning due to an inability to properly equate the number of within-modal and cross-modal features that participants were required to store and then retrieve from memory.

In the working memory literature, a paradigm that is widely used for both assessment and training purposes is the n-back task (Baddeley et al., 2010; He et al., 2022; Kirchner, 1958; W. Li et al., 2021; Pelegrina et al., 2015; Szmalec, Verbruggen, et al., 2011). In classic n-back tasks, continuous sequences of visual and/or auditory stimuli (*e.g.*, letters, digits, symbols) are presented to participants who are then asked to indicate whether the current stimulus matches the one presented *n* steps back in the sequence. To illustrate, in a 2-back task, individuals must remember the item presented two trials back in the sequence. Task difficulty (*i.e.*, working memory load) can generally be manipulated by increasing the *n*. N-back tasks demand continuous but temporary storage of relevant information in memory whilst suppressing ongoing interference from novel items (Redick & Lindsey, 2013). This process can be construed as somewhat akin to reading, in which efficient updating of (phonological, orthographic, and/or semantic) information as well as suppression of competing or no longer relevant information are similarly needed (Gernsbacher & Faust, 1991; Palladino et al., 2001). Indeed, previous research suggests that deficits in reading may stem from inefficient information updating skills (Palladino et al., 2001) as well as from an impaired ability to supress

irrelevant information (Chiappe et al., 2000; Pimperton & Nation, 2010). Because n-back tasks involve using dynamic information storage and simultaneous control processes, they are generally considered to have face validity as a measure of working memory performance (Kane et al., 2007; Meule, 2017; Owen et al., 2005), with few exceptions (Jaeggi et al., 2010). While some researchers have argued that working memory tasks might not be the best predictors of reading performance (Palmer, 2000), previous studies have shown that the n-back task is efficient in detecting working memory differences in children with typical reading ability from those with developmental dyslexia at the behavioural (Beneventi et al., 2010a, 2010b; Bogaerts et al., 2015; Lotfi et al., 2022; J. Wang et al., 2022, but see Sela et al., 2012) and neural levels (Sela et al., 2012)

The current study

In the present study, we examine within-modal and cross-modal working memory of novel bindings by asking adults with and without developmental dyslexia to perform three versions of a novel n-back task: auditory, visual, and audio-visual. Each n-back task version created is intended to assess the individual's associative memory for two features: visual-visual associations (within-modal binding), auditory-auditory associations (within-modal binding), and visuo-auditory associations (cross-modal binding). Auditory features comprise (a) pseudowords, following English phonotactics, and (b) speaker voice identity, whilst visual features comprise (c) Kanji characters, unfamiliar to the participants, and (d) colours. A visual-visual association would require binding a specific character to a specific colour; an auditory-auditory association would require binding a specific pseudoword to a specific speaker identity, and a visual-auditory association would require binding a specific character to a specific pseudoword. Thus, memory load (defined as the number of features) is

equated across conditions. The participant's task is to identify whether the current stimulus is identical or non-identical to a stimulus seen on the n-back trial.

If adults with dyslexia have difficulties with binding per se, we expect a group main effect that shows typically developed readers performing faster and more accurately in all three versions of the n-back task, consistent with previous observations that individuals with dyslexia demonstrate impaired working memory relative to their typical reader counterparts (Beneventi et al., 2010b; Chiappe et al., 2000; Reiter et al., 2005; J. Wang et al., 2022). However, in line with previous accounts of a cross-modal binding deficit in developmental dyslexia (Calabrich et al., 2021a; Garcia et al., 2019; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Toffalini et al., 2019) even in the absence of a spoken output (Calabrich et al., 2021b), here, we expect an interaction effect wherein performance in the *cross-modal* version of the n-back task is particularly impaired for those with dyslexia.

In addition to reporting accuracy rate and response times, we also report two other task performance indices, namely discrimination index (d') and response bias (C), calculated from subjects' hits and false alarms. Within a signal detection theory framework, discrimination index can be conceptualised as the ability to differentiate targets from non-targets, whereas response bias is the propensity to categorise stimuli as targets or nontargets (D. Green et al., 1966; Lynn & Barrett, 2014). In other words, discrimination index d' refers to how difficult or easy it is to detect that a target is present, whereas response bias C is the extent to which a given response is more probable than another. In both instances, the higher the value obtained, the better. When scrutinising performance in n-back tasks, such distinction is important because it allows us to also take commission and omissions errors into consideration, two frequently occurring errors that are often overlooked in n-back

tasks (Meule, 2017). Here, we hypothesise that individuals with a typical reading profile will demonstrate an overall superior ability to discriminate targets from nontargets by maximising 'hits' and 'correct rejections' and by minimising 'misses' and 'false alarms', relative to readers with dyslexia. Similarly, we also expect that readers with dyslexia will demonstrate a higher propensity to categorise nontargets as targets, relative to typical readers.

Methods

Participants

A total of 108 monolingual speakers of British English were recruited through Bangor University SONA system and Prolific (*www.prolific.co*). Of these, 12 were excluded due to technical issues and/or incomplete submissions, leaving us with 48 readers with dyslexia (32 females, 15 males, 1 other; M_{age} = 23.34, SD_{age} = 3.73) and 48 typical readers (25 females, 22 males, 1 unspecified; M_{age} = 21.09, SD_{age} = 3.30). None of the participants reported any history of psychiatric and/or neurological diseases, visual and/or hearing impairments, or any other risk factors. Participants initially self-referred to either the typical reader group or reader with dyslexia group, and this referral was verified *via* a battery of literacy tests (described below). Participants from both groups reported a similar level of education (Readers with dyslexia: M = 14.62 years, SD = 1.81; Typical readers: M = 14.19 years, SD = 1.82; p = .239). All participants provided informed consent, were unfamiliar with the experimental stimuli, and were naïve to the purpose of the experiment. Participants received course credits or payment for their time. The experiment was approved by Bangor University's Ethics Committee.

Literacy and General Cognitive Ability Measures

We administered a battery of twelve short tests to validate participants' group membership: (1) Adult Reading Questionnaire (Snowling et al., 2012); (2) word reading efficiency and (3) phonemic decoding efficiency subscales of the Test of Word Reading Efficiency (Torgesen et al., 1999), (4) Wide Range Achievement Test's (WRAT4) Word Reading subtest (Wilkinson & Robertson, 2006); (5) letter and (6) digit versions of the Rapid Automatized Naming (RAN) subtest from the Comprehensive test of Phonological Processing (Wagner et al., 1999); (7) Matrix Reasoning from the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) as an index of non-verbal IQ; (8) Similarities subtest from the Wechsler Adult Intelligence Scale (Wechsler, 1981) as an index of verbal intelligence quotient (IQ); computerised (9) forward and (10) backward digit span tests, each requiring keyboard responses to visually presented sequences of digits; and computerised (11) forward and (12) backward spatial span tests, each requiring mouse-click responses to visually presented blocks.

Stimuli

We created three 2-back task conditions for this study: (1) visual-visual; (2) auditory-only; and (3) visuo-auditory (*i.e.,* cross-modal). In each n-back condition, two stimulus features were arbitrarily but consistently matched across the course of the experiment on 80% of trials. These features consisted of (a) a Kanji character and a pseudoword in the visual-auditory condition, (b) a Kanji character and a colour in the visual-only condition, and (c) a pseudoword and a voice in the auditory-only condition. Colour and voice were introduced to the unimodal n-back conditions to equate the number of features that participants were required to store in memory in all three tasks. On the remaining 20% of trials, the experimental items had a non-consistent pairing of features (*i.e.*, a mismatch). In all n-back conditions, these

mismatches were bidirectional to prevent participants from attending to only one of the features whilst attempting to keep the items in their working memory (See Figure 15).

For the visual-only n-back condition, twenty 3-stroke Mandarin Chinese characters (*e.g.*, 大), unfamiliar to the participants, were arbitrarily matched to twenty colours. These colours were selected from an online tool that generates colour palettes that are suitable for different screen resolutions and are accessible for individuals with a wide range of visual capabilities (http://colorsafe.co). The colour-character pairings were consistent in 80% of the visual-only n-back trials (*e.g.,* Character 1 & Colour 1), whereas each character was matched with a different colour *(e.g.*, Character 1 & Colour 20) in the remaining trials. Colours used in dominant and non-dominant pairings were never of the same hue. Dominant and non-dominant pairings of each Kanji character were also carefully selected and revised by two independent researchers to minimise visual similarity between them. Pre- and post-tests checks (described below) were performed to ensure participants were able to adequately differentiate relevant pairs of colours and characters used in the present study.

For the auditory-only condition, twenty consonant-vowel-consonant (CVC) pseudowords (*e.g.*, /gɔp/) were arbitrarily matched to twenty distinct voices (10 females and 10 males, all native speakers of British English). Each auditory-only n-back pseudoword was recorded by one male ($N = 10$) and one female speaker ($N = 10$). The pseudoword-voice pairings were consistent in 80% of the auditory-only n-back trials (*e.g.*, Pseudoword 1 & Voice 1). In the remaining trials, each pseudoword was articulated by a speaker from the opposite gender *(e.g.*, Pseudoword 1 & Voice 20). The amplitude of each auditory file was multiplied on Praat (Boersma &

Weenink, 2021b), and its average intensity was used to homogenise sound intensity across different speakers.

For the cross-modal n-back condition, twenty consonant-vowel-consonant pseudowords (*e.g.*, /pɔb/) were arbitrarily matched to twenty 3-stroke Kanji characters (*e.g.*, 三). The visual and auditory stimuli used in the cross-modal condition were different from those used in the unimodal versions in order to obviate learning interference effects across conditions. The pseudoword-character pairings were consistent in 80% of the cross-modal n-back trials (*e.g.*, Pseudoword 21 & Character 21). In the remaining trials, each pseudoword was paired to a different character (*e.g.*, Pseudoword 21 & Character 40). A female native speaker of British English recorded all cross-modal n-back pseudowords. Pseudowords used in non-dominant mappings (*i.e.*, in 20% of the trials) did not share any phonemes with their dominant counterparts (*e.g.*, Character 1 mapped to 'bep' in 80% of the trials, but to 'fod' in the remaining trials).

All forty pseudowords followed English phonotactic rules, were generated with Wuggy — a multilingual pseudoword generator (Keuleers & Brysbaert, 2010), and were digitised at 44.1 kHz on Praat (P Boersma & Weenink, 2021). In all n-back conditions, visual stimuli were presented on a peach background (#EDD1B0), a dyslexia-friendly background colour suggested to enhance screen readability (Rello & Bigham, 2017). To ensure sufficient colour contrast levels between the background and the experimental characters, as well as between characters mismatched in colour in 2-back incongruent trials, we calculated Delta-E — a measure of distance between colours — *via* an online calculator (*http://colormine.org/delta-e-calculator*). All colour-contrasts included had a Delta-E value considerably higher than 2.0, a threshold below which humans fail to perceive

colour differences (Wyszecki & Stiles, 1982) (Visual version: $M = 80.38$, $SD = 23.19$; Cross-Modal version: $M = 82.79$, $SD = 28.77$). Similarly, a contrast ratio of at least 3:1 was used, following the Web Content Accessibility Guidelines to improve readability (Caldwell et al., 2008). Relative luminance, obtained *via* an online colour contrast ratio checker (*https://contrast-ratio.com*), ranged from .01 to .18 ($M = .11$, $SD = .07$).

Procedure

The experiment was programmed and deployed online on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). In all n-back conditions, each trial comprised an item (*i.e.*, a Kanji character in the visual and cross-modal conditions, or a black dot in the auditory condition) presented centrally on the screen, and occupying 60 × 60 units of Gorilla Experiment Builder's (Anwyl-Irvine et al., 2020) screen space (See Figure 15). In the auditory and cross-modal conditions, an audio clip with a CVC pseudoword played synchronously in the background. Participants were instructed to perform a key press in each trial to indicate whether the current item is identical (*i.e.*, a perfect match) to the one presented two trials earlier in the sequence. Items only matching partially to the one presented two trials earlier in the sequence (*e.g.*, same character but different pseudoword) required a 'mismatch' response. Henceforth, we refer to perfect matches as 'targets' and partial matches as 'pseudotargets'.

Each experimental item remained on the screen for 2000 ms or until a key press was detected. This was then followed by a 500-ms blank screen. Failure to provide a response in twenty consecutive trials – a strong indication that the participant's computer had been left unattended mid-task – would result in automatic removal of that participant from the experiment.

Figure 15

Examples of matches and mismatches in the (A) visual, (B) auditory, and (C) cross-modal n-back conditions
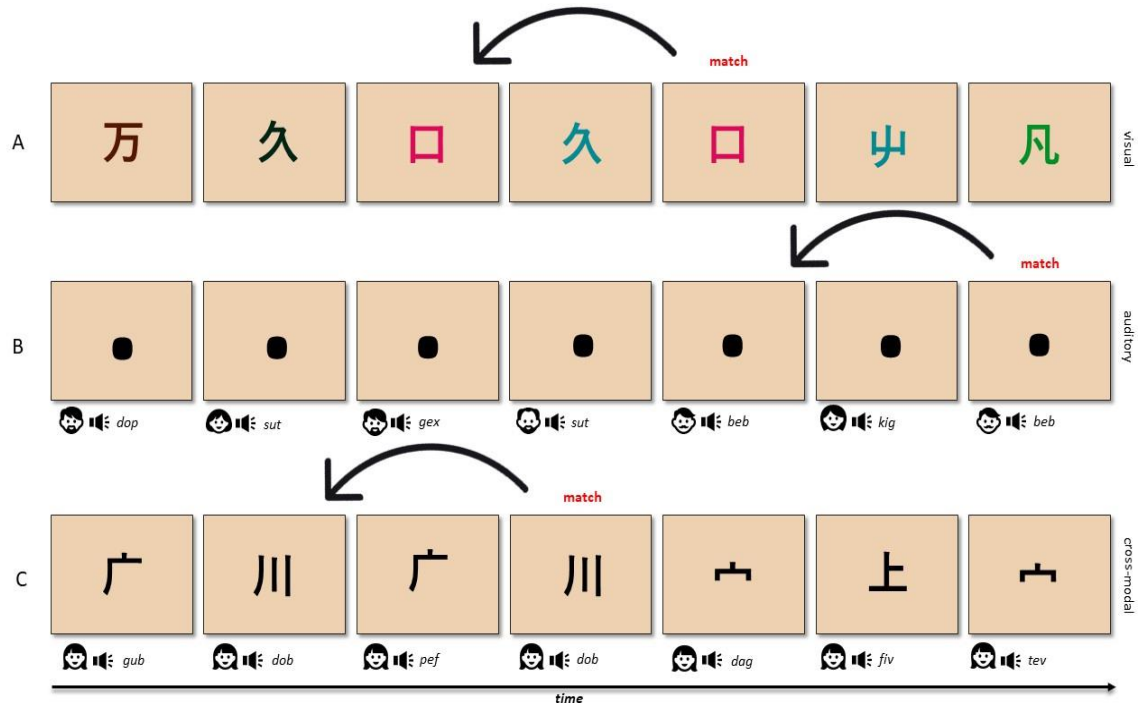
Note. Each square represents a trial (without the 500 ms inter-stimulus interval). To be considered a perfect match, the current stimulus must be identical to the one presented two trials earlier in two distinct features (*i.e.*, character and colour, pseudoword and voice, or character and pseudoword).

We created two lists to ensure key assignment ('f' and 'j') for 'match' and 'mismatch' was counterbalanced across participants in order to avoid laterality effects. There were four blocks per n-back condition per list, each comprising 100 trials, with an untimed break between blocks during which participants could rest. Each block contained 20 target trials (*i.e.*, match), 20 pseudotarget trials (*i.e.*, mismatch), and 60 filler trials (*i.e.*, mismatch). All experimental items appeared an equal number of times across the experiment. Trial presentation was pseudorandomised to ensure that 2-back matches and mismatches were equally

frequent within a block. N-back condition as well as block presentation within each condition were counterbalanced to avoid order effects.

Before each n-back condition, a video tutorial with eight trials representative of those used in the actual experiment were presented in order to demonstrate the procedure to the participants. Each tutorial was followed by a practice block with 16 trials with additional filler stimuli. Participants were provided with feedback after each practice trial and were given the option of repeating the practice block, and/or the video tutorial if needed.

The entire testing session lasted approximately 60 minutes, including background tests and experimental tasks. Any participant exceeding the 120-minute time limit would be automatically removed from the experiment. On average, participants sat 496.43 mm (SD =108.24) from their computer screens as estimated by the Virtual Chinrest task (Q. Li et al., 2020).

Pre- and Post-test checks

In order to screen out any participants potentially unable to discriminate the colours used in in pseudotarget trials, we administered a 'Colour Identification Task' (see Figure 16A) at the beginning of the experimental session. In this task, each trial started with a 1000-ms fixation cross followed by two colours occupying 50 × 80 units of Gorilla Experiment Builder's (Anwyl-Irvine et al., 2020) screen space, along with two response buttons at the bottom of the screen, each occupying 20 × 20 units of Gorilla Experiment Builder's screen space. Participants were instructed to indicate, *via* a mouse-click, whether the colours in each given pair ($N = 20$) were the 'same' or 'different'. Each trial terminated after 10000 ms or after a key press was detected. Trial presentation was randomised across participants. Participants with an

accuracy rate below 80% would be automatically filtered out of the experiment since poor performance in this task would constitute an indication of atypical colour detection. All participants passed this check with a mean accuracy rate of 19.75 (*SD* = 0.6).

Figure 16

Pre- and post-test checks
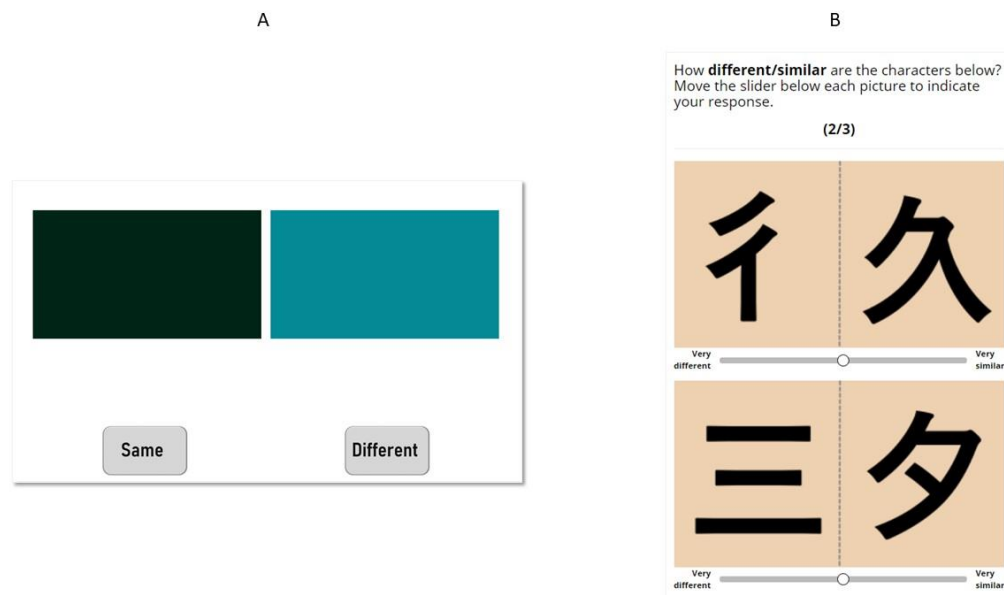
Note. Panel (A) shows a single trial in the 'Colour Identification Task' administered at the beginning of the experimental session, whereas Panel (B) shows two consecutive trials in the 'Shape Similarity Task', administered at the end of the session.

Similarly, we also administered an untimed 'Shape Similarity Task' (see Figure 16B) at the end of the experimental session to test participants' ability to differentiate the Chinese characters used in pseudotarget trials. To this end, we

135

displayed the characters in pairs (*N* = 45), on a black font, against a 'peach' background, and asked participants to rate each pair on a 5-point Likert scale ranging from 'Very different' to 'Very similar'. The items were displayed side-by-side in a 3-page scrollable list format, with each page showing 15 pairs. All items were presented in the same order to all participants. Any trials containing characters that had been rated as 'very similar' to another item by at least 80% of the participants would have been excluded from the analyses. All the items passed this check and were thus kept in all analyses.

No further analyses were conducted on the pre- and post-test checks.

Data Analyses

We conducted two separate logistic mixed effects regression analyses, *via* the glmer() function in the lme4 package (Bates et al., 2015), to model accuracy rate as a function of (1) reading ability (*Group, i.e.*, typical reader = 0, reader with dyslexia = 1) and n-back condition (*Condition, i.e.*, auditory = 0,0; cross-modal = 1,0; visual = 0,1), and as a function of (2) reading ability and presentation modality (*Modality, i.e.,* within-modal = 0, cross-modal = 1). We also ran two separate linear mixed effects regression analyses, *via* the lmer() function in the lme4 package (Bates et al., 2015), to model log transformed response time as a function of (1) reading ability and n-back condition, and as a function of (2) reading ability and modality, as described above. In all models, the predictors were treatment-coded and centred. All models include maximal random effects structures (Barr et al., 2013), reverting to a 'parsimonious' model in case of convergence errors (Bates et al., 2015).

In addition to accuracy rate and response time, we also computed signal detection theory indices, namely sensitivity index (*d'*) and response bias (*C*), to

probe the degree of uncertainty experienced by readers with and without dyslexia in the n-back tasks. To this end, we first calculated the hit rate (HR; *i.e.,* the proportion of correctly identified targets) and the false-alarm rate (FA; *i.e.,* the proportion of incorrectly identified targets in non-target trials). Then, we measured sensitivity (d') by using the formula $d' = zHR - zFA$, where z is the inverse of the standard normal cumulative distribution, and response bias (*C*) by using the formula $C = -1/2(zHR + zFA)$. We used the lm() function from the stats package (R Core Development Team, 2020) to model each signal detection theory metric as a function of reading ability and (1) n-back condition, and (2) modality, as described above. In both models, all predictors were treatment-coded and centred.

Results

Literacy and General Cognitive Ability Measures

Background measures for readers with dyslexia and typical readers are summarised in Table 8. In the Adult Reading Questionnaire, in which higher scores represent greater likelihood of reading disability (Snowling et al., 2012), readers with self-reported developmental dyslexia diagnoses scored significantly higher than those without such diagnoses. Overall, readers with dyslexia had a significant lower performance in word reading accuracy, as well as in word and pseudoword reading fluency measures relative to typical readers. Similarly, readers with dyslexia were significantly slower at naming digits and letters than typical readers. There were no significant group differences on non-verbal or verbal IQ measures, nor on forward and backward digit and spatial span measures.

Table 8

Group scores on literacy and general cognitive ability measures

Test	Measure	Group performance Dyslexic N = 48			Typical N = 48			t	p	Cohen's d
		Range	M	SD	Range	M	SD			
TOWRE	Word reading rate[a]	62-99	81.38	9.69	75-104	90.54	8	5.05	<.01	-1.03
	Pseudoword reading rate[a]	26-62	47.02	7.39	37-63	55.15	5.36	6.16	<.01	-1.26
WRAT	Word reading accuracy	31-52	42	5.01	37-54	46.58	3.83	5.03	<.01	-1.02
CTOPP	RAN digits[b]	11-25	17.02	3.37	10-20	13.9	2.6	-5.08	<.01	1.03
	RAN letters[b]	12-27	17.42	3.23	10-19	14.21	2.22	-5.67	<.01	1.16
WAIS	Verbal IQ	14-30	23.06	3.59	16-32	23.31	3.23	0.35	.721	-0.07
WASI	Non-verbal IQ	17-33	24.77	3.62	17-33	25.73	3.77	1.27	.207	-0.25
ARQ	Risk of reading impairment	15.5-36	25.36	4.7	7-25	15.01	4.6	-11.01	<.001	2.24
	Forward digit span[c]	2-9	5.81	1.35	3-9	6.4	1.61	1.93	.057	-0.39
	Backward digit span[c]	2-9	4.75	1.86	3-9	5.62	1.68	2.38	.019	-0.49
	Forward spatial span	4-9	7.15	1.2	4-9	6.94	1.36	-0.79	.428	.162
	Backward spatial span	4-8	6.31	1.15	4-8	6.21	0.99	-0.47	.635	.097

[a] Number of correctly read items within 45 seconds.
[b] Raw scores in seconds.
[c] Discontinue rule: two incorrectly typed responses in a row.

