Assessing and supporting working memory in children: the role of attention and the environment

Edward David John Berry

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy
University of Leeds
School of Psychology

September 2017

The candidate confirms that the work submitted is their own and that appropriate credit has been given where reference has been made to the work of others.

The data presented in Chapter 3 has been accepted for publication as:

Berry, E.D.J., Waterman, A.W., Baddeley, A.D., Hitch, G.J., & Allen, R.J. (2017). The limits of visual working memory in children: exploring prioritization and recency effects with sequential presentation Developmental Psychology.

Only material produced by Ed D. J. Berry has been used in the thesis.

Like the psychologists the physical scientist has to construe the world, and this business of construing in highly problematic. The world does not lie there before the scientist in ready-labelled pens and pre-constructed hierarchies: the scientist must start with a priori assumptions. The more philosophically sound these assumptions are the more empirically fruitful the work will be.

– James Russell, *Explaining Mental Life: Some Philosophical Issues in Psychology*

Acknowledgements

I would like to thank my supervisors Amanda Waterman, Richard Allen, and Mark Mon-Williams for their support throughout my PhD. Their encouragement and guidance has been essential to completing my studies and thesis in a timely manner. I am particularly grateful to Amanda and Richard for their rapid comments on my chapter drafts. Radka Jersáková should be thanked for tolerating my many questions about PsychoPy when I was first learning to program. Equally, I would not have been able to learn and apply Bayesian statistics without the patient support of Oscar Giles. Emma Portch was always on-hand for interesting discussions as well as breaks from work.

The large-scale schools testing that underpins this thesis would not have been possible without the support of many people. Katy Shire and Rebecca Sheridan 'taught me the ropes' when I first began my PhD, allowing me to move on to managing my own school testing. This early testing was also supported by the ever-cheery presence of Katie and Joyti. I would not have been able to effectively manage my own school testing without the support of the PAC Lab interns, Ella, Flora, Imogen, and Megan. Ella, in particular, efficiently managed testing for me on days where I was unavailable. I would also like the thank the many undergraduate project students who assisted with data collection for my PhD. Finally, this thesis would not have been possible without the kind cooperation of the schools where the data were collected.

During my PhD I have had the privilege to collaborate with Alan Baddeley and Graham Hitch. Both Alan and Graham brought a unique perspective to the work, as well as the field in general. I also had the opportunity to visit Bob Logie at the University of Edinburgh for a lab visit during my PhD. My discussions with Bob have always been interesting and illuminating, helping be clarify my thinking on a range of issues. The Experimental Psychology Society made this visit possible with their generous funding.

All the work in this thesis relies on free open source software. Without the work of Jon Peirce in creating PsychoPy, this thesis would be drastically diminished. The analysis, and indeed writing of this thesis, was all made possible by R and the amazing range of packages created for it. The authors of this software should be commended, as well as the fantastic R community.

Personally, I would like to thank my parents for always supporting me to follow my interests, wherever they took me. My dad, Miles, has been a calming influence to live with throughout my PhD, always keeping me well fed. Finally, I would not have been able to complete my PhD without the love and support of Miriam, who has helped me through all of my most stressful moments.

Abstract

Working memory (WM) – the ability to store information over short periods of time in support of complex cognition – is implicated in a range of cognitive processes and developmental milestones. Given the importance of WM, it is vital that tools exist to rapidly but effectively assess this set of abilities. In Chapter 2 the development of computerised set of measures is described that we designed to facilitate rapid group testing in a school setting. These aims were defined by links with the *Born in Bradford* longitudinal cohort study. The rest of the thesis investigates how WM might be supported in children, a critical line of research considering the developmental implications of WM difficulties. In Chapter 3 the first investigation of the ability of children to prioritise serial positions within a visual sequence is presented. Children were instructed to try especially hard to remember either the first or third item in three-item sequences of shapes. Adults are consistently able to do this, resulting in superior performance for the prioritised item, at a cost to other items. Unlike adults, children did not show an ability to prioritise a particular position, when instructed to do so. Chapter 3 also includes a novel individual difference analysis that further clarifies the automaticity of recency effects in visual WM. Following the absence of prioritisation effects in Chapter 3, an alternative approach informed by embodied theories of cognition was taken in Chapter 4. Participants were presented with a WM task where the task environment was either structured pseudorandomly or in a task-relevant manner. This task-relevant organisation was consistently beneficial for children with low WM, such that they performed better than when the environment was structured. Children's metacognitive understanding of the experimental manipulation was also investigated, highlighting the important of metacognitive factors to supporting WM in children.

Contents

List of Tables

List of Figures

Abbreviations

Term	Abbreviation
Attention Deficit Hyperactivity Disorder	ADHD
Backward Digit Recall	BDR
Executive Function	EF
Forward Digit Recall	FDR
Long-term memory	LTM
Odd-one-out	OOO
Short-term memory	STM
Socioeconomic status	SES
Working Memory	WM

Chapter 1

Introduction

This thesis presents a set of studies that investigate how working memory (WM) performance can be assessed and supported in primary school aged children. Chapter 2 describes a set of WM measures developed to facilitate the effective assessment of working memory, and presents an analysis of performance from a large sample of children. These measures are also employed in individual difference analyses in subsequent chapters. The experiments presented in Chapter 3 focus on children's ability to engage attention to prioritise a particular item within a visual sequence. Chapter 4 presents an investigation of the role of structure in the environment in supporting performance.

The general introduction begins by defining WM before detailing theoretical perspectives most relevant to the following chapters. Some of the common approaches to measuring WM are described followed by discussion of the importance of WM in development, as well as its relationship with executive functions. Finally WM difficulties in children are characterised and the future avenues to supporting WM in children investigated in this thesis are outlined.

1.1 Defining working memory

Working memory is a core component of the cognitive architecture, providing the means to store information over short periods of time in support of complex cognition. When defining working memory (WM) it is important to distinguish between definitions of what WM *is* and theoretical accounts of how WM operates (Cowan, 2016a). The definition of WM employed also has implications for what constitutes a measure of WM (see below). Definitions of WM differ in a number of ways but the most important axis of difference here concerns the role of storage versus processing. One definition of WM identifies it with temporary passive storage in combination with a range of processes to maintain, update and manipulate information (Cowan, 2016a). This is a multiple-component definition of WM in that it assumes that WM is constituted by separate components serving storage in different sensory codes, providing a broader replacement for the pre-existing idea of short-term memory (Baddeley and Hitch, 1974); all short-term storage involves the WM system proper, making the construct of short-term memory superfluous. However, as Cowan (2016a) argues, the notion of simple storage has re-emerged through work such as that of Daneman and Carpenter (1980). Here WM is defined as a system that handles storage and processing over short periods of time. Under this 'storage and processing' definition WM is involved only in cases where storage and processing are combined. The notion of short-term memory is then reintroduced to describe short-term passive storage in isolation (Cowan, 2016a). Further movement away from identifying the core of WM with storage comes from the perspective that WM captures the contribution executive attention makes to memory over short periods (Cowan, 2016a; Engle, 2002). Whilst none of the accounts that will frame the discussion here include long-term memory (LTM) proper within their definition of WM, this does apply to some accounts (Cowan, 2016a; Ericsson and Kintsch, 1995; Ericsson and Delaney, 1999). Ericsson and Kintsch (1995) extend WM to include structures in LTM that are accessed via retrieval cues in short-term memory. Such a view proposes a more fundamental role of LTM in WM than simply the activation of features stored in LTM (Cowan, 2005, see below).

The definition of WM that will be used here is what Cowan (2016a) calls his 'generic definition'. Such a definition is aimed to be as minimalist as possible, avoiding unnecessary theoretical baggage. This definition says that WM describes the limited-capacity cognitive system that holds information "temporarily in a heightened state of availability for use in ongoing information processing" (Cowan, 2016a, p. 2). The key difference between this definition of WM and a multiple-component definition is that executive processes are included in WM if and only if they support the maintenance of information in this state of heightened availability. The manipulation of that information is not, by itself, included within the definition of WM. Moreover, no assumptions are made about WM including distinct components.

Moving forward with this 'generic definition' of working memory, a number of *theoretical* distinctions can be drawn. These concern the types of storage that underpin WM, how information is stored within a particular 'type' of storage, and how it is maintained in WM. Firstly, we can distinguish between attentional and passive storage. Attentional storage occurs where executive resources are required to maintain the heightened activation of some information. Passive storage, on the other hand, occurs when information remains available for a period of time in absence of rehearsal or strategic mechanisms (Cowan, 2016a). Among theories that share a similar view about the *type* of storage that occurs in WM there is disagreement about how that storage is achieved. Some argue that items can be stored in fixed resolution slots (Cowan, 2005; Zhang and Luck, 2008), while others propose a continuous resource underlies attentional storage (Bays, 2015; Fallon et al., 2016; Ma et al., 2014). Further distinctions can be drawn between the proposed maintenance mechanisms in WM, such as attentional refreshing (Barrouillet et al., 2004), modality-specific rehearsal (Baddeley, 2007), or resisting interference (Engle, 2002; Oberauer et al., 2016).

1.2 Theories of working memory

Though something akin to WM has been described dating as far back as John Locke (Logie, 1996), Baddeley and Hitch (1974) are credited with providing the first theoretical account of WM. They attempted to demonstrate the need for multiple components in WM using dual-task paradigms. Such paradigms were used to ‘simulate’ the dissociations in performance observed with neurological patients in healthy young adults (Baddeley, 2012). Performance on two tasks completed separately is compared to performance when the tasks are performed concurrently, with drops in performance when the tasks are combined suggesting some overlap in the processes underpinning performance. Equally, minimal drops in performance in dual-task conditions are taken to indicate a dissociation in the underlying processes (Logie, 2011). The central proposal of the 1974 chapter was that a unified short-term store should be replaced with a working memory system that includes a passive phonemic store and a central processor for general processing, such as rehearsal mechanisms, and storage where necessary (Baddeley and Hitch, 1974). An equivalent passive store for visual information is suggested but none of the presented data directly speak to this proposal.

Much of the theory often attributed to the 1974 chapter in fact comes much later (Baddeley, 1986; Logie and Cowan, 2015). The 1986 Multiple Component Model had three components; two passive stores and a central executive. The phonological loop describes a passive store for verbal and auditory information. Information can be maintained in this store via active rehearsal from the articulatory loop. This active rehearsal process is controlled by the central executive. Thus, while the storage of information itself is passive there is an active element to maintaining its heightened availability (Cowan, 2016a). The second passive store, the visuospatial sketchpad, stores visual and spatial information. An 'inner scribe' rehearsal mechanism for this component has since been proposed (Logie, 1995, 2011). Unlike the Baddeley and Hitch (1974) formulation, the central executive was not assumed to have any storage capacity in Baddeley (1986). More recently this multiple component model has been expanded to include an episodic buffer (Baddeley, 2000). This domain-general component provides a store for integrated item, such as colour-shape bindings. It is also thought to provide the space for information from long-term memory to be integrated with information in the modality-specific stores. Storage in the episodic buffer was originally thought to be attentional rather than passive (Allen et al., 2006; Baddeley, 2000). However, later work suggested that maintaining bound items, though executively demanding, is not more demanding than maintaining features alone (Allen et al., 2006, 2014). Nevertheless, representations of bound items may be more fragile than simple features (Allen et al., 2006).

Whilst Baddeley's model is often referred to as *the* multiple component model, other multiple component models exist (Logie, 2011, 2016; Logie and Cowan, 2015). Logie (2016) rejects the need for a central executive or episodic buffer, instead arguing that tasks such as binding can be achieved from coordination between domain-specific stores. This view reasserts storage as the primary function of WM claiming it is adequately described by coordination between passive storage, without the need for a central executive (Logie, 2016).

Another influential account of WM has been provided by Cowan's embedded-processes model (Cowan, 1995, 2001, 2005, 2008, 2016a,b; Logie and Cowan, 2015). Like the account of Baddeley (2000), this framework includes both passive and attentional storage. Working memory is held to combine an activated portion of long-term memory in addition to a limited-capacity focus of attention. As information is perceived in the environment the features of items, such as colour, are automatically activated in long-term memory. These features are then more accessible than the rest of long-term memory. No differentiation is made within this passive form of storage for features from different sensory modalities. However, it has been suggested that similar features interfere more than dissimilar ones (Cowan, 2016b). More integrated items are stored in the capacity limited domain-general focus of attention, with its capacity typically estimated at approximately four items (Cowan, 2001, 2005). Within this account there are two routes by which information can enter the focus of attention (Cowan, 2016b). Firstly, an automatic route whereby attention is drawn to changes in the environment. Additionally information can deliberately enter the focus of attention through the engagement of executive processes (Cowan, 2016b). These executive processes are also crucial to overriding information that enters the focus of attention automatically.

Proactive interference provides a useful example to understand the interplay between the activated portion of long-term memory and the focus of attention (Cowan, 2016b). As a task proceeds the features of test items are activated in long-term memory. This results in a general familiarity with those features making the activation uninformative for responding to a task. In order to prevent this broad familiarity interfering with task performance participants must maintain integrated items for a specific trial in the focus of attention. Thus, performance deteriorates across a series of trials as participants have to increasingly rely on the focus of attention.

One important property of the focus of attention is its ability to ‘zoom’ in and out (Cowan, 2005). In some cases people can ‘zoom in’ on a single item to protect it from interference, while broadening the focus of attention to apprehend more information in other cases (Cowan, 1995, 2005). Others have drawn a distinction between a direct-access region and a focus of attention that is typically limited to one item (Oberauer, 2002, 2013). This view shares with Cowan the idea that WM is, in part, comprised of highly accessible information in long-term memory that can readily be brought into a focus of attention. It differs in separating Cowan’s focus of attention into a direct access region and a more limited focus of attention.

The direct access region holds a small number of temporary bindings but is limited due to the interference between items rather than an attentional resource. Thus, no fixed item limit is suggested, unlike for Cowan (Cowan, 2005). Within this direct-access region lies a focus of attention, which is also limited by interference. Again, this focus of attention is not strictly limited to a single item, but it has been shown to be typically limited in this way in computational implementations (Oberauer, 2013).

Whilst the difference between multiple component accounts and the embedded-processes framework might seem large, these differences are not as substantial as they seem (Cowan, 2005). This is particularly true with the addition of the episodic buffer to the multiple-component model (Baddeley, 2000; Cowan, 2005, 2016a). The episodic buffer does much of the work ascribed to the focus of attention within Cowan's view (Cowan, 2005). Moreover, Cowan does not deny the possibility of domain-specific storage altogether (Morey and Cowan, 2005). Instead the claim is that all storage in WM cannot be accounted for by domain specific passive storage alone (Cowan, 2005). This point has essentially been conceded by Baddeley with the addition of the episodic buffer to his multiple component model (Baddeley, 2000; Cowan, 2005). The difference between the two frameworks could then be seen as one of "level of analysis" (Cowan, 2005, pp. 48), or "emphasis and terminology" (Baddeley, 2012, pp. 20). However, some multiple component theories do remain in strong opposition to the idea of attentional storage in WM (Logie, 2011). Theories of WM such as Baddeley's or Cowan can be thought of as providing a general framework or 'map' (Baddeley, 2012) to organise our understanding of WM. Within them there are specific debates about particular processes or mechanisms that contribute to the overall performance of the system.

One difference that remains between embedded-processes and multiple-component theories concerns the status of verbal versus visual storage. For multiple-component models storage in different modalities is of essentially the same type, differing in the underlying sensory code. This is not the case for the embedded-processes model, which affords a special status to visual storage in always placing demands on domain general attention (Cowan and Morey, 2006; Morey and Bieler, 2013). Such a claim is supported by a body of work suggesting asymmetric dual task costs between verbal and visual storage (Morey and Cowan, 2004, 2005; Morey and Mall, 2012; Morey et al., 2013; Morey and Miron, 2016). The consistent finding is that visual storage is more vulnerable to a concurrent verbal load than vice versa. This would suggest that the scope of passive storage mechanisms to support visual storage is limited compared to verbal storage. As a result, the attentional demand that comes from concurrent verbal storage has a large detrimental effect on visual storage. Verbal storage, on the other hand, can be supported by more passive mechanisms when a demand is placed on central attention by visual storage.

An alternative account of WM to those discussed argues that a concept of short-term storage distinct from perception is unnecessary (Hughes et al., 2016; Macken et al., 2015, 2016). This account results from taking an embodied view of cognition (see below), which highlights the importance of the body and environment in thought. Rather than assuming stores for abstract representations, short-term memory performance can be explained by a detailed analysis of perceptual-motor systems (Hughes et al., 2016). Macken et al. (2015) identify three factors that contribute to performance on WM tasks, and therefore the observed capacity limits. Firstly, the *repertoire* describes the set of cognitive and perceptual-motor skills at the disposal of an individual. The *materials* refer to stimuli a given task must be completed with, such as spoken letters or coloured shapes. Finally, the *task* describes what a participant is required to do with the materials and repertoire of skills available to them, such as recall them after a short delay. This approach is described as object-orientated and that the objects of processing in a given situation result from the interaction between these three factors (Macken et al., 2015). The objects of processing can change by manipulating any one of these factors. For example, Hughes et al. (2016) manipulated the way that spoken sequences of letters were presented to alter the perceptual objects that participants form. This would be an example of a manipulation to the materials such that the objects of perception change. In another experiment Hughes et al. (2016) demonstrated how limiting the repertoire available to participants, through articulatory suppression, changed performance in line with a perceptual-motor account of short-term memory. Crucially, Hughes et al. (2016) reject the need for short-term storage distinct from the perceptual-motor processes arguing that a thorough analysis of the perceptual-motor affordances of a task, and the processes that contribute makes 'abstract' short-term storage superfluous (Hughes et al., 2016; Jones and Macken, 2015; Macken et al., 2015, 2016).

1.2.1 How is information maintained in working memory

In addition to the question of how information is stored in WM one can ask how it is maintained. Here the distinction is not between the particular set of constructs that serve storage (modality-specific, domain general, etc.), but the processes that keep information in those stores accessible. This question can be further divided to ask both why information must be maintained and then how that is achieved. Information must be maintained in WM in order to remain in a highly accessible state. However, there is ongoing debate about why exactly the accessibility of information in WM reduces over time. Oberauer et al. (2016) discuss three alternative mechanisms that place limits on WM: decay, resource, and interference. A decay account (e.g. Barrouillet et al., 2004) would say that over time representations of information in WM deteriorate and becomes less accessible Oberauer et al. (2016). Nevertheless, Oberauer et al. (2016) identify a range of phenomena with which decay accounts struggle, such as the improvement in recall observed with heterogeneous memoranda. Participants are better able to recall heterogeneous memoranda compared with homogeneous memoranda. Decay, *per* *se*, struggled to explain this observation because all items in memory are assumed to decay at the same rate.

Resource accounts of WM limitations (e.g. Ma et al., 2014) claim items in WM share a storage resource resulting in a reduction in the availability of previously presented material. As new material is presented the resource is shared among these items leading to previously presented items becoming less available. Finally, Oberauer et al. (2016) identify interference as a possible limit on WM. This view rejects the claim that items in WM share a common resource or decay over time. Instead, memory is limited because the features of items are confused, superposed, or overwritten (Oberauer et al., 2016). The specific capacity limits observed (e.g. 4-items, Cowan, 2005) result from the way interference operates in practice rather than a primary resource limitation (Oberauer, 2013). Importantly, full theories of WM often combine these factors. For example, Cowan's focus of attention represents a limited storage resource, but similar features in the activated portion of long-term memory are thought to interfere (Cowan, 2005, 2016b). Equally, models of WM that propose decay as the primary limitation also include interference (Hurlstone et al., 2014; Oberauer et al., 2016).

Given these limitations on WM, a number of mechanisms for maintaining information have been proposed. From its inception, Baddeley's multiple component model included articulatory rehearsal as a mechanism for maintenance (Baddeley and Hitch, 1974; Baddeley, 2012). Such articulatory rehearsal can be achieved both subvocally and overtly (Baddeley, 2012). Other multiple component theories have proposals for visuospatial rehearsal (Logie, 2011), but the exact mechanisms remain unclear (Baddeley, 2012). More recently refreshing has been proposed as a maintenance mechanism in the time-based resource sharing model (Barrouillet et al., 2004, 2009). Like other accounts of WM, the time-based resource sharing model assumes that representations in WM decay over time (Oberauer et al., 2016). Refreshing describes the situation where executive attention is used to increase the activation of information in WM. Thus, if attention is occupied by a demanding secondary task, information in WM will become less available over time (Barrouillet et al., 2004). Others argue that resisting interference and removing irrelevant information from WM are critical to maintenance (Engle et al., 1999; Engle, 2002; Oberauer, 2013; Oberauer et al., 2016).

1.3 Relationship of working memory to executive functions & intelligence

For some theorists, executive functions (EFs) such as planning, set-shifting or inhibition are included within their definition of WM (Baddeley, 2012; Cowan, 2016b). Under such a definition one could still ask how tasks that primarily involve storage in WM relate to executive abilities, measured by tasks with minimal storage requirements. Here the question would concern the internal relationship between components of WM. The ‘generic definition’ of WM (Cowan, 2016a) includes executive processes in WM if and only if they are being used for storage. Nevertheless, a relationship between WM and other cognitive abilities is implied in this definition as WM offers short-term storage in support of complex cognition.

Executive functions describe a broad range of abilities that underpin ‘higher cognition’ associated with frontal lobe activity (Smith and Jonides, 1999). Smith and Jonides (1999) identify a set of functions commonly held to constitute ‘executive functions’: attention and inhibition, task management/shifting, planning, monitoring, and coding. Attention and inhibition describe the functions that allow us to focus on relevant information, paralleling the description of executive control in the WM literature (e.g. Engle, 2002). Monitoring involves keeping track of the contents of WM to guide task performance by, for example, checking the next step in a task. Coding makes the most explicit reference to WM by describing the process of encoding information into WM. Updating, shifting, and inhibition have been described as more ‘low level’ executive functions (e.g. Miyake et al., 2000), featuring more frequently in work investigating the structure of EFs and their relationship to other abilities (Friedman, Naomi et al., 2006; Friedman et al., 2007, 2012; Miyake et al., 2000; Miyake and Friedman, 2012). While all storage in WM involves updating the contents of WM, the term is used more restrictively in the EF literature. The executive function ‘updating’ does not describe cases where the contents of WM are ‘updated’ by a prepotent response to some change in the environment, or by information becoming less accessible over time. Instead it describes engaging executive resources to actively update the contents of WM (Miyake and Friedman, 2012).

The relationship between the components that make up EFs can be represented in different complementary ways within a latent variable framework (Miyake and Friedman, 2012). One possibility is to represent updating, shifting and inhibition as distinct but highly correlated latent variables in a structural equation model. Another proposal captures the relationship between EF measures by having factors that represent updating-specific and shifting-specific variance (Miyake and Friedman, 2012). In addition a ‘Common EF’ factor captures the variance common to inhibition measures as well as related to all EF measures. This general inhibitory ability is highly relevant to WM tasks where participants have to inhibit competing representations or distracting stimuli. Thus, measures of WM could be described as reflecting specific updating abilities and ‘Common EF’. To successfully complete a WM task participants must be able to update the information that is represented in WM as well as inhibit a tendency to represent the ‘wrong’ information. This proposal has clear parallels with those theories of WM that emphasise the importance of domain-general executive control to explain the relationship between WM and general abilities (Engle et al., 1999); executive control involves engaging general cognitive resource to determine what does and does not enter WM.

One influential application of a latent variable approach distinguished between

WM and short-term memory (STM) in measuring their relationship to fluid intelligence (Engle et al., 1999). Engle et al. (1999) take the view that WM capacity describes the combination of short-term memory and the role of executive attention in, for example, resisting interference. They found that, while WM and STM were highly related, only WM showed a significant relationship with fluid intelligence. In subsequent analyses the executive control component of WM was more strongly related to intelligence than the component of WM reflecting storage (Engle et al., 1999). This suggests that executive attention drives the relationship between WM and fluid intelligence.