N-back tasks

A total of 2,017 n-back trials timed out (*i.e.*, no keyboard response was detected) and were thus not included in the statistical analyses. A further 2,634 trials with reaction time values below 100 ms were excluded, leaving the 110,549 trials summarised in Figure 17.

Figure 17

Subject-weighted mean accuracy rates
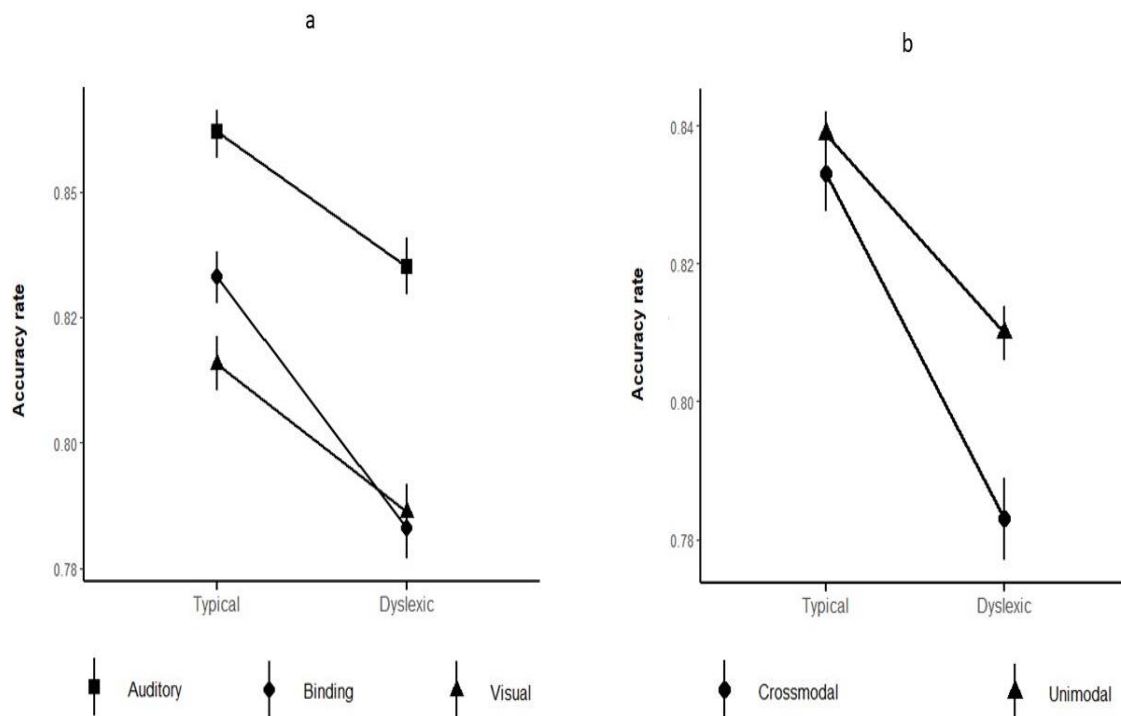
Note. Panel A shows the mean accuracy rates for the n-back conditions whereas Panel B shows the mean accuracy rates for the n-back modalities. Point ranges/error bars represent bootstrapped confidence intervals.

Accuracy rate

Contrary to our predictions, there was no statistically significant difference between typical readers and readers with dyslexia in their overall performance in the n-back

task (odds ratio: 0.80:1, $\beta_{ReaderType}$ = -0.22, SE = 0.16, p = .178). Individuals from both groups performed significantly lower in the visual-visual and audio-visual n-back conditions compared to the auditory-auditory condition (odds ratio: 0.84:1, $\beta_{NbackCondition}$ = -0.18, SE = 0.04, p < .001). This difference, however, was not observed between the two different reading groups (odds ratio: 1.05:1, $\beta_{NbackCondition\ x\ ReaderType}$ = 0.05, SE = 0.06, p = .399).

There was also no statistically significant difference in how typical readers and readers with dyslexia performed in within-modal and cross-modal versions of the n-back task (odds ratio: 0.96:1, $\beta_{ReaderType\ x\ NbackModality}$ = -0.04, SE = 0.11, p = .736). Indeed, all participants performed similarly within-modally and cross-modally (odds ratio: 0.90:1, $\beta_{NbackModality}$ = -0.10, SE = 0.10, p = .290).

Response Time

In addition to excluding timed-out trials and those with response times below 100 ms, we also excluded 20,039 error trials for the response time data analyses, thus leaving 90,510 trials containing accurate responses. Response times for these trials are summarised in Figure 18.

Figure 18

Subject-weighted mean response times in the n-back tasks
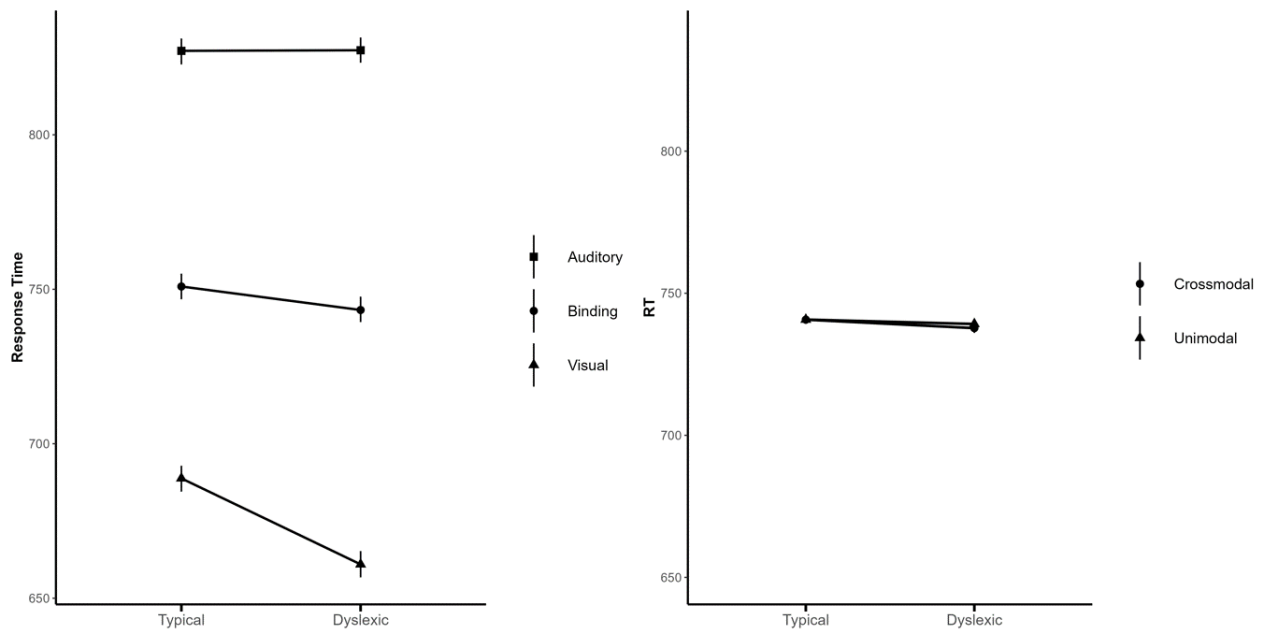
Note. Panel A shows the mean response times for the n-back conditions whereas Panel B shows the mean response times for the n-back modalities. Response times on the y axis are shown in milliseconds. Point ranges/error bars represent bootstrapped confidence intervals.

There was no statistically significant difference between typical readers and readers with dyslexia in how fast they responded in the n-back task ($\beta_{log(ReaderType)}$ = 0.03, *p* = .502), a pattern of results similar to that observed in the accuracy rate data. Individuals from both groups responded significantly faster in the visual and audio-visual n-back conditions compared to the auditory condition ($\beta_{log(NbackCondition)}$ = -0.12, *p* < .001). As observed in the accuracy rate data, the predictor n-back condition was not found to significantly interact with reader type in the response time analyses ($\beta_{log(NbackCondition\ x\ ReaderType)}$ = -0.03, *p* = .186).

Readers with dyslexia were shown to respond significantly faster than typical readers on both cross-modal and within-modal versions of the n-back task, whereas typical readers were faster in responding to the cross-modal compared with the within-modal conditions ($\beta_{log(NbackModality \times ReaderType)} = 0.02$, $p < .01$). None of the main effects approached statistical significance when looking at response times as a function of n-back presentation modality and reader type ($\beta_{log(NbackModality)} = 0.03$, $p = .836$; $\beta_{log(ReaderType)} = -0.10$, $p = .517$).

Signal detection theory

To obtain bias-free measures of performance, we also calculated d' prime and criterion, two signal detection theory indices. Participants' mean hit and false alarm rate are summarised in Table 9.

Table 9

Mean hit and false alarm rates on the n-back conditions and modalities

	Hit rate				False alarm rate			
	Reader with dyslexia		Typical reader		Reader with dyslexia		Typical reader	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
N-back condition								
Auditory	.92	.11	.94	.10	.41	.26	.38	.22
Visual	.87	.08	.88	.09	.36	.22	.38	.20
Audio-visual	.89	.10	.90	.06	.41	.23	.38	.20
N-back modality								
Within-modal	.89	.10	.91	.08	.38	.22	.38	.19
Cross-modal	.89	.10	.90	.06	.41	.23	.38	.20

There was no statistically significant difference in how typical readers and those with dyslexia performed in the n-back task, as measured by both d' prime and criterion (d' prime: $\beta_{ReaderType}$ = -0.17, *SE* = 0.12, *p* = .168; criterion: $\beta_{ReaderType}$ = 0.03, *SE* = 0.04, *p* = .433). Calculation of d' prime and criterion provided further evidence of an overall inferior performance by all participants in the visual-visual and audio-visual n-back conditions compared to the auditory-auditory condition (d' prime: $\beta_{NbackCondition}$ = -0.21, *SE* = 0.07, *p* = .003; criterion: $\beta_{NbackCondition}$ = 0.13, *SE* = 0.02, *p* < .001). Calculation of d' prime and criterion also demonstrated the absence of such difference across the two different reading groups (d' prime: $\beta_{NbackCondition \, x \, ReaderType}$ = 0.12, *SE* = 0.14, *p* = .419; criterion: $\beta_{NbackCondition \, x \, ReaderType}$ = 0.03, *SE* = 0.05, *p* = .556).

Calculation of d' prime and criterion also demonstrated the absence of statistically significant differences in how typical readers and readers with dyslexia performed in within-modal and cross-modal versions of the n-back task (d' prime: $\beta_{ReaderType \, x \, NbackModality}$ = -0.06, *SE* = 0.13, *p* = .619; criterion: $\beta_{ReaderType \, x \, NbackModality}$ = -0.11, *SE* = 0.09, *p* = .250). Indeed, all participants performed similarly within-modally and cross-modally (d' prime: $\beta_{NbackModality}$ = -0.10, *SE* = 0.26, *p* = .712; criterion: $\beta_{NbackModality}$ = -0.01, *SE* = 0.05, *p* = .842).

Discussion

The ability to create, store, and retrieve cross-modal bindings is crucial for proficient reading. This ability has been shown to be impaired in individuals with atypical reading skills, particularly those with developmental dyslexia, but research to date does not provide a well-controlled comparison of cross-modal and within-modal binding ability in a recognition task. Here, we sought to ascertain the specificity of a cross-modal binding deficit in dyslexia. We assessed visual-visual, auditory-auditory,

and visual-auditory (*i.e.*, cross-modal) working memory in adults with and without developmental dyslexia *via* three different versions of the n-back paradigm. In addition to measuring accuracy rate and response times, we also employed a signal-detection theory framework in our statistical analyses to obtain a fine-grained assessment of the mechanisms involved in encoding and updating novel within-modal and cross-modal bindings.

Contrary to our predictions, we did not find statistically significant evidence of a specific cross-modal binding deficit in adults with developmental dyslexia in the present study. Individuals with and without dyslexia performed similarly in all three versions of the n-back task, a finding that remained consistent whether we looked at the raw accuracy data or signal detection theory performance indices. Indeed, the pattern of hit and false alarm rates was similar for both reading groups across all n-back conditions (*i.e.*, auditory-auditory, visual-visual, and audio-visual) and modalities (*i.e.,* within-modal and cross-modal). This pattern of results is incompatible with previous accounts of a cross-modal binding deficit in readers with an atypical reading profile (Albano et al., 2016; Calabrich et al., 2021a; Garcia et al., 2019; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Toffalini et al., 2019). Our findings provide further support, however, to the claim that individuals with dyslexia might potentially only show a cross-modal binding deficit in tasks that demand a verbal output (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014).

When comparing response times within- and cross-modally, readers with dyslexia were found to be significantly faster at recognising targets and non-targets relative to typical readers. This finding is neither in the direction we expected in our original prediction nor in agreement with previous accounts of slower response times by individuals with an atypical reading profile, particularly to input presented cross-

modally (Calabrich et al., 2021b; M. Jones et al., 2018), and its magnitude is potentially too small to be deemed scientifically relevant. We propose that this finding should, therefore, be considered with caution.

Readers with and without dyslexia were more accurate but slower to respond in the auditory-auditory condition. This pattern of results points toward the emergence of the so-called speed-accuracy trade-off, wherein longer decision-making processes tend to yield higher accuracy rates (Heitz, 2014). Indeed, higher accuracy to the detriment of faster response times has been observed in previous working memory studies using the n-back paradigm (Amon & Bertenthal, 2018; He et al., 2022). Crucially, because participants took significantly longer to respond in the auditory condition, it is possible that the stimuli in that condition might have been kept for longer in the echoic buffer storage relative to how long the visual characters were stored in the visual system's iconic memory, which potentially aided processing of the former (see the Precategorical Acoustic Storage theory; Crowder & Morton, 1969). A similar pattern of results (*i.e.*, superior performance during auditory condition, relative to a visual condition) has also been observed in previous working memory tasks that involve sequential presentation of stimuli (Amon & Bertenthal, 2018), highlighting that, in the general population, distinct stimulus modality presentations will influence encoding and processing mechanisms differently (Amon & Bertenthal, 2018).

As a tentative secondary explanation for the overall higher performance in the auditory-auditory condition for both reading groups, as well as for the lack of evidence for a cross-modal deficit in readers with developmental dyslexia in the present study, it is plausible to consider that the visual characteristics of the Kanji characters used in the visual-visual and cross-modal versions of the n-back task

might have placed too much demand on participants' working memory as a whole. Despite our efforts to ensure that our visual and auditory stimuli were as comparable in complexity as possible, it is plausible to consider that we might have failed to achieve this. The visual characters used here were completely unfamiliar items, as opposed to the pseudowords used in the auditory stimuli which, despite also being novel, closely resemble real words in English. Computational modelling proposes that the underlying brain mechanisms needed for maintaining information that has no prior representation in the brain significantly differ from how working memory stores familiar information (Hasselmo & Stern, 2006). In essence, working memory for novel stimuli appears to require additional cellular mechanisms in the brain (Hasselmo & Stern, 2006).

Alternatively, it may also be possible that this recognition task – requiring a key press response as opposed to verbal production – is insufficiently sensitive to detect reader-group differences (Litt et al., 2019; Litt & Nation, 2014; Poulsen, 2011). A number of previous working memory studies employing the n-back paradigm were able to differentiate individuals with typical reading ability from those with developmental dyslexia (Beneventi et al., 2010a; Bogaerts et al., 2015; Lotfi et al., 2022; J. Wang et al., 2022). Our study, however, showed no behavioural difference between individuals with and without dyslexia (Sela et al., 2012). Our findings could be deemed indicative that recognition memory, in the absence of verbal output demands, potentially remains somewhat intact in dyslexia, at least for adults. However, a firm conclusion as to the manifestation of binding impairments in recognition paradigms is inconclusive, given that recent studies employing recognition tasks were sufficiently sensitive to detect reading group differences (Calabrich et al., 2021).

Despite all of the experimental controls put in place in this study, we cannot rule out the possibility that our task may have lacked the sensitivity to detect cross-modal working memory differences between the two reading groups due to the intrinsically noisier nature of online data collection. Further investigation in a controlled laboratory setting is needed to examine whether the current n-back paradigm can be effectively used to assess a potential link between reading ability and (cross-modal) working memory in adults.

Chapter 6

What the Hand in Motion Reveals about Reading: Children's Decision-Making Processes in Paired Associate Learning and its Relationship to Reading Outcomes

Abstract

A substantial literature documents the relationship between paired associate learning (PAL) and reading ability, yet the cognitive processes underpinning this relationship remains relatively unknown, particularly in younger readers. Here, we used mouse-tracking methodology to measure the degree of cognitive conflict (*i.e.*, indecision) in children's correct responses during a cross-modal, visual-to-phonological PAL task. Our data reveal curvatures in poorer readers' mouse trajectories that are indicative of greater indecision. Even accurate responses by poorer readers therefore belie relatively unstable and imprecise connections between visual and phonological items. Path analyses showed that PAL contributes to reading ability *via* its relationship with rapid, automatic retrieval of over-learned alphanumeric stimuli (*i.e.*, rapid automatised naming). Together, our findings suggest that the relationship between PAL and reading ability is underpinned by the early formation of visual-to-phonological codes. Compromised PAL performance may result in a cascaded effect on longer-term automatisation of print, with implications for word reading accuracy and fluency.

Keywords: Paired Associate Learning, Reading, Mouse Tracking, Rapid Automatised Naming, Path Analyses

What the Hand in Motion Reveals about Reading:

Children's Decision-Making Processes in Paired Associate Learning and its Relationship to Reading Outcomes

Learning to map letters to sounds is the fundamental principle of reading acquisition in alphabetic languages (Ehri, 2000; Harm & Seidenberg, 1999; Seidenberg, 2005; Seidenberg & McClelland, 1989), and children's skill in the beginning stages of mapping letter to letter-sound correspondence is often assessed using Paired Associate Learning (PAL) tasks. A substantial literature now shows that PAL is a significant, unique predictor of reading skill in children (Hulme et al., 2007; Toffalini et al., 2018; Warmington & Hulme, 2012) and in adults (Calabrich et al., 2021b; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013), and that performance on PAL reliably discriminates typically developing readers from poorer readers, such as those with developmental dyslexia (Kalashnikova et al., 2021; Litt & Nation, 2014; Messbauer & de Jong, 2003; Wimmer et al., 2000). Yet, the cognitive processes underpinning PAL – particularly those demarcating differences in good and poorer readers' ability – remains as yet underspecified. Here, we provide – for the first time – a window into children's cognitive processes as they complete a paired associate learning task. Specifically, we make novel use of mouse tracking procedures to assess differences in typical and poorer readers' confidence judgements as they *correctly* recognised newly learned character-to-pseudoword associations. We then also show how *response accuracy* in learning these novel visual-to-phonological associations uniquely contributes to reading outcomes.

Fluent reading in alphabetic languages requires the highly accurate and efficient mapping of letters to their corresponding letter sounds. Incremental learning of these initially arbitrary visual-to-phonological associations is a fundamental step in

the early stages of reading development and can be conceptualised as a type of 'cross modal paired associate learning', in which fluent, automatic representations are created of seemingly arbitrary bindings between visual and phonological stimuli (Litt et al., 2013; Scorpio et al., 2018). This description forms the logic of paired associate learning tasks, which require participants (children or adults) to memorise the association between novel characters (such as abstract shapes) and novel pseudowords (typically consonant-vowel-consonant structures, *e.g., dof*); thus recreating the very beginning stages of visual-to-phonological mapping processes during reading acquisition.

Poorer readers are consistently less accurate at PAL compared to their typical reading peers (Calabrich et al., 2021b; M. Jones et al., 2018; Litt & Nation, 2014; Messbauer & de Jong, 2003), revealing that connecting print to sound is compromised from the very first exposures, and which researchers have variously ascribed to a fundamental difficulty either in connecting visual-to-verbal information or a deficit linked with phonological processing (Clayton et al., 2018; Litt & Nation, 2014; Mayringer & Wimmer, 2000; Messbauer & de Jong, 2003). Recent PAL eye-tracking paradigms show that poorer readers' inefficient learning is underpinned by a dependency on recent, trial-by-trial episodic memories of stimulus exposure in order to accurately recall which character corresponds to which pseudoword. Whilst typical readers, who learn more efficiently, show a learning style indicative of long-range statistics tracking across the entire experiment (Calabrich et al., 2021b; M. Jones et al., 2018).

Experiments designed to dissect the component processes of PAL as characteristic of typical and poorer readers therefore identify several skills – phonological awareness, vision-to-phonology mapping, and statistical learning

characteristics – that are potentially key in mediating the relationship between PAL and reading ability. Converging evidence using a regression-based path analysis approach shows that PAL may influence reading indirectly *via* its relationship with phonology and rapid automatised naming. Indeed, PAL correlates moderately with indices of phonological awareness, and its phonologically-driven influence on reading is subtly different to the influence of phonological awareness skills (Warmington & Hulme, 2012; Windfuhr & Snowling, 2001). A relationship between PAL and rapid automatised naming (RAN) is perhaps more intuitive still, given that RAN reflects rapid access to a well-specified phonological representation bound to another visual element (*i.e.,* core PAL processes, but of highly automatised print such as letters and digits). RAN therefore reflects the speed with which over-learned visual-to-phonological associations are retrieved (Lervåg & Hulme, 2009). Whilst one study has shown little evidence of a RAN-mediated relationship between PAL and reading outcomes in Chinese (Georgiou et al., 2017), this hypothesis has yet to be tested in English. However, PAL is also found to contribute unique variance to reading outcomes when other component skills are controlled (Hulme et al., 2007; Lervåg & Hulme, 2009; Warmington & Hulme, 2012; Windfuhr & Snowling, 2001).

Here, we aim to (a) elucidate typical and poorer readers' online decision-making processing during PAL and to (b) examine the relationship between PAL and reading outcomes *within a single study*, aiming at an integrated picture of PAL cognition and its implication for reading skill. To this end, we used mouse tracking measures to assess primary-aged children's cognitive conflict in judging visual (*i.e.,* character) and phonological (*i.e.,* pseudoword) matches, based on exposure to character-pseudoword pairs in a PAL paradigm: a conservative estimate of PAL ability, given its emphasis on recognition – and phonological input – rather than

explicit output, production processes (Litt et al., 2019; Poulsen, 2011)[8]. On each trial, following a pseudoword probe, children executed a mouse movement, from bottom-centre screen to the correct character from a choice of two, located in top-left and top-right screen positions. In computerised mouse-tracking studies, researchers have used mouse trajectories – recording *x*-, *y*-coordinates as a function of temporality (milliseconds) – to distinguish between *straightforward* (*i.e.*, the initial commitment) and *switched* (*i.e.*, change of mind) decisions taken by participants in the decision-making process (Freeman et al., 2011; Hehman et al., 2015; Kieslich et al., 2020; Maldonado et al., 2019; Yamauchi et al., 2019). A switched decision essentially produces a larger area under the curve, the area between an idealised straight line trajectory and the observed response, and is considered to assess the degree of attraction towards an unselected response.

Thus, in the first part of this study, we examined mouse trajectories on correct trials, consistent with an experimental approach in which we consider whether indecision varies as a function of reading ability. We expected that if poorer readers' accurate responses were nevertheless characterised by less efficient mapping between the phonological representation and the visual character, they would exhibit a larger mouse curvature, indicative of a greater tendency toward a switched decision compared with readers who have a more typical reading profile. Alternatively, if the efficiency of recognising visual-and-phonological mappings is intact once readers reach a certain threshold enabling execution of a correct response, we might expect similar mouse tracking responses between groups (*i.e.*, *straightforward* responses), with slower responses nevertheless characterising

---

[8] We also included a phonological output (production) task in interspersed cued-recall trials during the main experiment and as a post test.

poorer readers. In the second part of this study, we implemented path analysis to examine the relationship between PAL response accuracy[9] and reading outcomes. If PAL exerts an influence on reading *via* the integrity of phonological representation and its mapping to another (here a visual) counterpart, from early exposures through to automatic retrieval, then we expect a RAN-mediated relationship between PAL and reading outcomes. Alternatively, if PAL is primarily linked with reading *via* its association with phonological processes, we expected a phonological awareness-mediated relationship between PAL and reading outcomes.

Open Practices Statement

The data and code for this study are publicly accessible at [tinyurl.com/HandInMotionDataRepository]. Data collection took place as part of a larger language and literacy project launched during the first UK lockdown (09/2020 to 10/2020) to mitigate the effects of the Covid-19 pandemic on children's reading development. As such, we had a narrow window in which to collect data before the end of the school year, and pre-registration and a priori power analyses were not conducted, as would be usual practice (P. Green & Macleod, 2016; Kraemer & Blasey, 2015; Kumle et al., 2021). Instead, we ran post-hoc power analyses (see Supplementary Materials section, Appendix D). Observed power calculations can be useful to determine whether replication studies might yield statistically significant results (Yuan & Maxwell, 2005).

---

[9] A PAL measurement commonly used in previous path analyses cited in the extant literature.

Method

Participants

A total of 88 children were recruited via Bangor University. Technical issues (*e.g.*, poor Internet connection) precluded eight data sets leaving 80 participants aged 7 – 12 years (age: M = 9:98, SD = 1:10; 45 females). All children were native British English speakers, with normal or corrected-to-normal hearing and visual acuity, who participated in a larger language and literacy instruction project (Remote Instruction of Language and Literacy; https://reshare.ukdataservice.ac.uk/855333/). Children verbally agreed to participate, and caregivers gave written informed consent without any compensation, financial or otherwise. Ethical approval was obtained from Bangor University Ethics Committee.

Stimuli

Twelve Kanji characters (*e.g.*, '夂') were arbitrarily matched to an equal number of monosyllabic consonant-vowel-consonant pseudowords (*e.g.*, '*dep*'; See Supplementary Materials, Appendix D). Experimental pseudowords followed English phonotactic constraints, and were generated with Wuggy (Keuleers & Brysbaert, 2010), a multilingual pseudoword generator. Auditory stimuli were recorded by a male native speaker of British English, and character-pseudoword were consistently paired, such that each Kanji character was presented with the same pseudoword on each exposure.

Procedure

Participants completed a 75-min session *via* Microsoft Teams on their personal desktop computers or laptops (all other devices such as smartphones and tablets were excluded). Each session was individually administered by one of eight trained

research assistants, randomly allocated to each child. In each session, children first completed a two-part paired associate learning (PAL) experiment on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), followed by a series of cognitive and literacy measures. The two-part PAL experiment, presented in the same order to all participants, included: (1) a 36-trial recognition task with 12-interspersed cued-recall trials, and (2) a 12-trial cued-recall task. The literacy measures included (1) word reading accuracy, (2) one-minute word and (3) pseudoword reading, (4) phoneme deletion, and rapid automatised naming of (5) letters and (6) digits, administered in the same order to all participants *via* Microsoft Teams' screen-sharing function. Research assistants used Audacity (A. Team, 2021), an open-source digital audio editor, to voice-record the entire testing session. All audio recordings were anonymised and immediately uploaded to each research assistant's channel on Microsoft Teams for subsequent offline scoring. To ensure participants' right to privacy, all voice recordings were immediately removed from the research assistants' personal computers. Nonverbal intelligence quotient (IQ) data was obtained in a different session. However, nonverbal IQ data for 18 participants were unavailable due to technical difficulties (*e.g.*, Microsoft Teams' login issues, poor Internet speed) and/or participants' inability to attend the remote session due to Covid-19 related impracticalities.

Paired Associate Learning Tasks

All tasks were programmed and hosted on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), and the session was conducted remotely by trained research assistants. Participants completed one of three counterbalanced versions of the experiment. In an effort to increase participants' engagement with the task, we embedded gamification elements, including a narrative (Toda et al., 2019), to

introduce an "Intergalactic Language Game" with characters (See

tinyurl.com/PALDemoVideo for a demonstration video).

The narrating character *Zop*'s voice was recorded by a native male speaker of

British English, and set to 80 Hz, formant shift ration 1.2 in Praat (Paul Boersma &

van Heuven, 2001) to allow for a robotic effect. A practice block of four filler trials

included feedback (which was absent in the experiment procedure) and could be

repeated.

Part 1: Recognition and Cued-Recall. Details of the trial sequence and

presentations times are presented in Figure 19A. A colourful robot face served as a

fixation cross (1000 ms) followed by presentation of two Kanji characters, in black on

a white background and occupying 20 x 20 units in Gorilla Experiment Builder's

screen space, presented at the top left and right of the screen. Each character was

highlighted in red (1800 ms; inter-stimulus interval: 250 ms) in a pseudorandomised

order, synchronously with the corresponding pseudoword played. Each item

(character-pseudoword pair) appeared six times over the course of the experiment –

three times as targets and three times as distractors. A blank screen (1000 ms)

ended the encoding phase followed by a visual backward masking phase (2*250 ms)

consisting of cartoon objects (*e.g.*, paintbrush, screwdriver) replacing the target

stimuli. Participants then clicked on a red button at the bottom-centre of the screen

which initiated re-presentation of the characters in the same screen position as in the

encoding phase, synchronously with auditory presentation of one of the

pseudowords presented during the trial. To assist the children in differentiating the

encoding phase from testing, the background colour changed from white (Hex code:

#FFFFFF) to light green (Hex code: #E8FAB9). Participants performed a mouse click

on the character that 'matched' the pseudoword. Manual initiation of the response

phase in each trial (*i.e.*, clicking on the red button at centre-bottom screen position) ensured that the start mouse position was homogenous across participants, allowing a more precise measurement of the mouse trajectory. The trial terminated automatically if no responses were detected within 5000 ms.

Figure 19

Timeline of a single recognition trial (A), a single interspersed cued-recall trial (B), and the post cued-recall task (C)
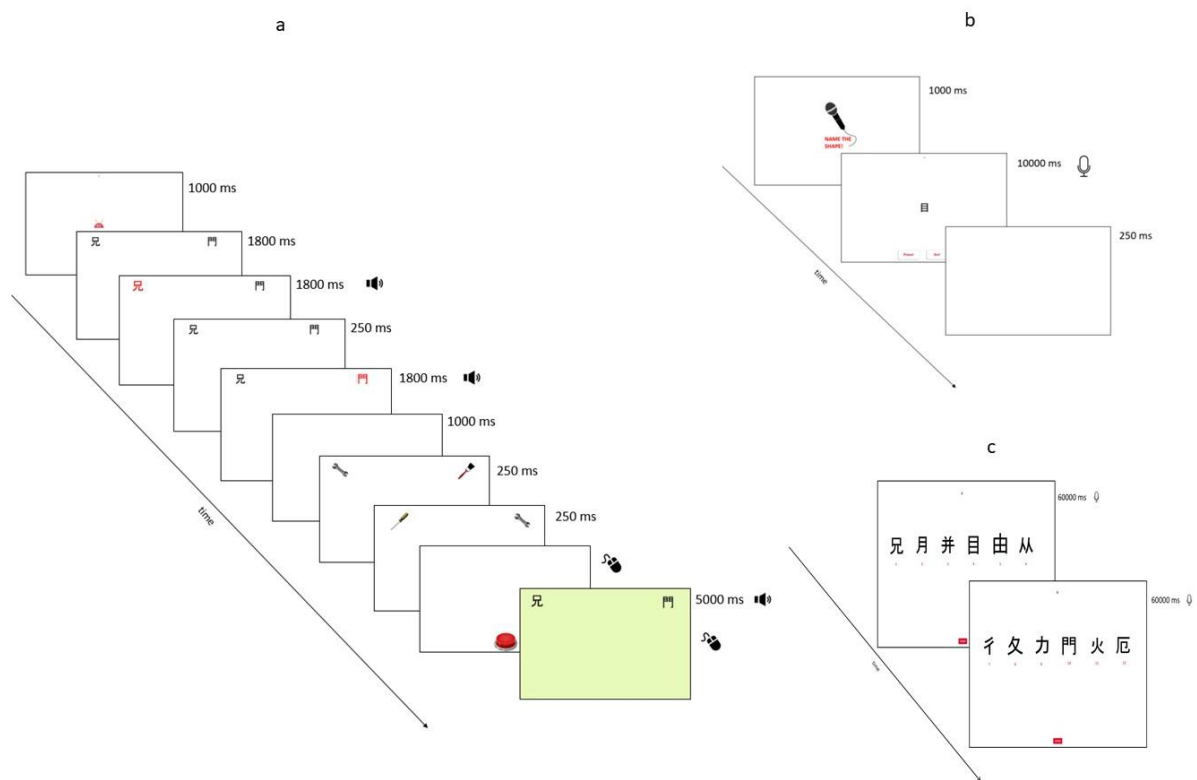
After every three recognition trials, positive feedback (*e.g.*, pictures and sound effects) was provided irrespective of performance, and immediately followed by a centrally presented image of a microphone, along with the instruction to "Name the

shape" (See Figure 19B). In these cued-recall trials, one character from the six previously presented trials appeared in centre-screen position and participants had 10 seconds to produce a verbal response. 'Pass' and 'Go!' buttons enabled skipping or earlier termination of the trial. Each character-pseudoword pair was cued once. Inclusion of interspersed cued-recall trials ensured explicit focus on character-pseudoword pairings and an opportunity for articulation in preparation for the subsequent cued-recall task at the end of the experiment.

A total of 48 experimental trials (36 recognition trials plus 12 interspersed cued-recall trials) were presented over 3 blocks. A break opportunity between blocks included a short animation (approximately 15 seconds long). Presentation order for the recognition trials was randomised across participants within each block. Presentation order for the interspersed cued-recall trials was fixed for all participants within each of the three lists created, allowing research assistants to score participants' responses live. A researcher not involved in the task administration double-scored the responses offline. Cohen's Kappa statistics obtained for this measure showed high interrater reliability (κ = .84)

Part 2: Cued-Recall. The recognition task was immediately followed by a cued-recall task. Six of the twelve Kanji characters were re-presented in black on a white background, displayed in a horizontal line in centre-screen position (See Figure 19C) and participants verbally recalled the pseudoword associated with each character (in no fixed order).[10] Each cued-recall trial terminated automatically after 1 minute or upon a participant-initiated button press, prompting the second trial of six characters.

---

[10] Since administration was remote, a number was shown below each character, which the participant also articulated in order that research assistants could accurately score the trial both online and offline (by second researcher not involved in the administration).

To finish the experiment, a 12-word message (for participants' enjoyment), created with each of the character-pseudoword items learned during the experiment, was presented (See Figure 20).

Figure 20

Message shown to participants in the fictional language Pofflish (A), followed by its 'translation' into English (B)
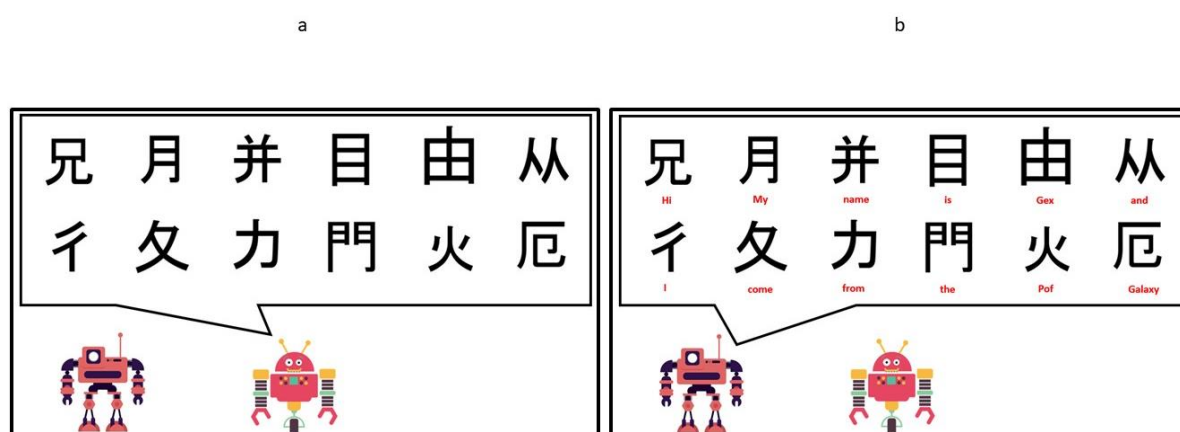
Cognitive measures

Word Reading. The WRAT-4 Word Reading subtest (Wilkinson & Robertson, 2006) assessed children's ability to accurately pronounce alphabet letters as well as words of increasing difficulty levels (*i.e.*, frequent to less frequent words, and transparent to opaque orthography). The test was discontinued after ten consecutive errors ($N = 70$ items), and accuracy was estimated by summing the number of correctly articulated items. This subtest was found to be reliable ($\alpha = .78$).

Word Reading Fluency. The MABEL One-Minute Word Reading Fluency subtest (Caravolas et al., 2018) assessed children's word reading efficiency,

consisting of 144 items ranging from one-letter words to three-syllable words, with a one-minute reading-aloud format. The test-retest reliability correlation for this test was .92. Accuracy was estimated by summing the number of correctly read items within one minute.

Pseudoword Reading Fluency. The MABEL One-Minute Pseudoword Reading Fluency subtest (Caravolas et al., 2018) assessed children's decoding efficiency. It consists of 144 items ranging from one-letter words to three-syllable words, with a one-minute reading-aloud format. The test-retest reliability correlation was .87. Accuracy was estimated by summing the number of correctly read items.

Phonological Awareness. The Phoneme Deletion test (Caravolas et al., 2018) comprised 40 items, with target phonemes in both onset and coda positions, and was used to measure phoneme awareness. This test was found to be highly reliable ($\alpha = .93$). Accuracy was estimated by summing the number of correctly repeated items *and* correctly deleted phonemes (*i.e.*, score of 2), and the number of incorrectly repeated items *but* correctly deleted phonemes (*i.e.*, score of 1).

Rapid Automatised Naming. RAN tests of letters and digits (Caravolas et al., 2018) were used to assess fluent naming of highly familiar alphanumeric stimuli. The intraclass correlation coefficient for the digits version was .94, and the test-retest reliability coefficient for the letters version was .84. Response speed was estimated by summing the naming time for all the items across the grid, per test.

Wide Range Intelligence. The WRIT Matrices subtest (Glutting et al., 2000) was used to assess non-verbal fluid abilities (*i.e.*, visual-spatial reasoning as well as abstract visual-perception relationships) comprising 44 items. This test was highly

reliable (Glutting et al., 2000). Accuracy was estimated by summing the number of correctly identified items.

Data Analyses

Mixed effects analyses were performed in RStudio version 4.0.0 (R. C. Team, 2020). We applied logistic mixed effects regression, *via* the lme4::glmer() function (Bates et al., 2015), in all paired associate learning accuracy data analyses, and linear mixed effects regression, *via* the lme4::lmer() function (Bates et al., 2015), in the response times data analysis. We modelled paired associate learning accuracy and log-transformed response times as a function of reading ability[11]. To provide a robust measure of reading ability, we extracted refined factor scores from a single-factor confirmatory factor analysis of the three reading measures: word reading accuracy, word reading fluency, and pseudoword reading fluency using Mplus 8.1 (Muthén & Muthén, 2018). S*ubject* and *item* were included as random effects in all models (See Supplementary Materials, Appendix D for additional analyses employing maximal random effects structure). Bootstrapped 95% confidence intervals (CIs) were calculated *via* the lme4::confint.merMod() function (Bates et al., 2015) based on 1,000 simulations for each measure. Recognition accuracy entailed the correct identification of the character that corresponded to an auditorily presented pseudoword. Cued-recall errors were defined as any mis-articulations deviating in at least one phoneme from the correct pseudoword.

We used the 'mousetrap' package (Wulff et al., 2021) to filter and analyse mouse trajectories detected in correctly answered PAL recognition trials. To enable

---

[11] To test whether participants' age might have influenced our pattern of results, we also ran an additional set of analyses including 'age' as a predictor in all linear models, which we report in the supplementary materials section, Appendix D.

comparisons of mouse movements across different screen sizes, we used

normalised coordinates, such that coordinate values of -0.5 and 0.5 would always

correspond to the centre of the screen irrespective of monitor size. In order to

optimise visualisation and analyses, we remapped mouse trajectories to ensure

trajectories from correctly answered trials ended on the left side of the screen.

Additionally, we aligned cursor start positions at 0,0 to account for varying start

positions across trials and participants. Finally, we resampled trajectories by

applying time-normalisation to ensure our analyses comprised of an identical number

of chronologically equidistant mouse movement data points (Spivey et al., 2005). We

then applied linear mixed effects regression, *via* the lme4::lmer() function (Bates et

al., 2015) to model three mouse curvature indices, namely maximum absolute

deviation (MAD), average deviation (AD), and area under the curve (AUC), as a

function of reading ability. As in the models described above, subject and item were

included as random intercepts in all three mouse curvature models. Bootstrapped

95% confidence intervals were calculated *via* the lme4::confint.merMod() function

(Bates et al., 2015) based on 1,000 simulations for each curvature measure.

We used the ggcorrmat() function from the 'ggstatsplot' package (Patil, 2021)

to run Pearson correlation analyses, and applied the Holm-Bonferroni method to

adjust for multiple comparisons. We then used Mplus 8.1 (Muthén & Muthén, 2018)

with Full Information Maximum Likelihood (FIML) to run path analyses. We

constructed separate path models with (1) word reading accuracy, (2) word reading

fluency, and (3) pseudoword reading fluency as endogenous variables. In all models,

we included phoneme deletion scores and our two composite measures of paired

associate learning and rapid automatised naming as predictor variables. All variables

of interest were standardised before conducting the path analyses. We used the

following criteria to assess model fit: (a) Chi-Square $p > .05$; (b) Root Mean Square Error of Approximation (RMSEA) $< .06$; (c) Standardised Root Mean Squared Residual (SRMR) $< .08$; (d) Comparative Fit Index (CFI) $> .95$; and (e) Tucker-Lewis Index (TLI) $> .95$ (Hu & Bentler, 1999). To assess potential mediated relationships, we followed Baron & Kenny's (1986) procedure. To test for indirect effects, we bootstrapped confidence intervals with 2000 bootstrap samples.

Results

Cognitive Measures

All cognitive and literacy test results are summarised in Table 10.

Table 10

Means and standard deviations for the experimental task and literacy measures.

	M	*SD*
Non-verbal IQ[a]	102.02	12.42
Paired associate learning[b]	36.47	6.34
Word reading accuracy	45.88	8.95
Word reading fluency[c]	87.69	22.7
Pseudoword reading fluency[c]	51.49	18.12
Phoneme awareness	31.65	6.16
RAN (digits and letters)[d]	89.97	27.78

[a] Standard scores.

[b] Calculation based on the total number of items presented ($N = 60$) across all three PAL tasks.

[c] Calculation based on the total number of items presented ($N = 144$).

[d] Measured in seconds across four trials.

Paired associate learning profiles as a function of reading ability

       Recognition and Cued-Recall. A total of 141 (4.89%) recognition trials timed out and were thus excluded from the accuracy analyses, leaving a remaining 2739 trials. Trials containing inaccurate responses ($N = 253$; 8% of the data) were also excluded from the response times analyses, leaving a remaining 2486 trials. On average, participants correctly recognised 31.08 ($SD = 4.17$) out of 36 visual-phonological associations ($M = .86$, $SD = .12$). Overall, participants took an average of 2263.41 milliseconds to correctly recognise the visual-phonological pairings ($SD = 830.75$).

       We modelled recognition accuracy and log-transformed response time (RT) as a function of reading ability. More skilled readers recognised novel character-pseudoword associations significantly more frequently than poorer readers (odds ratio: 1.54:1, $\beta_{ReadingAbility} = 0.43$ [0.11, 0.78], $SE = 0.16$, $p = .008$). Similarly, skilled readers' correct responses were significantly faster relative to poorer readers ($\beta_{ReadingAbility} = -0.08$ [-0.13, -0.2], $SE = 0.02$, $p = .005$).

       Four of the interspersed cued-recall trials timed out, leaving 956 trials for the analyses. In these trials, participants were shown to correctly assign a verbal label to an average of 3.66 characters ($SD = 2.21$) out of the 12 distinct newly taught visual-phonological associations ($M = .31$, $SD = .18$). Logistic mixed effects regression analyses showed that more skilled readers correctly recalled the novel visual-

phonological associations more frequently than poorer readers (odds ratio: 1.41:1,

$\beta_{ReadingAbility}$ = 0.35 [0.08, 0.62)], SE = 0.13, p = .010).

Mouse Trajectories. Mouse movement data were unavailable for two

participants. Incorrectly answered trials and trials lacking mouse movement data (N

= 407; 14% of the data) were excluded, leaving 2401 trials for the mouse curvature

analyses. All means and standard deviations for the curvature indices are reported in

Table 11.

Table 11

Means and standard deviations for the mouse movement curvature indices.

	M	*SD*
Maximum Absolute Deviation	0.25	0.31
Average Deviation	0.08	0.11
Area Under the Curve	0.11	0.16

As described in the Method, we applied linear mixed effects regression to model

three mouse movement curvature indices as a function of reading ability. The

maximum absolute deviation of the observed trajectory from an idealised direct path

was significantly lower for more skilled readers ($\beta_{ReadingAbility}$ = -0.07 [-0.09, -0.04], SE

= 0.01, p < .001). More skilled readers also showed a significantly lower average

deviation of the observed mouse trajectory and the direct path ($\beta_{ReadingAbility}$ = -0.02 [-

0.03, -0.01], SE = 0.01, p < .001). Similarly, the area under the curve was shown to

be significantly smaller for more skilled readers ($\beta_{ReadingAbility}$ = -0.02 [-0.04, -0.01], SE

= 0.009, p = .017). For illustration purposes, we performed a median split on our

composite measure of reading to categorise participants as poorer and more skilled

readers (See Figure 21).