An alternative approach to investigating how WM and other cognitive abilities covary has been to compare low and high capacity participants (Conway et al., 2001; Engle, 2002; Kane et al., 2001; Kane and Engle, 2003). (Conway et al., 2001) presented low and high capacity adults with a task where two streams of words were presented over headphones. Participants were instructed to listen to and repeat one of the streams aloud. After a period of shadowing the participant's name was presented in the unattended stream. 65% of low span participants reported hearing their name compared to only 20% of high span individuals. This result suggests that low span individuals struggle to deal with distraction. This conclusion is further supported by the finding that low WM participants are less able to resist attention-drawing distractors (Kane et al., 2001). Similarly, low WM adults are less able to resist response competition and maintain goal information (Kane and Engle, 2003). Taken together these results suggest that what differs between high and low WM individuals is not storage capacity directly (Engle, 2002). Rather, the observed differences on measures that require storage results from differences in the attentional control mechanism required to store information in the face of interference (Engle et al., 1999; Engle, 2002; Kane et al., 2001).

1.4 Measures of working memory

Various approaches have been taken in measuring WM. The measures used within experimental settings are numerous, ranging from verbal immediate serial recall (e.g. Baddeley and Hitch, 1974) to visual precision measures (e.g. Bays et al., 2009). Here the focus will be on measures common within the individual differences literature. Chapter 2 provides further details on measures of WM, particularly those used with children.

A common distinction is made between measures that emphasise storage versus those that combine storage and processing (Baddeley, 2007; Case et al., 1982; Cowan, 2005, 2016a; Daneman and Carpenter, 1980; Engle et al., 1999; Engle, 2002). As previously discussed, many argue that a processing component is crucial for WM measures to predict domain-general abilities (Engle et al., 1999; but see Cowan et al., 2006). While acknowledging the role of attentional control in short-term storage, 'simple' and 'complex' WM measure will be distinguished here. 'Simple WM' will be used to describe tasks where the primary demand is the storage of information. For example, a forward digit recall task (see Chapter 2) requires participants to remember and recall a series of digits in the order they are presented. Performance on such tasks primarily reflects the ability to store information over short periods, placing minimal demands on executive attention (Engle, 2002; Engle et al., 1999; St Clair-Thompson, 2010). Participants can successfully perform such tasks by making use of domain-specific rehearsal mechanisms or chunking (Engle et al., 1999; Engle, 2002). In contrast 'complex WM' tasks combine storage with some concurrent processing such as counting objects (Case et al., 1982) or reading (Daneman and Carpenter, 1980). The two original complex span measures were reading and listening span (Daneman and Carpenter, 1980). For reading span participants were required to read a series of sentences, recalling the final word of each sentence at the end. This measure was then adapted such that participants had to judge whether the sentences were true or false in additional to remembering the final word (Daneman and Carpenter, 1980, Experiment 2). Listening span also involved participants judging the veracity of sentences prior to recalling the final word, but by listening to the sentences. This procedure has also been applied to counting to create a counting span measure (Case et al., 1982). Here participants have to count the instances of a target shape in an array of distractors before recalling the individual array counts in order. Operation span describes a procedure where participants have to indicate whether mathematical equations are correct while remembering a word presented alongside each equation (Engle et al., 1999). What all these measures share is that participants are required to carry out some concurrent processing while remembering information. The critical factor to the usefulness of these measures in predicting domain-general abilities is the need to maintain information in face of interference, rather than the particular nature or difficulty of the processing task (Engle et al., 1999; Engle, 2002). Better performance on complex span measure reflects a "greater ability to control attention, not a larger memory store" (Engle, 2002, p. 20).

It is noteworthy that the tasks described above all involve participants storing verbal material. While complex visual tasks do exist (e.g. Russell et al., 1996), the necessity of concurrent processing for a measure to tap executive attention is challenged by research in the visual domain (Cowan and Morey, 2006; Morey and Bieler, 2013; Zokaei et al., 2014). The suggestion that storing visual information always involves general attentional resources would mean that simple visual measures should still relate to general abilities. Indeed, analysis presented in Chapter 2 confirms this prediction. This idea that all visual WM tasks involve general attention is easy to account for using the perceptual-motor approach of Jones and colleagues (Hughes et al., 2016; Macken et al., 2015). Verbal sequences afford grouping in terms of rhythm or sound-source that support storage in the absence of concurrent processing. This is not the case for the standard visual tasks as spatial cues dominate grouping in vision (Hughes et al., 2016). Moreover, when spatial or featural cues are provided visual WM performance is improved (Morey et al., 2015; Wang et al., 2016; Woodman et al., 2003).

While it is typically argued that complex WM measures are required to observe relationships with domain-general abilities, ‘simple’ storage measures have been shown to relate to IQ alongside measures of attentional control (Cowan et al., 2006). This is explained by arguing that when mechanisms such as chunking and rehearsal are effectively blocked even simple storage can reflect a domain general scope of attention (Cowan, 2005; Cowan et al., 2006). For example, Cowan et al. (2006) showed that measures of WM that include only a storage component predicted intelligence in children too young to effectively rehearse or make use of chunking strategies. This relationship was observed even where measures of attentional control were included in the analysis. For adults, in contrast, a dual task measure of attentional control and a visual measure of attentional storage predicted intelligence, whereas simple storage measures did not (Cowan et al., 2006).

1.5 Working memory and executive function development

1.5.1 Working memory development

The importance of WM in development is demonstrated by its utility in predicting a range of outcomes, most notably educational attainment (Alloway and Alloway, 2010; Cowan, 2013; Cragg and Gilmore, 2014; Gathercole et al., 2004b; LeFevre et al., 2013; Monette et al., 2011). Alloway and Alloway (2010) showed that verbal WM measured at 4 to 5-years-old, alongside IQ, predicted measures of literacy and numeracy at 10 to 11-years-old. Furthermore, WM explained more variance in academic attainment than IQ, though the WM and IQ measures were not very closely matched.

The relationship between WM and academic attainment can be characterised in at least two ways. Firstly, at the level of cognitive mechanisms, working memory is engaged in knowledge acquisition, such as for language (Baddeley, 2007; Gathercole and Baddeley, 1993; Gathercole, 1999). Secondly, working memory is involved in general success in the classroom (Gathercole and Alloway, 2008). For example, working memory is needed to hold in mind and follow a series of instructions, such as those given by a teacher (Allen and Waterman, 2015; Gathercole et al., 2008; Jaroslawska et al., 2016; Yang et al., 2014, 2015b). Gathercole et al. (2008) developed WM tasks that would serve as analogues of classroom activities to investigate the involvement of WM in success in the classroom. They showed that children with low WM performed worse on these tasks, even when controlling for IQ and other processing abilities. Complex working memory was also related to a task requiring participants to follow sequences of instructions in an individual difference analysis. This suggests a fundamental role for WM in classroom activities.

Performance on WM tasks improves across childhood only reaching an asymptote in adolescence, with performance on complex measures reaching a plateau later in development than simple measures (Gathercole, 1999; Gathercole and Alloway, 2004). A range of processes have been said to contribute to the increases in WM storage capacity with development (Cowan, 2016b; Gathercole, 1999) such as rehearsal mechanisms (Hitch et al., 1988), executive control strategies (Morey et al., 2017), or processing efficiency (Case et al., 1982). Case et al. (1982) demonstrated relationships between word span and word repetition speed, and counting span and counting speed. These relationships between memory and processing did not remain when the efficiency of processing was controlled for. This suggests a role for the efficiency of underlying processing abilities in accounting for the increases in WM span observed with age (Case et al., 1982). These results challenge a view that accounts for all capacity increases in terms of rehearsal, though it may still play a role (Case et al., 1982). Case et al. (1982) suggest that with development, increases in processing efficiency (and therefore span) might free up cognitive resource for metacognitive strategy selection.

The role of rehearsal in WM development has been most widely investigated in the verbal domain (Elliott et al., 2016; Flavell et al., 1966; Hitch et al., 1988). Flavell et al. (1966) demonstrated that younger children (below 7-years-old) show little evidence of verbal rehearsal in memory tasks. Hitch et al. (1988) used a range of manipulations to demonstrate that 5-year-olds rely far more heavily on visual storage for nameable drawings, compared to 10-year-olds. With increasing age children appear to rely on verbal recoding, possibly underpinned by rehearsal, to store visual stimuli. More recently the importance of rehearsal in the development of verbal WM has been questioned from a number of directions (Jarrold and Hall, 2013). Some of the markers of rehearsal assumed to be absent in children, such as a correlation between articulation rate or word-length effects, have been observed in children under 6 years old (Hitch et al., 1989; Jarrold et al., 2000). However, some methodological issues remain outstanding due to limitations on what measures are possible to take with young children (Jarrold and Hall, 2013).

Cowan (2016b) reviews alternative accounts for the observed increases in WM capacity with development, including processing speed and rehearsal as well as increases in knowledge, filtering abilities, and the use of context. While these factors may play a role in performance on WM tasks they fail to explain genuine capacity increases (Cowan, 2016b). For example, while it may be that processing becomes more efficient with age, this fact cannot explain all the observed development in WM capacity. Ultimately improvements in WM with age will result from the development of task-specific processes and domain-general abilities. The contribution of multiple mechanisms to WM limitations in children is exemplified by the duplex-mechanism account of interference in verbal STM (Elliott et al., 2016; Hughes, 2014); in verbal STM interference occurs due to the disruption of task-specific processes or by attentional diversion. Elliott et al.

(2016) showed that children's more pronounced susceptibility to distraction is driven by attentional diversion, though task-specific interference also occurred. The underdevelopment of executive control in children has been described by drawing a comparison with low WM adults (Cowan et al., 2006). When instructed to selectively attend to one of two streams of memoranda, 10 year-olds showed similar performance levels for the attended and unattended streams, suggesting a difficulty with appropriately controlling attention. The minimal difference in performance between the attended and unattended stimuli suggests that 10 year-olds struggle to direct attention to improve memory.

Specifically within the visual domain, there has been much debate about how WM capacity develops (Cowan, 2016b; Simmering, 2016, for a review), and its capacity at different ages. This debate stems from differences in capacity estimates when using different measures of visual WM. Some researchers have estimated that infants have a visual WM capacity close to the typical adult capacity of four items (Ross-Sheehy et al., 2003). These studies make this claim using change-preference measures of visual WM. As young children cannot verbally report memory, other techniques must be used to infer the contents of their memory. With change-preference measures researchers make use of the fact infants tend to look longer at novel stimuli. If two stimuli are presented, one that is novel and one that is 'in memory', then they should look at the novel stimuli for longer. While Ross-Sheehy et al. (2003) demonstrated that 10-month-olds show preferential looking to novel stimuli for up to four items they only claimed this finding represented a 'relative' difference between younger and older children. However, this reluctance from the original authors to claim infants have a capacity of four items has not been reflected in later accounts of the study (see Simmering, 2016, for discussion).

With older children, capacity can be estimated using the same change-detection procedures employed with adults. Typically, participants are presented with a simultaneous array of coloured shapes and required to indicate whether a second array is the same or different to the first array (Rouder et al., 2011). Using an adaptation of this procedure suitable for younger children, capacity was estimated as two items at 3-years-old increasing to four items by 7 to 8-years-old (Simmering, 2012, 2016). Using a more conventional procedure the capacity estimate for 9 to 10-year-olds was 2.4 and 3.6 for adults (Cowan et al., 2006).

A number of attempts have been made to account for this divergence in capacity estimates; it seems implausible to think that visual WM capacity *decreases* with age. Some accounts attempt to argue that change-preference measures are not really 'proper' measures of visual WM capacity. This account struggles with the finding that measures of visual change-preference predict later WM performance and IQ (see Simmering, 2016, for discussion). A more unified account of the two approaches to estimating capacity is offered by Simmering (2016) specifying a single computational model that can explain the experimental data with the same set of mechanisms. This approach is preferable in that only a single set of mechanisms needs to be invoked rather than making *ad hoc* suggestions of the mechanisms that contribute to different tasks. While the computational details of this model are not germane here, a number of key features are useful to highlight what develops with respect to visual WM. This model takes as a starting point that the various changes in, for example, encoding speed, inhibition, and attention that occur with development result in increasingly robust memory representations (Simmering, 2016). Moreover, the approach is 'dynamic' in the sense that stability emerges from the interaction between a number of underlying processes. In contrast to accounts that propose a single source for WM errors, dynamic accounts hold that errors can arise from encoding, maintenance, comparison, and response processes (Simmering, 2016). Importantly, capacity is not fixed within dynamic models as the number of items that the computational architecture can represent varies with a range of factors (Simmering, 2016). This approach shares features with the view of Macken et al. (2015) in centring the range of factors that contribute to WM performance rather than searching for some ultimate capacity estimate. What makes the model of Simmering (2016) useful is that it explains how various factors contribute to performance within a unified computational framework.

In addition to reporting improvements in WM performance with development the *structure* of WM has been investigated in children (Alloway et al., 2004; Gathercole et al., 2004a; Gray et al., 2016, for a review). Gathercole et al. (2004a) investigated the structure of WM in children testing a set of models informed by Baddeley's multiple-component model. They found that in children as young as 6 to 7 years-old WM performance was best described by a three-factor structure capturing phonological storage, visuospatial storage, and tasks involving storage and processing. These factors were interpreted as corresponding to the phonological loop, visuospatial sketchpad and central executive, respectively. This structure provided a good fit of the data for children up to 15-years-old, despite WM performance improving across development (Gathercole et al., 2004a). One noteworthy feature of the analysis was the pattern of covariance between the factors. The covariance between the central executive and phonological loop factors varied between .73 and .92. For the phonological loop and visuospatial sketchpad factors the covariance was markedly smaller ranging from .32 to .41. Finally, the covariance between the central executive and visuospatial sketchpad factors varied from .68 to .85. This pattern broadly supports the claim that phonological and visuospatial storage are reasonably distinct with the central executive controlling both stores. The large covariance between the central executive and phonological loop likely partly reflects the fact that two of the storage-and-processing tasks involved verbal storage.

Gray et al. (2016) report a similar study with a group of 5-year-olds, except that the structural models tested were informed by both the multiple-component and embedded processes models. A range of WM measures were used to test 3 alternative models: (i) Cowan's embedded processes model with constructs for the focus of attention, central executive and domain-specific phonological processes;

(ii) a multiple-component model with phonological loop, visuospatial sketchpad, and central executive components; and (iii) an updated multiple-component model including the episodic buffer (Baddeley, 2000). The embedded processes model provided the best fit of the data, though the 3-factor multiple-component model also provided a good fit. This is unsurprising given the similarity between the two models (Gray et al., 2016). Nevertheless, differences between the models highlighted by Gray et al. (2016) are useful to understanding the different strands of experimental work they have inspired. Within the multiple-component model, phonological and visual storage have the same status, in that both are served by modality specific stores under the control of the central executive. This is not the case for the embedded processes model (Gray et al., 2016). Visual storage is held to always be attentional demanding (Morey et al., 2013) whereas phonological storage can be automated through rehearsal processes. The differential 'status' of visual and verbal storage in working memory was supported by the superior fit of the embedded-processes model (Gray et al., 2016).

As with adults, WM and intelligence are related in children (e.g. Alloway and Alloway, 2010; Gray et al., 2016). For example, Gray et al. (2016) showed that latent variables that reflect executive processing and, in particular, attentional storage in WM related to both fluid intelligence and visual processing. Here attentional storage refers to effortful storage in the focus of attention as opposed to the automatic activation of features in long-term memory. This role for attentional storage and central executive processes highlighted by Gray et al. (2016) runs counter to earlier work with adults that identified executive control as driving the relationship with intelligence (e.g. Engle et al., 1999). The importance of attentional storage for predicting intelligence in children is further supported by Cowan et al. (2006). They found that measures of verbal storage, in the absence of concurrent processing, predicted intelligence in children but not in adults. A visual array measure of storage in the focus of attention predicted intelligence in both adults and children (Cowan et al., 2006). Finally, a measure of attentional control, reflecting the ability to selectively attend to a given stream, predicted intelligence in adults but not children. Cowan et al. (2006) distinguish between the ‘scope’ and ‘control’ of attention in describing these results. The scope of attention describes the amount of information that can be apprehended and held in a state of heighted activation. Cowan et al. (2006) also refer to this ability as the capacity or focus of attention. The control of attention, in contrast, involves using attention to resist interference from competing information within a trial, or proactively from previous trials. These results bolster the idea that storage can be related to intelligence where rehearsal mechanisms are unavailable due to underdevelopment or the specific memoranda (e.g. visual arrays) (Cowan et al., 2006).

1.5.2 The development of executive functions

Concurrent with changes in WM capacity, executive functions (EFs) develop across childhood and into adolescence (Friedman et al., 2012; Jurado and Rosselli, 2007; Miyake and Friedman, 2012; Zelazo et al., 2013). This has already been suggested in highlighting the role of executive control in WM development. Executive control involves engaging executive functions to influence the contents of WM. Indeed, as many models of WM include a notion of attentional storage, development of attentional abilities is intimately tied to the development of storage capacity. Regardless of the particular taxonomy of EFs employed it is clear that these abilities develop across childhood, not reaching adult-like levels until late adolescence (Jurado and Rosselli, 2007; Waszak et al., 2010). While EFs develop across childhood, stable individual-differences emerge early in life and persist into adolescence (Friedman et al., 2007, 2012). Within the developmental literature updating, shifting, and inhibition are frequently studied (e.g. Blakey et al., 2016; Friedman, Naomi et al., 2006; Friedman et al., 2007, 2012). For example, Friedman et al. (2012) used measures of inhibition, updating, and shifting to investigate the relationship between early self-restraint and later executive functions, observing strong relationships, particularly for inhibition.

Davidson et al. (2006) used a range of tasks to orthogonally manipulate inhibitory requirements and memory load with a sample of children from 4 to 13 years, and young adults. They found that the ability to inhibit prepotent responses and the irrelevant features of a stimulus increased with age. In the youngest children (4-6 years) the reaction time costs to increasing inhibitory demands were greater than the costs associated with increasing memory load, suggesting a primary role for inhibition at this age. Despite these inhibitory abilities developing in early childhood there remained tasks where even 13 year-olds were not performing at adult levels. In addition the developmental trajectory observed for a given task depended on the combination of memory and inhibitory demands. This is important to highlighting how the prolonged development of executive functions will vary considerably with the particular task setting.

As with WM, developmental changes in the structure of executive functions have been investigated. Executive functions are consistently found to be undifferentiated in 3 year-olds (Fuhs and Day, 2011; Wiebe et al., 2008, 2011). When the performance of 3-year-olds on a range of executive function tasks is analysed specific factors for updating, shifting, and inhibition do not emerge. The structure of executive function then takes a two-factor form in childhood where updating and shifting can be distinguish (Lee et al., 2012, 2013; der Ven et al., 2012). Finally, by adolescence the structure begins to resemble the 3-factor structure observed in adults (Lee et al., 2013; Rose et al., 2011; Wu et al., 2011).

Alongside developments in executive functions a role for metacognition or ‘meta-control’ has been proposed (Chevalier et al., 2014, 2015; Chevalier, 2015b). This approach takes as a starting point the distinction between reactive and proactive control strategies (Braver, 2012). Reactive control describes an approach to a task where participants respond based on the information that is currently presented. This contrasts with proactive control where responses are organised in anticipation of future events. Chevalier et al. (2014) contrasted children's use of reactive and proactive control strategies with a working memory task. They compared the response times for the first item in a sequence with later items for children aged 3 to 10 years old, as well as young adults. Children under 6 years were quicker to respond for the first item than for subsequent items. This pattern was reversed for older children and adults. These results highlight a shift from reactive to proactive control strategies with age (Chevalier et al., 2014). Younger children respond reactively by quickly recalling the first item but are then slower as subsequent items must be reactively recalled in turn. Older children and adults, on the other hand, proactively plan their recall of a number of items in the sequence such that recall of the first item is slower (Chevalier et al., 2014).

Chevalier (2015a) extended this distinction between reactive and proactive control by exploring the effect of task demands on strategy choice. Children were presented with a switching task where objects had to be sorted based on two categories. In one condition the sorting category was cued prior to the target presentation, as well as concurrent with the target. In another condition the cue was presented only prior to the target. 10-year-olds demonstrated a proactive strategy for a range of outcome measures wherever it was possible (i.e. when a cue was presented prior to the target). 5-year-olds only showed a proactive approach when the sorting category was not available concurrently with the target. When the cue information was present alongside the target younger children simply responded reactively. These results challenge the idea that young children fail to engage optimal executive control strategies due to a general under-development of their executive functions. Instead they suggest that task demands influence the strategy choice made by children of different ages (Chevalier, 2015a,b). The selection of optimal strategies will develop with age as more complex proactive strategies become less effortful to implement and greater experience with different strategies has occurred (Chevalier, 2015a,b). These 'meta-control' abilities are related to the general development of executive functions in that executive abilities will influence the relative costs of different strategies (Chevalier, 2015b). 5-year-olds were able to engage a proactive control strategy where a reactive control strategy was made more costly with the omission of a cue alongside the target object. This work highlights that, while executive functions influence the repertoire of strategies, metacognitive factors are important in accounting for when certain strategies are spontaneously used.

1.5.3 Working memory difficulties

A number of children have difficulties with WM, which can have a profound impact. These difficulties can manifest in several ways (Gathercole and Alloway, 2008) and are potentially mistaken as inattentiveness by teachers (Alloway et al., 2009a). For example, a class could be instructed to 'finish what they are writing, put their pencils back in the pots, put their books away, and sit down on the carpet'. A child with poor working memory might finish what they are writing, notice everyone else is sat on the carpet, and go and sit with them, having forgotten the other instructions (Gathercole and Alloway, 2008). Such behaviour could result from a memory failure rather than simply 'not paying attention'. Children with poor WM also perform poorly on a range of executive function tasks (Alloway et al., 2009a; Holmes et al., 2014; St Clair-Thompson, 2011). This speaks both to the importance of WM developmentally and the interelatedness of WM and executive functions. Furthermore, Alloway et al. (2009a) observed that deficits for low WM children on measures of reading and maths increase over time, despite WM scores remaining stable. These findings suggest that WM difficulties have cascading effects on development such that low WM children become increasingly impaired over time (Alloway et al., 2009a). This suggestion reflects a neuroconstructivist (Karmiloff-Smith, 2009) or neo-Piagetian (Case, 1987; Cowan, 2016b) approach to development. A neuroconstructivist view argues that the brain is a-modular at birth with broad regions processing a wide range of inputs (Karmiloff-Smith, 2009). Over development, regions of the brain become specialised to process particular inputs, resulting in the modularity observed in adults (Karmiloff-Smith, 2009). This specialisation of the brain is highly interactive such that a deficit in one process can have cascading effects over time. Thus, a difficulty with working memory will affect related cognitive systems. A neo-Piagetian view argues that cognitive systems such as attention and memory develop over time. This development then allows for more complex learning and knowledge acquisition (Cowan, 2016b). Children with poor WM who are unable to hold the information required to complete, for example, complex maths problems will then struggle to acquire this knowledge.

Comparisons can be drawn between the cognitive profile of children with poor WM and those with ADHD (Holmes et al., 2014). Holmes et al. (2014) investigated these similarities by comparing children with ADHD, low WM children and average WM children on a range of measures. Whilst key differences do emerge, particularly with respect to those abilities most closely related to the symptoms of ADHD, a number of similarities exist between low WM and ADHD children (Holmes et al., 2014). Perhaps most crucial to highlighting the significance of WM difficulties is the similarity between teachers' descriptions of children with ADHD and low WM. Holmes et al. (2014) asked teachers to rate the behaviour and executive functions of low WM children and children with ADHD. Low WM and ADHD children were rated as significantly more distractable and inattentive than average WM children. While ADHD children were rated as more impulsive than low WM children the two groups did not differ in teacher ratings of inattentiveness (Holmes et al., 2014). This result suggests that WM difficulties result in behavioural issues of a similar magnitude to children diagnosed with an attentional disorder. This suggestion was supported by cognitive measures where low WM and ADHD children showed a similar pattern of deficits for measures of WM and executive functions (Holmes et al., 2014).

1.5.4 Improving and supporting WM in children

Given the implications of WM difficulties a key question is how children with poor WM can be supported. St Clair-Thompson et al. (2010) suggest two possibilities. Firstly, it may be possible to directly train WM. Early research addressing this question proved promising (e.g. Holmes et al., 2009). However, more recent reviews have identified some limitations with early investigations into WM training (Shipstead et al., 2012; but see Gathercole et al., 2012). A meta-analysis of training studies found that training does result in some immediate improvements in WM, but these do not persist over time (Melby-Lervåg and Hulme, 2013). Moreover, there are no immediate or long-term far-transfer effects – that is, benefits to outcomes such as IQ, maths or verbal ability. Melby-Lervåg & Hulme also found that studies with more rigorous designs observed smaller training effects. More recent meta-analyses have suggested that N-back training results in small improvements in fluid intelligence (Au et al., 2014). It has also been suggested that WM and EF training result in small far transfer effects in older adults (Karbach and Verhaeghen, 2014). However, a re-analysis of these two studies argues that the claimed effects are not, in fact, supported (Melby-Lervåg and Hulme, 2015). Au et al. (2015) responded by arguing against a number of decisions made by Melby-Lervåg and Hulme (2015) in their re-analysis. Despite these contentions over meta-analytic methodology, Au et al. (2015) acknowledge that the effects of current training programs are possibly too small to have practical significance. The most recent meta-analysis of WM training supports Melby-Lervåg and Hulme (2013)'s findings of near transfer in the absence of far transfer (Sala et al., 2017). Sala et al. (2017), like Melby-Lervåg and Hulme (2013), report an inverse relationship between study quality and the magnitude of reported training effects.

If we assume that the standard training procedures do not work, or at least have small effects, we can ask why these effects might be absent. Amso and Scerif (2015) suggest that, hitherto, training regimes have proceeded on an analogy between cognitive systems and muscles. Training a muscle ‘automatically’ transfers to tasks that involve that muscle. Similarly, it was assumed that training WM would result in transfer to tasks that involve WM. Amso and Scerif (2015) propose an alternative analogy where cognitive training should be compared to training a dancer. Learning to dance may well involve strengthening critical muscles but this will occur as part of a broader coordination between the systems that contribute to the activity. When applied to training this would suggest that training activities need to be embedded within the activities to which training should transfer (Amso and Scerif, 2015). For example, training participants to apply improvements in WM to practical tasks may be vital to maintaining genuine transfer. Investigations with adults have suggested that training results in the development of task-specific strategies, perhaps explaining why far-transfer tends not to be observed (Dunning and Holmes, 2014). Dunning and Holmes (2014) suggest that training needs to encourage the development of general strategies that can be applied to a range of tasks. Related concerns with training have been raised by those arguing for a more ecological approach (Moreau and Conway, 2014). Moreau and Conway (2014) suggest that the problem with a lot of training regimes is that they involve repeating the same task. This encourages the development of task-specific strategies rather than the desired general improvements. They suggest three components that cognitive training tasks should include to potentially achieve improvements in domain general processing: complexity, novelty and diversity. Moreau and Conway (2014) suggest that this can be achieved through a range of activities people already engage in, such as sport or playing musical instruments, both of which require cognitive and perceptual-motor processing. This combination of cognitive and perceptual-motor demands with skill development results in a complex and diverse task more likely to develop general cognitive abilities (Moreau and Conway, 2014).

St Clair-Thompson et al. (2010) also identify adapting tasks and teaching strategies as a route to supporting children with WM difficulties. As already suggested this approach could be complementary to direct training of WM (Amso and Scerif, 2015; Dunning and Holmes, 2014). If a task can be adapted to reduce the WM load then memory failures for low WM children become less likely. For example, a teacher might reduce the number of instructions that are given simultaneously (Gathercole and Alloway, 2008). The load of a task has also been found to interact with the efficacy of particular strategies for supporting WM (Waterman et al., 2017). Waterman et al. (2017) found that by reducing the number of possible items in an instruction-recall task participants benefited from acting out a sequence of instructions at encoding. When the demands of the task were increased enacting a sequence at encoding impaired performance.

Tasks could also be designed to allow information to be meaningfully grouped or chunked, or external aids could be provided (Gathercole and Alloway, 2008). This approach has been applied in a study where a group of teachers were trained to apply these strategy and task-adjustment procedures (Elliott et al., 2010). There was limited evidence that this approach transferred to improvements in maths or reading. However, there was some suggestion that the use of appropriate strategies by teachers related to academic attainment in the treatment and control groups (Elliott et al., 2010). Below, two novel extensions of this approach are discussed. Firstly, to anticipate Chapter 3, the use of selective attention to support WM performance in children is considered. Secondly, the role of the organisation of task-relevant objects in the environment is discussed in light of embodied or grounded accounts of cognition. Chapter 4 describes an attempt to apply this approach with children.

1.6 Future avenues

1.6.1 Selective attention

One near-ubiquitous claim about WM is that it is severely capacity-limited (Baddeley, 2007; Cowan, 2005; Oberauer et al., 2016; but see Macken et al., 2015). Coping with a complex changing world in the face of these capacity limits requires an ability to selectively attend to a subset of the information one is faced with. Without this ability, WM would be constantly overflowing with irrelevant information making it impossible to appropriately organise behaviour. As previously discussed, the ability to engage executive resources to manage the contents of WM is central to the relationship between WM and other cognitive abilities (e.g. Engle et al., 1999). Additionally, it is something that children struggle with, developing over a prolonged period (Cowan et al., 2006, 2010). However, as these abilities develop children may be able to use them to support performance by focusing on a subset of the information they are faced with. Some of the literature from adults will first be reviewed, before turning to work with children.

Classically it has been held that storage in WM is served by a small number of fixed resolution slots (Cowan, 2001; Luck and Vogel, 1997; Zhang and Luck, 2008). It follows from this proposal that items are either in memory or not, and, moreover, that all items in memory are represented with the same fidelity. Under such a view, selective attention would play a role primarily in determining what accesses one of these slots (Cowan et al., 2006). Additionally, once items are in WM, selective attention might be engaged to 'zoom in' on a particular item protecting it from interference (Cowan et al., 2006). A more fundamental role of attention has been proposed by those that argue storage in WM is served by a continuous mnemonic resources (Bays and Husain, 2008; Bays et al., 2009, 2011; Bays, 2015; Fallon et al., 2016; Ma et al., 2014). This view is supported by studies showing that cues can effect a change in the precision with which items are represented in WM. For example, Bays et al. (2011, Experiment 3) presented participants with two-item arrays for 1000ms, followed by a retro-cue to one item, a variable post-cue display time, and a mask. For post-cue display durations greater than 200ms, the retro-cued items were more precisely represented in WM than uncued items. The initial 1000ms display ensured that both items were in WM when the retro-cue was displayed, meaning memory resources were flexibly reallocated between items already in memory. These results suggests a role for attention beyond simply determining what enters into WM. The reallocation of resources occurs when items are in memory, meaning it is not simply an effect of processing at encoding. Crucially the reallocation of resource to one item entails costs to other items, assuming a single fixed-limit resource. Such costs have been frequently observed (Bays and Husain, 2008; Bays et al., 2011; Gorgoraptis et al., 2011; Ma et al., 2014, for a review). The observation of costs to prioritisation is important when considering if it could be effectively applied as a strategy to support WM.

Further evidence for the influence of attention on the allocation of resources in visual WM comes from contrasting effects of memory load and probe likelihood (Emrich et al., 2017). Emrich et al. (2017) orthogonally manipulated memory load and the likelihood of a cued item being probed. They found that variability in responses (i.e. the inverse of precision) was better predicted by the probe likelihood than memory load. The more likely a cued item was to be probed the more precisely it was represented in WM. By fitting a range of models to their data Emrich et al. (2017) found that a variable precision model, as a function of probe likelihood, provided the best fit by a substantial margin.

An alternative account of prioritisation effects in WM have been proposed where it does not always involve costs to other items (Myers et al., 2017b,a). This work is interesting when thinking about supporting WM difficulties as it implies a way to increase total capacity. Myers et al. (2017a) presented participants with four-item arrays of shapes where two shapes were probed for recall of their orientation. Following a mask, a cue was presented indicating one item that would be probed. On other trials, neutral cues were presented which gave no information about which item would be probed. By probing two items, Myers et al. (2017b) examined the costs of prioritising a retro-cued item independent of expectations about which item would be probed. Across four experiments they observed no costs to prioritising a cued item, provided expectations were controlled. This cost-free prioritisation could result from the prioritised item entering a more accessible state for driving the next action made (see also Myers et al., 2017b). It may also reflect the prioritised item entering a single-item focus of attention distinct from a direct-access region (Myers et al., 2017a; Oberauer, 2013). Myers et al. (2017b) suggest a mechanism by which a prioritized item results in a task-specific transformation of the representation of a set of stimuli optimised to guide behaviour. Once this representation is in place, sustained internal attention is not required to maintain it. Thus, there is no cost in terms of attentional resources to a change in task-setting that improves performance (Myers et al., 2017b).

Adults also show an ability to focus on specific items in sequential tasks. This has been demonstrated using sequential visual tasks where participants are instructed to try especially hard to remember a particular item in a sequence. For example, they might be told to try especially hard to remember the first item. Adults are able to prioritise a particular position in a sequence, resulting in large boosts in performance (Hu et al., 2014, 2016). This ability is vulnerable to cognitive load, highlighting the role of attention in resource allocation in WM (Hu et al., 2016). The reduction of the prioritisation effect under cognitive load has also been demonstrated in the context of cueing (Janczyk and Berryhill, 2014). Participants' ability to prioritise an item in a visual array following a retro-cue was reduced by performing a concurrent tone judgement task.

The executive limitations of children make it implausible that they would be able to allocate attention (or memory resources) as flexibly as adults. Nevertheless, children do seem able to use some task factors to selectively allocate attention to items in WM (e.g. Cowan et al., 2010; Shimi et al., 2013, 2014). Shimi et al. (2013) had adults and 10-year-olds perform a task where four-item arrays were presented alongside pre-cues, retro-cues, or neutral-cues. Adults showed similar boosts in performance for prospective and retrospective cues. Children, in contrast, showed a larger benefit for pre-cues than retro-cues. This result, combined with EEG analysis, suggests that children are less able to engage attentional control to selectively maintain a subset of an array following a retro-cue (Shimi et al., 2013).

Shimi et al. (2014) contrasted central and peripheral cues (arrows versus boxes) presented prior to and following, a memory array. Young adults and 11 year-olds showed similar benefits for pre- and retro-cues, whether they were presented centrally or peripherally. Seven year olds, on the other hand, showed a larger benefit for pre-cues than retro-cues. Again, this highlights potential developmental changes in the ability to retrospectively maintain specific items within an array. While younger children clearly cannot engage attention as flexibly as adults they were able to respond – as were adults and 11-year-olds – to a reduction in cue validity to 50%, such that cuing benefits disappeared. This suggests that children as young as 7-years-old are able to selectively *ignore* cues when they no longer become informative. What they struggle with is retrospectively maintaining a subset of an array in the face of interference from other items (Shimi et al., 2014). The suggestion that responding to retro-cues involves resisting interference was supported by the finding that performance on these trials was most strongly related to individual differences in visuospatial WM.

Cowan et al. (2010) equally demonstrated both the capabilities and limitations of young children's attention control. Participants were cued as to which subset of an array was likely to be probed, with cue validity varying between 100% and 20%. The authors could then estimate how much capacity was allocated to a cued shape based on this validity manipulation. For arrays of four items, seven to eight year-olds, 12 to 13 year-olds, and young adults allocated less capacity to items that were less likely to be probed. This pattern emerged despite overall capacity being lower for younger children. However, when WM was overloaded by using sets of 6 items, young children showed a sub-optimal allocation of capacity with respect to probe likelihood. Whilst children can allocate capacity adaptively in some cases this effect is sensitive to task demands.

To summarise, using attention to select which items enter memory, and/or how memory resources are allocated, is an important ability that adults possess. While children are able to engage attention to prioritise items within WM, this ability is limited by particular task demands. Furthermore, this ability has only been demonstrated using simultaneous displays and cues on individual trials. Chapter 2 describes an attempt to investigate children's ability to prioritise with a sequential visual task. In addition, rather than cuing different individual items children were instructed to prioritise a particular position in a sequence throughout a whole set of trials. If children demonstrate the ability to prioritise in this way it would represent an avenue to overcoming WM limitations by selectively prioritising the most relevant information in a situation.

1.6.2 Using the environment

Embodied accounts of cognition emphasise the role that the body and environment play in thought (Clark, 1999, 2008, 2016; Shapiro, 2010; Wilson, 2002). Such accounts begin from the uncontroversial observation that human beings are embodied, and embedded within environments (Clark, 1999). The question is then what implications these facts should have for our account of cognition. Clark (1999) identifies two responses, one 'simple' and one 'radical'. A simple approach to embodied cognition states that our embodiment and embeddedness should place constraints on cognitive theories. A radical approach instead argues that the dominant approach to understanding cognition should be overhauled and replaced (Clark, 1999). The dominant approach to which embodied cognition contrasts itself can be described as a ‘cognitivist’ view of cognition (Shapiro, 2010). The critical feature of such accounts is that cognition is thought to proceed by logical operations on abstract representations.

One influential approach to embodied cognition is the ‘extended mind’ account of Clark and Chalmers (Clark and Chalmers, 1998; Clark, 2008, 2016), a view that has developed recently to encompass work on predictive coding (Clark, 2013, 2016). Broadly speaking, the claim is that the ‘mind’ is, at times, best described as a system that spans brain, body, and environment. Thus, the view is not merely that the body and environment play a casual role in cognition, but, more strongly, that they are constituents of the thinking system (for discussion of this distinction see Shapiro, 2010). To illustrate this point, Clark (2008) asks us to consider working with the aid of pen and paper. In the process of working out some problem we would usually write down partial solutions, which would then feedback into later thinking. What the extended mind view argues is that situations like these — where information is flexibly moving between internal and external stores — are best described by a single system spanning brain, body, and environment. We could uphold the brain as the sole seat of cognition through asserting that once any information reaches the paper it simply is not part of the thinking system any more. Such information might influence later thinking when some partial solution on paper causes later thinking, but anything outside of the head only ever plays a causal rather than constitutive role. The problem for such a view is that it often leads to an unparsimonious account of what is going on. Functionally speaking it’s not clear how much the brain really respects the sort of strong skin-skull divides drawn to reject the extended mind (Clark, 2008). Moreover, much of the resistance to the extended mind view turns on an insistence that nothing outside of the brain can be part of the mind, despite this being the very thing the extended mind view challenges (Clark, 2008; Shapiro, 2010). It is worth stressing that the 'extended mind' view is not committed to some promiscuous inclusion of the environment in all cognition. Rather it applies to cases where some resource in the environment is used with the same flexibility and automaticity as internal resources.

This embodied perspective has been applied to experimental psychology in a range of studies. As discussed above, (Section 1.2), a body of work within the WM literature exists applying this perspective (e.g. Hughes et al., 2016; Macken et al., 2016; see Macken et al., 2015, for a review). Equally, the approach has proved influential in developmental psychology (Smith and Thelen, 2003; Smith and Gasser, 2005; Smith and Sheya, 2010; Thelen and Smith, 1996; Yu and Smith, 2017). One paper has taken an embodied perspective in interpreting the relationship between motor control, working memory, and inhibition (Gottwald et al., 2016). Infants aged 18 months completed complex and simple inhibition tasks, as well as tasks measuring prospective motor control and working memory. Both simple inhibition and working memory were related to prospective motor control. This was interpreted within an embodied perspective as highlighting the importance of basic perceptual motor skills to high-order cognition. However, the relationships were very weak with a small number of outliers driving the effects. In line with an embodied account of cognitive development, others have found that spatial exploration in infancy is related to visuospatial WM at four and, in particular, six years (Oudgenoeg-Paz et al., 2014). Spatial exploration was measured by parents reporting the engagement in self-locomotion and exploratory spatial play by their child from birth to two years. Visuospatial WM was measured using a dot matrix task where participants were required to remember sequences of locations within a spatial matrix. The relationship between early spatial exploration and later spatial memory was significant at age six and marginally significant at age four (Oudgenoeg-Paz et al., 2014). Both these studies focused on individual differences. To the best of my knowledge, no experimental investigation of working memory in children informed by embodied accounts of cognition has been carried out. That said, some effects observed in children with WM can certainly be interpreted as supporting embodied account of cognition (see Chapter 4 for discussion). However, the implications of such work for embodied accounts of cognition is rarely stressed by the original authors.

More generally, experimental investigations to test predictions of embodied cognition in children have been carried out. For example, Smith and Thelen (2003) review a dynamic systems approach to the classic A-not-B error. They demonstrate that performance on the task emerges from the dynamic interplay between specific features of the task and the perceptual-motor system of the infant. There is no need to invoke abstract epistemic structures such as 'Object Permanence' to account for the observed patterns of behaviour. Instead the error can be made to disappear with small adjustments to the perceptual-motor affordances of the task (Smith and Thelen, 2003). This reframing is useful for highlighting the differences between an embodied perspective and a Piagetian approach. At first look a Piagetian approach to development might appear to be embodied; more complex understanding is built on lower level structures more closely related to action. However, such a view is mistaken in failing to acknowledge the primacy of logical epistemic structures in Piagetian theory (Case, 1987). For example, the understanding of 'Object Permanence' is held to be a piece of knowledge about the essential logic of the world that, once acquired, transfers to all settings. The difference between the approaches can be made starker when considering that within Piaget's system this abstract knowledge can constrain the actions a child will perform and, therefore, motor development (Case, 1987). Embodied cognition, at least in its most radical form (Clark, 1999), would reject outright the notion of abstract domain-general epistemic structures, such as Object Permanence. Indeed, some reject the notion that development involves "building representations at all!" (Thelen and Smith, 1994, p. 338).

Chapter 4 describes an attempt to apply this embodied perspective to a WM task with children. The aim was to investigate the effect of task-relevant structure in the environment on WM performance with a particular focus on children with poor WM. In addition, children's metacognitive insight into such effects is explored. While some research has investigated the influence of spatial configurations on WM performance in adults (Morey et al., 2015; Woodman et al., 2003), the focus has been on the spatial properties of visual arrays. Chapter 4 uses a task that combines verbal storage with visuospatial recall to create a more 'ecologically valid' task.

In conclusion, despite being relatively neglected, the body and environment are crucial to understanding how cognition works in the world (Clark, 1999, 2008; Shapiro, 2010). Spatial configurations and dynamics can influence cognition such that processing requirements are reduced (Kirsh, 1995). This provides an avenue to supporting children with low WM in reducing the strain placed on their limited processing/storage capacity. Classrooms are complex environments where successful performance might emerge from complex interactions between classically cognitive constructs (memory, attention), the body, and features of the environment. If performance on a WM task is constrained by a range of factors rather than monolithic developmental milestones then improving performance would be more tractable.

1.7 Thesis outline

1.7.1 Chapter 2: Assessing working memory in children

Clearly in order to support children with WM difficulties those individuals must be identified. Chapter 2 describes the development of a computerised battery of WM tasks that have been used with ~600 children for this thesis. One aim for the battery was to provide a set of measures that could be administered quickly in a school setting, while remaining reliable measures of the construct of interest for a wide age range. This was so that the measures could be used in the *Born in*

Bradford cohort study, a longitudinal study following over 13,000 children (Wright et al., 2013). A subset of the measures presented in Chapter 2 are currently being used with thousands of children in that study. This data will not be accessible until after the completion of data collection in July 2018, and is, therefore, not presented here. Developing measures that were relatively quick to administer also facilitated the collection of large samples for the experimental work presented in Chapters 3 and 4 (80 – 100 participants per experiment).

Four WM measures are presented in Chapter 2:

1. Forward digit recall: a verbal measure of simple WM or short-term memory (e.g. Gathercole and Alloway, 2004).
2. Backward digit recall: a verbal measure of complex WM (Gathercole and Alloway, 2004).
3. Corsi blocks: a simple visuospatial measures of WM (Gathercole and Alloway, 2004).
4. Odd-one-out: a complex measure of visuospatial WM (Alloway et al., 2006; Russell et al., 1996).

For each WM measure a range of analyses are presented to interrogate the success of the measure. For example, with the working memory measures, serial position, list length, and age effects are analysed to ensure that the expected effects in each case emerge. In addition to the WM measures, a measure of processing speed was developed, as well as an inhibition task. These measures were developed to meet the particular requirements of the *Born in Bradford* project and are not directly relevant to the experimental work presented in the thesis. For this reason the data from these measures are not included in Chapter 2 or in subsequent chapters.

1.7.2 Chapter 3: Using attention to support working memory performance

Chapter 3 presents the first attempt to investigate, in a rigorous experimental setting, the ability of children to prioritise a particular serial position within a visual sequence. This work involves the novel application of a procedure developed in adults to children aged 7 to 10 years old (Hu et al., 2014, 2016). Three experiments are described (total N = 210) where children were instructed to 'try especially hard' to remember either the first of third item in a three-item sequence. In addition to the primary experimental task, the working memory measures described in Chapter 2 were also used. A novel individual difference analysis with these measures and performance on the experimental task is presented to investigate the contribution of executive resources to WM across serial positions.

In terms of expected outcomes, some previous work indicates an ability to flexibly allocate attention within visual WM (e.g. Cowan et al., 2010; Shimi et al., 2014). However, allocating attention within a temporal sequence may be more complicated, meaning that children would not demonstrate an ability to prioritise when instructed to do so. We expected that large recency effects would be observed in line with previous investigations in children and adults (Hu et al., 2014, 2016; Walker et al., 1994). In addition we predicted that performance at the first and second position within a sequence would be related to additional WM measures, whereas performance at the third position would not. This prediction flows from the proposed automaticity of recency effects in WM.

1.7.3 Chapter 4: Using the environment to support working memory performance

The approach taken in Chapter 3 was to investigate the ability of endogenous attentional resources to effect boosts in working memory performance in response to specific instructions. Following the absence of these effects an alternate approach was taken for Chapter 4, influenced by embodied perspectives described above. Rather than draw on internal resources, the potential to leverage structure in the task environment to support performance was investigated. In two experiments (total N = 166) participants were presented with verbal sequences of colours and required to recall them by arranging physical blocks in the appropriate serial order. The blocks used for recall were either grouped by colour or arranged randomly. We expected that participants would successfully recall longer sequences when these coloured blocks were ordered in a task-relevant arrangement. In addition we hypothesised that this benefit for structure in the environment would be more pronounced for low WM children if, as embodied theories of cognition claim, we are able to offload processing requirements onto our environment. For the second experiment in this chapter, participants were given the opportunity to select their own arrangement for the blocks, in addition to rating the difficulty of the two primary conditions. Firstly, we expected that participant would rate the ordered arrangement as easier than a random arrangement. Our expectations were less clear with respect to participants' chosen arrangements for the blocks. Research on children's insight into task difficulty is limited, and a range of factors, including WM and EF development, likely contribute to translating an awareness of difficulty into implementing an effective strategy. Finally we suggest that children with poor WM may particularly struggle with identifying more difficult arrangements, and implementing effective strategies.