Figure 21

Depiction of the mouse movement curvature indices investigated in the current study

(A) and mouse movement curvature observed in the current study (B)
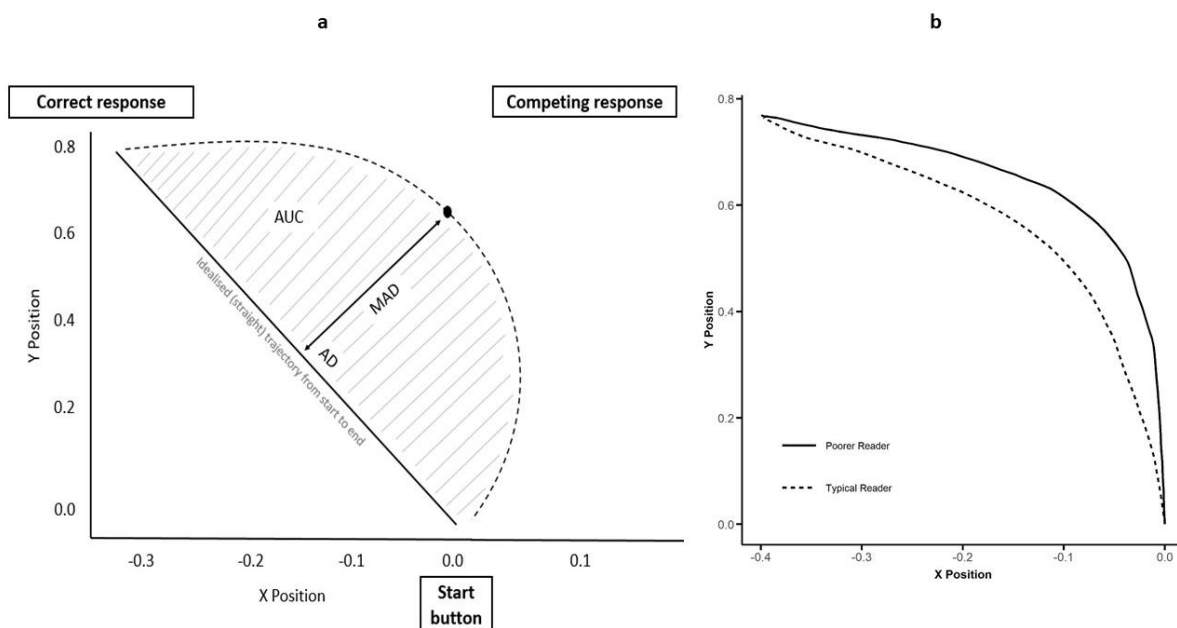
Note. The Maximum Absolute Deviation (MAD) represents the distance between the

furthest point in the observed mouse trajectory (*i.e.*, dashed line) and the line that

represents the shortest possible path (*i.e.*, solid line) connecting the start point and

the correct response option. The Average Deviation (AD) is the mean deviation of

the actual trajectory and the idealised direct path. Finally, the Area Under the Curve

(AUC; striped grey area) is the geometric area between the idealised and the actual

mouse trajectories. For illustration purposes, we performed a median split on our

composite measure of reading to categorise participants as poorer and more skilled

readers.

Cued-Recall. Data for the separate-cued recall task were unavailable for 4 participants due to Internet connection related issues. A total of 912 trials were included in these analyses. In the separate cued-recall task, participants correctly assigned a verbal label to an average of 1.8 characters (*SD* = 1.68) out of the 12 distinct associations taught in the previous task (*M* = .15, *SD* = .14). As in the two measures of paired associate learning reported above, recall of the newly taught visual-phonological associations increased as a function of reading skill (odds ratio: 1.68:1, $\beta_{ReadingAbility}$ = 0.52 [0.15, 0.91], *SE* = 0.19, *p* = .006).

Composite Measure of PAL. Accuracy scores obtained in the paired-associate learning tasks were aggregated in order to create a composite measure of paired associate learning. Data from the four participants who did not have any separate cued-recall data were not included. The PAL composite score (*M* = 0.61, *SD* = 0.11) was then converted to z-scores for inclusion in the path analyses.

The relationship between PAL and reading correlates

Pearson's correlation analyses evaluated patterns of association between paired associate learning and the other correlates of reading. Participants' overall paired associate learning performance, as measured by our experimental paradigm, was found to positively and moderately correlate with several literacy measures, such as word reading accuracy, and word and pseudoword reading fluency (See Table 12). We also observed a moderate but negative correlation with rapid automatised naming speed of digits and letters. We found no significant correlation between paired associate learning and phoneme awareness.

Table 12

Correlation coefficients for the experimental task and literacy measures.

	1	2	3	4	5	6
1. Paired associate learning	*	.39	.28	.35	.08	-.38
2. Word reading accuracy	.003	*	.73	.71	.40	-.55
3. Word reading fluency	.036	<.001	*	.82	.40	-.69
4. Pseudoword reading fluency	.007	<.001	<.001	*	.39	-.63
5. Phoneme awareness	.457	.003	.003	.003	*	-.15
6. Rapid automatised naming speed	.003	<.001	<.001	<.001	.343	*

Note. Pearson correlation (*r*) coefficients displayed to the right of the asterisks, and *p* values displayed to the left

We used path analyses to discern relationships among paired associate learning accuracy and other reading measures. As described in the Method section, we constructed path models containing (1) word reading accuracy, (2) word reading fluency, and (3) pseudoword reading fluency as endogenous variables. We followed an iterative approach to examine the effect of PAL on reading. First, we specified a baseline model that included direct paths from the predictors, namely PAL, phoneme awareness, and rapid automatised naming, to the reading variable. Then, we fit a model including not only direct paths from the three predictor variables but also indirect paths from PAL through phoneme awareness and rapid automatised naming. We dropped all paths that did not approach significance, leaving three simplified path models (*i.e.*, one per endogenous variable).

Word reading accuracy

We assessed the role of (a) paired associate learning, (b) phoneme awareness, and (c) rapid automatised naming as predictors of (1) word reading accuracy. All direct paths in our baseline model reached significance (all ps < .05). We next tested potential mediations by adding paths from PAL to phoneme awareness and RAN. In this model, $\chi^2(1)$ = 1.44, p = .231; RMSEA = .08; SRMR = .05; CFI = .99; and TLI = .95, there was no significant loss of fit, χ^2_{diff} = 1.44, Δ_{df} = 1, p = .230. The path from paired associate learning through phoneme awareness was not significant (p < .05 in all other paths). We, therefore, dropped that path but kept the indirect path from PAL through RAN as it had reached statistical significance. The resulting word reading accuracy path model fit the data well: $\chi^2(1)$ = 1.44, p = .231; RMSEA = .08; SRMR = .05; CFI = 0.99; and TLI = 0.96. Whilst adding this mediation reduced the direct effect of PAL on reading, it did not completely degrade the relationship, suggesting partial mediation. The total standardised indirect effects (PAL → RAN → Reading were small, but significant (.17, p = .010). This model accounted for 43% of the variance in word reading accuracy. Figure 22A shows the final path model for word reading accuracy.

Figure 22

Path models constructed to assess the effects of paired associate learning,

phoneme awareness, and rapid automatised naming on word reading accuracy (A),

word reading fluency (B), and pseudoword reading fluency (C)
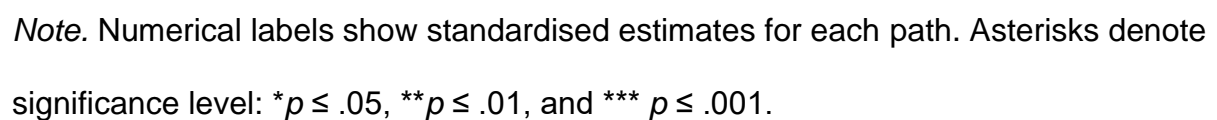
Note. Numerical labels show standardised estimates for each path. Asterisks denote

significance level: *$p \le .05$, **$p \le .01$, and *** $p \le .001$.

Word reading fluency

We assessed the role of (a) paired associate learning, (b) phoneme awareness, and (c) rapid automatised naming, as predictors of (2) word reading fluency. In the baseline model, the direct path from PAL did not reach statistical significance (p = .894; all other ps < .001). Note, though, that when not controlling for phoneme awareness or RAN, there is a significant (p = .009) direct relationship between PAL and reading fluency. In the second model, we therefore only regressed PAL onto phoneme awareness and rapid automatised naming, χ^2(2) = 1.46, p = .483; RMSEA = .00; SRMR =.05; CFI = 1.00; and TLI = 1.00, in which there was no significant loss of fit, χ^2_{diff} = 1.46, Δ_{df} = 2, p = .482, although the path from PAL through phoneme awareness was not significant (p = .460; p < .01 in all other paths). We therefore removed that path before running a third model, in which the only remaining mediator was RAN. The resulting word reading fluency path model fit the data well: χ^2(2) = 1.46, p = .483; RMSEA = .00; SRMR = .05; CFI = 1.00; and TLI = 1.00. Given that there was no significant path from PAL to reading, this model demonstrates complete mediation. The total standardised indirect effects (PAL → RAN → Reading were significant (.25, p = .001). This model accounted for 55% of the variance in word reading fluency. Figure 22B shows the final path model for word reading fluency.

Pseudoword reading fluency

We assessed the role of (a) paired associate learning, (b) phoneme awareness, and (c) rapid automatised naming, as predictors of (3) pseudoword reading fluency. As in the word reading fluency model, the direct path from PAL was not significant in the baseline model (p = .187; all other ps < .001). However, when phoneme awareness or RAN was not controlled, there was a significant (p = .001) direct relationship

between PAL and reading fluency. We next created paths from PAL to phoneme awareness and RAN. In this model, $\chi^2(2) = 3.15$, $p = .207$; RMSEA = .09; SRMR =.05; CFI = 0.98; and TLI = 0.94, there was no significant loss of fit, $\chi^2_{diff} = 3.15$, $\Delta_{df} = 2$, $p = .207$, and the PAL to phoneme awareness path was not significant ($p = .458$; $p < .01$ in all other paths). Thus, this path was dropped from the next, and final, model. The resulting path model, depicted in Figure 22C, fit the data largely well: $\chi^2(2) = 3.15$, $p = .207$; RMSEA =.09; SRMR = .06; CFI = 0.98; and TLI = 0.95. Given that there was no significant path from PAL to reading, this model demonstrates complete mediation. The total standardised indirect effects (PAL → RAN → Reading were significant (.23, $p < .001$). This model accounted for 47% of the variance in pseudoword reading fluency.

Discussion

A substantial body of work shows that paired associate learning ability is related to reading skill (Warmington & Hulme, 2012) and discriminates good and poor readers (Litt & Nation, 2014). But we are yet to arrive at a comprehensive picture of both the cognitive processes underpinning PAL task performance in real time *and* its relationship to reading outcomes, within the same sample of participants. Elucidating these two elements and the overall picture they show can help show how this fundamental skill links with reading development. To date, surprisingly few studies – none of which include child participants – attempt to observe the online cognitive processes that take place during PAL. We first took an experimental approach using mouse tracking, in addition to accuracy and reaction time measures, to provide a window on to the cognitive processes underpinning paired associate learning decision-making processes on *correct* trials, and how this differs as a function of reading ability. Children's mouse trajectories showed that, even when responses

were correct, poorer readers showed evidence of more switched decisions, indicative of lower confidence in their nascent visual-to-phonological representations and compromised paired associate learning efficiency. Next, we examined the relationship between paired associate learning efficiency as measured by *accuracy* and reading outcomes. Path analyses showed that PAL performance accuracy is linked with reading outcomes *via* rapid automatised naming (*i.e.*, visual-phonological representations that are highly automatised). Phoneme awareness did not show such mediating effects. Overall, our findings suggest that the integrity of new visual-to-phonological representations is key to unlocking the relationship between PAL and reading, and that compromised PAL efficiency leads to greater indecision even when the response is accurate.

A detailed look at our findings shows that overall, poorer readers were more inaccurate, and their responses were slower. Poorer readers verbally recalled fewer items compared with more skilled readers on both an interspersed cued-recall task and a post-test cued-recall task. Behavioural measurement therefore showed compromised recognition and recall accuracy in poorer readers, consistent with the extant literature (Litt & Nation, 2014; Messbauer & de Jong, 2003; Wimmer et al., 1998). Observations of readers' online decision-making processes *via* mouse tracking measurements were also taken in order to examine tentativeness and switches in decision in the case of correct responses (Freeman et al., 2011; Hehman et al., 2015; Kieslich et al., 2020; Maldonado et al., 2019). The curvature indices reported here, namely MAD, AD, and AUC, index the magnitude of activation and tentative commitment to each response alternative as the decision-making process unfolds over time (Hehman et al., 2015). On accurate trials, poorer readers showed mouse trajectories indicative of greater indecision between the response options,

suggesting an inefficiency in forming an attachment between the phonological

representation and the correct character. This inefficiency manifested in both

phonological input and output (recognition and recall) tasks, suggesting that poorer

readers have difficulty establishing phonological representations during PAL, as well

as retrieving them (Litt et al., 2019).

We next examined the indirect effects of PAL accuracy on reading measures,

in the context of other reading correlates. Whilst previous studies show that PAL

directly relates to reading ability (Warmington & Hulme, 2012; Windfuhr & Snowling,

2001)[12] our focus is the *indirect* relationships between PAL and reading in order to

elucidate the how learning new visual-verbal connections relates to more

consolidated, reading-related skills. In all models, PAL was indirectly related to

reading skill via RAN and the strength of these relationships was broadly similar

across models, suggesting that the skills underpinning PAL may represent a

precursor to those underpinning RAN (Warmington & Hulme, 2012). Specifically,

efficient formation and access to a well-specified phonological representation, bound

to another – visual – element which, over repeated exposures, leads to increased

specificity of these representations, including the automaticity and precision with

which they are accessed. Notably, our findings showed no evidence that phonemic

awareness mediates the relationship between PAL and reading (Georgiou et al.,

2017; Warmington & Hulme, 2012; Windfuhr & Snowling, 2001). We tentatively

suggest that phonological processing in PAL is important *via* the efficiency with

---

[12] Post-hoc Monte-Carlo simulations (see supplementary materials, Appendix D) revealed that the current models were not well powered enough to detect direct effects of PAL on reading. However, crucially, these simulations did reveal that the models had sufficient power to detect indirect effects between PAL and reading, via RAN.

which a phonological representation is bound to another element; here, its corresponding visual character.

It is important to comment on a few limitations that may have potentially influenced our findings. Firstly, it is possible that uncontrollable variables, such as our participants' fine motor development and visual-motor integration skills, may have affected the results of the mouse trajectories reported here. A future replication of this study should, therefore, seek to provide a fuller picture of participants' motor and cognitive development. Secondly, and relatedly, it is unclear whether the higher indecision shown by poorer readers during paired associate learning of novel visual-phonological associations, as indexed by their mouse trajectories, relate to their difficulty in establishing cross-modal associations or to poor executive functions in general. Finally, due to the data having been collected *via* convenience sampling (*i.e.*, our participants were taking part in a larger language and literacy instruction project, as noted above), the age range of the participants reported here crosses multiple stages of reading (Chall, 1983). We cannot rule out the possibility that paired associate learning skills may be more important for younger readers, who are still developing their basic reading skills, than for older readers, who have already developed these skills. It is important to note, however, that the pattern of results found here remains the same even after controlling for age, as described in the supplementary analyses (See Appendix D).

Here, we sought to elucidate the processes underpinning PAL in typical and poorer child readers, and the relationship of PAL accuracy to reading outcomes. We show that poorer reading is characterised by weaker, less confident representations at the early learning stages, which may compromise longer term automatisation of these visual-phonological representations, with consequences for a number of

reading outcomes, including word level accuracy and fluency. Our findings require

replication with a longitudinal design in order to examine whether the current

characterisation of PAL is valid in the context of longer-term reading outcomes.

Chapter 7

General Discussion

1. Chapter Overview

In this thesis, I focused on some of the cognitive mechanisms that underpin cross-modal binding, a fundamental process to reading development. In four empirical studies, I simulated the process of letter-sound mapping acquisition – a type of cross-modal binding – in order to examine episodic memory and statistical learning, working memory, and decision-making in the context of acquisition of novel visual-phonological associations in readers with and without developmental dyslexia. As outlined in Chapter 1, the primary research questions addressed in this thesis are as follows:

RQ1 How do task-irrelevant episodic details modulate visual-phonological binding performance in individuals with and without dyslexia?

RQ2 Is there a specific cross-modal binding deficit in adults with developmental dyslexia, even in the absence of spoken output demands?

RQ3 Can online measures elucidate Paired Associate Learning mechanisms in beginning readers? And what is the relationship between paired associate learning response accuracy and reading outcomes in these children?

In this chapter, I will summarise how findings from the four empirical studies designed to address these three questions, along with existing body of literature, contribute to our understanding of cross-modal binding in typical and atypical reading. In concluding this chapter, I will offer suggestions for future research directions.

7.1. How do task-irrelevant episodic details modulate visual-phonological binding performance?

Recent work in paired associate learning of novel cross-modal bindings shows that typical readers are sensitive to long-range statistical regularities (*i.e.*, stimuli screen location) detected over multiple exposures to novel visual-phonological associations, a sensitivity that aids their subsequent recall of these associations, whereas readers with dyslexia are not (M. Jones et al., 2018). Here, my goal was to examine the extent to which statistical properties of the stimuli, namely the consistency of their spatial (*i.e.*, screen location) but also their contextual (*i.e.*, item co-occurrences) properties, would modulate the initial formation, and subsequent recall *and* recognition, of novel visual-phonological associations in adults with typical and atypical reading profiles.

Research Question 1 was addressed in two of the four empirical chapters reported in this thesis, namely Chapters 3 and 4. A unique methodological aspect of the studies reported in these two chapters is the use of webcam-based eye-tracking technology. To the best of my knowledge, webcam-based eye-tracking had not been used in previous paired associate learning studies and/or in those employing a looking-at-nothing paradigm, making this thesis a valuable contribution to the field. Below is a summary of the main behavioural and eye-movement findings reported in Chapters 3 and 4.

Behavioural findings: In both Chapters 3 and 4, participants were exposed to novel visual-phonological associations involving Chinese characters and English-like pseudowords, and were then asked to recognise – *via* mouse-clicks, and to recall – *via* verbal responses, these novel associations. Study 1, reported in Chapter 3

involved typical readers only, whereas study 2, reported in Chapter 4, involved readers with developmental dyslexia as well.

In study 1, the consistency of the episodic cues modulated both recognition *and* recall of the bindings for the typical readers. In sum, items that had been presented in both the same screen location and with the same co-occurring items were recognised and recalled more accurately by typical readers than those items which had only one or no consistent feature. This pattern of results is consistent with episodic memory formation in that features that are present at the time of encoding can aid memory retrieval (Tulving, 1972).

In study 2, typical readers consistently and significantly outperformed those with dyslexia in all measures of recall and recognition, in line with the paired associate learning literature (Albano et al., 2016; Hulme et al., 2007; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Toffalini et al., 2019; Warmington & Hulme, 2012). This finding provides further empirical evidence for a *persistent* cross-modal binding deficit in individuals with dyslexia. Availability of multiple episodic cues (*i.e.,* stimulus screen location and intra-trial co-occurrences) aided recall and recognition for individuals with typical and atypical reading profiles. Those with dyslexia, however, were shown to be particularly reliant on consistent contextual information (*i.e.*, consistent item co-occurrences) when prompted to retrieve the items from memory. Findings from study 2, therefore, provide evidence that individuals with dyslexia appear to be able to track statistical regularities in the input when these are highly salient cues that can be beneficial for item learning. However, individuals with dyslexia's long-term reliance on such cues are indicative of fragile memory representations, which might lead this group to a tendency to disproportionately – and thus inefficiently – depend on declarative memory when reading.

Contrary to previous accounts that individuals with dyslexia only show a cross-modal binding deficit in tasks that require a spoken output (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014), here, through our recognition tasks, we showed that individuals with dyslexia do not perform as well as their typical reading counterparts even in the absence of verbal output demands. Indeed, our findings showed that individuals with dyslexia require substantially more repetition than their typical reading counterparts when learning novel cross-modal bindings. This pattern of results is relevant because it suggests that a cross-modal binding deficit might potentially emerge in the very early stages of acquisition of the alphabetic principle, when children with dyslexia are first learning to bind graphemes and phonemes together, and persists well into adulthood, becoming apparent whenever they are required to learn a novel orthography.

Eye-tracking findings: In studies 1 and 2, participants' eye movements were tracked remotely *via* their personal webcams during exposure to the novel visual-phonological associations. The goal was to examine a potential emergence of looking-at-nothing behaviour during memory retrieval of novel visual-phonological associations, as previously demonstrated in a similar paradigm by Jones and colleagues (2018), as well as the extent to which this oculomotor behaviour, if observed, would modulate recognition of the newly learnt items.

In study 1, the looking-at-nothing effect was not shown to modulate memory retrieval of the novel visual-phonological associations. This finding is not in consonance with Jones and colleagues' (2018) study that showed that, for typical readers, fixating a target's now-empty screen location enhances recall of newly learnt visual-phonological associations, a pattern that appeared to emerge for items that had been presented in the same location across different trials. Findings from

study 1 are also not in agreement with previous research that showed that directing our gaze to now-empty location has the potential to aid memory retrieval (Johansson & Johansson, 2020; Kumcu & Thompson, 2020; Richardson & Spivey, 2000). It is important to note, however, that study 1 was a small-scale study meant to be used for pilot purposes. For that reason, it is possible that the small sample size, and, as a result, the lack of statistical power, prevented proper detection of the looking-at-nothing effect. Using the pilot data from study 1 to simulate power, and re-running the paradigm with a larger sample size resulted in the effect being detected in study 2.

In study 2, we observed an interesting pattern of looking-at-nothing behaviour. When participants *did not* direct their gaze towards now-empty regions of interest, accuracy rates were higher for items that had been presented consistently in the same screen location across different trials. It is possible that when the visual-phonological associations were consistently presented in the same screen location, participants encoded this spatial information along with the items, a common occurrence in episodic memory formation (Tulving, 1972), and did not need to perform an oculomotor movement to retrieve those items from memory. Interestingly, participants were increasingly less accurate as they re-fixated empty regions of interest previously occupied by items that had been presented in *in*consistent locations throughout the task. It is likely that this behavioural pattern emerged due to participants' inability to decide where exactly on the screen they should be looking at to retrieve these items from memory. It is important to note here that this decision is likely an unconscious one due to the automatic nature of the looking-at-nothing behaviour (De Groot et al., 2016; Ferreira et al., 2008; Johansson & Johansson, 2020; Richardson & Spivey, 2000).

Despite the emergence of a higher-order interaction in our statistical analyses involving reading ability and looking-at-nothing behaviour, lack of power prevents us from drawing any conclusions from this finding. A replication study with a larger sample size is, therefore, needed to further investigate differences in how typical readers and those with dyslexia engage in looking-at-nothing behaviour when learning visual-phonological items encoded with and without task-irrelevant episodic cues.

Overall, the behavioural differences observed between typical readers and those with dyslexia in our study are consistent with the paired associate learning literature. Our results provide further empirical evidence for a persistent cross-modal binding deficit in developmental dyslexia, even in the absence of phonological output task demands. An interesting finding in our study is that individuals with dyslexia appear to rely heavily on episodic cues, particularly on consistent contextual information, when retrieving visual-phonological associations from memory, a behavioural pattern that is indicative of fragile memory representations. Further replications are needed to properly explore specific differences between individuals with and without dyslexia in how they engage their oculomotor system during acquisition of novel visual-phonological associations.

7.2. Is there a specific cross-modal binding deficit in adults with developmental dyslexia, even in the absence of spoken output demands?