Chapter 2

Developing a working memory battery

This chapter describes the development of a set of computerised working memory (WM) measures. These measures were developed for use in the *Born in Bradford* cohort study (Wright et al., 2013), a large longitudinal study following thousands of children. The time constraints inherent in testing thousands of children necessitated developing a set of measures that could be quickly administered to groups of children in a school setting. Presenting tasks on tablet computers was an effective way to save time compared to pencil-and-paper measures, as well as removing the possibility for experimenter errors. Scoring for computerised measures is also considerably easier than with pencil-and-paper. For the tasks described below, the scoring was carried out as participants completed the task. With pencil-and-paper measures, manual scoring would likely be required. There is also a greater risk of data loss using pencil-and-paper methods. All these concerns with pencil-and-paper measures are addressed by the tasks described in this chapter.

A wide range of measures have been taken for the The Born in Bradford study, including participant-level biology, family characteristics, and neighbour-level demographics (Raynor et al., 2008; Wright et al., 2013). Making measures of cognitive ability available alongside these other measures is crucial to understanding how WM relates to life outcomes *compared to* other factors. For example, studies of the relationship between WM and educational attainment typically focus on WM alone, and possibly IQ (e.g., Alloway and Alloway, 2010; Gathercole et al., 2004b; St Clair-Thompson and Sykes, 2010). In contrast, a study like Born in Bradford allows one to investigate how WM relates to, say, educational attainment compared to a range of non-cognitive factors. The set of tasks developed in this chapter will provide the data needed to answer such questions. As the data collection by Born in Bradford using the measures described in this chapter will not been completed until later in 2018, this question cannot be answered here.

As noted in Chapter 1, the range of measures used to answer particular theoretical questions is vast. The focus here will be on measures used in an individual difference setting with children, reflecting the aim of this chapter to develop measures for a longitudinal cohort study. A more general theoretical overview of measures of WM can be found in Chapter 1.

The chapter begins with a brief summary of some of the WM measures that are commonly used with children. An overview of the tasks described in the chapter is then provided, followed by general methodological information and an outline of the analyses that will be reported. Each of these tasks was an adaptation of measures previously used in the literature. The complex visuospatial WM task presented here addresses the substantial shortcomings of widely used measures with a novel redesign of an existing task. Details on each individual tasks are provided alongside analyses of their internal structure – that is, the effects of aspects of the task structure itself (e.g. serial position) on the outcomes measure of interest. Finally the relationship between the measures in the chapter is reported, as well as how the measures predict additional outcomes beyond WM.

2.1 Measures of working memory

Here the terms simple and complex working memory are used to describe tasks that involve storage only, or storage alongside additional processing (e.g. Gathercole, 1999). This terminology is preferred to a distinction between short-term memory and working memory measures (e.g. Alloway et al., 2008), as all short-term storage involves the whole WM system.

2.1.1 Verbal measures of working memory

Here a measure is classed as verbal if the storage requirements involve participants remembering information that can be verbally produced. Typically, this would involve either words or digits, although non-words are also used for some verbal measures. For those who take a multiple component view of WM (Baddeley, 2007; Logie, 2011), verbal tasks all involve storage in a specialised phonological store.

2.1.1.1 Simple verbal measures

Verbal measures of WM have a relatively long history, dominating early canonical work (e.g. Baddeley and Hitch, 1974). Typical measures involve presenting a series of words, such as digits, for immediate serial recall (Alloway et al., 2006; Alloway, 2007; Alloway et al., 2008, 2009a,b; Baddeley and Hitch, 1974; Hughes et al., 2016; Gathercole, 1999; Gathercole et al., 2004a, 2016; Pickering, 2001). Digit span or digit recall measures are some of the most commonly used in an individual difference setting, particularly with children. Popular test batteries such as the Wechsler Memory Scale-III (WMS-III, Wechsler, 1997), Wechsler Adult Intelligence Scale-IV (WAIS-IV, Wechsler, 2008), Wechsler Intelligence Scale for Children-IV (WISC-IV, Wechsler, 2003), and Automated Working Memory Assessment (AMWA, Alloway, 2007) include digit span measures. These measures involve presenting participants with sequences of digits for serial recall. For example, a participant might hear (or read) ‘3, 9, 2’ and have to response with ‘3, 9, 2’. Performance on such tasks can be scored using a number of procedures (see below; Conway et al., 2005). For the WISC-IV span scoring is used, such that participants are presented with sequences of increasing length until they cannot be recalled accurately. The maximum sequence length that a participant can recall accuracy (e.g. by getting more than two thirds of trials correct) is taken to be their span score. It is unclear how exactly the digit span tasks in the AWMA are scored, as this information in not included in papers using the tasks (e.g. Alloway et al., 2008; Gathercole et al., 2016). Alternatively simple verbal WM tasks exist where words or non-words are used as the stimuli, rather than digits (Alloway et al., 2006; Alloway, 2007; Alloway et al., 2008).

2.1.1.2 Complex verbal measures

One common adaptation of digit span tasks requires participants to recall the sequence in backward order (Alloway et al., 2008, 2009a,b; Brown, 1974; Gathercole, 1999; Gathercole et al., 2004a, 2016). This additional processing requirement is taken to transform the task into a complex WM measure, in contrast to simple digit recall (Gathercole, 1999). Unlike more canonical storage-and-processing WM tasks (e.g. Kane et al., 2004), with backward digit recall (BDR) the processing occurs after the information has been encoded into WM. Initially participants encode a simple verbal sequence before rearranging the sequence in memory for backward recall. This feature of the task has led some to question whether it should be called a complex measure of WM, or whether it, in fact, reflects only simple storage (St Clair-Thompson, 2010; St Clair-Thompson and Allen, 2013). Using confirmatory factor analysis St Clair-Thompson (2010) argued that BDR reflects WM in children, but STM (i.e. ‘simple WM’) in adults. As the aim here was the develop a battery of tasks suitable for children this finding does not undermine the use of BDR as a measure of complex verbal WM.

A number of storage-and-processing WM tasks involve verbal storage requirements. With counting span tasks (Alloway et al., 2006; Case et al., 1982) participants are presented with a sequence of arrays within which a number of target items have to be counted. At the end of a sequence the tally of target items for each array must be recalled. Thus, while the processing involves visuospatial search, digits are ultimately what is stored and recalled. Such tasks differ from BDR in that processing and storage are interleaved. Similar tasks, such as reading span (Daneman and Carpenter, 1980; Engle et al., 1999; Stone and Towse, 2015) and operation span (Engle et al., 1999; Turner and Engle, 1989; Stone and Towse, 2015), involve storing words or digits while carrying out some concurrent processing. For reading span the processing involves reading sentences and, in some implementations, indicating whether the sentences are true or meaningful (e.g. Stone and Towse, 2015). A version of this task where the sentences are presented auditorily is referred to as listening span (Daneman and Carpenter, 1980), and is commonly used with children (Alloway et al., 2006; Alloway, 2007; Gathercole et al., 2004a; St Clair-Thompson, 2010). Operation span requires participants to complete (Turner and Engle, 1989) or verify (Stone and Towse, 2015) simple mathematical equations alongside storing words or digits.

2.1.2 Visuospatial measures of working memory

A measure is classed as visuospatial if the stimuli are presented visually and require participants to remember some spatial or visual information. For example, visuospatial tasks might involve remembering locations or orientations. With visuospatial tasks there is always a concern that participants are able to verbally recode the information in some way, such that it is not stored visuospatially. Some have argued that a shift to verbal recoding of visual information is important in development (e.g. Hitch et al., 1988), though this data could be accounted for by changes in attentional control, instead (Morey et al., 2017).

Within experimental studies articulatory suppression is often used to reduce the possibility of verbal recoding (Baddeley, 2007). Interestingly, recent work has suggested that opportunities for verbalisation are unimportant to performance on some visual tasks (Sense et al., 2017). A more appropriate alternative to articulatory suppression for designing individual difference measures is to use stimuli that cannot be easily verbally recoded.

2.1.2.1 Simple visuospatial measures

One common measure of simple visuospatial WM is the Corsi block recall task (Alloway et al., 2006; Alloway, 2007; Alloway et al., 2009a,b; Gathercole, 1999; Gathercole et al., 2004a; Isaacs and Vargha-Khadem, 1989; Milner, 1971; Shimi and Scerif, 2015). Here participants are presented with a pseudorandom arrangement of blocks, and a series of blocks are touched by an experimenter. Participants are required to recall the sequence by touching the blocks in the same order as the experimenter. Another simple visuospatial measure is the matrix span or dot matrix task (Alloway et al., 2006; Alloway, 2007; Alloway et al., 2008; Kane et al., 2004; Stone and Towse, 2015). Here participants are presented with a matrix of, say, 4 rows and 4 columns. A series of cells are marked in the matrix and their locations must be recalled at the end of a trial (Alloway et al., 2006; Alloway, 2007; Gathercole et al., 2016). Both these measures require participants to remember a sequence of spatial locations. Maze recall, instead, requires remembering an entire spatial path through a maze, for subsequent recall (Gathercole et al., 2004a; Pickering and Gathercole, 2001). A visual patterns task (Della Sala et al., 1997; Gathercole et al., 2004a; Pickering, 2001) involves participants remembering a complex spatial configuration, instead of a sequence of simple locations. Participants are presented with a pattern made up of filled and unfilled squares, which must be recalled after a short delay (Pickering, 2001). Performance can then be expressed as the number of individual filled squares people are able to recall (Pickering, 2001).

2.1.2.2 Complex visuospatial measures

Common complex visuospatial measures in the developmental literature include spatial recall, odd-one-out, and Mr X. (Alloway, 2007; Alloway et al., 2008, 2009a,b; Gathercole et al., 2016; Russell et al., 1996; Shimi et al., 2014; Shimi and Scerif, 2015). For spatial recall participants are presented with two abstract shapes and have to indicate whether they are the same. In addition, they must remember the location of a red dot on one of the shapes, for subsequent recall (Alloway et al., 2008). Mr X. resembles spatial recall in that participants are presented with two cartoon characters and have to indicate whether they are holding a ball in the same hand or not. The arms of the characters can be positions at one of six compass points; these locations must be recalled for one character following a series of judgements (Alloway, 2007). Odd-one-out also involves interleaved processing judgements prior to the recall of spatial locations. The task was originally developed by Russell et al. (1996), building on a previous task. Russell et al. (1996) presented a 3x4 grid of squares where three stimuli were presented in each column in turn. For each column participants had to indicate which of three shapes presented differed from the other two. At the end of the trial participants then had to recall the positions of the odd-ones-out in order, starting from the first column. Alloway et al. (2006) describe a similar task where participants are presented with three shapes arranged in a row and are required to touch the odd-one-out. Similarly, at the end of a trial they then have to recall the locations of the odd-ones-out.

All the visuospatial tasks described here have the potential to be verbally recoded, particularly odd-one-out. In the case of Alloway et al. (2006), the positions could be verbally represented as ‘left’, ‘right’, and ‘centre’. Participants could encode no visuospatial information in WM, instead simply remembering a sequence of verbal labels. Below a novel developmental of the odd-one-out task is described which addresses this concern.

With adults, a number of complex visuospatial tasks have been used where verbal recoding is unlikely. For example, for symmetry span participants have to remember a sequence of grid locations, while making interleaving processing judgements as to whether a complex pattern is symmetrical or not (Kane et al., 2004; Stone and Towse, 2015). With rotation span, participants are required to remember the size and length of a sequence of arrows while concurrently judging whether interleaved letters are mirror-reversed or not (Kane et al., 2004; Shah and Miyake, 1996; Stone and Towse, 2015). While these measures are unlikely to afford verbal recoding, they would potentially not be suitable for children.

2.1.3 Scoring methods

Despite the prevalence of span scoring, it has a number of disadvantages, particularly in an individual difference context (Conway et al., 2005). Span scores are necessarily limited to a small number of possible values (see Figure 2.4), limiting their ability to discriminate between participants. This issue is compounded in children, where the set of likely span scores is smaller than for adults. One alternative to span scores is all-or-nothing scoring (Conway et al., 2005), where the proportion of correct whole sequences within a set of trials is taken as the measures of performance. However, a partial-credit approach is preferable, where correct items within a sequence are scored even where the whole sequence was not correctly recalled (Conway et al., 2005). Scoring at the item level provides an alternative whereby scores are better able to discriminate between participants. With item-level scoring the proportion of, say, letters correctly recall is scored rather than simply whether a whole sequence is correctly recalled. For example, suppose a participant is presented with a sequence of ‘3, 7, 2, 4’ and responds ‘3, 5, 2, 4’. With all-or-nothing scoring this sequence would be scored as incorrect (i.e. 0), whereas with partial credit it would be scored as 0.75 (3/4 items correct). This partial credit approach is used for all the tasks here to maximise discriminability between participants. Item-level proportion correct

Table 2.1: *Overview of tasks described in this chapter organised by the modality of the stimuli, and whether the task including a processing component.*

	Modality	
	Verbal	Visuospatial
Simple	Forward digit recall	Corsi block recall
Complex	Backward digit recall	Odd-one-out

has also been demonstrated as a better predictor of academic attainment than the total number of correct sequences (St Clair-Thompson and Sykes, 2010).

2.2 Overview of the tasks

The tasks described in this section underwent considerable development. For the sake of brevity, only the final piloted versions will be described. Any changes following piloting are noted, though these were minimal. The tasks outlined below were designed to meet a number of aims that reflected both practical and theoretical concerns. Most importantly, the measures had to be quick for teams of research assistants to run with groups of children, due to the practical constraints on testing thousands of children in a cohort study (Wright et al., 2013). Nevertheless, we wanted to create a set of measures that mirrored those commonly used in the developmental literature, while remaining short enough to administer in a single session. In addition, we wanted to capture the contributions of executive processes and different sensory modalities to WM. Table 2.1 summarises the tasks described in this chapter by dividing them along two axes. 'Modality' refers to the information that participants had to remember; numbers for verbal tasks and spatial locations for visuospatial tasks.

'Simple' and 'complex' refer to whether the tasks required additional processing on top of simply having to remember some information. The odd-one-out task included a processing component embedded within the presentation of stimuli –

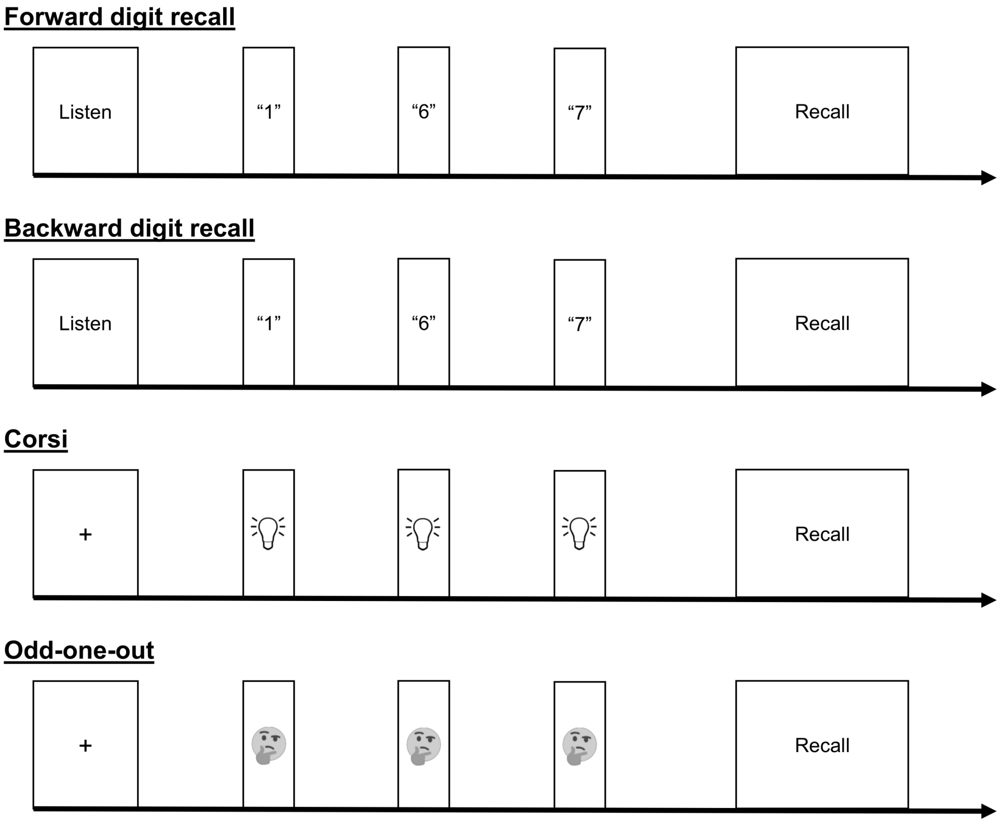


Figure 2.1: *Overview of tasks described in this chapter. The width of the components of each task are roughly proportional to their duration. See Figures 2.3, 2.13, & 2.19 for more detailed task schematics.*

processing had to be carried out while concurrently encoding spatial locations (see Figure 2.1). In addition, the backward digit recall task involved processing during the retention interval or at recall (see below). Figure 2.1 shows, in broad terms, how each task progressed, demonstrating the tasks were well matched in terms of overall structure.

For both verbal tasks, digit recall measures were used. Digits were preferable to words and non-words for a number of reasons. Firstly, a considerable proportion of the target sample are bilingual or have English as an additional language. Digits are far less likely to effected by linguistic factors than word or even non-words. In addition, on-screen response boxes rather than verbal recall was used for the verbal measures, in order to facilitate group testing. Visually presenting words or non-words at recall would require participants to identify the written representation of the stimuli following auditory presentation. Performance would, therefore, be far more likely to be confounded by reading ability. As the use of response boxes at recall was somewhat atypical, we wanted to ensure that the required mapping between auditory presentation and visual representation was as simple as possible. Digits clearly meet this criteria, especially for children. In summary, digits are widely used and relatively uncontroversial stimuli, compared to constructing a set of word or non-word suitable for our diverse target sample.

The visuospatial measures were selected in order to be easily computerised and understandable to children. We also wanted to have a pair of measures that mirrored the close relationship between forward and backward digit recall. Backward digit recall is identical to forward digit recall except that the stimuli must be recalled in reverse order. We wanted to have a similarly close relationship between our visuospatial tasks to aid interpretation. If complex and simple measures differ in a number of ways it is less clear whether the additional complexity itself is driving differences in outcomes. Corsi and our novel redesign of odd-one-out, therefore, represented a suitable pair of tasks in being similar in structure and suitable for children. Backward Corsi was not used as our complex task because the observed differences in performance between forward and backward Corsi are very small compared to digit recall tasks (e.g. Isaacs and Vargha-Khadem, 1989), suggesting that it is not a complex task, even in children. Alternately, it could be that forward Corsi is sufficiently tasking by itself that backward recall results in minimal additional load.

2.3 General methods

Table 2.2 summarises the studies that contributed data to this chapter, as well as the number of participants tested for each study. Data were collected from five different studies and are combined for the analysis. The initial piloting study was

Table 2.2: *Overview of the studies that contributed data to this chapter, grouped by age and task. The values in the cells are the number of participants contributed by the study.*

Experiment	Age Group	FDR	BDR	Corsi	OOO
Pilot	6-7 y.o.	56	55	57	55
Pilot	7-8 y.o.	57	53	57	57
Pilot	8-9 y.o.	53	53	53	53
Pilot	9-10 y.o.	60	58	59	60
Ch. 3, Exp 1	7-8 y.o.	29	29	28	28
Ch. 3, Exp 1	8-9 y.o.	26	26	24	25
Ch. 3, Exp 1	9-10 y.o.	32	32	32	32
Ch. 3, Exp 2	7-8 y.o.	29	28	29	28
Ch. 3, Exp 2	8-9 y.o.	29	29	29	29
Ch. 3, Exp 2	9-10 y.o.	29	28	28	28
Ch. 3, Exp 3	7-8 y.o.	29	29	29	29
Ch. 3, Exp 3	8-9 y.o.	28	28	28	28
Ch. 3, Exp 3	9-10 y.o.	29	29	27	26
Ch. 4, Exp 1	8-9 y.o.	41	41	41	42
Ch. 4, Exp 1	9-10 y.o.	58	58	57	58

[1] FDR = forward digit recall; BDR = backward digit recall; OOO = odd-one-out

designed to collect data from the measures to ensure they ‘worked’, where ‘working’ involves exhibiting a number of properties expected for successful WM measure, such as sequence length effects, and minimal ceiling and floor effects. The piloting stage also provided an opportunity to update the measures if, for example, they appeared to be too easy or difficult. In practice the piloting was very successful, and minimal changes were made to the measures. Thus, the data from the pilot study are combined with the data from four subsequent experimental studies for this chapter. The experimental studies for which the measures were used are described in greater detail in Chapters 3 and 4.

Table 2.3: *Sample sizes for each age group combining across experiments. The values in the Ns column show the minimum and maximum number of participants tested in each age group, as the sample sizes were not identical between tasks.*

Age group	Ns	Mean age	SD age
6-7 y.o.	55 - 57	6.95	0.302
7-8 y.o.	139 - 144	7.98	0.325
8-9 y.o.	175 - 177	9.01	0.31
9-10 y.o.	203 - 208	10	0.312

2.3.1 Participants

Participants were recruited from primary schools in Bradford, UK. The majority of the children tested were from low SES British-Pakistani neighbourhoods. Table 2.3 summarises the number of participants that were tested in each age group. Six to seven year-olds were only tested for the pilot study, explaining the low sample size in Table 2.3.

Exclusions. The sample sizes reported in Tables 2.2 and 2.3 are prior to any exclusions being made. Participants were excluded from the subsequent analyses for three reasons: (i) they had special educational needs; (ii) notes from the experimenters indicated that they were distracted/not trying during testing; or (iii) they scored at or below chance. Table 2.4 summarises the exclusions made for each task. The counts are presented such that SEN children are not included in the 'Distracted' column even if they were also distracted during testing. In addition, only those children scoring below chance who were not distracted or SEN are counted in the 'Chance' column.

SEN children were excluded for a number of reasons. The inclusion of SEN children could potentially either inflate or suppress the observed relationship between tasks. If SEN children have a specific difficulty with, say, verbal material, then that may inflate the relationship between verbal measures; they would perform poorly on both measures for reasons unrelated to the structure of WM.

Table 2.4: *Overview of the number of participants excluded for each task.*

Task	Total N	Distracted	SEN	Chance	Final N
BDR	576	9	94	3	470
Corsi	578	9	97	0	472
FDR	585	6	101	2	476
OOO	578	18	94	35	431

[1] SEN = special educational needs; Chance = performing at chance

The relationship between tasks could also be suppressed if SEN children had an equivalent global difficulty with all tasks. This could then mask performance differences and specific relationships between tasks. We also did not have information on the particular reason that individual children were categorised as SEN. Consequently, interpretation of any outcomes for SEN children would be fundamentally incomplete. In practice, the majority of SEN children were either distracted, scoring below chance, or had incomplete data.

2.3.2 Materials

All the tasks were written using PsychoPy (Peirce, 2007) and presented on touchscreen tablets (screen size: 25.7 x 14.4cm; resolution: 1920 x 1080). The same pseudorandom stimulus sequences were use for all participants constrained such that, for example, the same stimuli were not presented twice within a trial (see below for further details).

2.3.3 Procedure

Participants were always tested one-on-one with an experimenter to ensure they remained engaged with the tasks. To ensure efficient testing of large numbers of children, teams of experimenters worked within the same space, each testing a single child at a time. Testing sessions ranged from 20 to 40 minutes, with the measures being completed alongside experimental tasks for the studies described in Chapter 3 and 4.

2.3.4 Analysis plan

For all the tasks the outcome measure was accuracy at the item level. To respect the fact that accuracy data is not normally distributed binomial logistic regression was used (Jaeger, 2008). Bayesian binomial logistic regression models were estimated in R (R Core Team, 2017) using *rstanarm* (Gabry and Goodrich, 2017). Visualisation of the posterior estimates will be used to evaluate the absence or presence of different effects. This will be supplemented by posterior odds and posterior contrasts where appropriate. With Bayesian estimation a single estimate of some parameter is not made, as with maximum likelihood techniques. This is because Bayes Rule cannot be solved analytically for anything but the simplest models. Instead, sampling techniques are used to determine the most likely parameter values given the data. A sampling procedure is constructed to explore the posterior distribution of $probability(parameter|data)$ such that the frequency with which different values appear in the set of samples approximates the analytic posterior distribution. This provides the distribution of likely parameter values given the data. All the models were run for 20,000 iterations, split across 4 chains. An additional 20,000 iterations were used to 'warm-up' the sampling algorithm, prior the 20,000 draws from the posterior used to make the inferences below. The default, very weak, priors in *rstanarm* (Gabry and Goodrich, 2017) were used for this model. These are shown in Figure 2.2.

Serial position, sequence length and age. A single model is used to estimate effects of serial position, sequence length, and age group, though they will be discussed separately. For the sake of computational ease interactions between the predictors were not included. In addition, preliminary analysis of the data did not support interactions between these factors. The effect of these

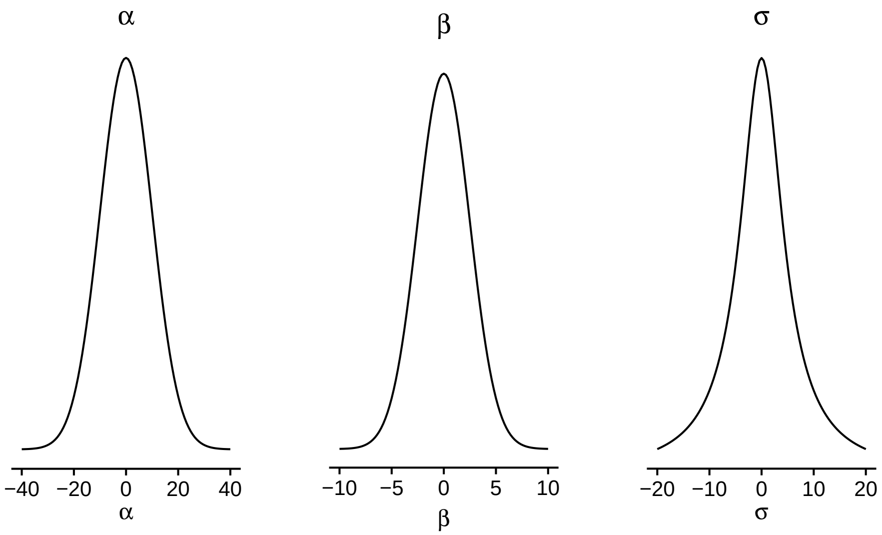


Figure 2.2: *Default priors for rstanarm Version 2.15.3 for the models used here. The figures show the priors for the intercept, coefficients, and variance parameters, respectively.*

factors on accuracy was investigated with a Bayesian binomial logistic regression. Preliminary analyses of the response times for the tasks indicated that there were not enough trials to make precise estimates of the effects of serial position or sequence length. This preliminary analysis revealed a small reduction in response times with age, but for the sake of brevity an analysis of reaction times is not included here.

Relationship between tasks. The relationship between the tasks is investigated by predicting each measure from the other measures, again using a Bayesian logistic regression.

Relationship to additional outcomes. School-based measures of academic attainment were obtained for a subset of the sample. The relationship between the measures described and academic attainment is investigated using Bayesian linear regression.

2.4 Task details and analysis

2.4.1 Forward digit recall

Forward digit recall could be described as a simple measure of verbal WM or, alternately, a measure of verbal STM. The former description is preferred in highlighting the view taken in this thesis that all short-term storage involves the WM system proper. Previous research can allow concrete predictions to be made about the expected effects of serial position, sequence length, and age on performance. Firstly, we would expect large primacy effects with performance reducing for later serial position (Archibald and Gathercole, 2007; Pickering et al., 1998; St Clair-Thompson and Allen, 2013). Small recency effects for the final item in the sequence should also be observed, particularly for longer sequence. Increasing sequence length should result in drops in performance, particularly for later serial positions (Pickering et al., 1998). Finally, as with almost any cognitive measure, we would expect performance to improve with age. However, these increases may begin to plateau for the oldest children for a simple verbal measure such as FDR (Alloway et al., 2006; Gathercole, 1999; but see, Gathercole et al., 2004a).

2.4.1.1 Task details

Stimuli and timings. Figure 2.3 provided a schematic of the FDR task. Trials began with the word ‘Listen’ being presented in the middle of the screen followed by a blank screen. A sequence of digits was then presented over headphones in a neutral female voice, with a 1000ms inter-stimulus interval. The presentation time for the digits varied between 350-550ms due to differences in utterance length for different digits. No digit was presented more than once within a given sequence. A 1250ms retention interval followed the presentation of the final digit, at which point a row of response boxes was displayed on screen. The response

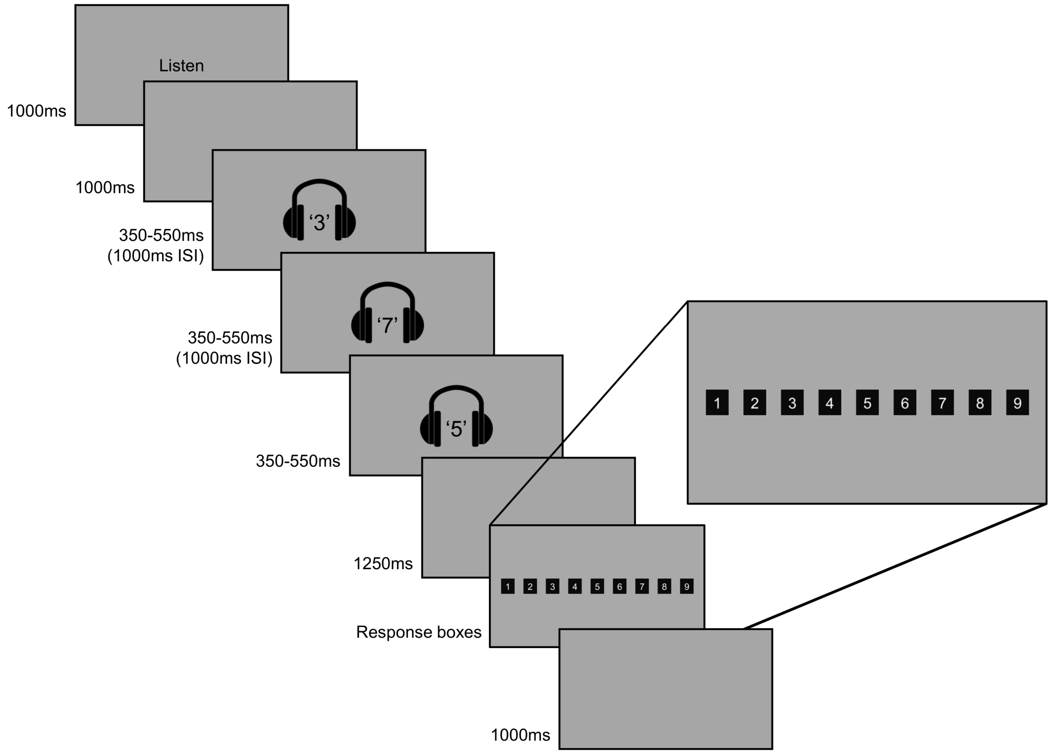


Figure 2.3: *Schematic of the timings and stimuli for the forward digit recall task.*

boxes were presented in a row from one to nine; participants had to touch the boxes in order for each digit that was presented (e.g. the correct response would be 3, 7, 5 in Figure 2.3). The boxes measured 1.7 by 1.9cm, and were dark blue with the internal numbers displayed in white. When pressed a box would briefly turn yellow, and the same box could not be pressed twice in a row. This response method marks the major difference between this implementation of FDR and commonly used versions, where oral recall is typical. Indeed, Chapter 4 (Experiment 1) describes an experiment where a pencil-and-paper version of FDR was employed that included oral recall. The response method chosen here has a number of advantages for large scale testing. Response boxes remove the need for an experimenter to manually record responses, allowing participants to be tested with limited supervision. Avoiding oral recall also makes for a less distracting environment for group testing. This allows for large samples to be collected with relative ease, exemplified by the sample sizes reported in this chapter.

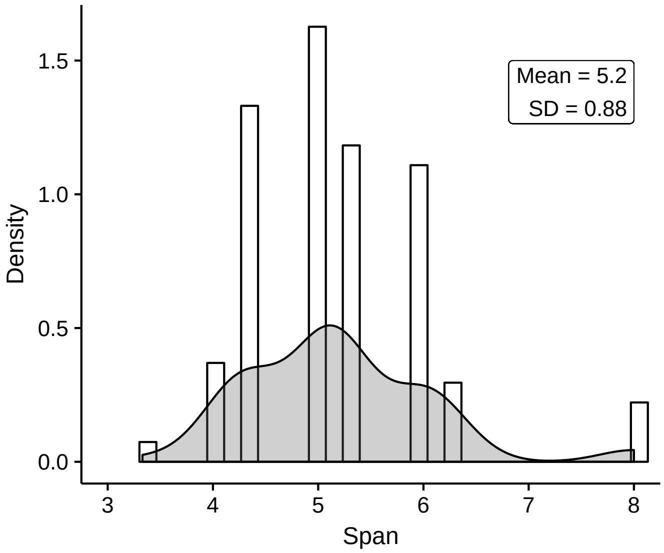


Figure 2.4: *Histogram of span scores from the pencil-and-paper forward digit recall task in Experiment 1 of Chapter 4 with an overlaid density estimate.*

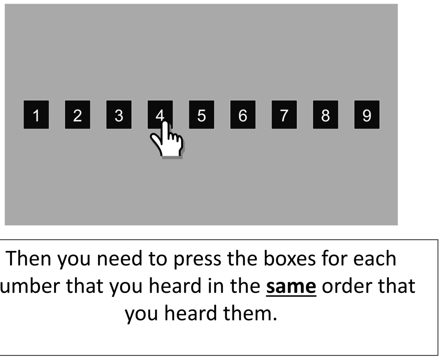
Trial structure. Data for the pencil-and-paper FDR task used in Chapter 4 (Experiment 1) was collected prior to the completion of the development of the computerised tasks in this chapter. That data was, therefore, used to inform the sequence lengths chosen for the computerised task. Figure 2.4 shows the distribution of span scores for the sample of ~80 nine to ten year-olds tested for Experiment 1 in Chapter 4. In the pencil-and-paper task participants were presented with three trials beginning at sequences of three digits. Sequences would be increased by one digit if participants got two or more trials correct at a given sequence length. For example, if a participant got two or three trials correct for sequence of three digits they would proceed to complete three trials with sequences of four digits. A participant's span was the longest sequence length where two or more sequences were correctly recalled. An additional 0.33 was added to a participants' span if they got one trial correct for sequences one item longer than their span. For example, if a participant got 3/3 trials correct at span-five and 1/3 trials correct at span-six their span would be 5.33. Importantly, a participant can achieve a span of six despite having incorrectly recalled some items in reaching that span.

Table 2.5: *Trial structure for the computerised forward digit recall task.*

Block	Sequence Length	N Trial	N Item
1	3	4	12
2	4	4	16
3	5	4	20
4	6	4	24

For the computerised version of FDR, sequences from three to six items were selected, as the aim was to avoid floor effects for younger children as well as avoiding ceiling effects. Children older than ten years have not been tested with the measures presented here, making 10 year-olds (and a span of six) the upper limit on performance. Table 2.5 provided full details of the trial structure for the FDR task. Across 16 trials 72 items are presented providing many more possible scores than a span approach.

Instructions and practice. Participants were always guided through the instructions with the aid of an experimenter. In earlier versions of the task instructions proceeded at a fixed pace. For the piloted version 'Next' and 'Back' boxes were added to allow the experimenter to control the pace of instructions. Experimenters were provided with a manual explaining how to administer the tasks, and were trained prior to testing to ensure consistency. Building instructions into the task and providing a training manual meant that the tasks could be administered by experimenters with relatively little experimental experience. The instructions were followed by three practice trials; one with a sequence of two digits, and two with sequence lengths of three. Following the practice trials, participants were given the opportunity to ask question prior to the test trials beginning. After each block of test trials a message was displayed stating 'Now you will have to remember one more number'.



Then you need to press the boxes for each number that you heard in the **same** order that you heard them.

Figure 2.5: *An example of one of the instruction images presented at the start of the forward digit recall task.*

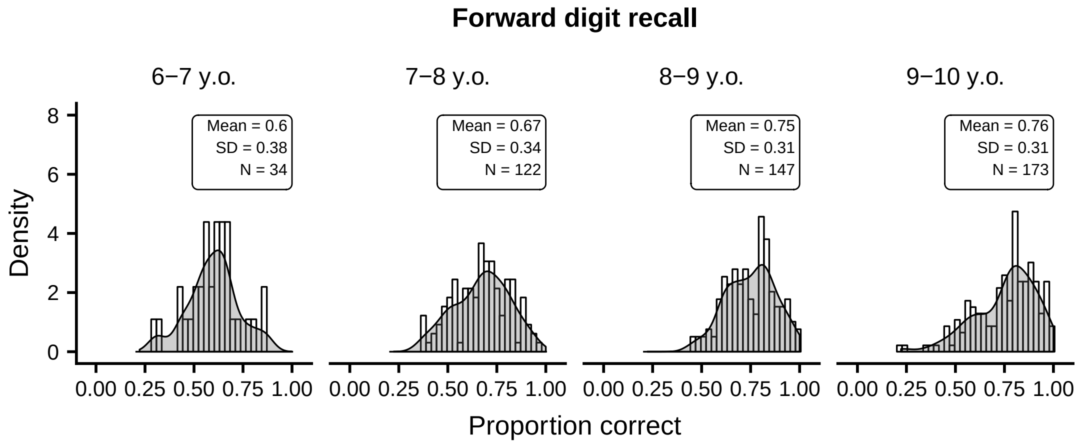


Figure 2.6: *Histogram of overall proportion correct by age group, with overlaid density estimates.*

2.4.1.2 Analysis

Overall performance. Firstly, we can consider the overall distribution of accuracy scores. Figure 2.6 shows the distribution of scores for each age group tested, as well as the sample sizes, means and SDs. This, and all other plots, were created after the exclusions (described in Table 2.4).

It is clear from the histograms of overall performance that there is a substantial amount of variability between participants. This is desirable for an individual difference measures aimed at comparing participants. It is also noteworthy that

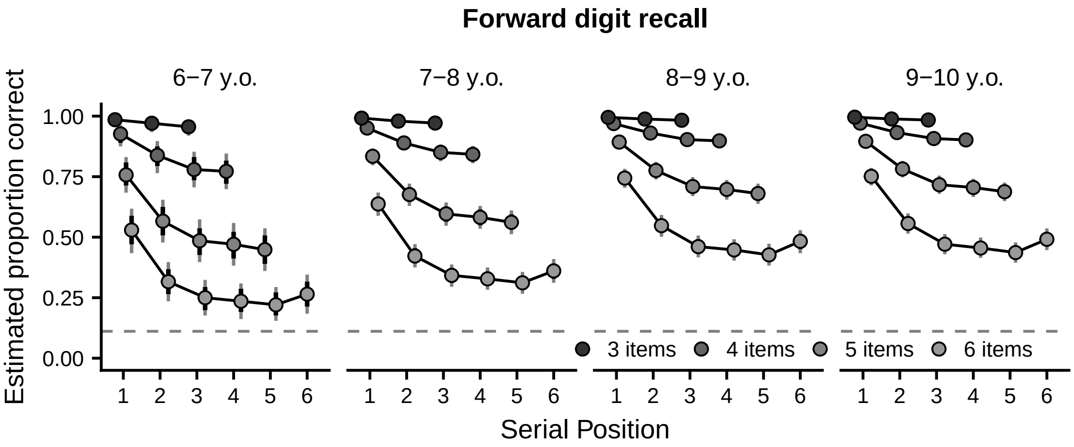


Figure 2.7: *Posterior estimates for accuracy grouped by sequence length, serial position, and age group. The black lines either side of the point estimates denote the 80% Bayesian credible interval. The thinner grey lines show the 95% credible interval. The horizontal dotted line shows chance performance.*

there are many more possible values when using a proportion correct outcome measure. The histograms in Figure 2.6 do not show the discontinuity typical of span measure (see Figure 2.4).

Sequence length and serial position effects. Figure 2.7 shows Bayesian posterior estimates for the proportion correct at different span lengths and serial positions, grouped by age. The points show the most likely estimate from the model, and the error bars show the 80% and 95% posterior credible intervals. Non-overlap between these intervals can be used to conclude a genuine difference between two estimates. Firstly it is clear that there are substantial sequence length effects at all serial positions and for all age groups. In terms of serial position effects, performance at the first serial position is consistently superior to subsequent positions. These serial position effects are particularly pronounced for longer sequence lengths, where participants are no longer performing at ceiling. There also appear to be small recency effects for sequences of six items.

To investigate these recency effects further, posterior odds were used. The odds in support of superior performance at the final position with sequences of six items

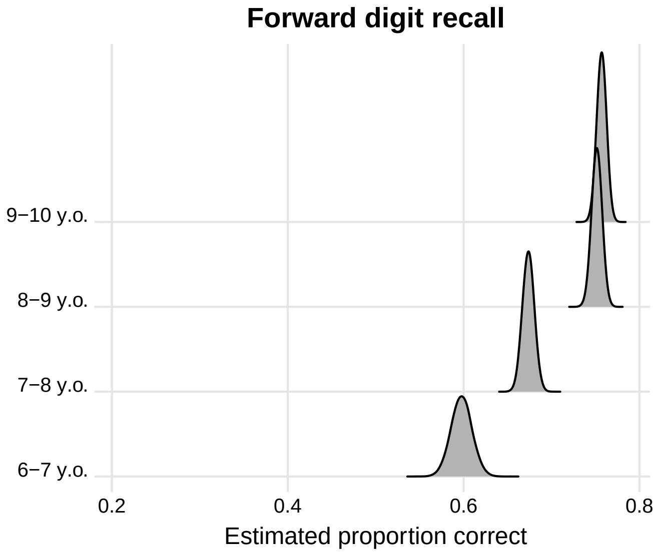


Figure 2.8: *The distribution of posterior estimates of overall accuracy for each age group.*

varied from 4:1 for the youngest children to 26:1 for the oldest. Thus, the support is strong for primary effects and less clear for recency effects.

Age effects. From Figure 2.7, there appear to be improvements in performance with age, particularly at longer sequence lengths. Figure 2.8 shows the distribution of posterior estimates of the overall proportion correct for each age group, averaging across sequence length and serial position. Performance appears to improve with age up to 8-9 years, after which there is then only a small improvement with age.

2.4.1.3 Discussion

Overall the forward digit recall task showed the expected patterns in accuracy. Recency, and particularly primacy, effects were observed, in line with similar tasks in the literature (Archibald and Gathercole, 2007; Pickering et al., 1998; St Clair-Thompson and Allen, 2013). In addition performance improved with age, plateauing towards the upper end of our age range. This is, again, a typical profile for a simple verbal WM measure. For example, Gathercole (1999) summarised data showing that performance on forward digit recall tasks improves rapidly between 4 and 8 years, with only small improvements thereafter. However, other studies have shown that digit recall performance improves at a similar rate between 6 and 10 years (Gathercole et al., 2004a).

One concern for this task might have been that sequences of six items would result in ceiling effects, due to the spans observed with similarly aged children using pencil-and-paper method (see Figure 2.4). These concerns did not materialise; minimal ceiling effects were observed, even for the oldest children. Equally, there were no floor effects for the youngest children, with only a very small number performing at chance. Why was performance lower on this measure than a pencil-and-paper span measure? It is likely that when experimenters are reading out sequences for pencil-and-paper measures they naturally chunk them to some degree, even when trying not to. This would result in improved performance for pencil-and-paper measures compared to computerised measures, where presentation is controlled. The response method used here may also infer small costs by requiring participants to translate (presumably) verbal representation into a spatial response. However, the observation of typical response profiles for verbal STM suggests that this difference of response method does not undermine the validity of the task. Taken together, the results suggest that the FDR task effectively captured the construct of interest, despite the methodological changes made to maximise ease of testing.

2.4.2 Backward digit recall

The backward digit recall (BDR) task was identical to FDR, except that participants were instructed to recall the sequence of letters in backward order. Shorter sequences were also used to accommodate the difficulty of the additional processing requirements.

As with FDR, we would expect to see performance on BDR reduce with increasing sequence length. Serial position curves for backward recall often look like the reverse of those observed for forward recall (Brown, 1974; Farrand and Jones, 1996; Hitch et al., 1988; St Clair-Thompson and Allen, 2013). With backward recall the first items to be presented is the last to be recalled, meaning accuracy for these items is low. In contrast, later items in a sequence are recalled first, and accuracy would be expected to be higher (St Clair-Thompson and Allen, 2013). Thus, large recency effects should be observed for BDR. St Clair-Thompson and Allen (2013) observed primacy effects for backward recall, in addition to large recency effects. However, Brown (1974) did not find primacy effects with a sample of 8 year-old children. It is, therefore, less clear whether primacy effects should be expected.

Finally, performance on BDR should improve with age. Gathercole et al. (2004a) report considerable improvements in BDR between six and ten years (see also, Alloway et al., 2006). Between these ages performance improves by approximately 1 SD in their sample. They also do not report any plateau in performance, meaning that considerable differences between even the oldest age groups should be expected here.

2.4.2.1 Task structure

Stimuli and timings. The stimuli and timings for BDR were identical to FDR (see Figure 2.3).

Trial structure. As with FDR, data from a pencil-and-paper version of BDR was used to inform the chosen sequence lengths. Figure 2.9 shows the distribution of BDR scored for the same group of 9 to 10 year-olds tested for the first experiment in Chapter 4.

Average BDR span for nine to ten year-olds was approximately two items lower than for FDR. However, reducing sequence lengths by two, relative to FDR, would involve presenting sequences of length one. Clearly, a sequence of length one

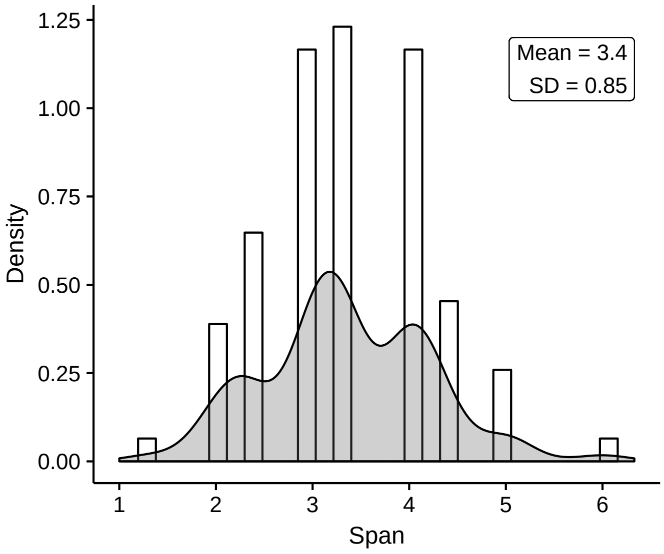


Figure 2.9: *Histogram of span scores from the pencil-and-paper backward digit recall task in Experiment 1 of Chapter 4 with an overlaid density estimate.*

Table 2.6: *Trial structure for the computerised backward digit recall task.*

Block	Sequence Length	N Trial	N Item
1	2	4	8
2	3	4	12
3	4	4	16
4	5	4	20

cannot be recalled in backwards order. Thus, sequence lengths were one item shorter than FDR instead, as Table 2.6 shows. Across the four blocks participants were presented with a total of 56 items.