The ability to establish cross-modal associations has been shown to correlate strongly with reading ability, discriminating readers with dyslexia from typical readers

(Hulme et al., 2007; Warmington & Hulme, 2012). This binding deficit, however, has been suggested to emerge only in tasks in which a spoken output is required (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014). Contrary to this claim, findings from study 2, reported in Chapter 4 of this thesis, provide evidence that a cross-modal binding deficit appears to persist in the adult population with dyslexia even in the absence of verbal output demands (Calabrich et al., 2021b). To further explore this issue, I adapted the well-known n-back paradigm and created unimodal and cross-modal conditions of the task. This novel adaptation ensured that task difficulty was equated across conditions (*i.e.,* visual-visual, auditory-auditory, visual-auditory), allowing us to probe within-modal and cross-modal working memory in the absence of verbal output demands, an approach which, to the best of my knowledge, had never been adopted before. Below is a summary of the main findings reported in Chapter 5.

Behavioural findings: In Chapter 5, adults with and without developmental dyslexia were exposed to novel associations involving English-like pseudowords and voices (*i.e.*, auditory-auditory condition), Chinese characters and colours (*i.e.*, visual-visual condition), and English-like pseudowords and Chinese characters (*i.e.*, audio-visual condition). These associations were presented sequentially in a continuous stream, and participants were asked to identify, *via* button presses, whether the association presented in the current trial matches the one presented two trials back, a classic n-back paradigm. In addition to calculating accuracy rate and mean reaction times, a signal detection theory approach was also employed to properly take into account the four different types of errors that can be made in an n-back task (*i.e.,* hits, misses, false alarms, and correct rejections). Results from both statistical approaches converged: adults with and without dyslexia were shown to perform at a

comparable rate in the n-back task. Our findings are at odds with previous n-back paradigms that showed individuals with dyslexia performing significantly poorer than their typically developing reading counterparts (Beneventi et al., 2010b; Bogaerts et al., 2015; Lotfi et al., 2022; J. Wang et al., 2022). It is important to note that our n-back manipulation may not have been sensitive enough to differentiate typical readers from those with dyslexia. This lack of sensitivity is not unsurprising though given that, in the PAL literature, recognition tasks have been shown to not be as sensitive as recall tasks in detecting individual differences in how visual-phonological associations are learned by readers with and without dyslexia (Litt et al., 2019; Poulsen, 2011).

We also failed to replicate findings from Chapter 4 in showing deficient recognition memory for novel cross-modal bindings in adults with dyslexia (Calabrich et al., 2021b). Contrary to our predictions, the pattern of results detected in our study aligns closely with previous accounts that show no cross-modal binding deficit in developmental dyslexia in tasks that do not require a spoken output (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014).

The auditory-auditory condition was shown to be easier than the visual-visual and auditory-visual ones for participants in both reading groups. A tentative explanation for this performance might relate to how the brain is assumed to store familiar and unfamiliar information: the complete lack of familiarity with the visual stimuli used in the visual-visual and visual-auditory condition (*i.e.*, written characters that do not closely resemble any of the letters used in the English language) might have rendered these conditions more challenging for the participants, which, in turn, possibly recruited a comparatively superior number of cellular mechanisms in the brain relative to those recruited to store the less *un*familiar stimuli used in the

auditory-auditory condition (*i.e.*, pseudowords that closely resemble real words in English) (Hasselmo & Stern, 2006).

Surprisingly, readers with dyslexia were shown to be faster at correctly identifying matches and mismatches in the auditory-auditory condition of the n-back task relative to typical readers. This pattern of results, combined with individuals with dyslexia's higher accuracy rates in the same condition, indicates emergence of the so-called speed-accuracy trade-off, wherein longer response times are sometimes associated with higher accuracy rates (Heitz, 2014), a behavioural pattern that is not uncommon in n-back studies (He et al., 2022).

In sum, we found that, in the n-back task, individuals with dyslexia performed as well as their typically developing counterparts, inconsistent with previous research that showed poorer working memory performance in this group. A cross-modal binding deficit was, therefore, not behaviourally detectable through our manipulation, which may potentially have been due to insufficient task-sensitivity (see Poulsen, 2011).

7.3. Can online measures elucidate Paired Associate Learning mechanisms in beginning readers? And what is the relationship between paired associate learning response accuracy and reading outcomes in these children?

Paired associate learning ability is consistently shown to correlate with reading proficiency (Albano et al., 2016; Calabrich et al., 2021b; M. Jones et al., 2018; Litt et al., 2013; Litt & Nation, 2014; Toffalini et al., 2018, 2019; Warmington & Hulme, 2012) However, whilst performance in PAL tasks reliably discriminates skilled and

poorer readers (Litt & Nation, 2014), there is not enough evidence on how the ability to learn novel visual-phonological associations relates to more consolidated reading-related skills, such as rapid automatised naming. In Chapter 6, we examined a potential relationship between PAL response accuracy and three reading outcomes, namely word reading accuracy, word reading fluency, and pseudoword reading fluency. Specifically, we looked at whether PAL would exert an indirect influence on these reading outcomes *via* rapid automatised naming ability and/or *via* phonological awareness. Innovatively, we employed mouse-tracking technology to investigate the potential emergence of cognitive conflict (*i.e.*, indecision) during paired associate learning of novel cross-modal bindings, and the extent to which said conflict would vary as a function of reading ability. Below is a summary of the main behavioural and mouse trajectory findings reported in Chapter 6.

Behavioural findings: In Chapter 6, an unselected group of school-aged children were exposed to novel cross-modal bindings involving Chinese characters and English-like pseudowords. In a gamified remotely administered task, these children were prompted to recognise, *via* mouse-clicks, and recall, *via* verbal responses, the novel visual-phonological associations.

Previous research shows that children and adults with less proficient reading abilities, such as those with developmental dyslexia, exhibit deficient cross-modal binding skills (Calabrich et al., 2021b; Clayton et al., 2018; Garcia et al., 2019; Hulme et al., 2007; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Litt et al., 2013, 2019; Litt & Nation, 2014; Toffalini et al., 2019). Here, we also demonstrated that school-aged children with poorer reading skills do not perform as well as more skilled readers in tasks that involve either recall or recognition of novel cross-modal bindings. Overall, our pattern of results further highlights a link between inefficient

paired associate learning ability and low reading proficiency (Calabrich et al., 2021b; Garcia et al., 2019; Hulme et al., 2007; M. Jones, Branigan, et al., 2013; Litt et al., 2013; Litt & Nation, 2014; Toffalini et al., 2018, 2019).

Our experimental paradigm also allowed us to detect a RAN-mediated relationship – but no phonological awareness-mediated relationship – between paired associate learning and our three measures of reading outcomes. This pattern of results suggests that the accurate formation of visual-phonological associations in the early stages of literacy acquisition (*i.e.*, paired association learning ability) may, over repeated exposures, lead to automatic and precise retrieval of these associations (*i.e.*, rapid automatised naming ability) (Kirby et al., 2010; Warmington & Hulme, 2012), ultimately automatising the act of reading.

Mouse-tracking findings: In a remotely administered gamified task, participants' mouse trajectories were recorded during paired associate learning of novel visual-phonological associations. Participants were instructed to indicate, *via* a mouse-click, the visual characters that corresponded to auditorily presented English-like pseudowords. Mouse trajectories associated with their accurate responses were then analysed to examine, for the first time, decision-making processes during paired associate learning.

Overall, poorer readers showed a greater tendency towards switching decisions during recognition of novel visual-phonological associations. Their mouse trajectories indicated that, in accurate trials, the incorrect response option was momentarily considered significantly more frequently than for children with higher reading skills. This pattern of results highlights, for the first time, poorer readers' lower confidence levels when engaged in paired associate learning of novel cross-modal bindings. Employing a mouse-tracking approach in this study had a major

advantage over looking at accuracy rates and response time measures alone. Specifically, because mouse trajectories provide a continuous record of the decision-making process, here we were able to gain insights into how our participants weighed up different alternative options before making a final decision.

In sum, our results highlight a connection between inefficient – and hesitant – paired associate learning ability and lower reading proficiency. Our findings also indicated that rapid automatised naming ability, but not phonological awareness, mediates the relationship between paired associate learning and reading outcomes.

7.4. Theoretical Implications

The findings of the studies discussed in this thesis hold significant theoretical implications for our understanding of typical and atypical reading. Overall, the behavioural differences observed between typical readers and those with atypical reading profiles are consistent with the paired associate learning literature. Specifically, our results provide further empirical evidence for a persistent cross-modal binding deficit in developmental dyslexia (Aravena et al., 2013; Blomert, 2011; Fraga González et al., 2015; Garcia et al., 2019; Hulme et al., 2007; M. Jones et al., 2018; M. Jones, Branigan, et al., 2013; Litt et al., 2019; Toffalini et al., 2019, 2018; H. C. Wang et al., 2017; Žarić et al., 2014, 2015), even in the absence of phonological output task demands in some cases, challenging previous findings in the field (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014).

In our webcam-based eye-tracking study, individuals with dyslexia were shown to (over-)rely on episodic cues, a finding that further supports the idea that features present during encoding can assist in memory retrieval (Tulving, 1972). This finding also resonates with the claim that individuals with neurodevelopmental

disorders, such as those with developmental dyslexia, might rely more on compensatory strategies during memory retrieval (Ullman & Pullman, 2015). Crucially, we demonstrate that individuals with dyslexia may be able to track statistical regularities in the input when these are highly salient cues that can be beneficial for item learning, as previously observed by Jones and colleagues (2018). However, individuals with dyslexia's long-term reliance on such cues are indicative of fragile memory representations, which might lead this group to a tendency to disproportionately depend on declarative memory when reading (Ullman & Pullman, 2015). Further research is needed to better understand the relationship between statistical learning and declarative memory in individuals with dyslexia.

The results of our webcam-based eye-tracking study challenge the notion that cross-modal binding deficits in individuals with dyslexia are restricted to tasks demanding a spoken output (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014). Instead, here we show that these deficits may persist even in the absence of verbal output demands, shedding new light on the nature of cross-modal binding difficulties in dyslexia. These observations indicate that the binding deficits observed in individuals with dyslexia may not be strictly tied to phonological processes, raising intriguing questions about the potential root causes of dyslexia, as well as the specificity of the phonological deficits in relation to cross-modal binding tasks (Frith, 1997; Snowling et al., 2020; Vellutino et al., 2004). It is important to note, however, that it is unlikely that there will ever be a single "root cause" of dyslexia. Our findings suggest that cross-modal binding deficits are just one piece of the puzzle that contributes to the complex nature of developmental dyslexia. It is important to note that, while our findings contribute to a deeper understanding of paired associate

learning impairments in individuals with developmental dyslexia, having used a cross-sectional approach inherently limits our ability to establish causality.

Here, we also explored the relationship between paired associate learning and reading outcomes. Contrary to traditional beliefs that PAL primarily relies on phonological awareness (Clayton et al., 2018; Litt et al., 2019), our results indicated a stronger association between PAL and rapid automatised naming as mediators to reading outcome measures. This finding challenges existing models and theories that have historically prioritised the role of phonology in the development of reading skills (Frith, 1997; Snowling et al., 2020; Vellutino et al., 2004) and highlights the growing recognition of rapid automatised naming as another critical skill in reading development (Georgiou et al., 2005; Kirby et al., 2010; Landerl & Wimmer, 2000; Wimmer, 1993; Wimmer et al., 1998, 2000; Zugarramurdi et al., 2022)

An important aspect to consider when interpreting the findings of this thesis is the inter-individual variability in the severity of dyslexia and reading difficulties. Crucially, variations in dyslexia severity may relate to distinct patterns of paired associate learning, rapid automatised naming, phonological awareness, or a combination of these factors. It is, therefore, of the utmost importance that future studies consider the heterogeneity within the dyslexic population and its implications for reading development.

7.5. Directions for Future Research

In this section, I will outline potential paths for future research on paired associate learning in individuals with typical and atypical reading profiles. These paths will

consider the limitations that were identified in the current thesis as well as outstanding research questions that could not have been addressed here.

While efforts were made to ensure high data quality, as thoroughly described above, it is important to note that data collected in online settings may be inherently noisier due to a range of different factors such as distractions, variable environmental conditions (*i.e.*, lighting, computer position), and technical difficulties (Finley & Penningroth, 2015). It is, therefore, crucial that the findings reported in this thesis be replicated in a controlled laboratory setting in order to confirm their validity. This is particularly important considering that the field of experimental psychology is currently facing a replicability crisis, where a high number of published studies and classic findings do not appear to be reproducible (Ioannidis, 2005).

In Chapter 4, we showed a persistent cross-modal binding deficit in developmental dyslexia, even in the absence of phonological output task demands. When attempting to retrieve visual-phonological associations from memory, the individuals with dyslexia in our study were shown to rely on episodic cues, particularly item co-occurrences. An interesting follow-up study would be to investigate the extent to which sleep promotes the acquisition of cross-modal bindings in individuals with and without dyslexia. Additionally, examining the reliance on episodic cues by individuals with dyslexia following a night of sleep would provide valuable insights into their memory consolidation processes. Integrating polysomnography sensors into the study design would allow us to control for potential confounding factors, such as individual differences in sleep duration and sleep quality. This research direction will allow us to examine how episodic cues are integrated into memory for those with dyslexia and whether sleep plays a compensatory role in this process.

Another potentially intriguing avenue for future research as a follow-up for the study reported in Chapter 4 is exploring whether typical readers, when receiving lower levels of exposure to the novel visual-phonological associations, would exhibit a similar learning pattern to those observed in readers with dyslexia. Employing a learning to criterion approach will allow us to investigate whether individuals with dyslexia and those with a typical reading profile differ in their *rate of learning* or in their *ability to learn* novel visual-phonological associations. If typical readers do not behave like those with dyslexia after lower levels of exposure, this would be indicative of a qualitative rather than a quantitative difference between these two reading profiles. In essence, a *quantitative* difference would be a difference in the amount of exposure required to learn the novel visual-phonological associations, whereas a *qualitative* difference would relate to the type of cognitive processing required to learn the novel associations. Findings from this type of investigation could have important implications for the development of targeted interventions for individuals with atypical reading profiles, such as those with dyslexia.

One question that arose in Chapter 5 of this thesis is whether our n-back manipulation was indeed not sensitive enough to detect differences between typical readers and those with dyslexia, or whether our results actually provide further support for the claim that a cross-modal binding difficulty is only apparent in the presence of spoken output demands (Clayton et al., 2018; Litt et al., 2019; Litt & Nation, 2014). Here, I propose three follow-up studies to further examine and expand our findings: 1) Given that participants from both reading groups scored equally well in all 2-back versions of the task, possibly due to the experimental task not being challenging enough, a replication study with a higher memory load (*e.g.,* 3-back) is needed. This will potentially maximise the chances that the n-back task will be

sufficiently sensitive to potential group differences in cross-modal and within-modal working memory performance; 2) Surprisingly, participants scored higher in the auditory-auditory version of the n-back task relative to the visual-visual and auditory-visual conditions. It is possible – but not entirely clear – that employing Chinese characters in the n-back task may have added an unexpected layer of complexity relative to the more familiar auditory stimuli. It is important to note that while Chinese characters were also used in all other studies reported in this thesis, the main difference here is that, in an n-back task, the items are presented in a continuous stream, which may have potentially affected how the encoding of the novel stimuli took place. A replication study that uses strictly controlled sets of artificial characters, such as the Brussels Artificial Character Sets (Vidal et al., 2017), could be attempted to examine whether our choice of stimuli introduced a confounding variable to our n-back manipulation. This suggestion should, however, be treated with caution, and use of real-world letters in experiments that simulate letter learning should be preferred whenever possible since letter shapes have been suggested to not be an entirely arbitrary cultural decision (Dehaene, 2009); 3) While our manipulation was potentially not sensitive enough to detect working memory differences between *adults* with and without developmental dyslexia, it might be worth testing our n-back paradigm with *children* at early stages of reading development.

Another question that arose relates to the findings from our school-aged children study. As aforementioned, the age range of the participants reported in Chapter 6 crosses multiple stages of reading (Chall, 1983). While the pattern of results that we found remained the same even after controlling for age (*i.e.*, more hesitant decision-making by poorer readers during acquisition of novel visual-phonological associations), it would be fruitful to attempt to replicate this study using

separate age groups – which we were unable to do here due to our small sample size – or a longitudinal cohort. In a future replication, I would also suggest including measures of fine motor development and visual-motor integration skills, as well as executive function assessments, to allow an adequate evaluation of the decision-making process related findings, as indexed by the mouse trajectories, and how they relate to paired associate learning ability.

The cross-sectional approach taken in this thesis limited our ability to examine whether deficient paired associate learning is likely a symptom or cause of dyslexia. To address this causality issue, a longitudinal design would be preferred, as it would more objectively help us disentangle whether PAL impairments are secondary outcomes of the reading difficulties experienced by those with dyslexia, or a causal factor. Assessing PAL performance in pre-readers would allow us to investigate the extent to which PAL abilities predict individual differences in later reading skill development. Crucially, an early intervention study in which we track pre-readers' progress would allow us to further our understanding of whether enhancing PAL abilities can prevent or mitigate the onset of reading difficulties.

Last but not least, any follow-up studies stemming from this thesis should also consider taking a transdiagnostic approach to participant recruitment. In this thesis, dyslexia was predominantly treated as a separate and distinct disorder (*e.g.*, Chapters 4 and 5), as has traditionally been done in most – if not all – paired associate learning studies in the field of reading research. It is crucial, however, to acknowledge the limitations of this approach. By dichotomising participants into readers with and without dyslexia, we might inadvertently oversimplify the complex nature of this condition and of reading difficulties in general. In chapter 6 of this thesis, we made a deliberate shift toward treating reading as a continuous variable,

recognising that reading ability may exist on a continuum within the general

population (Snowling et al., 2020). Reading impairments are reportedly not separate

and distinct from typical reading abilities, but rather exist on a continuum within the

normal distribution with no clear cutoff point that separates individuals with

impairments from those without (Harlaar, Spinath, Dale, & Plomin (2005). Whilst

looking at reading as a continuous variable has the potential to allow us to capture

the full spectrum of reading abilities, it is important to note that it requires

considerably larger and more diverse participant samples to account for the inherent

variability in reading skills in the general population. Increasingly, reading

researchers have recently started moving beyond dichotomising participants into

those with and without developmental dyslexia, and have started recognising that the

difficulties experienced by those that generally receive a dyslexia diagnosis may be

related to not only the reading domain, but to more general cognitive and neural

mechanisms as well (Astle et al., 2022; S. D. Jones et al., 2023; Snowling et al.,

2020; Snowling & Hulme, 2021; Thomas, 2020). Considering that the binary

approach to dyslexia has traditionally provided a clear demarcation for research and

educational purposes, we believe that integrating both perspectives, whenever

possible, and thus considering dyslexia as a point on a continuum within a

transdiagnostic framework may be a promising way forward. By recognising that

reading difficulties may be related to more general cognitive and neural mechanisms,

we can contribute to the development of comprehensive theories that encompass a

wide range of reading profiles, thereby influencing the design of reading

interventions that are tailored to individual needs.

By addressing the limitations of the current research and exploring our

outstanding questions, the proposed studies can provide valuable insights into the

nature of cross-modal binding deficits, the role of episodic cues in memory retrieval, and the relationship between paired associate learning, rapid automatised naming, and phonological awareness. Crucially, the proposed studies have the potential to advance existing theories, refine models, and provide valuable insights into the cognitive processes underlying reading development and developmental dyslexia.

7.6. Final Summary

My overall aim in this thesis was to examine paired associate learning of novel visual-phonological associations in individuals with typical and atypical learning profiles. This overarching aim was addressed *via* three research questions: 1) How do task-irrelevant episodic details modulate visual-phonological binding performance in individuals with and without dyslexia?, 2) Is there a specific cross-modal binding deficit in adults with developmental dyslexia, even in the absence of spoken output demands?, 3) Can online measures elucidate Paired Associate Learning mechanisms in beginning readers? And what is the relationship between paired associate learning response accuracy and reading outcomes in these children?

Over four empirical chapters, I established that: 1) when retrieving newly learnt visual-phonological associations from memory, individuals with dyslexia appear to rely heavily on episodic cues, particularly on consistent contextual information, a behavioural pattern that suggests that this group may have more fragile memory representations than their typical reading counterparts; 2) the evidence for a persistent cross-modal binding deficit in adult developmental dyslexia in the absence of phonological output remains somewhat unclear. Different experimental paradigms employed in this thesis yielded different results, thus

highlighting the need for further investigation on this issue; 3) the relationship between paired associate learning and reading appears to be mediated by rapid automatised naming, an ability that involves cross-modal associative mechanisms that are similar to those employed in PAL, albeit with overlearned stimuli; 4) even when accurate, individuals at the lower end of the reading spectrum are less confident about their knowledge of newly learnt visual-phonological associations.

In concluding this chapter, I outlined potential future research avenues to address the limitations and outstanding questions from this thesis.

References

References

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences*, *11*(11). https://doi.org/10.1016/j.tics.2007.08.015

Albano, D., Garcia, R. B., & Cornoldi, C. (2016). Deficits in working memory visual-phonological binding in children with dyslexia. *Psychology and Neuroscience*, *9*(4). https://doi.org/10.1037/pne0000066

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development*, *77*(6), 1698–1716.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm." *Cognition*, *93*(2). https://doi.org/10.1016/j.cognition.2004.02.005

Altmann, G. T. M. (2017). Abstraction and generalization in statistical learning: Implications for the relationship between semantic types and episodic tokens. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0060

Amon, M. J., & Bertenthal, B. I. (2018). Auditory Versus Visual Stimulus Effects on Cognitive Performance During the N-back Task. *CogSci*.

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, *53*, 1407–1425.

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1). https://doi.org/10.3758/s13428-019-01237-x

Aravena, S., Snellings, P., Tijms, J., & van der Molen, M. W. (2013). A lab-controlled simulation of a letter-speech sound binding deficit in dyslexia. *Journal of Experimental Child Psychology*, *115*(4). https://doi.org/10.1016/j.jecp.2013.03.009

Aravena, S., Tijms, J., Snellings, P., & van der Molen, M. W. (2018). Predicting Individual Differences in Reading and Spelling Skill With Artificial Script–Based Letter–Speech Sound Training. *Journal of Learning Disabilities*, *51*(6). https://doi.org/10.1177/0022219417715407

Asbjørnsen, A. E., Jones, L. Ø., Eikeland, O. J., & Manger, T. (2021). Can a questionnaire be useful for assessing reading skills in adults? Experiences with the Adult Reading Questionnaire among incarcerated and young adults in Norway. *Education Sciences*, *11*(4), 154.

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry*, *63*(4), 397–417.

Baddeley, A. (1986). *Working memory (vol. 11).* Oxford: Oxford University Press, Clarendon Press.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A., Allen, R., & Hitch, G. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*, *49*(6), 1393–1400.

Baddeley, A., Allen, R. J., & Hitch, G. (2010). Investigating the episodic buffer. *Psychologica Belgica*, *50*(3), 223–243.

Baddeley, A., & Della Sala, S. (1996). Working memory and executive control.

Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, *351*(1346), 1397–1404.

Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive Psychology*, *4*(2), 207–228.

Baron, R. M., & Kenny, D. A. (1986). The moderator--mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

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

Beneventi, H., Tønnessen, F. E., Ersland, L., & Hugdahl, K. (2010a). Executive working memory processes in dyslexia: behavioral and fMRI evidence. *Scandinavian Journal of Psychology*, *51*(3), 192–202.

Beneventi, H., Tønnessen, F. E., Ersland, L., & Hugdahl, K. (2010b). Working memory deficit in dyslexia: behavioral and FMRI evidence. *International Journal of Neuroscience*, *120*(1), 51–59.