Instructions and practice. The instructions for BDR were very similar to FDR except that the need to recall the sequences in backward order was emphasised. Experimenters were specifically trained on how to explain this aspect of the task to young children, as it is something they can struggle to grasp. Three practice trials followed the instructions, all with sequences two items in length.

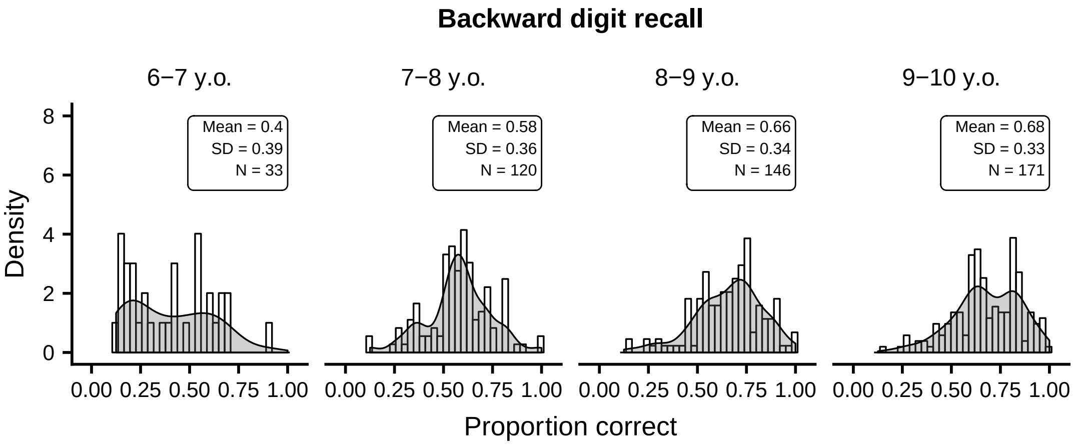


Figure 2.10: *Histogram of overall proportion correct by age group with overlaid density estimates.*

2.4.2.2 Analysis

Overall performance. The distribution of overall accuracy for each age group is shown in Figure 2.10. There appears to be greater variability in performance between participants than for FDR, with a number scoring close to chance. This is in addition to the 3 participants who were excluded for scoring at chance.

Sequence length and serial position effects. Figure 2.11 shows the posterior estimate for performance on the BDR task at different (presented) serial positions and sequence lengths, grouped by age group. Serial position 1 on the figure was the first item to be presented but the last to be recalled, due to requirement for backward recall. There are large recency effects across all sequence lengths for each age group. Unlike FDR there is not any evidence of primacy effects. Performance decreased reasonably steadily across positions, and appears to be at chance for sequences of 5 items in the youngest children.

Figure 2.11 also shows a noticeable improvement in accuracy after the second position for sequences of 4 and 5 items. This could reflect the fact that 3 digits are relatively easy to recall in reserve order as only the first and second digits need to be swapped (e.g. recalling ‘592’ as ‘295’). Participants may able to apply

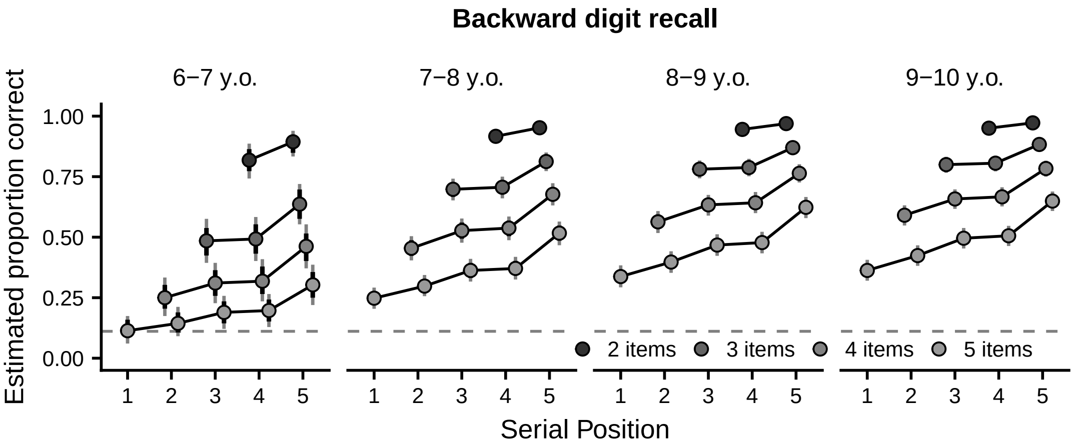


Figure 2.11: *Posterior estimates for accuracy grouped by sequence length, serial position, and age group. The black lines either side of the point estimates denote the 80% Bayesian credible interval. The thinner grey lines show the 95% credible interval. The horizontal dotted line shows chance performance.*

this recall strategy to the last three items, which are recalled first.

Age effects. To further demonstrate the age effects shown in Figure 2.11 the estimates were averaged over serial position and sequence length (Figure 2.12). There are large increases in overall accuracy between 6-7 years and 7-8 years of almost 20%. The increases in accuracy reduce with increasing age, but the two oldest age groups appear to be more differentiated than the same age groups for FDR. In addition, performance for each age group is lower on the BDR task than the FDR task (see Figure 2.8).

2.4.2.3 Discussion

For BDR the expected effects of serial position, age group, and sequence length were observed. The main differences with FDR were that overall performance was lower, differences between age groups were larger, and serial position curves were ‘reversed’. The serial position curves, including the absence of primacy effects, replicated previous work with children (Brown, 1974). The absence of primacy effects contrasts to work with adults, where small primacy effects have

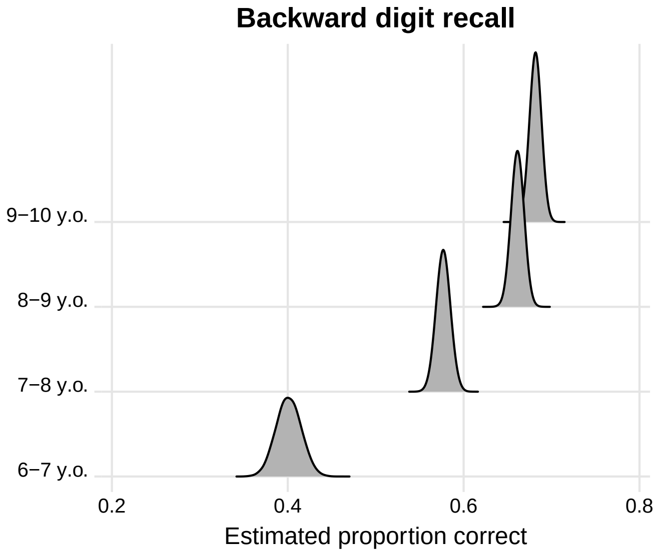


Figure 2.12: *The distribution of posterior estimates of overall accuracy for each age group.*

been observed (St Clair-Thompson and Allen, 2013). Farrand and Jones (1996) compared performance on a forward and backward letter recall tasks in adults, showing that the serial position curves for direction of recall did not differ if the letters were displayed on screen at recall. This manipulation resembles the presentation of response boxes in this task, possibly explaining the similarity of the FDR serial position curves and the reverse of the serial position curves for BDR (i.e. going from position 6 to 1 in Figure 2.11).

2.4.3 Corsi block recall

Originally developed for use with neurological patients (Milner, 1971), Corsi block recall has been widely used in the developmental literature (Alloway et al., 2006; Gathercole, 1999; Gathercole et al., 2004a; Pickering, 2001). It offers a simple measure of visuospatial WM where participants are required to recall whole sequences of spatial locations. The task is, therefore, similar to FDR in that it requires whole-sequence serial recall. Rather than use a physical set of blocks, the locations were presented on a tablet here to allow for more controlled presentation.

In adults, large primacy and small recency effects have been observed using a Corsi block recall task (Hurlstone and Hitch, 2015). Unsurprisingly given the size of the effects, recency effects are not consistently observed (Martin et al., 2017). Pickering et al. (1998) report large primacy effects, for early serial positions and small recency effects for 5 and 8 year-olds. However, recency effects were not observed for the oldest children with sequences of six items. Galera and Souza (2010) also did not observe recency effects for their sample of 8 to 10 year-olds. In addition, primacy effects were limited to the first position, at least for some sequence lengths (Galera and Souza, 2010). Thus, we would expect primacy effects for the earlier items in the sequence and, potentially, small recency effects. As with almost any WM tasks, performance on Corsi reduces with increased sequence length (Galera and Souza, 2010; Pickering et al., 1998), meaning such effects would be predicted here. Performance should also increase with age, in-line with previous work. These improvements with age should be reasonably consistent as block recall performance does not begin to plateau until 13 years-olds (Gathercole, 1999; Gathercole et al., 2004a).

2.4.3.1 Task structure

Stimuli and timings. Figure 2.13 provided a schematic of the Corsi block task. The timings were intentionally very similar to FDR and BDR to facilitate comparison between the tasks. A sequence of dark blue boxes lit up yellow for 500ms before participants were required to recall the sequence by touching the boxes in the same order, after a 1250ms retention interval. The same pseudorandom arrangement of nine boxes was used throughout the set of trials. The same boxes did not light up twice within a given trial. Only the stimuli differed from FDR with the timings and required response (serial recall) remaining the same.

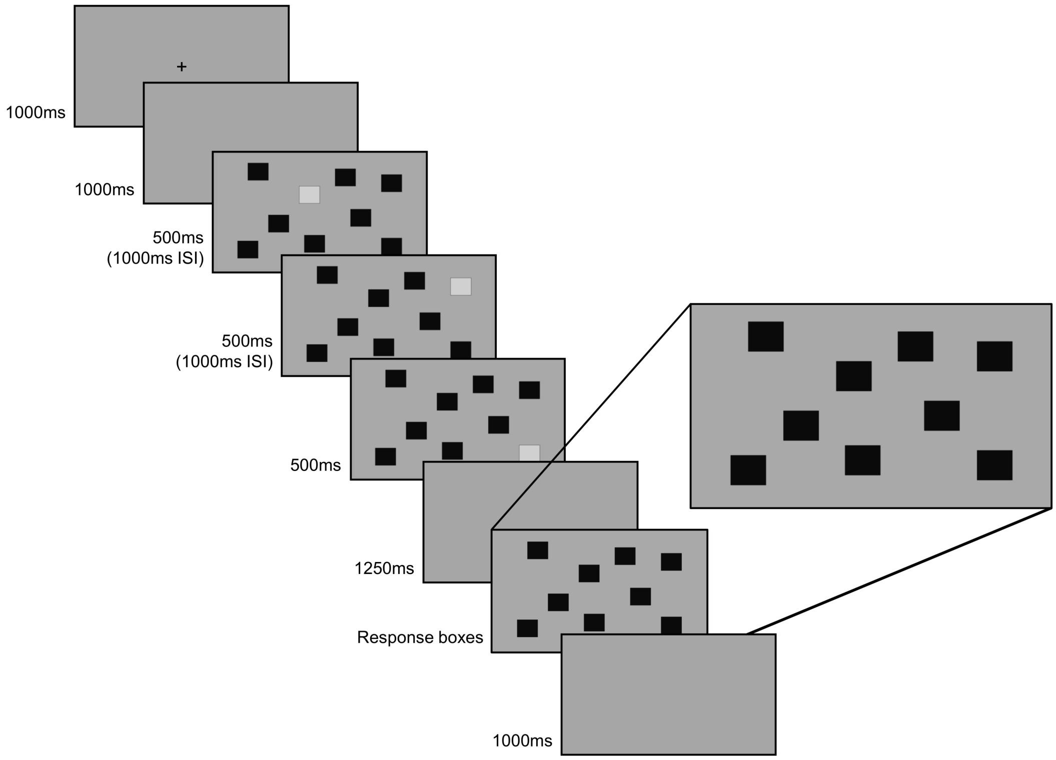


Figure 2.13: *Schematic of the timings and stimuli for the Corsi block recall task.*

Table 2.7: *Trial structure for the computerised Corsi block recall task.*

Block	Sequence Length	N Trial	N Item
1	3	4	12
2	4	4	16
3	5	4	20
4	6	4	24

Trial structure. As with FDR and BDR, data had been previously collected for a pencil-and-paper version of the block recall task. Figure 2.14 shows the distribution of span scores for the sample of 9 to 10 year-olds used in Experiment 1 of Chapter 4. While the distribution of span scores was lower than FDR, the most common span was also five. To facilitate comparison with FDR, sequences of the same lengths as FDR were used. Table 2.7 summarises the trial structure of the block recall task. As with FDR, this resulted in a total of 72 items being presented.

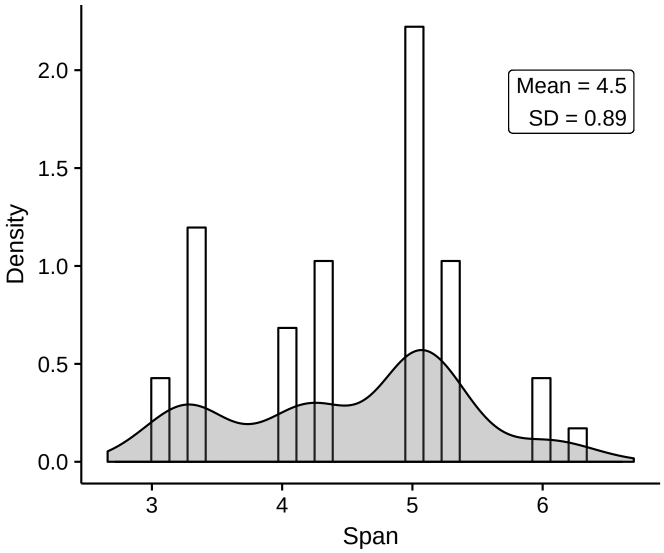
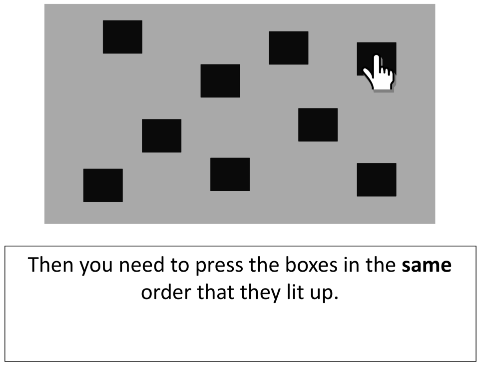


Figure 2.14: *Histogram of span scores from the pencil-and-paper Corsi block recall task in Experiment 1 of Chapter 4 with an overlaid density estimate.*

Instructions and practice. The instructions for Corsi were structured similarly to FDR and BDR. Again, experimenters were given a manual and training on how explain the on-screen instructions (see Figure 2.15). At the end of each block a message was displayed stating 'Now one more box will light up.'



Then you need to press the boxes in the **same** order that they lit up.

Figure 2.15: *An example of one of the instruction images presented at the start of the Corsi block recall task.*

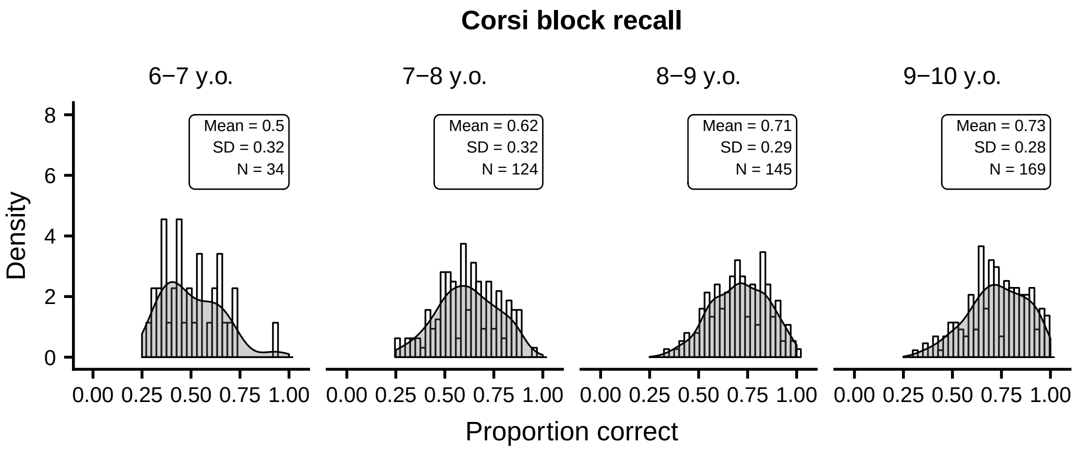


Figure 2.16: *Histogram of overall proportion correct by age group with overlaid density estimates.*

2.4.3.2 Analysis

Overall performance. The raw distribution of accuracy scores are shown in Figure 2.16. The distributions resemble FDR in showing minimal floor and ceiling effects. That said, there is a suggestion of some ceiling effects for the oldest participants. As with previous measures there is considerable variability in performance between participants of the same age.

Sequence length and serial position effects. For Corsi, the effects of serial position were less pronounced than BDR or FDR. Besides small primacy effects, the serial position curves are reasonably flat. The differences in performance for different sequence lengths were also reduced, as a result of the shape of the serial position curves; for FDR and BDR the differences between sequence lengths were driven by later serial positions. Nevertheless, the credible interval for each position did not overlap with those for other sequence lengths, except for 6-7 year-olds, where the sample size was small.

Age effects. As with FDR and BDR there were considerable age effects when looking at participants' accuracy scores. The pattern more closely resembles BDR than FDR in that, while the magnitude of the effects reduce with increasing age,

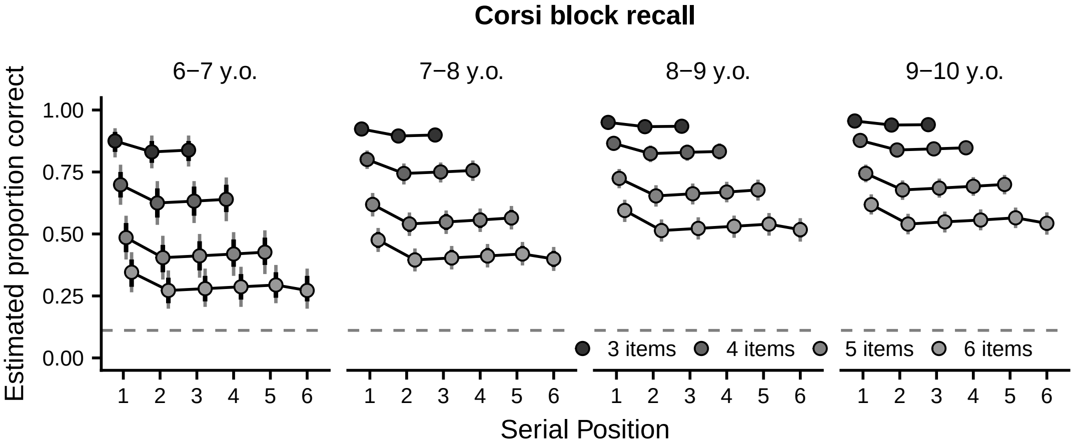


Figure 2.17: *Posterior estimates for accuracy grouped by sequence length, serial position, and age group. The black lines either side of the point estimates denote the 80% Bayesian credible interval. The thinner grey lines show the 95% credible interval. The horizontal dotted line shows chance performance.*

there is some differentiation between the oldest age groups.

2.4.3.3 Discussion

Corsi successfully showed effects of age, serial position, and sequence length. The serial position curves were generally flatter than those observed previously (Galera and Souza, 2010; Pickering et al., 1998). That said, even for studies that observed greater curvature in serial position curves, the profile was flatter than those for verbal tasks (Pickering et al., 1998). Floor and ceiling effects were minimal, except for the oldest children. The task could be adapted for use with older children by replacing the 3 item block with one where 7 item sequences are used, as all ages were close to ceiling for 3 item sequences, suggesting their utility might be minimal. The increases in performance with age began to plateau for the oldest age groups, in contrast to some (Gathercole, 1999; Gathercole et al., 2004a), though not all (Pickering, 2001), previous studies. Including longer sequences could result in larger age differences by removing 3 item sequences, where most participants are at ceiling. Taken together, the analyses suggest that Corsi was effective in showing

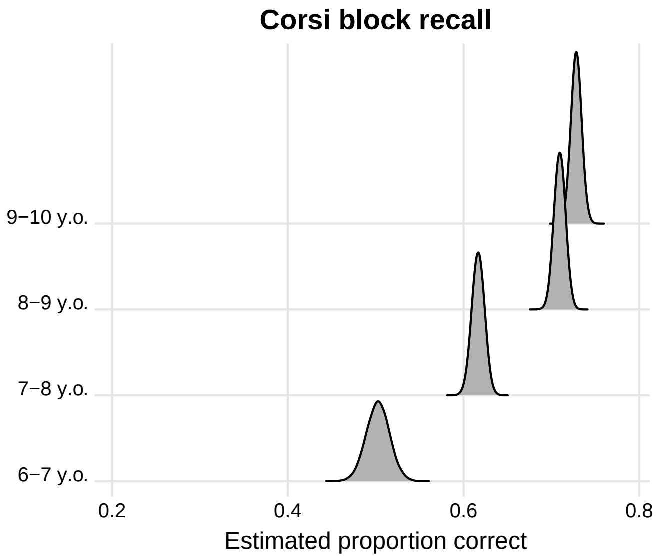


Figure 2.18: *The distribution of posterior estimates of overall accuracy for each age group.*

the expected effects of sequence length and age. While the serial position curves were flatter than expected, primacy effects were still observed for the first serial position.

2.4.4 Odd-one-out

This odd-one-out task (OOO) was an adaptation on those that exist in the literature, aimed at addressing some perceived weaknesses, particularly the possibility of verbal recoding. This issue was addressed here by using five of a possible nine pseudorandom locations to present the stimuli on each trial. These locations did not afford easy verbal labelling, making it far more likely that participants are genuinely encoding visuospatial information. The locations used were also the same as those used for Corsi, making the basic storage requirements of the two tasks as similar as possible.

As the OOO was a novel adaptation of previous tasks, predictions with respect to serial position were less easy to make. The basic storage requirements of OOO are very similar to Corsi, meaning that a similar profile of primacy effects with small recency effects was expected. Alloway et al. (2006) report performance on their odd-one-out task between 4 and 10 years-old. Improvements with age are rapid at first, slowing between 8 and 10 years. However, as noted, this could reflect the recruitment of a verbal recoding strategy. This is particularly plausible considering that performance on complex WM measures typically continues to improve into adolescence (Gathercole, 1999). We would, therefore, expect performance to improve with age, and for even the oldest age groups to differ from one another, as the possibility of verbal recoding was reduced here.

2.4.4.1 Task structure

Stimuli and timings. Figure 2.19 provided a schematic of the odd-one-out (OOO) task. The task combined storage of visuospatial information with concurrent processing. Items could be presented at nine possible locations – the same nine locations used for Corsi. Within a given trial only five of these nine locations were used. For the processing component of trials, three shapes were simultaneously displayed, with one shape differing from the other two (e.g. two circles and a square, see Figure 2.19). Participants were required to touch the shape that differed from the other two (the ‘odd-one-out’), as well as remembering its location. Upon touching a shape it would turn yellow for 200ms, followed by a 1000ms blank screen prior to the next set of shapes. The location of the odd-one-out was not repeated within a trial. Eight different shapes (circle, diamond, hexagon, inverted triangle, pentagon, square, trapezium, triangle) were used for the stimuli, displayed at 2.5cm^2. The shapes were paired to create eight possible combinations, with each shape serving as the distractor and odd-one-out once. The shape pairs were selected manually by trying to match the visual similarity of the two shapes within each pair.