Berninger, V. W., Abbott, R. D., Thomson, J., Wagner, R., Swanson, H. L., Wijsman, E. M., & Raskind, W. (2006). Modeling phonological core deficits within a working memory architecture in children and adults with developmental dyslexia. *Scientific Studies of Reading*, *10*(2), 165–198.

Blau, V., Reithler, J., Atteveldt, N. Van, Seitz, J., Gerretsen, P., Goebel, R., &

Blomert, L. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: A functional magnetic resonance imaging study of dyslexic children. *Brain*, *133*(3). https://doi.org/10.1093/brain/awp308

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced Neural Integration of Letters and Speech Sounds Links Phonological and Reading Deficits in Adult Dyslexia. *Current Biology*, *19*(6). https://doi.org/10.1016/j.cub.2009.01.065

Blomert, L. (2011). The neural signature of orthographic-phonological binding in successful and failing reading development. *NeuroImage*, *57*(3). https://doi.org/10.1016/j.neuroimage.2010.11.003

Blythe, H. I., Dickins, J. H., Kennedy, C. R., & Liversedge, S. P. (2020). The role of phonology in lexical access in teenagers with a history of dyslexia. *PloS One*, *15*(3), e0229934.

Boersma, P, & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.50. In *retrieved 23 March 2020*. www.praat.org

Boersma, Paul, & van Heuven, V. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, *5*(9–10).

Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P. A., Woumans, E., & Duyck, W. (2015). Increased susceptibility to proactive interference in adults with dyslexia? *Memory*, *23*(2), 268–277.

Bott, N. T., Lange, A., Rentz, D., Buffalo, E., Clopton, P., & Zola, S. (2017). Web camera based eye tracking to assess visual memory on a visual paired comparison task. *Frontiers in Neuroscience*, *11*(JUN). https://doi.org/10.3389/fnins.2017.00370

Brockmole, J. R., & Franconeri, S. L. (2009). Introduction to the special issue on binding. *Visual Cognition*, *17*, 1–9.

Bundt, C., Ruitenberg, M. F. L., Abrahamse, E. L., & Notebaert, W. (2018). Early and late indications of item-specific control in a Stroop mouse tracking study. *PloS One*, *13*(5), e0197278.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5). https://doi.org/10.1038/nrn3475

Calabrich, S. L., Oppenheim, G., & Jones, M. (2021a). Episodic memory cues in the acquisition of novel visual-phonological associations: a webcam-based eyetracking study. *43rd Annual Conference of the Cognitive Science Society: Comparative Cognition: Animal Minds*.

Calabrich, S. L., Oppenheim, G. M., & Jones, M. (2021b). Audiovisual Learning in Dyslexic and Typical Adults: Modulating Influences of Location and Context Consistency. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.754610

Calabrich, S. L., Oppenheim, G. M., & Jones, M. (2021c). Episodic memory cues in acquisition of novel visual-phonological associations: a webcam-based eye-tracking study. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

Caldwell, B., Cooper, M., Reid, L. G., Vanderheiden, G., Chisholm, W., Slatin, J., & White, J. (2008). Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)*, *290*, 1–34.

Caravolas, M., Mikulajová, M., Defior, S., & Seidlová Málková, G. (2018). *MABEL. Tests*. https://www.eldel-mabel.net/test/

Carroll, J. M., Solity, J., & Shapiro, L. R. (2016). Predicting dyslexia using prereading skills: The role of sensorimotor and cognitive abilities. *Journal of Child Psychology and Psychiatry*, *57*(6), 750–758.

Chall, J. S. (1983). *Stages of reading development*.

Chang, Y.-N., Taylor, J. S. H., Rastle, K., & Monaghan, P. (2020). The relationships between oral language and reading instruction: Evidence from a computational model of reading. *Cognitive Psychology*, *123*, 101336.

Chiappe, P., Siegel, L. S., & Hasher, L. (2000). Working memory, inhibitory control, and reading disability. *Memory \& Cognition*, *28*(1), 8–17.

Clayton, F. J., Sears, C., Davis, A., & Hulme, C. (2018). Verbal task demands are key in explaining the relationship between paired-associate learning and reading ability. *Journal of Experimental Child Psychology*, *171*, 46–54.

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.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception \& Psychophysics*, *5*(6), 365–373.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, *10*(3), 277–299.

De Groot, F., Huettig, F., & Olivers, C. N. L. (2016). Revisiting the looking at nothing phenomenon: Visual and semantic biases in memory search. *Visual Cognition*, *24*(3), 226–245.

Dehaene, S. (2009). Reading in the brain. *New York*.

Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized'naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471–479.

Di Pietro, S. V, Karipidis, I. I., Pleisch, G., & Brem, S. (2023). Neurodevelopmental trajectories of letter and speech sound processing from preschool to the end of elementary school. *Developmental Cognitive Neuroscience*, 101255.

Egan, C., Payne, J. S., & Jones, M. W. (2023). The impact of phonological relatedness on semantic congruency judgements in readers with dyslexia: Evidence from behavioural judgements, event related potentials and pupillometry. *Neuropsychologia*, 108548.

Egner, S., Reimann, S., Hoeger, R., & Zangemeister, W. H. (2018). Attention and information acquisition: Comparison of mouse-click with eye-movement attention tracking. *Journal of Eye Movement Research*, *11*(6).

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading*, *18*(2). https://doi.org/10.1111/j.1467-9817.1995.tb00077.x

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders.*

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2). https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, *18*(1), 5–21.

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing*, *7*(3). https://doi.org/10.1007/BF03162082

El-Kalliny, M. M., Wittig, J. H., Sheehan, T. C., Sreekumar, V., Inati, S. K., & Zaghloul, K. A. (2019). Changing temporal context in human temporal lobe promotes memory of distinct episodes. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-018-08189-4

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11). https://doi.org/10.1016/j.tics.2008.07.007

Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Frontiers in Psychology*, *3*, 335.

Finley, A., & Penningroth, S. (2015). Online versus in-lab: Pros and cons of an online prospective memory experiment. *Advances in Psychology Research*, *113*, 135–162.

Fraga González, G., Žarić, G., Tijms, J., Bonte, M., Blomert, L., & van der Molen, M. W. (2015). A randomized controlled trial on the beneficial effects of training letter-speech sound integration on reading fluency in children with dyslexia. *PloS One*, *10*(12), e0143914.

Francks, C., MacPhie, I. L., & Monaco, A. P. (2002). The genetic basis of dyslexia. *The Lancet Neurology*, *1*(8), 483–490.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, *42*(1), 226–241.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*, 59.

Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: Split-second social perception. *Trends in Cognitive Sciences*, *20*(5), 362–374.

Frith, U. (1997). Brain, mind and bahaviour in dyslexia. *Dyslexia, Biology, Cognition and Intervention*.

Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2009). The long road to automation: neurocognitive development of letter--speech sound processing. *Journal of Cognitive Neuroscience*, *21*(3), 567–580.

Garcia, R. B., Tomaino, A., & Cornoldi, C. (2019). Cross-modal working memory binding and learning of visual-phonological associations in children with reading difficulties. *Child Neuropsychology*, *25*(8). https://doi.org/10.1080/09297049.2019.1572729

Georgiou, G., Liu, C., & Xu, S. (2017). Examining the direct and indirect effects of visual--verbal paired associate learning on Chinese word reading. *Journal of Experimental Child Psychology*, *160*, 81–91.

Georgiou, G., Parrila, R., & Papadopoulos, T. C. (2005). Predictors of word reading accuracy and fluency in English and Greek: A cross-linguistic comparison. *11th Biennial Conference of the European Association for Research on Learning and Instruction, Nicosia, Cyprus*.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: a

component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 245.

Ghetti, S., & Bunge, S. A. (2012). Neural changes underlying the development of episodic memory during middle childhood. *Developmental Cognitive Neuroscience*, *2*(4), 381–395.

Glutting, J., Adams, W., & Sheslow, D. (2000). *Wide range intelligence test: WRIT*. Wide Range Wilmington, DE.

Gooch, D., Hulme, C., Nash, H. M., & Snowling, M. J. (2014). Comorbidities in preschool children at family risk of dyslexia. *Journal of Child Psychology and Psychiatry*, *55*(3), 237–246.

Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, *16*(1), 43–54.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10.

Green, D., Swets, J., & others. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4). https://doi.org/10.1111/2041-210X.12504

Grootswagers, T. (2020). A primer on running human behavioural experiments online. *Behavior Research Methods*, *52*, 2283–2286.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3).

https://doi.org/10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662.

Hasselmo, M. E., & Stern, C. E. (2006). Mechanisms underlying working memory for novel information. *Trends in Cognitive Sciences*, *10*(11), 487–493.

He, Y., Guo, Z., Wang, X., Sun, K., Lin, X., Wang, X., Li, F., Guo, Y., Feng, T., Zhang, J., & others. (2022). Effects of Audiovisual Interactions on Working Memory Task Performance—Interference or Facilitation. *Brain Sciences*, *12*(7), 886.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes \& Intergroup Relations*, *18*(3), 384–401.

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 150.

Herten, N., Otto, T., & Wolf, O. T. (2017). The role of eye fixation in memory enhancement under stress--an eye tracking study. *Neurobiology of Learning and Memory*, *140*, 134–144.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Huestegge, L., Rohrßen, J., van Ermingen-Marbach, M., Pape-Neumann, J., & Heim, S. (2014). Devil in the details? Developmental dyslexia and visual long-term memory for details. *Frontiers in Psychology*, *5*(JUL).

https://doi.org/10.3389/fpsyg.2014.00686

Huettig, F., & Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia—evidence from eye-tracking. *Dyslexia*, *21*(2), 97–122.

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology*, *96*(2). https://doi.org/10.1016/j.jecp.2006.09.002

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412.

Jaffe-Dax, S., Frenkel, O., & Ahissar, M. (2017). Dyslexics' faster decay of implicit memory for sounds and words is manifested in their shorter neural adaptation. *ELife*, *6*. https://doi.org/10.7554/elife.20557

Jaffe-Dax, S., Lieder, I., Biron, T., & Ahissar, M. (2016). Dyslexics' usage of visual priors is impaired. *Journal of Vision*, *16*(9). https://doi.org/10.1167/16.9.10

Jaffe-Dax, S., Raviv, O., Jacoby, N., Loewenstein, Y., & Ahissar, M. (2015). Towards a computational model of Dyslexia. *BMC Neuroscience*, *16*(S1). https://doi.org/10.1186/1471-2202-16-s1-o12

Johansson, R., & Johansson, M. (2020). Gaze position regulates memory accessibility during competitive memory retrieval. *Cognition*, *197*. https://doi.org/10.1016/j.cognition.2019.104169

Jones, M., Ashby, J., & Branigan, H. P. (2013). Dyslexia and fluency: Parafoveal and foveal influences on rapid automatized naming. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2). https://doi.org/10.1037/a0029710

Jones, M., Branigan, H., & Kelly, M. L. (2009). Dyslexic and nondyslexic reading fluency: Rapid automatized naming and the importance of continuous lists. *Psychonomic Bulletin \& Review*, *16*(3), 567–572.

Jones, M., Branigan, H., Parra, M., & Logie, R. (2013). Cross-modal binding in developmental dyslexia. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(6). https://doi.org/10.1037/a0033334

Jones, M., Kuipers, J., Nugent, S., Miley, A., & Oppenheim, G. (2018). Episodic traces and statistical regularities: Paired associate learning in typical and dyslexic readers. *Cognition*, *177*. https://doi.org/10.1016/j.cognition.2018.04.010

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

Jones, S. D., Jones, M., Koldewyn, K., & Westermann, G. (2023). *Rational inattention: A new theory of neurodivergent information seeking*.

Kalashnikova, M., Burnham, D., & Goswami, U. (2021). The Role of Paired Associate Learning in Acquiring Letter-Sound Correspondences: A Longitudinal Study of Children at Family Risk for Dyslexia. *Scientific Studies of Reading*, *25*(6), 504–518.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working

memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3). https://doi.org/10.3758/BRM.42.3.627

Kibby, M. Y., Lee, S. E., & Dyer, S. M. (2014). Reading performance is predicted by more than phonological processing. *Frontiers in Psychology*, *5*, 960.

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis 1. In *A handbook of process tracing methods* (pp. 111–130). Routledge.

Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., & Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods*, *52*(1), 317–341.

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

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352.

Kraemer, H. C., & Blasey, C. (2015). *How many subjects?: Statistical power analysis in research*. Sage Publications.

Kumcu, A., & Thompson, R. L. (2020). Less imageable words lead to more looks to blank locations during memory retrieval. *Psychological Research*, *84*(3). https://doi.org/10.1007/s00426-018-1084-6

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528–2543.

Kuperman, V., Van Dyke, J. A., & Henry, R. (2016). Eye-movement control in RAN and reading. *Scientific Studies of Reading*, *20*(2), 173–188.

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

Lallier, M., Acha, J., & Carreiras, M. (2016). Cross-linguistic interactions influence reading development in bilinguals: a comparison between early balanced French-Basque and Spanish-Basque bilingual children. *Developmental Science*, *19*(1), 76–89.

Lallier, M., & Carreiras, M. (2018). Cross-linguistic transfer in bilinguals reading in two alphabetic orthographies: The grain size accommodation hypothesis. *Psychonomic Bulletin \& Review*, *25*, 386–401.

Lallier, M., Carreiras, M., Tainturier, M.-J., Savill, N., & Thierry, G. (2013). Orthographic transparency modulates the grain size of orthographic processing: behavioral and ERP evidence from bilingualism. *Brain Research*, *1505*, 47–60.

Lallier, M., Donnadieu, S., Berger, C., & Valdois, S. (2010). A case study of developmental phonological dyslexia: Is the attentional deficit in the perception of rapid stimuli sequences amodal? *Cortex*, *46*(2), 231–241.

Lallier, M., Thierry, G., Barr, P., Carreiras, M., & Tainturier, M.-J. (2018). Learning to read bilingually modulates the manifestations of dyslexia in adults. *Scientific Studies of Reading*, *22*(4), 335–349.

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., O'Donovan, M., Williams, J., Bartling, J., Bruder, J., & others. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, *54*(6), 686–694.

Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics*, *21*(2), 243–262.

Landi, N. (2010). An examination of the relationship between reading comprehension, higher-level and lower-level reading sub-skills in adults. *Reading and Writing*, *23*(6), 701–717.

Leavett, R., Nash, H. M., & Snowling, M. J. (2014). Am I dyslexic? Parental self-report of literacy difficulties. *Dyslexia*, *20*(4), 297–304.

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, *20*(8), 1040–1048.

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific Reports*, *10*(1), 1–11.

Li, W., Zhang, Q., Qiao, H., Jin, D., Ngetich, R. K., Zhang, J., Jin, Z., & Li, L. (2021). Dual n-back working memory training evinces superior transfer effects compared to the method of loci. *Scientific Reports*, *11*(1), 1–10.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). *The alphabetic principle and learning to read.*

Lin, Y.-C., & Lin, P.-Y. (2020). Reading minds in motion: Mouse tracking reveals

transposed-character effects in Chinese compound word recognition. *Applied Psycholinguistics*, *41*(4), 727–751.

Litt, R. A., de Jong, P. F., van Bergen, E., & Nation, K. (2013). Dissociating crossmodal and verbal demands in paired associate learning (PAL): What drives the PAL--reading relationship? *Journal of Experimental Child Psychology*, *115*(1), 137–149.

Litt, R. A., & Nation, K. (2014). The nature and specificity of paired associate learning deficits in children with dyslexia. *Journal of Memory and Language*, *71*(1). https://doi.org/10.1016/j.jml.2013.10.005

Litt, R. A., Wang, H.-C., Sailah, J., Badcock, N. A., & Castles, A. (2019). Paired associate learning deficits in poor readers: The contribution of phonological input and output processes. *Quarterly Journal of Experimental Psychology*, *72*(3), 616–633.

Lotfi, S., Ward, R. T., Mathew, A. S., Shokoohi-Yekta, M., Rostami, R., Motamed-Yeganeh, N., Larson, C. L., & Lee, H.-J. (2022). Limited visual working memory capacity in children with dyslexia: An ERP study. *NeuroRegulation*, *9*(2), 98.

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological Science*, *25*(9), 1663–1673.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, 1–14.

Madar, A. D., Ewell, L. A., & Jones, M. V. (2019). Pattern separation of spiketrains in hippocampal neurons. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-41503-8

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into

decision making. *Behavior Research Methods*, *51*(3), 1085–1101.

Mayringer, H., & Wimmer, H. (2000). Pseudoname learning by German-speaking children with dyslexia: Evidence for a phonological learning deficit. *Journal of Experimental Child Psychology*, *75*(2), 116–133.

McDougall, S., Hulme, C., Ellis, A., & Monk, A. (1994). Learning to read: The role of short-term memory and phonological skills. *Journal of Experimental Child Psychology*, *58*(1), 112–133.

Menghini, D., Carlesimo, G. A., Marotta, L., Finzi, A., & Vicari, S. (2010). Developmental dyslexia and explicit long-term memory. *Dyslexia (Chichester, England)*, *16*(3). https://doi.org/10.1002/dys.410

Messbauer, V. C. S., & de Jong, P. F. (2003). Word, nonword, and visual paired associate learning in Dutch dyslexic children. *Journal of Experimental Child Psychology*, *84*(2). https://doi.org/10.1016/S0022-0965(02)00179-0

Meule, A. (2017). Reporting and interpreting working memory performance in n-back tasks. *Frontiers in Psychology*, *8*, 352.

Moll, K., Snowling, M. J., Göbel, S. M., & Hulme, C. (2015). Early language and executive skills predict variations in number and arithmetic skills in children at family-risk of dyslexia and typically developing controls. *Learning and Instruction*, *38*, 53–62.

Moll, K., Snowling, M. J., & Hulme, C. (2020). Introduction to the special issue "comorbidities between reading disorders and other developmental disorders." In *Scientific Studies of Reading* (Vol. 24, Issue 1, pp. 1–6). Taylor \& Francis.

Muthén, & Muthén. (2018). *Mplus User's Guide (Eighth Edition). Los Angeles, CA: Muthén & Muthén*.

Neta, M., Berkebile, M. M., & Freeman, J. B. (2021). The dynamic process of ambiguous emotion perception. *Cognition and Emotion*, *35*(4), 722–729.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology*, *63*, 427–452.

O'Brien, G., & Yeatman, J. D. (2021). Bridging sensory and language theories of dyslexia: Toward a multifactorial model. *Developmental Science*, *24*(3), e13039.

Oakhill, J., & Kyle, F. (2000). The relation between phonological awareness and working memory. *Journal of Experimental Child Psychology*, *75*(2), 152–164.

Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2014). Visual and auditory digit-span performance in native and non-native speakers. *International Journal of Bilingualism*, *18*(6), 663–673.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory \& Cognition*, *29*(2), 344–354.

Palmer, S. (2000). Phonological recoding deficit in working memory of dyslexic teenagers. *Journal of Research in Reading*, *23*(1), 28–40.

Papoutsaki, A., Daskalova, N., Sangkloy, P., Huang, J., Laskey, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. *IJCAI International Joint Conference on Artificial Intelligence*, *2016-Janua*.

Patil, I. (2021). Visualizations with statistical details: The'ggstatsplot'approach. *Journal of Open Source Software*, *6*(61).

Pelegrina, S., Lechuga, M. T., Garc\'\ia-Madruga, J. A., Elosúa, M. R., Macizo, P., Carreiras, M., Fuentes, L. J., & Bajo, M. T. (2015). Normative data on the n-back task for children and young adolescents. *Frontiers in Psychology*, *6*, 1544.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385–413.

Perry, C., Ziegler, J. C., Braun, M., & Zorzi, M. (2010). Rules versus statistics in reading aloud: New evidence on an old debate. In *European Journal of Cognitive Psychology* (Vol. 22, Issue 5). https://doi.org/10.1080/09541440902978365

Pimperton, H., & Nation, K. (2010). Suppressing irrelevant information from working memory: Evidence for domain-specific deficits in poor comprehenders. *Journal of Memory and Language*, *62*(4), 380–391.

Poulsen, M. (2011). Do dyslexics have auditory input processing difficulties? *Applied Psycholinguistics*, *32*(2), 245–261.

R Core Development Team. (2020). R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing,*

Vienna, Austria. URL Http://Www.R-Project.Org/.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841–865.

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin \& Review*, *20*(6), 1102–1113.

Reiter, A., Tucha, O., & Lange, K. W. (2005). Executive functions in children with dyslexia. *Dyslexia*, *11*(2), 116–131.

Rello, L., & Bigham, J. P. (2017). Good background colors for readers: A study of people with and without dyslexia. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 72–80.

Richardson, D. C., & Spivey, M. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition*, *76*(3). https://doi.org/10.1016/S0010-0277(00)00084-6

Rozin, P., & Gleitman, L. R. (1977). The structure and acquisition of reading II: The reading process and the acquisition of the alphabetic principle. *Toward a Psychology of Reading*, *66*.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4). https://doi.org/10.1006/jmla.1996.0032

Samuels, S. J. (2002). Reading fluency: Its development and assessment. *What Research Has to Say about Reading Instruction*, *3*, 166–183.

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences*, *10*(4), 251.

Scholz, A., Klichowicz, A., & Krems, J. F. (2018). Covert shifts of attention can account for the functional role of "eye movements to nothing." *Memory \& Cognition*, *46*, 230–243.

Scholz, A., Mehlhorn, K., Bocklisch, F., & Krems, J. F. (2011). Looking at Nothing Diminishes with Practice. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*.

Scholz, A., Mehlhorn, K., & Krems, J. F. (2016). Listen up, eye movements play a role in verbal memory retrieval. *Psychological Research*, *80*, 149–158.

Scorpio, K. A., Islam, R., Kim, S. M., Bind, R., Borod, J. C., & Bender, H. A. (2018). Paired-Associate Learning. In John, C. B. K. J. S., & DeLuca (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 2554–2558). Springer International Publishing. https://doi.org/10.1007/978-3-319-57111-9_1137

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*, *14*(5), 238–242.

Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*, *96*(4). https://doi.org/10.1037/0033-295X.96.4.523

Sela, I., Izzetoglu, M., Izzetoglu, K., & Onaral, B. (2012). A working memory deficit among dyslexic readers with no phonological impairment as measured using the n-back task: an fNIR study. *PloS One*, *7*(11), e46527.

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in

cognitive science: A first look. *Behavior Research Methods*, *50*(2). https://doi.org/10.3758/s13428-017-0913-7

Seymour, P. H. K., Aro, M., Erskine, J. M., & with COST Action A8 Network, C. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.

Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological Psychiatry*, *57*(11), 1301–1309.

Sheehan, T. C., Sreekumar, V., Inati, S. K., & Zaghloul, K. A. (2018). Signal complexity of human intracranial EEG tracks successful associative-memory formation across individuals. *Journal of Neuroscience*, *38*(7). https://doi.org/10.1523/JNEUROSCI.2389-17.2017

Snowling, M., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia*, *18*(1). https://doi.org/10.1002/dys.1432

Snowling, M., & Hulme, C. (1994). The development of phonological skills. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *346*(1315), 21–27.

Snowling, M., & Hulme, C. (2021). Annual Research Review: Reading disorders revisited--the critical importance of oral language. *Journal of Child Psychology and Psychiatry*, *62*(5), 635–653.

Snowling, M., Hulme, C., & Nation, K. (2020). Defining and understanding dyslexia: past, present and future. *Oxford Review of Education*, *46*(4), 501–513.

Snowling, M., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., Hulme, C., Language, W., & Team, R. P. (2019). Developmental outcomes for children at

high risk of dyslexia and children with developmental language disorder. *Child Development*, *90*(5), e548--e564.