After completing a sequence of processing judgements, five black outlines were

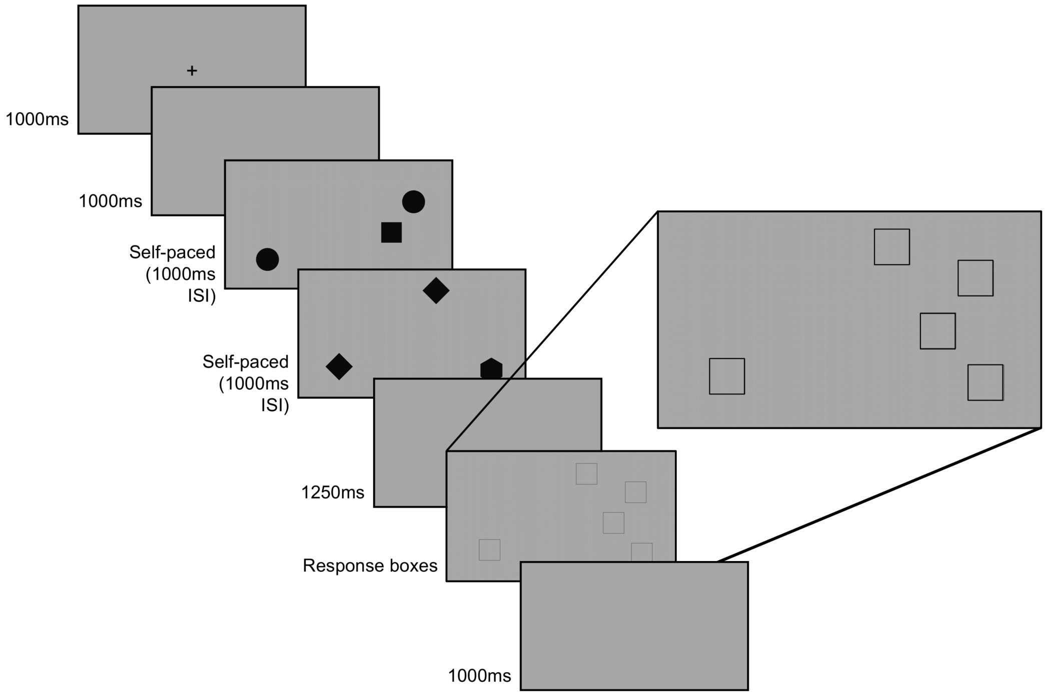


Figure 2.19: *Schematic of the timings and stimuli for the odd-one-out task.*

displayed at the locations used within that trial. Participants were required to recall the locations of the odd-ones-out in order by touching their locations on screen. As with the shapes, the location would briefly turn yellow when pressed so that participants knew their response had been successfully recorded. The timings of the task were matched as closely as possible to the other tasks, with the caveat that the processing component was self-paced.

Trial structure. For the odd-one-out (OOO) task the trial structure was selected to match BDR. We expected the task to be sufficiently difficult to avoid ceiling effects with primary school aged children, while also minimising floor effects. Table 2.8 details the trial structure for OOO.

Instructions and practice. Figure 2.20 shows an example of one of the instruction images for OOO. At the end of each block participants were made aware of the increase in sequence length with a message stating: 'Now there will be one more odd-one-out to remember'. Participants completed three practice

Table 2.8: *Trial structure for the computerised odd-one-out task.*

Block	Sequence Length	N Trial	N Item
1	2	4	8
2	3	4	12
3	4	4	16
4	5	4	20

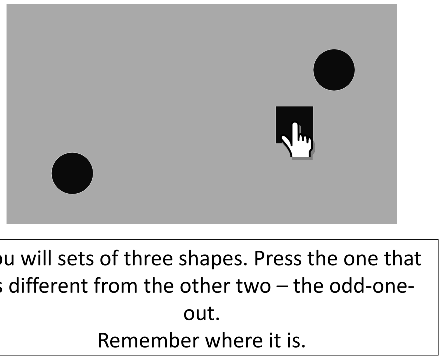

You will sets of three shapes. Press the one that is different from the other two – the odd-one-out.
Remember where it is.

Figure 2.20: *An example of an instruction image from the odd-one-out task.*

trials after the instructions all with sequence of two sets of shapes.

2.4.4.2 Analysis

Overall performance. Overall accuracy for OOO was highly variable with a number of participants performing close to chance (Figure 2.21). For OOO chance performance was 0.2 as only five locations were used within a given trial. For the youngest age group a number of children were excluded due to performing at chance. Those that were left were generally still performing very close to chance.

Sequence length and serial position effects. Figure 2.22 shows the posterior estimates for accuracy at each serial position and sequence length, for each age group. For sequences of four items or more, the youngest children appear to be performing at chance. Performance is also lower in general than Corsi, with no ceiling effects, even for sequences of two items. Primacy effects are also smaller than those observed for Corsi. Finally, the serial position curves are flat for all

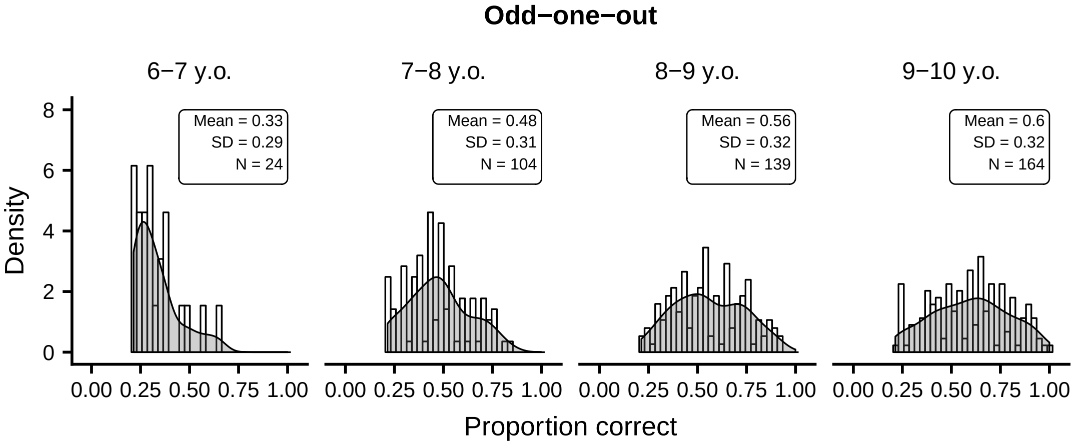


Figure 2.21: *Histogram of overall proportion correct by age group with overlaid density estimates.*

age groups, as with Corsi.

Age effects. Estimates of overall performance for each age group (Figure 2.23) were lower than any other task, including BDR (Figure 2.12). Even for the oldest age group the most credible estimate for overall accuracy was ~60%. The age groups were also more consistently differentiated than other tasks – the difference between the oldest age groups was largest for OOO.

2.4.4.3 Discussion

The results suggest that OOO was possibly too difficult for the youngest children. A number of 6-7 year-olds were scoring at chance and, anecdotally, some struggled to grasp the aim of the task. The serial position curves were similar to those observed for Corsi, though the primacy effects were smaller. Reductions in performance for longer sequences were of a similar magnitude to Corsi, but smaller than the verbal tasks. Finally, as predicted, performance improved with age such that there was a notable difference in accuracy between even the oldest age groups. With the exception of some floor effects for the youngest children, OOO was generally successful. The absence of a plateau in performance for the

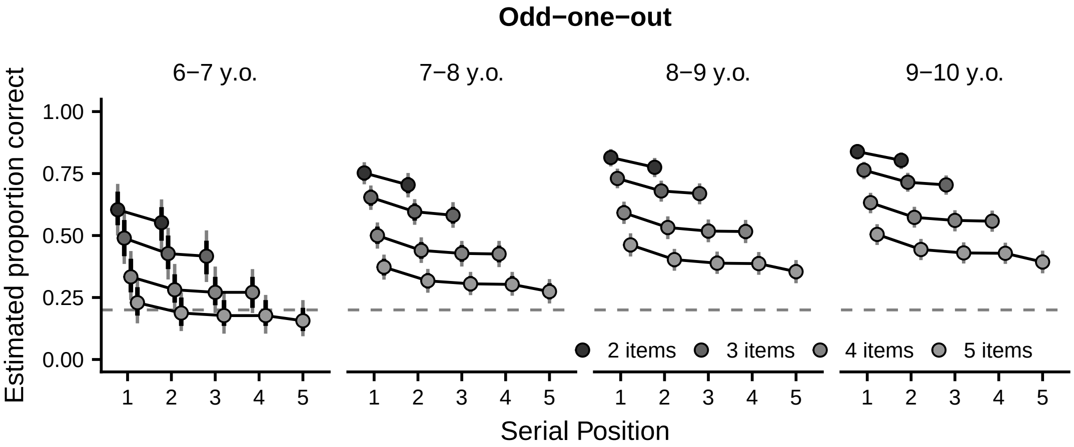


Figure 2.22: *Posterior estimates for accuracy grouped by sequence length, serial position, and age group. The black lines either side of the point estimates denote the 80% Bayesian credible interval. The thinner grey lines show the 95% credible interval. The horizontal dotted line shows chance performance.*

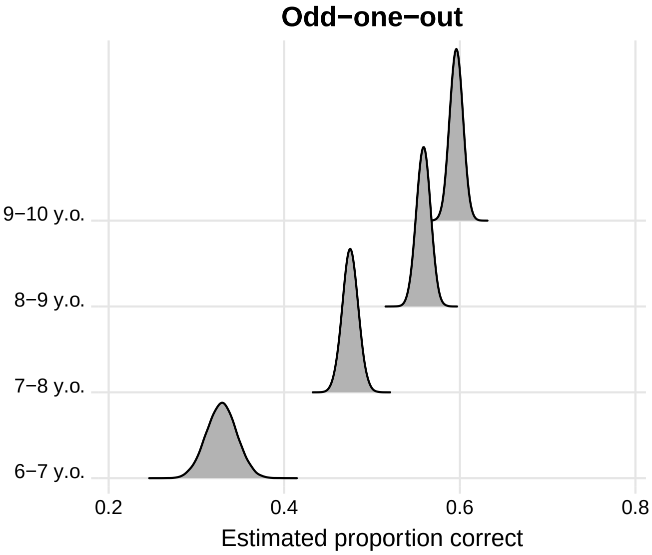


Figure 2.23: *The distribution of posterior estimates of overall accuracy for each age group.*

Table 2.9: *The average estimates of how long each task took to administer.*

Task	Average Length (minutes)
Forward digit recall	5
Backward digit recall	5
Corsi block recall	5
Odd-one-out	10

oldest children could reflect the fact that verbal recoding was more difficult with this version of OOO than those used previously (Alloway et al., 2006).

2.4.5 General discussion of task structure

All the tasks were successful as short measures of WM. Indeed, the only changes made after the initial piloting stage were bug fixes. Improvements in performance with age were more prolonged for complex measures compared to simple WM measures, in line with previous work (Alloway et al., 2006; Gathercole, 1999; Gathercole et al., 2004a). The serial position curves were broadly as expected, particularly for the verbal measures. For the visuospatial measures the curves were generally flatter than others have observed (Galera and Souza, 2010; Pickering et al., 1998). Performance reduced for longer sequences for all tasks. There was also some indication that 6-7 year-olds found OOO too difficult; a number were performing at chance, and the performance of the rest of that age group was low.

As shown in Table 2.9 the tasks were generally quick to administer, although OOO tended to take over 10 minutes with the youngest children. Thus, the aim of producing a set of WM measures that could be administered quickly was met, especially with FDR, BDR, and Corsi.

2.5 Relationship between tasks

The success of the measures outlined above can be further investigated by looking at their inter-relationship. Two broad predictions can be made for the relationships between the tasks. Firstly, relationships within modalities should be larger than those between modalities – FDR and BDR should be more strongly related than, say, FDR and Corsi or OOO. This prediction flows from the ubiquitous observation that, other things being equal, tasks where similar information has to be remembered are more closely related than tasks where the memoranda are dissimilar (Alloway et al., 2006; Bayliss et al., 2003; Gathercole et al., 2004a, 2016; Gray et al., 2016; Pickering et al., 1998). Importantly, the evidence of stronger within modality relationships is not contested. Where approaches to WM differ is whether they take these observations as evidence for modality specific components (e.g. Gathercole et al., 2004a), or simply a reflection of domain specific processes peripheral to WM (Gray et al., 2016). The second prediction is that complex measures should be more strongly related to measures in other modalities than simple measures – the relationship between BDR and Corsi should be stronger than the relationship between FDR and Corsi. Two plausible assumptions underlie this prediction: (i) complex measures reflect domain-general attention resources as well as storage (Engle et al., 1999), and (ii) domain-general attention is implicated in all WM tasks to some degree by, for example, coordinating modality specific stores (Baddeley, 2007) or resisting interference (Engle, 2002). The relationship between, say, BDR and Corsi would then reflect the fact both tasks draw on general executive resources, to some degree. One complication to these predictions stems from the claim that visual WM tasks are always attentionally demanding, even in the absence of processing requirements (Cowan and Morey, 2006; Morey and Bieler, 2013; Zokaei et al., 2014). If this is the case we would expect Corsi to be more strongly related to verbal WM than FDR with visuospatial WM.

Table 2.10: *Pearson pairwise correlations between the 4 tasks using pairwise complete observations. All correlations are significant with $p < .001$.*

	FDR	BDR	Corsi	OOO
FDR	1.000	0.605	0.452	0.427
BDR	0.605	1.000	0.556	0.490
Corsi	0.452	0.556	1.000	0.623
OOO	0.427	0.490	0.623	1.000

2.5.1 Analysis

For each task a logistic regression was used with overall accuracy on the other tasks and age as predictors. For each predictor, coefficients, 95% Bayesian credible intervals, and posterior odds are reported. Posterior odds describe the numbers of estimates for the parameter of interest that were positive versus negative. For example, posterior odds of 20,000 to 1 would mean that all 20,000 posterior estimates of the parameter of interest were positive. Odds of 20 to 1 would mean that 20 estimate were positive for every estimate that was negative. These ratios can be directly interpreted as the odds in favour of a relationship between a given predictor and the outcome. Diagnostics for each parameter are also reported in the form of effective sample sizes and $\hat{R}$ (R-hat in the tables) value. The effective sample size can be said to measure the unique information in the posterior estimates, where larger values are preferable (Kruschke, 2014). $\hat{R}$ is a measure of convergence between the chains that make up the posterior estimates, where lower values are preferable, and 1 is the minimum possible value. To aid readers unfamiliar with Bayes methods or more comfortable with simple correlations, a correlation matrix is provided in Table 2.10.

2.5.1.1 Predicting forward digit recall

For FDR, as expected BDR was most strongly related to performance. Table 2.11 shows the coefficients for each predictor, expressed as the change in log-odds

Table 2.11: *Coefficients for predicting performance on FDR expressed in log-odds units. For each predictor the coefficients are the change in log-odds predicted from a 1 SD change in the predictor.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	1.010	0.985, 1.035	20000:1	17484	1
BDR	0.354	0.323, 0.385	20000:1	17090	1
Corsi	0.073	0.039, 0.107	19999:1	16264	1
OOO	0.064	0.031, 0.097	6666:1	16163	1
Age	0.049	0.021, 0.077	20000:1	19284	1

[1] Mean refers to posterior distribution

[2] 95% CI is the 95% Bayesian credible interval

[3] Odds are the posterior odds

[4] ESS = effective sample size

expected from a 1 SD change in proportion correct for the predictor. These log-odds values can then be translated back into probabilities with an inverse logistic function to get the changes in accuracy associated with a 1 SD change in BDR. A 1 SD increase in accuracy for BDR was associated with an expected 6.3% increase in accuracy for FDR. While the coefficient was largest for BDR, relationships for all the other predictors were supported with odds of 6666 to 1 or greater.

2.5.1.2 Predicting backward digit recall

For BDR relationships between all the predictors and performance were supported, with posterior odds of 20,000 to 1 in each case. The strongest relationship was for FDR where a 1 SD change in accuracy was associated with a 7.22% increase in expected accuracy on BDR. Despite being a ‘simple’ measure, Corsi had a stronger relationship with BDR than OOO.

Table 2.12: *Coefficients for predicting performance on BDR expressed in log-odds units. For each predictor the coefficients are the change in log-odds predicted from a 1 SD change in the predictor.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	0.561	0.534, 0.587	20000:1	20000	1
FDR	0.329	0.299, 0.36	20000:1	19493	1
Corsi	0.217	0.182, 0.252	20000:1	14019	1
OOO	0.143	0.108, 0.178	20000:1	14832	1
Age	0.085	0.056, 0.114	20000:1	19178	1

[1] Mean refers to posterior distribution

[2] 95% CI is the 95% Bayesian credible interval

[3] Odds are the posterior odds

[4] ESS = effective sample size

2.5.1.3 Predicting Corsi block recall

The relationships for FDR, BDR, OOO, and age with Corsi are summarised in Table 2.13. As expected OOO has the strongest relationship with Corsi, with a 1 SD change in OOO being associated with a 6.84% change in accuracy. BDR showed a stronger relationship with Corsi than FDR.

2.5.1.4 Predicting odd-one-out

Finally, Table 2.14 shows the coefficients for predicting OOO. Corsi showed the largest relationship with OOO, as expected. A 1 SD chance in performance on Corsi was associated with a 4.87% change in the expected accuracy on OOO. In line with our predictions, BDR was more strongly associated with OOO than FDR.

2.5.2 Discussion

In general, the expected relationships between the tasks were observed; within-modality relationships were larger than those between modalities.

Table 2.13: *Coefficients for predicting performance on Corsi expressed in log-odds units. For each predictor the coefficients are the change in log-odds predicted from a 1 SD change in the predictor.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	0.800	0.775, 0.824	20000:1	19296	1
FDR	0.068	0.038, 0.097	20000:1	15676	1
BDR	0.187	0.155, 0.218	20000:1	15098	1
OOO	0.344	0.316, 0.374	20000:1	17109	1
Age	0.088	0.062, 0.114	20000:1	18848	1

[1] Mean refers to posterior distribution

[2] 95% CI is the 95% Bayesian credible interval

[3] Odds are the posterior odds

[4] ESS = effective sample size

Table 2.14: *Coefficients for predicting performance on OOO expressed in log-odds units. For each predictor the coefficients are the change in log-odds predicted from a 1 SD change in the predictor.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	1.090	1.069, 1.11	20000:1	18556	1
FDR	0.034	0.009, 0.059	293:1	15656	1
BDR	0.114	0.087, 0.141	20000:1	14604	1
Corsi	0.278	0.252, 0.303	20000:1	16665	1
Age	0.065	0.043, 0.087	20000:1	17889	1

[1] Mean refers to posterior distribution

[2] 95% CI is the 95% Bayesian credible interval

[3] Odds are the posterior odds

[4] ESS = effective sample size

Additionally, complex measures were more related to measures in other modalities than simple measures, with the exception that Corsi was more strongly related to FDR and BDR than OOO. The relationship between Corsi – ostensibly a simple WM measure – and verbal WM speaks to the claim that visuospatial WM is always attentional demanding (e.g. Morey and Bieler, 2013). If even simple visuospatial measures draw on general attention, then this would explain a relationship with verbal WM, assuming that attentional resources are implicated, to at least some degree, in all WM tasks. The pattern of relationships observed here further support the conclusion that this set of measures represent valid measures of the constructs of interest.

2.6 Predicting academic attainment

One of the most consistent findings in the WM literature is that performance on WM measures is related to academic attainment in children (Alloway and Alloway, 2010; Cowan, 2013; Cragg and Gilmore, 2014; Gathercole et al., 2004b; LeFevre et al., 2013; Monette et al., 2011). Thus, if these measures had successfully tapped WM they would be expected to relate to academic attainment.

Here school-based measures of academic attainment are used. These measures are routinely used by schools and are central to monitoring the progress of children. The three core domains reflected in these measures are maths, reading, and writing. In comparison to experimental measures of academic attainment, these measures are rather complicated in that they encompass a wide range of tasks. Thus, there is a limit to how precise conclusions and predictions can be when using these measures; it is not possible to comment on specific mechanisms that contribute to, say, maths attainment. Nevertheless, school-based measures have the advantage of being easily accessible and important to schools. Furthermore, the complexity of the measures (in capturing a range of abilities) allows for the investigation of how measures relate to academic ability *in general*. Using school-based measures also marries with the aims of this chapter to develop WM measures for use in a longitudinal cohort study. School-based attainment measures are often used in such studies where it may be infeasible to test thousands of children with experimental attainment measures.

While different measures of academic attainment tend to be highly correlated, two predictions can be made based on the past literature. Firstly, complex measures of WM should be more strongly related to maths attainment than simple measures (Alloway and Alloway, 2010; Gathercole et al., 2016; Raghubar et al., 2010; St Clair-Thompson and Sykes, 2010; Van de Weijer-Bergsma et al., 2014). A recent meta-analysis suggests that these relationships with maths should be similar for verbal and visuospatial complex tasks (Peng et al., 2015). Secondly, verbal measures should be more strongly related to reading than visuospatial measures (Alloway et al., 2005; Gathercole et al., 2016; Hall et al., 2015; St Clair-Thompson and Sykes, 2010; Swanson et al., 2009). That said, recent work has found similar magnitude relationships with reading for visual and verbal WM measures (Swanson, 2017). For writing, relationships with WM are varied, likely due to the range of processes involved in writing. Relationships have been observed only for complex measures (Alloway et al., 2005; Swanson and Berninger, 1996), particularly for visuospatial WM (Bourke et al., 2013), and for both complex and simple verbal and visuospatial measures (St Clair-Thompson and Sykes, 2010). Thus, it is hard to make strong predictions for writing, except that there will be a relationship with WM, as results appear to vary for the different processes underpinning writing attainment.

2.6.1 Analysis

Academic attainment data was obtain for the pilot study, Experiment 1 of Chapter 3, and Experiment 2 of Chapter 4. However, the form of the data was not consistent between the studies due to changes in the attainment data that schools

Table 2.15: *Standardised coefficients and diagnostics for a linear regression predicting maths.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	9.983	9.753, 10.214	20000:1	13634	1
FDR	0.246	-0.054, 0.544	17:1	20000	1
BDR	0.510	0.195, 0.828	999:1	20000	1
Corsi	0.263	-0.054, 0.584	18:1	20000	1
OOO	0.392	0.071, 0.715	118:1	14655	1
Age	1.538	1.291, 1.787	20000:1	18004	1

[1] Mean refers to posterior distribution
[2] 95% CI is the 95% Bayesian credible interval
[3] Odds are the posterior odds
[4] ESS = effective sample size

were required to collect by the Department of Education, UK. For the pilot study, data was obtained that indicated whether a children was performing at, above, or below the expected level for their year group. For the other two experiments, information on the level at which the child was performing was obtained. The pilot data is not analysed given that a more precise measure of academic attainment was obtained for the other two experiments. The sample used for the attainment analysis included 154 children, after excluding participants who were distracted or had special educational needs. The mean age of the sample was 9.33 (SD = 0.83; Range = 7.50 - 10.49).

2.6.1.1 Predicting maths

Table 2.15 shows the coefficients for a Bayesian linear regression predicting maths, combining the data from Experiment 1 of Chapter 3, and Experiment 2 of Chapter 4 (N = 154). The posterior odds support relationships with maths for BDR, OOO and age. Even ‘controlling’ for age, a 1 SD change in BDR was associated with a 0.5 SD change in academic attainment. The estimated R^2 for this model was 0.70 (95% credible interval: 0.62 - 0.76), showing that the set of predictors explained the majority of the variance in academic attainment.

Table 2.16: *Standardised coefficients and diagnostics for a linear regression predicting reading.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	10.424	10.179, 10.668	20000:1	14713	1
FDR	0.366	0.053, 0.676	84:1	20000	1
BDR	0.351	0.024, 0.677	55:1	20000	1
Corsi	0.340	0.008, 0.671	44:1	20000	1
OOO	0.258	-0.077, 0.594	15:1	16513	1
Age	1.520	1.252, 1.783	20000:1	20000	1

[1] Mean refers to posterior distribution

[2] 95% CI is the 95% Bayesian credible interval

[3] Odds are the posterior odds

[4] ESS = effective sample size

2.6.1.2 Predicting reading

As expected, both verbal measures were related to reading attainment, with similar coefficients (see Table 2.16). In addition, a relationship between Corsi and reading was supported by odds of 44:1. The estimated R^2 for this model was high at 0.62 (95% credible interval: 0.58 - 0.73), meaning the predictors again explain the majority of variance in this attainment measure.