Spivey, M. (2007). Uniting and Freeing the Mind. In *The Continuity of Mind*. https://doi.org/10.1093/acprof:oso/9780195170788.003.0011

Spivey, M., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398.

Squire, L. R., & Zola, S. M. (1998). Episodic memory, semantic memory, and amnesia. In *Hippocampus* (Vol. 8, Issue 3). https://doi.org/10.1002/(SICI)1098-1063(1998)8:3<205::AID-HIPO3>3.0.CO;2-I

Stanovich, K. E. (1994). Annotation: Does dyslexia exist? *Journal of Child Psychology and Psychiatry*, *35*(4), 579–595.

Stark, S. M., & Stark, C. E. L. (2016). Chapter 67 – Introduction to Memory. In *Neurobiology of Language* (Issue 1991).

Stein, J. (2023). Theories about developmental dyslexia. *Brain Sciences*, *13*(2), 208.

Stein, J., & Talcott, J. (1999). Impaired neuronal timing in developmental dyslexia—the magnocellular hypothesis. *Dyslexia*, *5*(2), 59–77.

Szmalec, A., Loncke, M., Page, M., & Duyck, W. (2011). Order or disorder? Impaired Hebb learning in dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1270.

Szmalec, A., Verbruggen, F., Vandierendonck, A., & Kemps, E. (2011). Control of interference during working memory updating. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 137.

Team, A. (2021). *Audacity(R): Free Audio Editor and Recorder* (2.4.1). https://audacityteam.org/

Team, R. C. (2020). *R: A language and environment for statistical computing*.

Thomas, M. S. C. (2020). Developmental disorders: few specific disorders and no specific brain regions. *Current Biology*, *30*(7), R304--R306.

Toda, A. M., Palomino, P. T., Oliveira, W., Rodrigues, L., Klock, A. C. T., Gasparini, I., Cristea, A. I., & Isotani, S. (2019). How to gamify learning systems? an experience report using the design sprint method and a taxonomy for gamification elements in education. *Journal of Educational Technology \& Society*, *22*(3), 47–60.

Toffalini, E., Marsura, M., Garcia, R. B., & Cornoldi, C. (2019). A Cross-Modal Working Memory Binding Span Deficit in Reading Disability. *Journal of Learning Disabilities*, *52*(2). https://doi.org/10.1177/0022219418786691

Toffalini, E., Tomasi, E., Albano, D., & Cornoldi, C. (2018). The effects of the constancy of location and order in working memory visual–phonological binding of children with dyslexia. *Child Neuropsychology*, *24*(5). https://doi.org/10.1080/09297049.2017.1329411

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). Test of word reading efficiency (TOWRE). *Austin, TX: Pro-Ed*.

Treisman, A., & Gelade, G. (1980). Treisman (1980) a Feature-Integration Theory of Attention.Pdf. In *Cognitive Psychology* (Vol. 12).

Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & Cognition*, *34*(8), 1704–1719.

Tulving, E. (1972). Chapter 10: Episodic and sematic memory. In *Organisation of memory*.

Tulving, E. (1982). Synergistic ecphory in recall and recognition. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *36*(2). https://doi.org/10.1037/h0080641

Twomey, K. E., Ranson, S. L., & Horst, J. S. (2014). That's More Like It: Multiple Exemplars Facilitate Word Learning. *Infant and Child Development*, *23*(2). https://doi.org/10.1002/icd.1824

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. In *Neuroscience and Biobehavioral Reviews* (Vol. 51). https://doi.org/10.1016/j.neubiorev.2015.01.008

Vaughan-Evans, A., Liversedge, S. P., Fitzsimmons, G., & Jones, M. W. (2020). Syntactic co-activation in natural reading. *Visual Cognition*, *28*(10), 541–556.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40.

Vidal, C., Content, A., & Chetail, F. (2017). BACS: The Brussels Artificial Character Sets for studies in cognitive psychology and neuroscience. *Behavior Research Methods*, *49*, 2093–2112.

Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology*, *108*(2). https://doi.org/10.1016/j.jecp.2010.09.011

Vogel, S. A. (2003). Introduction to the Special Issues on International Perspectives on Dyslexia and the International Adult Literacy Survey. *Dyslexia: An*

International Journal of Research and Practice.

von der Malsburg, T. (2019). *saccades: Detection of Fixations in Eye-Tracking Data. R package version 0.2.1.* https://github.com/tmalsburg/saccades

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (1999). *CTOPP: Comprehensive test of phonological processing*. Pro-ed Austin.

Wang, H. C., Wass, M., & Castles, A. (2017). Paired-Associate Learning Ability Accounts for Unique Variance in Orthographic Learning. *Scientific Studies of Reading*, *21*(1). https://doi.org/10.1080/10888438.2016.1231686

Wang, J., Huo, S., Wu, K. C., Mo, J., Wong, W. L., & Maurer, U. (2022). Behavioral and neurophysiological aspects of working memory impairment in children with dyslexia. *Scientific Reports*, *12*(1), 1–13.

Wantz, A. L., Martarelli, C. S., & Mast, F. W. (2016). When looking back to nothing goes back to nothing. *Cognitive Processing*, *17*(1). https://doi.org/10.1007/s10339-015-0741-6

Warmington, M., & Hulme, C. (2012). Phoneme Awareness, Visual-Verbal Paired-Associate Learning, and Rapid Automatized Naming as Predictors of Individual Differences in Reading Ability. *Scientific Studies of Reading*, *16*(1). https://doi.org/10.1080/10888438.2010.534832

Wechsler, D. (1981). *WAIS-R : Wechsler adult intelligence scale-revised*. New York, N.Y. : Psychological Corporation, [1981] ©1981. https://search.library.wisc.edu/catalog/999605091402121

Wechsler, D. (1999). Manual for the Wechsler abbreviated intelligence scale (WASI). In *WASI.*

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test (WRAT4).* Lutz, FL: Psychological Assessment Resources.

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics*, *14*(1), 1–33.

Wimmer, H., Mayringer, H., & Landerl, K. (1998). Poor reading: A deficit in skill-automatization or a phonological deficit? *Scientific Studies of Reading*, *2*(4), 321–340.

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology*, *92*(4), 668.

Windfuhr, K. L., & Snowling, M. J. (2001). The relationship between paired associate learning and phonological skills in normally developing readers. *Journal of Experimental Child Psychology*, *80*(2), 160–173.

Wolf, M. (2008). *Proust and the Squid: The Story and Science of the Reading Brain.(2008).*

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*(3), 415.

Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J., & Schulte-Mecklenbeck, M. (2021). *Movement tracking of cognitive processes: A tutorial using mousetrap.*

Wyszecki, G., & Stiles, W. S. (1982). *Color science* (Vol. 8). Wiley New York.

Yamauchi, T., Leontyev, A., & Razavi, M. (2019). Mouse Tracking Measures Reveal

Cognitive Conflicts Better than Response Time and Accuracy Measures. *CogSci*, 3150–3156.

Ye, W., & Damian, M. F. (2023). Effects of conflict in cognitive control: Evidence from mouse tracking. *Quarterly Journal of Experimental Psychology*, *76*(1), 54–69.

Yuan, K.-H., & Maxwell, S. (2005). On the post hoc power in testing mean differences. *Journal of Educational and Behavioral Statistics*, *30*(2), 141–167.

Žarić, G., Fraga González, G., Tijms, J., van der Molen, M. W., Blomert, L., & Bonte, M. (2014). Reduced neural integration of letters and speech sounds in dyslexic children scales with individual differences in reading fluency. *PLoS One*, *9*(10), e110337.

Žarić, G., González, G. F., Tijms, J., van der Molen, M. W., Blomert, L., & Bonte, M. (2015). Crossmodal deficit in dyslexic children: Practice affects the neural timing of letter-speech sound integration. *Frontiers in Human Neuroscience*, *9*(JUNE). https://doi.org/10.3389/fnhum.2015.00369

Zugarramurdi, C., Fernández, L., Lallier, M., Valle-Lisboa, J. C., & Carreiras, M. (2022). Mind the orthography: Revisiting the contribution of prereading phonological awareness to reading acquisition. *Developmental Psychology*, *58*(6), 1003.

Appendices

Appendix A

Pre-Registration for Chapter 4

Loc&ContConsInNovelCross-ModalBindings:ReadingDysl&Eyetracking (#55653)

Created: 01/11/2021 12:56 PM (PT)

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Based on previous research (Jones, Kuipers, Nugent, Miley, & Oppenheim, 2018; Toffalini, Tomasi, Albano, & Cornoldi, 2018), we predict that typical readers obtain better overall (a) recall and (b) recognition accuracy than their dyslexic peers when learning new shape-sound pairs (cross modal bindings), and improve more quickly over a fixed number of exposures (*i.e.*, blocks). Based on evidence that dyslexics use chunking as a memory retrieval strategy (Ullman & Pullman, 2015) and benefit from certain features being kept consistent during encoding (Toffalini et al., 2018), we will further test whether consistency in screen location and/or in (neighbouring) context during encoding will increase accuracy and reduce latencies for correct

responses for individuals with dyslexia (vs. typical readers) during (a) recall and (b) recognition of the bindings.

As a sub-analysis, will individuals with dyslexia benefit in terms of behavioural accuracy (compared with typical readers' accuracy) from executing anticipatory looks at empty screen locations previously occupied by targets during a recognition task?

3) Describe the key dependent variable(s) specifying how they will be measured.

Task 1 – Recognition task with interspersed recall trials.

Stimuli comprise consistently paired combinations of a single symbol and a CVC pseudoword. In each trial, participants are first presented with three of these shape-sound combinations. The shapes disappear from the screen, and participants then hear one of the three pseudowords again. After a 1000 ms blank screen, the shapes re-appear, and participants are prompted to click on the matching shape. We will assess accuracy and log-transformed response times for the 216 trials (6 blocks) of this task.

Interspersed recall trials ($N = 36$) will occur at regular intervals: participants will see a shape and be prompted to articulate the associated pseudoword. While we will assess and report accuracy for these recall trials, we include them only to ensure participants are: a) properly attending to the phonological information of the audio-visual pairings, and b) attempting to store the bindings beyond the temporal boundaries of each individual trial.

Task 2 – Separate recall task

This comprises a single block with 36 trials administered after Task 1. Participants view a single shape (from Task 1) in the centre of the screen, and will be prompted to articulate the pseudoword associated with the shape they just saw. We will assess accuracy, classifying as an error any oral production deviating from the target by one phoneme or more.

Task 3 – Separate recognition task

In this task, which will also comprise of a single block with 36 trials, and which will be administered after Task 2, participants will hear a pseudoword and then see three shapes (from Task 1). They will be prompted to click on the matching shape. We will assess accuracy and log transformed response times.

4) How many and which conditions will participants be assigned to?

The only between-subject condition is 'ReaderType' (typical reader vs. dyslexic), distinguished on the basis of a self-reported confirmation from the participant of a prior affirmative assessment of dyslexia, and subjective verification of this diagnosis via a positive response to the questions 'do you think you are dyslexic?' (Snowling, Dawes, Nash, & Hulme, 2012).

Two within-subjects factors include Location (consistent vs. inconsistent) and Context (consistent vs. inconsistent). Consistent location involves shapes appearing in the same screen location across trials; consistent context involves shapes appearing with the same distractor items across trials.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

All analyses will use linear (for RTs) or logistic (for error rates) mixed effects regressions, including maximal random effects structures but reverting to a 'parsimonious' approach in the case of convergence errors. For Task 1, we w model error rate for recognition trials as a function of reading impairment ("ReaderType", *i.e.*, typical versus self-reported dyslexia), location consistency (*i.e.*, target appearing in fixed or variable screen locations), context consistency (*i.e.*, target co-occurring with the same or variable distractors), and target repetition ("Block", over six blocks, log-transformed). Following Jones et al. (2018), we also include two eyetracking-related binomial predictors: 1. "FixatedAnyROI", indicating whether we identified any ROI fixations during the blank screen immediately preceding the testing phase, and 2. "PrimaryFixation", a nested predictor indicating whether they fixated the target's former location more than the distractors'), to probe whether looks at such locations facilitate recognition.

Based on previous data and theory, we expect typical readers' error rates to decrease more quickly than dyslexics' over repetitions, especially when both the target's screen locations and the distractors that co-occurred with it are kept constant across repetitions, but our power analysis suggests insufficient power to statistically detect this pattern as a 4-way interaction (ReaderType * log(Block) * LocationConsistency * ContextConsistency). We do, however, anticipate sufficient power to detect the following effects:

- (a) error rates will decrease with consistent context and location (LocationConsistency x ContextConsistency).

- (b) over repetitions, error rates will decrease faster for typical than dyslexic readers (ReaderType x log(Block)).

-       (c) compared with typical readers, dyslexics will err more when fixating screen locations previously occupied by distractors (ReaderType x Primary Fixation).

For the short single-block recognition task, and for the separate single-block recall task analyses, all the same predictors will be included in the error rate model, except target repetition (*i.e.* Block) and those related to eye movements. In both models, we expect to detect:

    1) A main effect of reading impairment (*i.e.* ReaderType), showing that typical readers recognise and recall binding more accurately than dyslexics.

    2) A LocationConsistency x ContextConsistency interaction showing that both typical and dyslexic readers were more accurate when the target consistently occurred in the same screen location and in with the same distractor stimuli.

In both recognition tasks, we will also use linear mixed effects regressions to model log-RTs for correct responses using the same predictors as for error rates,, with the same predictions.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will exclude:

1.) Trials in which response time is below 100 ms, as these are likely to be fast guesses.

2.) Trials in which face convergence values are below 0.5 will be included in the behavioural data analysis but excluded from the eyetracking analyses because the eyetracking data for such trials is considered unreliable.

3.) Participants who exceed a three hour limit to complete the experiment. Based on a pilot study, we expect the whole experiment to take approximately 2 hours; an extra hour will be allowed as a 'grace period' so participants, especially those with dyslexia, can take moderately long breaks if needed, but we assume that exceeding this allowance indicates a failure to attend to the task.

4.) Participants who fail to pass three 'attention check' trials in a row in Task 1, since that will be an indication that their computer has been left unattended during the task.

5.) Participants whose audio files recorded during the recall task are excessively noisy (*e.g.*, loud music/TV, people talking in the background), or which provide any indication that they were not actually on task (*e.g.*, participants discernibly answering their phone or talking to other people).

6.) Participants who fail to complete the experiment due to issues with their Internet connection, with Gorilla Experiment Builder's servers, or due to any other hardware related malfunctioning.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We used data collected from a pilot study to run a power analysis on the Location x Context interaction mentioned above. We used the powerCurve function in the simR package (Green & MacLeod, 2016; Green, MacLeod, & Alday, 2016), which indicated that 35 participants per group should be sufficient to detect a beta value of -0.87, which is what we detected in the pilot study ($N = 14$). The typical readers group ($n = 35$) will comprise individuals with no self-reported literacy, language or

sensory impairments. The dyslexic group (n = 35) will comprise individuals who have self-identified as having dyslexia, as screened by the Adult Reading Questionnaire (Snowling et al. 2012), but who have no other comorbid difficulties (*e.g.*, SLI or sensory impairments). All participants will be monolingual speakers of British English, and range in age from 18 to 40. Note that our goal of 35 participants per group does not include any participants whose data may need to be excluded, as described in the Exclusions section. Therefore, if we need to exclude any, then we will recruit additional participants as replacements.

8) Anything else you would like to pre-register? (*e.g.*, secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

In order to validate our 'typical' and 'dyslexic' ReaderType classifications, we will also collect and report the following literacy and cognitive group validation measures: 1.) forward digit span, 2.) backward digit span, 3.) nonverbal IQ (WRIT matrix reasoning), verbal IQ (WAIS Similarities subtest), 4.) word and nonword reading (TOWRE), and 5.) Rapid Automatized Naming Test (CTOPP: Letters and Digits).

We will also collect mouse-tracking data in the recognition trials, and eye-tracking data in the separate recall and recognition tasks for exploratory purposes.

Available at https://aspredicted.org/1MV_TX5

Appendix B

Supplementary Materials for Chapter 4

Audiovisual Learning in Dyslexic and Typical adults: Modulating Influences of Location and Context Consistency

Supplementary Materials

Response time data

For completeness, we present the response time analyses which were not part of our a priori hypotheses. In a linear mixed effects regression analysis, log transformed response times were analysed as a function of group membership, repetition (*i.e.*, Block), location consistency, context consistency, fixations to any regions of interest, and primary fixation (see summary in Supplementary Table 1) for the training and recognition task. We also ran a similar analysis on log transformed response times as a function of group membership, location consistency, and context consistency for the post-training recognition test (see summary in Supplementary Table 2). These analyses exclude incorrect responses as well as response times of less than 100ms; *p*-value estimations use the Wald approximation method.

Training

Overall, readers from both groups responded faster as a function of repetition (-50 ms; $\beta_{\log(Block)} = -0.050$, *p* <.001) in the training and recognition task. In general, while participants responded slower over time for items encoded under the *in*consistent context condition (26ms; $\beta_{\log(Block)} = 0.026$, *p* = .018), that effect was stronger for readers with dyslexia (52ms; $\beta_{Group \times \log(Block) \times Context} = 0.052$, *p* = .016).

Supplementary Table 1

Summary of a linear mixed effects regression analysis of log-transformed recognition response times (Training). Millisecond effect estimates are generated by back-transforming the effect at the intercept.

	Coef (β)	Lower (β)	Upper (β)	p	in ms
(Intercept)	7.470	7.433	7.507	-	1754.61
Group (typical, dyslexic)	-0.001	-0.062	0.060	.976	-1
Location (consistent, inconsistent)	0.037	-0.007	0.081	.097	37
Context(consistent, inconsistent)	0.020	-0.024	0.063	.376	20
FixatedAnyROI (no,yes)	0.005	-0.017	0.026	.673	5
PrimaryFixation (target, distractor)	0.005	-0.029	0.039	.775	5
log(Block)	-0.050	-0.076	0.024	<.001	-50
Group x Location	-0.002	-0.027	0.022	.863	-2
Group x Context	-0.017	-0.039	0.005	.121	-17
Group x FixatedAnyROI	0.023	-0.019	0.066	.279	23
Group x PrimaryFixation	0.029	-0.039	0.098	.401	29
Group x log(Block)	0.016	-0.035	0.067	.548	16
Location x Context	-0.021	-0.108	0.066	.638	-21

Location x FixatedAnyROI	-0.001	-0.035	0.034	.973	-1
Location x PrimaryFixation	0.008	-0.057	0.073	.814	8
Location x log(Block)	-0.017	-0.039	0.005	.122	-17
Context x FixatedAnyROI	0.009	-0.030	0.047	.658	9
Context x PrimaryFixation	-0.008	-0.073	0.058	.820	-8
Context x log(Block)	0.026	0.004	0.048	.018	26
FixatedAnyROI x log(Block)	-0.011	-0.055	0.032	.613	-11
PrimaryFixation x log(Block)	-0.031	-0.094	0.032	.329	-31
Group x Location x Context	0.013	-0.031	0.056	.561	13
Group x Location x FixatedAnyROI	0.065	-0.004	0.135	.063	65
Group x Location x PrimaryFixation	-0.031	-0.161	0.099	.642	-31
Group x Location x log(Block)	0.002	-0.041	0.044	.944	2
Group x Context x FixatedAnyROI	-0.057	-0.134	0.020	.148	-57
Group x Context x PrimaryFixation	0.075	-0.055	0.204	.260	75
Group x Context x log(Block)	0.052	0.010	0.094	.016	52
Group x FixatedAnyROI x log(Block)	-0.008	-0.095	0.078	.847	-8
Group x PrimaryFixation x log(Block)	-0.029	-0.159	0.101	.660	-29
Location x Context x FixatedAnyROI	0.007	-0.061	0.075	.834	7

Location x Context x PrimaryFixation	0.002	-0.127	0.131	.976	2
Location x Context x log(Block)	-0.010	-0.059	0.039	.683	-10
Location x FixatedAnyROI x log(Block)	-0.034	-0.113	0.045	.396	-34
Location x PrimaryFixation x log(Block)	-0.079	-0.200	0.043	.204	-79
Context x FixatedAnyROI x log (Block)	0.005	-0.064	0.074	.879	-5
Context x PrimaryFixation x log(Block)	-0.093	-0.209	0.024	.118	-93

Supplementary Figure 1

Subject-weighted mean response time for accurate responses in the training and recognition task.
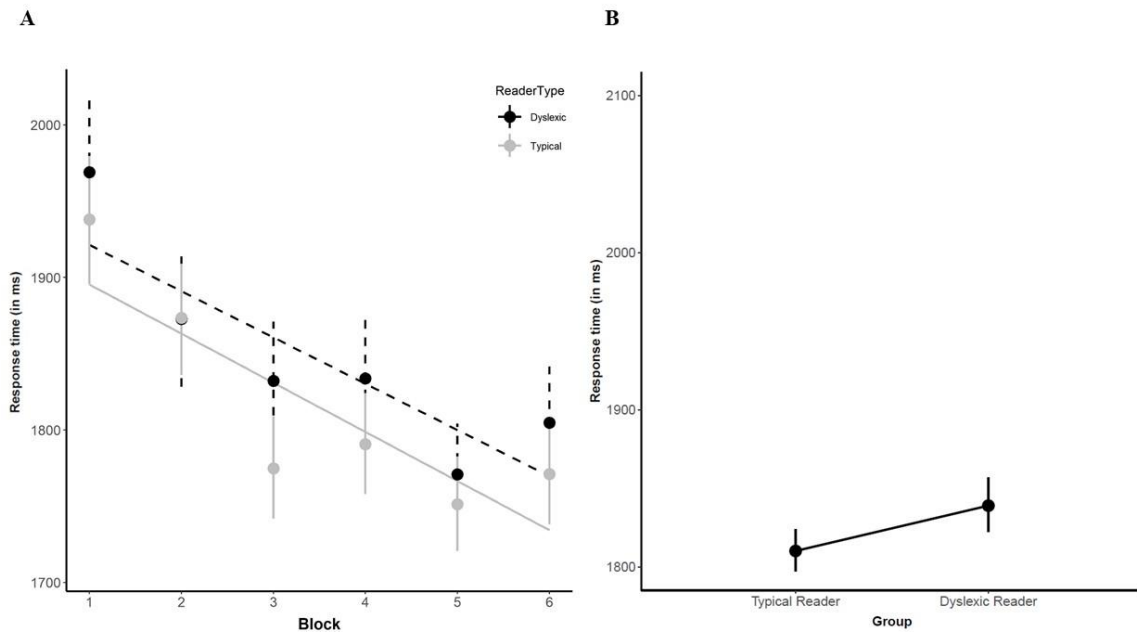
Note. Panel (A) shows response time data per block, while Panel (B) shows the overall response time pattern. In both plots, point ranges represent bootstrapped confidence intervals.

Post-training

In the post-training recognition task, readers with dyslexia responded significantly slower than typical readers (126ms; $\beta_{\text{Group}} = 0.126$, $p = .015$).

Supplementary Table 2

Summary of a linear mixed effects regression analysis of log-transformed recognition response times (Post-Training). Millisecond effect estimates are generated by back-transforming the effect at the intercept.

	Coef (β)	Lower (β)	Upper (β)	p	in ms

(Intercept)	7.515	7.453	7.577	<.001	1835.37
Group (typical, dyslexic)	0.126	0.024	0.227	.015	126
Location (consistent, inconsistent)	0.072	-0.005	0.149	.066	72
Context(consistent, inconsistent)	0.035	-0.041	0.111	.369	35
Group x Location	0.000	-0.058	0.058	.998	0
Group x Context	0.032	-0.023	0.086	.254	32
Location x Context	-0.133	-0.285	0.020	.088	-133
Group x Location x Context	-0.034	-0.142	0.074	.537	-34

Supplementary Figure 2

Subject-weighted mean response time for accurate responses in the post-training recognition test.
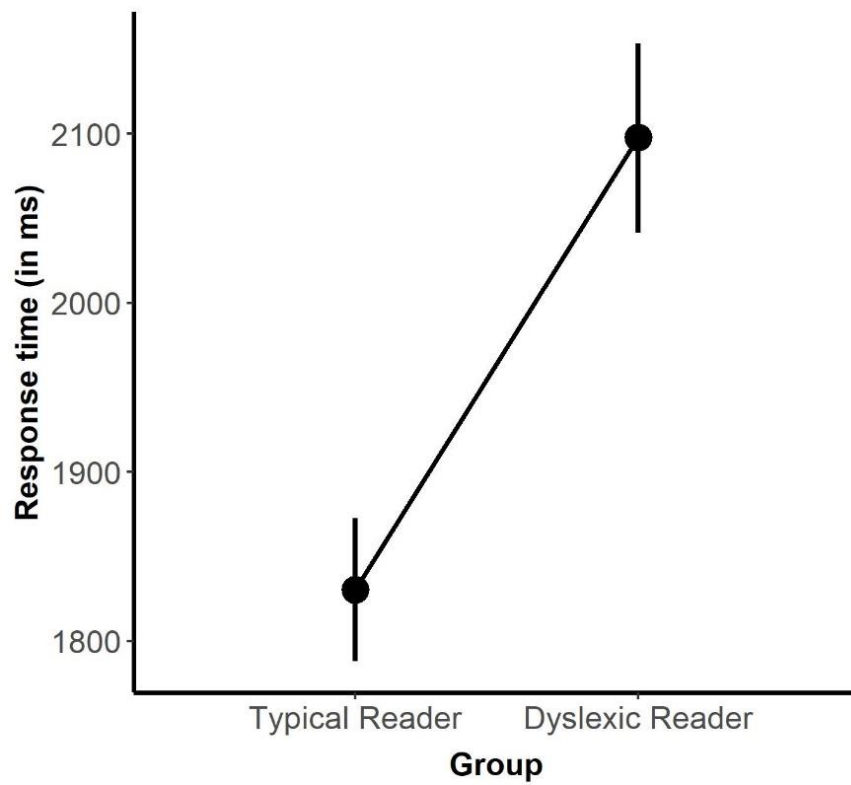
Note. Point ranges represent bootstrapped confidence intervals.

Error data

Below we present the full output for each error analysis model reported in the main manuscript.

Training

Recognition

Supplementary Table 3

Summary of a logistic mixed effects regression analysis of recognition error frequency (Training)

	Coef(β)	SE(β)	p	OR (exp (β))
(Intercept)	-2.18	0.12	<.001	0.11
log(Block)	-1.13	0.08	<.001	0.32
Group (typical, dyslexic)	1.00	0.22	<.001	2.72
Location (consistent, inconsistent)	0.19	0.13	.153	1.20
Context (consistent, inconsistent)	0.30	0.13	.018	1.35
PrimaryFixation (target, distractor)	0.36	0.23	.121	1.43
FixatedAnyROI (no, yes)	-0.17	0.12	.154	0.84
log(Block) x Group	0.26	0.15	.069	1.30
log(Block) x Location	0.05	0.11	.631	1.05
Group x Location	-0.09	0.15	.525	0.91
log(Block) x Context	0.05	0.12	.706	1.05
Group x Context	0.08	0.14	.575	1.08
Location x Context	-0.12	0.26	.650	0.89

log(Block) x PrimaryFixation	0.09	0.33	.783	1.09
log(Block) x FixatedAnyROI	0.19	0.19	.314	1.21
Group x PrimaryFixation	0.08	0.47	.864	1.08
Group x FixatedAnyROI	0.02	0.26	.936	1.02
Location x PrimaryFixation	-0.13	0.46	.772	0.87
Location x FixatedAnyROI	0.24	0.25	.356	1.27
Context x PrimaryFixation	-0.04	0.46	.937	0.96
Context x FixatedAnyROI	0.23	0.24	.341	1.26
log(Block) x Group x Location	-0.07	0.19	.704	0.93
log(Block) x Group x Context	-0.12	0.21	.587	0.89
log(Block) x Location x Context	-0.01	0.22	.957	0.99
Group x Location x Context	0.17	0.29	.563	1.18
log(Block) x Group x PrimaryFixation	-1.00	0.75	.183	0.37
log(Block) x Group x FixatedAnyROI	0.29	0.36	.426	1.34
log(Block) x Location x PrimaryFixation	-0.19	0.65	.775	0.83
log(Block) x Location x FixatedAnyROI	0.85	0.36	.018	2.33
Group x Location x PrimaryFixation	-0.44	0.93	.637	0.64
Group x Location x FixatedAnyROI	-0.39	0.53	.464	0.68

log (Block) x Context x PrimaryFixation	-0.36	0.65	.576	0.69
log (Block) x Context x FixatedAnyROI	0.52	0.35	.145	1.68
Group x Context x PrimaryFixation	0.52	0.93	.575	1.68
Group x Context x FixatedAnyROI	-0.24	0.51	.643	0.79
Location x Context x PrimaryFixation	0.83	0.96	.389	2.29
Location x Context x FixatedAnyROI	-0.94	0.48	.049	0.39
log(Block) x Group x Location x Context	0.06	0.38	.882	1.06
log(Block) x Group x Location x PrimaryFixation	-1.26	1.49	.399	0.28
log(Block) x Group x Location x FixatedAnyROI	-0.82	0.70	.243	0.44
log(Block) x Group x Context x PrimaryFixation	2.60	1.49	.081	13.45
log(Block) x Group x Context x FixatedAnyROI	-0.15	0.69	.824	0.86
log(Block) x Location x Context x PrimaryFixation	-0.73	1.31	.576	0.48
log(Block) x Location x Context x FixatedAnyROI	-0.55	0.71	.438	0.58
Group x Location x Context x PrimaryFixation	-2.19	1.94	.258	0.11
Group x Location x Context x FixatedAnyROI	0.66	1.01	.517	1.93
log(Block) x Group x Location x Context x PrimaryFixation	-1.92	2.99	.520	0.15
log(Block) x Group x Location x Context x FixatedAnyROI	3.80	1.38	.006	44.78

Cued-recall

Supplementary Table 4

Summary of a logistic mixed effects regression analysis of cued-recall error

frequency (Training)

	Coef (β)	SE (β)	*p*	OR (exp (β))
(Intercept)	0.33	0.13	.011	1.39
Group (typical, dyslexic)	0.82	0.19	<.001	2.28
Location (consistent, inconsistent)	-0.05	0.19	.789	0.95
Context(consistent, inconsistent)	0.19	0.19	.320	1.21
Group x Location	0.02	0.18	.898	1.02
Group x Context	0.26	0.17	.144	1.30
Location x Context	-1.04	0.39	.007	0.35
Group x Location x Context	-0.22	0.37	.549	0.80

Post-training

Cued-recall

Supplementary Table 5

Summary of a logistic mixed effects regression analysis of cued-recall error

frequency (Post-training)

	Coef (β)	SE (β)	p	OR (exp (β))
(Intercept)	0.44	0.16	.007	1.56
Group (typical, dyslexic)	1.25	0.28	<.001	3.50
Location (consistent, inconsistent)	-0.20	0.20	.334	0.82
Context(consistent, inconsistent)	0.05	0.19	.773	1.06
Group x Location	0.08	0.19	.648	1.09
Group x Context	0.39	0.19	.047	1.48
Location x Context	-0.17	0.38	.658	0.84
Group x Location x Context	-0.03	0.41	.948	0.97

Recognition

Supplementary Table 6

Summary of a logistic mixed effects regression analysis of recognition error

frequency (Post-training)

	Coef (β)	SE (β)	p	OR (exp (β))

(Intercept)	-2.23	0.18	<.001	0.11
Group (typical, dyslexic)	0.99	0.33	.003	2.71
Location (consistent, inconsistent)	-0.17	0.20	.379	0.84
Context(consistent, inconsistent)	0.26	0.20	.191	1.30
Group x Location	-0.22	0.27	.419	0.81
Group x Context	-0.08	0.27	.761	0.92
Location x Context	-0.28	0.40	.481	0.76
Group x Location x Context	1.01	0.54	.060	2.74

Appendix C

Pre-Registration for Chapter 5

CONFIDENTIAL - FOR PEER-REVIEW ONLY

Unimodal and cross-modal
working memory (#82276)

Created: 12/07/2021 08:27 AM (PT)

This is an anonymized copy (without author names) of the pre-registration. It was created by the author(s) to use during peer-review.

A non-anonymized version (containing author names) should be made available by the authors when the work it supports is made public.

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

Do typical readers and readers with dyslexia differ in their working memory performance, as measured by an n-back task, when exposed to novel unimodal and cross-modal bindings?

What is the nature of cross-modal working memory binding differences between typical and dyslexic readers? Specific hypotheses regarding this aim are: do readers with dyslexia have a specific difficulty in temporarily storing novel cross-modal features (*e.g.*, an arbitrary character paired with an aurally presented pseudoword), relative to typical readers? Alternatively, do they have a more general difficulty in storing any two novel features in their working memory (*e.g.*, an arbitrary character paired with a colour; or an arbitrary voice paired with an aurally presented pseudoword), relative to typical readers?

3) Describe the key dependent variable(s) specifying how they will be measured.

Participants will be exposed to novel unimodal and cross-modal bindings in three separate n-back tasks as follows:

- Task 1: unimodal/visual; *i.e.*, character paired with a colour (N = 20).

- Task 2: unimodal/auditory; *i.e.*, pseudoword paired with a voice (N = 20).

- Task 3: cross-modal/visual plus auditory; *i.e.*, pseudoword paired with a character (N = 20).

In tasks 1 and 3, the stimuli are Chinese characters, unfamiliar to the participants (*i.e.*, monolingual speakers of English with and without a dyslexia diagnosis). In tasks 2 and 3, the pseudowords are consonant-vowel-consonant strings following English phonotactics. No items will be repeated across tasks.

Each item will be presented 20 times over the course of each n-back task (*i.e.*, 5 times per block). In 80% of the time, the stimuli will be consistently presented with the same two features. In the remaining 20%, there will be a mismatch involving one of the features. The mismatches will be bi-directional to account for potential attentional bias towards only one of the features.

Responses provided in the three n-back tasks will be classified as 'hit', 'false alarm', 'miss', and 'correct rejection'.

- In Tasks (1), (2), and (3), a 'hit' will be a correct identification that (1) the character plus its corresponding colour; (2) the pseudoword plus its corresponding

voice; and (3) the pseudoword plus its corresponding character, presented in a given trial either matches or does not match the item presented two trials back. A 'correct rejection' will consist in correct dismissal of a mismatch in trials in which only one of the two features was presented two trials back (*e.g.*, a character identical to the character presented two trials back, but matched with the incorrect colour; a pseudoword identical to the pseudoword presented two trials back, but matched with the incorrect voice; a pseudoword identical to the pseudoword presented two trials back, but matched with the incorrect character). A 'false alarm' will consist in an incorrect identification that an item matches the one presented two trials back when one or both of its features differ. Finally, a 'miss' will consist in failure to identify that an item presented in a given trial matches the one presented two trials back.

We will be measuring the following:

-Accuracy rate (*i.e.*, number of hits and correct rejections);

-Log-transformed reaction time (for hits and correct rejections only);

-Sensitivity [*i.e.*, d'prime; Z(PHit) – Z(PFalseAlarms)] and bias [*i.e.*, c; - (Z(PHit) + Z(PFalseAlarms)/2)).

4) How many and which conditions will participants be assigned to?

The conditions are as follows:

-Between-subject: reading ability status (*i.e.*, adult readers with dyslexia versus adult typical readers).

-Within-subject factor: modality (*i.e.*, visual, auditory, visual plus auditory).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

-We will run logistic (for accuracy data) and linear (for RTs; log-transformed) mixed effects regressions, modelling each outcome variable as a function of reading ability (*i.e.*, adult readers with dyslexia versus typical readers) and modality (*i.e.*, a. visual versus auditory versus cross-modal; and b. unimodal versus cross-modal).

-We will also perform signal detection theory analyses to ascertain whether adult readers with dyslexia and typical readers significantly differ in their sensitivity to detect the difference between the signal present and signal absent distributions. We will then use linear mixed effects regression to model d' prime as a function of reading skill status (*i.e.*, adult readers with dyslexia and typical readers) and modality (*i.e.*, a. visual versus auditory versus

cross-modal; and b. unimodal versus cross-modal).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We will exclude:

a. Trials in which decision time is below 100 ms, as these are likely to be fast guesses.

b. Participants who exceed a 2-hour limit to complete the experiment, which is expected to take approximately 70 minutes.

c. Participants who fail to complete the experiment due to issues with their hardware, Internet connection, or with Gorilla Experiment Builder's servers.

d. Participants who are not monolinguals and whose first language is not British English. Similarly, we will also reject participants with previous knowledge of Kanji characters.

e. Participants who do not provide a response in twenty consecutive trials. These participants will be automatically rejected from the experiment, as this behaviour will be an indication that their computers have been left unattended mid task.

Similarly,

f. A "Colour Identification task" will be administered at the beginning of the experiment to ensure all participants are able to distinguish the pairs of colours used in the visual condition of the n-back task. Participants who score below 80% in this task will be filtered out as this could be an indication of a potential colour deficiency and/or visual impairment.

g. Participants will be shown pairs of characters and asked to rate how similar they perceive the items to be on a Likert scale ranging from 1 ('very different') to 5 ('very similar') in a "Shape Similarity Task" to be administered at the end of the experiment. The items shown in this task will be pairs of characters which appeared in mismatched trials in the visual and cross-modal conditions of the n-back task. Trials containing characters that have been rated as 'very similar' by at least 80% of the participants will be excluded from the analyses.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We intend to recruit 96 adults (48 readers with dyslexia + 48 typical readers) aged between 18 and 30 years old.

The dyslexic group will comprise individuals who have self-identified as having dyslexia, as screened by the Adult Reading Questionnaire (Question 15; Snowling et

al. 2012), but who have no other comorbid difficulties (*e.g.*, SLI or sensory impairments).

Our goal of 96 participants does not include any participants whose data may need to be excluded. Therefore, if we need to exclude any participants, then we will recruit additional participants as replacements.

8) Anything else you would like to pre-register? (*e.g.*, secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We will collect basic demographic information such as gender, age, handedness, language background, and years of education. We will also administer the following background tests:

1. Rapid Automatized Naming (digits and letters);

2. Word reading accuracy;

3. Word reading fluency;

4. Pseudoword reading fluency;

5. Non-verbal IQ;

6. Verbal IQ;

7. Adult Reading Questionnaire;

8. Forward and Backward Digit Span.

9. Forward and Backward Spatial Span

Available at https://aspredicted.org/H8M_PCH

Appendix D

Supplementary Materials for Chapter 6

What the Hand in Motion Reveals about Reading: Children's Decision-Making Processes in Paired Associate Learning and its Relationship to Reading Outcomes

Supplementary Materials

The experimental pseudowords and Kanji characters used in the study are listed in Supplementary Table 7.

Supplementary Table 7

Experimental stimuli

Item number	Pseudoword	Kanji character
1	bix	兄
2	fip	月
3	beb	并
4	fum	目
5	gex	由
6	gop	力
7	saf	从
8	dep	夂
9	sut	彳
10	dod	門
11	pag	厄
12	pof	火

We conducted an observed power analysis via the powerCurve() function in the simr package (P. Green & Macleod, 2016b) on all linear models reported in the

paper. In each analysis, we ran 1000 simulations to calculate power at four sample sizes (*i.e.*, 70, 80, 90, 100 participants). The results for each analysis can be found in Supplementary Table 8.

Supplementary Table 8

Observed power calculation results for the reading ability predictor

	Observed Power [95% Confidence Interval]			
	N = 70	*N* = 80	*N* = 90	*N* = 100
Recognition Accuracy	77.80% [75.09, 80.34]	80.50% [77.91, 82.91]	86.40% [84.12, 88.46]	88.90% [86.79, 90.78]
Recognition Response Time	78.70% [76.03, 81.20]	81.80% [79.27, 84.15]	90.60% [88.62, 92.34]	92.50% [90.69, 94.06]
Cued-recall Trials	73.30% [70.44, 76.02]	76.80% [74.06, 79.38]	79.50% [76.86, 81.96]	83.40% [80.95, 85.66]
Cued-recall Task	78.60% [75.93, 81.10]	84.10% [81.68, 86.31]	88.30% [86.14, 90.23]	91.00% [89.05, 92.70]
Maximum Absolute Deviation	99.90% [99.44, 100.0]	100% [99.63, 100.0]	100% [99.63, 100.0]	100% [99.63, 100.0]
Average Deviation	96.40% [95.05, 97.47]	97.70% [96.57, 98.54]	99.10% [98.30, 99.59]	99.00% [98.17, 99.52]
Area Under the Curve	63.30% [60.23, 66.29]	66.90% [63.89, 69.81]	72.00% [69.11, 74.76]	77.50% [74.78, 80.05]

To test whether participants' age might have influenced our pattern of results, we also ran an additional set of analyses including 'age' as a predictor in all linear

models. Age was not found to significantly affect any of our results. All corresponding *β* and *p* values are reported in Supplementary Tables 9, 10, and 11.

Supplementary Table 9

Summaries of logistic mixed effects regression analyses of accuracy frequency

	Coef. (*β*)	SE (*β*)	*p*	OR [exp (*β*)]
Recognition				
(Intercept)	2.64	0.15	<.001	14.07
Age	0.01	0.01	.180	1.01
Reading	0.35	0.17	.038	1.42
Cued-recall (interspersed trials)				
(Intercept)	-0.94	0.17	<.001	0.39
Age	-0.01	0.01	.118	0.99
Reading	0.42	0.14	.003	1.53
Cued-recall (separate task)				
(Intercept)	-2.16	0.27	<.001	0.12
Age	0.00	0.01	.905	1.00
Reading	0.53	0.20	.009	1.69

Supplementary Table 10

Summary of linear mixed effects regression analyses of recognition response times

	Coef. (*β*)	SE (*β*)	*p*
(Intercept)	7.67	0.02	<.001
Age	-0.00	0.00	.199
Reading	-0.06	0.03	.027

Supplementary Table 11

Summary of linear mixed effects regression analyses of mouse curvature measures

	Coef. (β)	SE (β)	p
Maximum Absolute Deviation			
(Intercept)	0.25	0.02	<.001
Age	0.00	0.00	.438
Reading	-0.07	0.01	<.001
Average Deviation			
(Intercept)	0.08	0.00	<.001
Age	0.00	0.00	.648
Reading	-0.02	0.00	<.001
Area under the Curve			
(Intercept)	0.11	0.01	<.001
Age	0.00	0.00	.236
Reading	-0.02	0.00	<.001

We also conducted supplementary logistic and linear mixed effects regression analyses to explore the extent to which the inclusion of a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) in all models would affect our results. The pattern of results shown in Supplementary Tables 12, 13, 14 are virtually identical to the results reported in the main body of text. However, due to the higher complexity of the maximal random effects models, we opted to adopt a parsimonious approach in order to improve the balance between Type 1 error rates and power (Matuschek et al., 2017).

Supplementary Table 12

Summaries of logistic mixed effects regression analyses of accuracy frequency using a maximal random effects structure approach

	Coef. (β)	SE (β)	p	OR [exp (β)]
Recognition				
(Intercept)	2.52	0.15	<.001	12.43
Reading	0.41	0.16	.011	1.51

Cued-recall (interspersed trials)				
(Intercept)	-0.94	0.17	<.001	0.39
Reading	0.35	0.14	.009	1.42
Cued-recall (separate task)				
(Intercept)	-2.19	0.27	<.001	0.11
Reading	0.58	0.22	.008	1.80

Supplementary Table 13

Summary of linear mixed effects regression analyses of response time using a maximal random effects structure approach

	Coef. (β)	SE (β)	*p*
Recognition			
(Intercept)	7.67	0.02	<.001
Reading	-0.08	0.03	.005

Supplementary Table 14

Summary of linear mixed effects regression analyses of mouse curvature measures using a maximal random effects structure approach

	Coef. (β)	SE (β)	*p*
Maximum Absolute Deviation			
(Intercept)	0.25	0.02	<.001
Reading	-0.07	0.01	<.001
Average Deviation			
(Intercept)	0.08	0.00	<.001
Reading	-0.02	0.00	<.001
Area under the Curve			
(Intercept)	0.11	0.01	<.001
Reading	-0.02	0.01	.017

Monte Carlo Simulation Studies to Test Path Models' Power

We conducted a series of Monte Carlo simulation studies to examine the adequacy of our path models given that we had a relatively small sample ($N = 76$). To do so, we used the Monte Carlo facilities in Mplus 8.4 (Muthén & Muthén, 2018). The simulations were run on 76 observations, and we used the structure and the estimates from each of the final models of word reading accuracy, word reading fluency, and pseudoword reading fluency, respectively (see Figure 22). There was no missing data, and all variables were continuous and approximately normally distributed.

We report the findings from our simulations on the parameters of interest in Supplementary Table 9. The adequacy of the sample size was evaluated using the criteria proposed by (Muthén & Muthén, 2002). These criteria state that (a) the parameter and standard error biases must not be larger than 10% of any parameter, (b) standard error bias of the parameter being assessed are no greater than 5% of the parameters of interest, (c) coverage estimates should be between .91 - .98, (d) power should be greater than .8. In all three models, we found parameter biases were not larger than 10% than any parameter, nor were standard error biases for each parameter of interest larger than 5%. We also found coverage was between 0.91 and 0.98 (see Supplementary Table 15). Turning to power, the simulation study of the word reading accuracy model revealed that there was adequate power to detect a significant effect across all parameters in all models apart from the direct path from PAL to reading in the reading accuracy model.

Taken together, these simulation studies demonstrate that the sample size we were able to access in the present study was adequate for unbiased parameter

estimates, unbiased standard errors, and good coverage. Furthermore, overall, the current sample size provided enough power to detect significant indirect effects of PAL on rapid naming and rapid naming on reading. However, the sample size was not large enough to detect direct effects of PAL on word reading.

Supplementary Table 15

Parameter Bias, Standard Error Bias, Coverage, and Power Estimates for the Final Path Models of Word Reading Accuracy, Word Reading Fluency, and Pseudoword Reading Fluency

	Parameter Bias	Standard Error Bias	Coverage	Power
Word Reading Accuracy				
Reading on				
Phoneme awareness	0.013	-0.02	0.94	0.83
Rapid naming	-0.010	-0.034	0.93	0.96
Paired associate learning	-0.014	-0.013	0.95	0.37
Rapid naming on				
Paired associate learning	-0.013	-0.032	0.93	0.91
Word Reading Fluency				
Reading on				
Phoneme awareness	0.014	-0.012	0.94	0.81
Rapid naming	-0.005	-0.044	0.93	1.00
Rapid naming on				
Paired associate learning	-0.015	-0.040	0.93	0.91
Pseudoword Reading Fluency				
Reading on				
Phoneme awareness	0.013	-0.014	0.94	0.80
Rapid naming	-0.006	-0.041	0.93	1.00
Rapid naming on				
Paired associate learning	-0.014	-0.038	0.93	0.90