2.6.1.3 Predicting writing

Finally, Table 2.17 shows the coefficients for predicting writing. The posterior odds strongly supported relationships with writing for BDR and Corsi. A relationship between FDR and writing was also supported by odds of 21 to 1. Again, the set of predictors explained the majority of variance in the outcomes with an R^2 of 0.63 (95% credible interval: 0.54 - 0.71).

Table 2.17: *Standardised coefficients and diagnostics for a linear regression predicting writing.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	10.086	9.813, 10.36	20000:1	13606	1
FDR	0.307	-0.053, 0.66	21:1	20000	1
BDR	0.451	0.083, 0.822	112:1	20000	1
Corsi	0.379	0, 0.752	39:1	20000	1
OOO	0.285	-0.089, 0.659	14:1	14939	1
Age	1.528	1.227, 1.825	20000:1	16876	1

[1] Mean refers to posterior distribution
[2] 95% CI is the 95% Bayesian credible interval
[3] Odds are the posterior odds
[4] ESS = effective sample size

2.6.2 Discussion

Both predictions with respect to maths and reading were supported: complex measures were more strongly related to maths, whereas verbal measures were more strongly related to reading. For writing the predictions were less clear but ultimately relationships were observed for BDR and Corsi. St Clair-Thompson and Sykes (2010) used similar school-based measures of attainment and found that both a Corsi block recall and BDR task were related to writing in 7 year-olds. As with all previous analyses, further support for the validity of these measures is provided by the relationships with attainment.

2.7 General Discussion

The aim of the work in this chapter was to develop a set of working memory measures that would be both comprehensive and easy to administer to groups of children in a school setting. Such measures are important for identifying children with WM difficulties, before any support can be offered. The measures were generally successful in this aim, particularly FDR, BDR, and Corsi. The complex visuospatial measure (odd-one-out) produced data that conformed to prior predictions, except that is was possibly too difficult for the youngest children. The item-level scoring method used here achieved the aim of allowing for greater discriminability between participants, due to a larger set of possible scores. It is not possible to compare this result to other implementations of similar tasks as participant-level data are rarely reported or provided. A novel response method was employed for the verbal tasks to aid with efficient group testing. This appears to not have undermined the verbal nature of the tasks, as the serial position curves resembled those for verbal recall (Pickering et al., 1998).

Performance on all measures reduced with increasing sequence length. For the verbal tasks, these changes with sequence length were particularly evident for longer sequences. Considerable age effects were observed for all measures. Performance began to plateau for older children on the simple measures, particularly FDR. These age effects were generally driven by performance on longer sequences, as most age groups were at ceiling for two and three item sequences.

The relationships between the tasks also conformed to predictions based on previous literature. FDR and BDR were strongly related, reflecting the verbal storage aspects of both tasks. Equally, Corsi and OOO were strongly related; both tasks required participants to recall spatial locations. The effects were not isolated to modality, however. Both FDR and BDR were related to the visuospatial measures, though the coefficients were considerably smaller for FDR. The relationships between FDR and visuospatial WM may speak to the fact that children are less able to automate verbal processes, meaning domain general attentional resources are involved in even simple verbal storage (Cowan et al., 2006; Cowan, 2016b). BDR includes a processing component that draws on attentional resources, at least in children (St Clair-Thompson, 2010). This processing component likely drives the stronger relationships with visuospatial WM, compared to FDR.

The results for the visuospatial measures might seem less clear in that Corsi was more strongly related to BDR than OOO. However, a considerable amount of work now suggests that visual WM tasks are more attentionally demanding than verbal tasks, always requiring general attention resources, even for simple storage in adults (Morey and Cowan, 2004, 2005; Morey and Mall, 2012; Morey et al., 2013; Morey and Bieler, 2013; Morey and Miron, 2016; Zokaei et al., 2014). Thus, Corsi may draw on general attentional resources to a similar extent to, say, BDR. While the additional processing load of OOO will be more attentionally demanding this may, paradoxically, reduce its ability to predict BDR. A small number of participants were scoring very close to chance on the task, such that the predictive power of the task was reduced. This could then explain why the relationship between Corsi and BDR was larger than that for OOO, as there were minimal floor effects with Corsi.

While the literature on academic attainment is complicated by the range of measures used (Peng et al., 2015), the key predictions with respect to this outcome were supported. Both complex WM measures were related to maths attainment, particularly BDR. For reading, both verbal measures and Corsi were related to attainment, with similar coefficients in each case. Finally, complex verbal WM and simple visuospatial WM were related to writing performance. The focus here was to look at the *general* relationships between the tasks and the attainment measures used by schools to monitor progress in children. Thus, the aim was not to make specific claims about particular processing underpinning attainment, as school-based measures of attainment will reflect a range of factors that contribute to attainment. More specific experimental measures would be needed to comment on the various subcomponents of maths, reading, or writing ability (e.g. Bourke et al., 2013; García-Madruga et al., 2014; Gathercole et al., 2006; Raghubar et al., 2010; Swanson and Berninger, 1996). Nevertheless, in as much as the pattern of results reflected those common in the literature, we have further evidence that this set of measures successfully tapped the key aspects of WM. Furthermore, the differences in relationships for each attainment measures show that the WM tasks would provide useful information in the context of large-scale cohort studies (Wright et al., 2013) making use of school-based attainment measures.

One factor which may limit the generalisability of this work is that the majority of participant were of low socioeconomic status (SES). While the evidence is mixed as to whether SES relates to WM ability as such (Arán-Filippetti, 2013; Hackman et al., 2014; Noble et al., 2007; but see, Alloway et al., 2014; Engel et al., 2008; Noble et al., 2005), it is relevant to the development of general cognitive abilities (Farah et al., 2006; Hackman and Farah, 2009). However, it is unlikely that the fundamental *structure* of WM would be affected by SES. Instead, performance levels may be lower than those that would be observed in other populations, particularly for tasks with substantial executive requirements (Arán-Filippetti, 2013). That said both Alloway et al. (2014) and Engel et al. (2008) found no evidence of a difference in performance on BDR between low and high/middle SES children. In addition, Noble et al. (2005) found no evidence of a difference between low and middle SES groups on a spatial WM task, though they later found that SES explained a significant proportion of the variance in a different spatial WM measure (Noble et al., 2007). The aim of this chapter was to determine whether this set of measures showed the expected serial position, age, and sequence length effects, as well as being related to one another and academic attainment. The generalisability of the conclusion based on these criteria is unlikely to be undermined by the use of a low SES sample, though performance levels may be higher for some tasks in other samples. Furthermore, the tasks presented in this chapter were designed for use in the Born in Bradford cohort study, where the majority of participants are in the most deprived decile on the Indices of Multiple Deprivation (Wright et al., 2013). Thus, it was crucial that the tasks were validated with a similar low SES sample.

2.7.1 Applications

Following the success of this set of measures, one could ask how they could be used. Most obviously, they would be effective for use with large-scale studies such as Born in Bradford where quick and effective measures of WM are required. For such projects the time allotted to collect cognitive measures is often very limited. Therefore, if cognitive factors are to be investigated at all, the measures used must be quick to administer, as those presented here are.

As noted in the introduction, one interesting use of these measures could be to understand the relationship between WM and academic attainment alongside a range of other non-cognitive factors that contribution to educational outcomes. This would allow for benchmarking cognitive factors against other factors, rather than simply observing that WM is important to attainment. The battery of WM tasks presented here are also easy enough to administer that they could be included alongside experimental tasks in a wide range of studies. As Chapters 3 and 4 show, including additional individual difference measures in this way provides novel and important insights into the core experimental tasks of interest. Clearly if these tasks were to be used with adults the difficulty would need adjusting. It also may not be appropriate to use BDR as a measure of complex verbal WM with adults (St Clair-Thompson, 2010).

2.7.2 Future directions

One obvious way to extend this set of tasks would be to include more measures. As it stands, each construct of interest is captured by a single measure. Taking multiple measures of a single construct allows more powerful analysis techniques to be used, such as structural equation modelling (e.g. Gathercole et al., 2004a; Gray et al., 2016). Broadening the stimuli used would also facilitate investigating more fine-grained questions about, for example, the differences between visual and spatial tasks (e.g. Pickering et al., 2001). However, such changes would represent a movement beyond the original aims of these tasks to provide a short but reasonably comprehensive set of measures for use by *Born in Bradford* (Wright et al., 2013).

2.7.3 Conclusions

The measures described in this chapter provide a way to quickly assess the core aspects of working memory in primary school aged children. The computerisation of the measures means that large scale testing can be carried out with groups of children, where all responses are automatically recorded. Three of these measures (FDR, BDR, and Corsi) are currently being used to test thousands of children in schools as part of the *Born in Bradford* cohort study. In addition to the practical advantages of presenting tasks on a tablet, the presentation of stimuli can be more carefully controlled. Computerising response recording also removes the possibility of human error present in pencil-and-paper measures. The item-level scoring method employed here also infers benefits over the span scoring approach used in some commercial batteries (e.g. Wechsler, 2003). The following chapters both include experiments where these measures were used alongside novel experimental tasks. The inclusion of individual difference measures provided novel insights into the experimental effects observed.

Chapter 3

Using attention to support working memory performance

3.1 Introduction

Given the apparent limitations in WM capacity and executive control, and the central role these functions play in broader cognition and scholastic attainment, it is important to understand how children's WM performance might be optimised. This could be achieved through both automatic beneficial processes and identifying controlled strategic approaches that the children are able to employ. Indeed, recent perspectives on WM training suggest that the development and utilisation of beneficial strategies might be one way in which such training could prove useful (e.g. Dunning and Holmes, 2014; Gathercole, 2014; Peng and Fuchs, 2017). Several factors are likely to be relevant in considering whether children will show similar strategic benefits to adults. For example, identification and implementation of strategic approaches may be effortful and resource-demanding, with children not having the same degree of resources available, relative to adults. Developments in metacognition or 'meta-control' may also be important (Chevalier, 2015a,b); in addition to understanding a task, and having the ability to engage in a strategy, children must have the metacognitive ability to select appropriately among the strategies available to them.

Classic investigations of strategy use in memory have explored whether, for example, children spontaneously support their performance using verbal rehearsal (e.g. Flavell et al., 1966; Hitch et al., 1988; Jarrold and Hall, 2013). In an analogous exploration in the visuospatial domain, Morey et al. (2017) identified that children aged 5-7 years show strategic sequential looking in rehearsing spatial locations, though this appears to be reactive in nature, whereas older children (8-11 years) and adults demonstrate a more proactive rehearsal approach. Recent work has also started to explore the extent to which children are able to direct their attention towards aspects of a task environment that are particularly goal-relevant (e.g. Cowan et al., 2006, 2010; Shimi and Scerif, 2015; Shimi et al., 2014). For example, Shimi et al. (2014) presented 7 year-olds, 11 year-olds and young adults with simultaneous four-object arrays, followed by a single recognition probe, while manipulating the timing (before versus after encoding) and location (central versus peripheral) of visual cues orienting participants to a particular item in the array. All three age groups showed a similar boost in performance for pre-cues, whereas 7 year-olds showed a smaller advantage from retro-cues. They also found that individual differences in visuospatial short-term memory and, especially, visuospatial WM predicted performance on retro-cued trials, even controlling for age and baseline performance. Thus, children show an effect that may share some features with the prioritisation effect in adults.

Cowan et al. (2010) also demonstrated that children can use context to adjust how much they attend to items. They presented children and young adults with a change detection task manipulating the frequency with which a type of item was probed. With set sizes of 4, all age groups showed the same pattern of performance; capacity allocated to an object reduced as its likelihood of being probed reduced. However, the youngest group (7 to 8 year-olds) failed to appropriately optimise their performance in response to probe frequency under high cognitive load, while performance remained adult-like for 12 to 13 year-olds. This suggests children can adjust what information accesses the focus of attention in response to regularities such as probe frequency, though not to the same extent as adults. Cowan et al. (2006) further highlighted the difference in attentional control between children and adults. Ten year-olds, unlike adults, showed little difference in performance when recalling attended versus unattended streams of information. This suggests that children differ from adults in being less able to use attention to selectively store a subset of the information they are presented with.

One important topic in understanding how limited WM capacity may be allocated concerns the way that participants can be directed to prioritise certain items within a set. This ability has been demonstrated in adults with sequentially presented visual stimuli (Hu et al., 2014, 2016). When presented with 4-item sequences of coloured shapes, adults show a boost in performance for the prioritised item when instructed to try especially hard to remember the item in a specified serial position, at a cost to non-prioritised items (Hu et al., 2014, 2016). This prioritisation effect is reduced under cognitive load, suggesting it is executive in nature (Hu et al., 2016). In achieving the boost in performance prioritised items appear to enter a vulnerable state; prioritised items are also particularly vulnerable to interference from a to-be-ignored item drawn from the stimulus set, termed a visual suffix, presented briefly following the final item in a sequence (Hu et al. 2014, Experiment 4).

An informative feature of sequential presentation is the ability to investigate performance by serial position. This allows the separation of different mechanisms that contribute to visual WM. A robust finding from such analysis is a large recency effect for the final item in a list (Allen et al., 2006, 2014; Gorgoraptis et al., 2011; Hu et al., 2014, 2016). Additionally, performance on the final item in a sequence is vulnerable to suffix interference, yet is largely unaffected by cognitive load (Allen et al., 2014; Hu et al., 2014, 2016). Together these findings suggest that the recency effect for the final item is relatively automatic in nature. The relative absence of a cognitive load effect suggests that endogenous executive mechanisms, such as attentional refreshing (Barrouillet et al., 2004, 2009), are not required to maintain this item in visual WM. The selective suffix effect for later items suggests that perceptually driven processing of a to-be-ignored item displaces the final item from a state where it is otherwise automatically maintained.

Following the attention literature (e.g., Yantis and Jonides, 1990; Yantis, 2000), this work suggests that there are two ‘routes’ to boosting performance on sequential visual WM tasks (Allen et al., 2014; Hu et al., 2014), and more broadly in determining what enters and remains active and accessible in the focus of attention (Cowan, 2005, 2016b). One involves ‘top-down’ goal-directed executive control and applies most strongly to earlier items in a sequence, while the other involves perceptually-driven heightened activation of the most recently encoded item. Once an item is within the focus of attention, or held in the episodic buffer (Baddeley, 2000, 2012), executive resources are needed to continue to maintain it in this accessible and privileged state, in the face of retroactive interference from further incoming stimuli and processing (Allen et al., 2014; Hu et al., 2014).

Within the embedded processing account Cowan identifies two routes into the focus of attention, one ‘automatic’ and one ‘deliberate’. The automatic route is driven by attention being drawn to changes in the environment whilst the deliberate route is executively controlled (Cowan, 2016b). This distinction within the embedded processes account is clearly reflected in Allen et al. (2014)‘s proposal of two ’routes’ into visual WM. Both accounts also make similar claims about the relationship between attentional control and automatic processing. Allen et al. (2014) suggest items pass through a ‘perceptual filter’ that executive attention is able to adjust. Support for this claim comes from the differential effect of plausible and implausible suffixes in visual WM. A suffix is plausible if it resembles the items in an experimental set. For example, a coloured shape following coloured shapes stimuli. An implausible suffix might be an irregular shape in a colour distinct from the experimental set (Hu et al., 2014). The reduced interference for implausible suffixes suggests that participants change how they filter incoming items based on the context of the experimental set. Equally, Cowan identifies an important role for the deliberative route to the focus of attention in overriding automatic processing of changes in the environment. Hu et al. (2016) suggest a role for executive control in setting the parameters of a perceptual filter through which items must pass to enter WM. Once an item has passed through this 'gateway' further executive resources are needed to maintain it in WM.

Yantis and Jonides (1990) identify two criteria for automaticity common in the attention literature: the *load-insensitivity criterion* and the *intentionality criterion*. Load insensitivity means that the drawing of attention to some stimulus is unaffected by the presence of cognitive load. The intentionality criterion is met if an individual is unable to voluntarily prevent attention being drawn to a stimulus. An alternative framing of this distinction, attributed to Kahneman and Treisman (1984), describes a process as strongly automatic when it is neither facilitated by focusing attention on it, nor inhibited by focusing attention elsewhere (Yantis and Jonides, 1990). In Kahneman and Triesman's terminology a process is partially automatic if these criteria are sometimes not met. The 'automatic' recency effect in visual working would thus be a case of partial automaticity; performance at the final serial position can be boosted by focusing attention on it through prioritisation.

The ability to use goal or task information to bias processing is identified with the central executive component of working memory (Allen et al., 2014). This component captures a range of abilities that are required to perform complex cognitive tasks, such as inhibition, set-shifting or maintaining goal information (Baddeley, 2012; Logie, 2016). Within the executive functions literature, biasing the processing of incoming items has been included under the common executive function (EF) factor by Miyake and Friedman (2012), a broad ability that supports other executive functions and is useful in predicting "clinically and societally important behaviours" (p. 4).

Whereas there is a growing body of work investigating prioritisation effects in adults, children's ability to prioritise within serial memory has not been explored to date. While previous developmental work has guided attention to a target using external cues or regularities, our prioritisation instruction directly encourages participants to increase the attention allocated to one item while also processing the other items in the set. This requires explicit engagement of 'top-down' attentional resources without the possibility that participants might simply be responding to regularities in a task in a way that does not involve executive resources. Furthermore, previous studies exploring visual WM in children have typically used arrays of multiple objects encountered simultaneously in a single display. The present study differs by using sequential presentation, allowing for informative analyses of performance by serial position. Assessing performance for prioritised and non-prioritised items encountered in sequence potentially provides a means of examining how visual WM changes over time, and of more clearly differentiating between items that vary in their reliance on different forms of attentional control. It therefore also allows an exploration of whether children show potentially automatic recency benefits for the final item in a sequence.

There has been limited research on children's memory for visual information across serial positions to date. Memory for an entire list of sequentially presented visual items has been explored in children using nameable line drawings (Hitch et al., 1988) and spatial sequences (Pickering et al., 1998). In both cases primacy effects were observed alongside less pronounced recency effects, though the primacy effects were not always observed for 5-year-olds (Hitch et al., 1988). Others have observed recency effects with children, in the absence of primacy effects, using coloured shape stimuli and single item probed recall (Walker et al., 1994), or orientation judgements (Burnett Heyes et al., 2012). Burnett Heyes et al. (2012) presented three coloured bars sequentially, and participants (aged 9 to 13) were required to recall their orientation. Older children performed better at each serial position and overall precision for the final serial position was better than the 1st and 2nd positions, which did not differ. While performance at the final position can be thought of as partially automatic it, nevertheless, appears to improve with age. However, these previous studies did not examine children's ability to prioritise specific items in a sequence, or how performance across the sequence varies with working memory ability.

Here the ability to prioritise in sequential visual WM was investigated with children aged 7 to 10 years. These age ranges were selected to reflect previous investigations of visual WM in children (e.g., Cowan et al., 2010; Shimi et al., 2014; Walker et al., 1994) with similar tasks. These are also ages at which significant changes are occurring in the development of executive functions (EFs). After age 6, performance on measures of executive functions takes a two-factor structure where inhibition and shifting can be distinguished (e.g. Lee et al., 2013). The emergence of more specialised inhibitory abilities, reflected in this change in structure, will likely be relevant to visual WM performance. However, it is also critical to highlight that the structure of EFs continues to change well into adolescence, presumably as further specialised abilities emerge (Lee et al., 2013). Performance on measures of EFs also increase well beyond the age ranges tested here, in line with the prolonged development of the frontal lobes (Jurado and Rosselli, 2007; Zelazo et al., 2013). One executive ability crucial to performance on our task involves representing an attentional set in WM that is used to bias lower-level processing (Yantis, 2000). Yantis (2000) suggests pre-frontal areas as the obvious candidate for where such task-goals are represented (see also, Myers et al., 2017b). Seven-years-old also reflects an age at which children often begin to spontaneously engage proactive control strategies, such as those needed to complete this task (Braver, 2012; Chevalier et al., 2014; Chevalier, 2015b; Chevalier et al., 2015). More specifically in the WM literature, 7 years old reflects an age where rehearsal processes become more efficient (Tam et al., 2010) and children begin to engage attentional refreshing (Camos and Barrouillet, 2011).

In the present study, two 'routes' to supporting WM performance in children are investigated within a controlled experimental setting. Firstly, we investigated whether children demonstrate the ability to prioritise a particular position in a visual sequence. This ability in adults offers a way to focus on the most important information in a given situation, maintaining it in a more accessible state. Children aged 7 to 10 were presented with 3-item sequences of coloured shapes before being probed to recall the colour of one of the items. All participants completed a baseline condition where they were instructed to try equally hard to remember each item, and a prioritisation condition in which they were either instructed to try especially hard to remember the first (Experiments 1 & 2) or third item (Experiment 3). Research with young adults (Hu et al., 2014, 2016) found that their ability to prioritise items in visual WM required executive resources. Given that executive resources develop over childhood and into adolescence, we might expect to see a reduced ability to prioritise in children.

A second route to supporting WM stems from the large recency effects typically observed for the final item in a visual sequence (Walker et al., 1994). By exploiting recency effects presentation order can be used to ensure some information is more accessible than others. An outcome from the work with adults (e.g. Allen et al., 2014) is the suggestion that performance boosts at the final serial position are relatively automatic. Items at earlier positions, on the other hand, require resources to be maintained in the face of interference or decay. As well as expecting to observe large recency effects, a novel investigation into the automaticity of these effects is presented. The relationship between four WM measures described in Chapter 2 and performance at each serial position is reported. If recency effects are indeed relatively automatic then individual differences in WM should relate to performance at the first two serial positions, but not the third. This prediction flows from the observation in adults that performance at early serial positions, but not the last, is vulnerable to cognitive load (Allen et al., 2014; Hu et al., 2016). This suggests that performance for the final item is boosted ‘for free’ without drawing on executive resources. In contrast, performance for early items in a sequence draws on executive resources such that it would be expected to relate to individual differences in WM.

Given that measures of simple and complex working memory are highly related in children (Alloway et al., 2006; Gathercole and Alloway, 2004), we would expect both our simple and complex measures to relate to performance at early serial positions. However, complex measures involve a processing component, in addition to storage, that draws on domain-general executive resources. Thus, if a task is executively demanding, complex measures – in combining storage and processing – would be expected to relate more strongly to it than simple storage measures. The domain-general component of complex measures also means that they tend to relate to tasks in other sensory modalities more strongly than simple measures. In addition, given the visual nature of our primary task, we would expect measures that rely on visuospatial storage to relate more strongly to performance than those that rely on phonological storage, assuming storage in working memory is served by modality specific sub-components (Baddeley, 2007).

3.2 General Methods

3.2.1 Participants

Participants for all three experiments were 7 to 10 years old and recruited from primary schools in Bradford, UK in a predominantly Pakistani British low-SES neighbourhood. Participants were drawn from three consecutive Year Groups, Years 3, 4 & 5, which correspond to ages 7 to 8 (hereafter, “8-year-olds”), 8 to 9 (hereafter, “9-year-olds”), and 9 to 10 (hereafter, “10-year-olds”), respectively. A different group of children participated in each experiment. Consent was obtained from the school in addition to verbal assent from individual participants. The study was approved by the School of Psychology Ethics Committee, University of Leeds. Participants were excluded if they had Special Educational Needs (SEN), were distracted on the primary task, had missing data, or performed below chance on the primary task.

3.2.2 Materials & procedure

All tasks were created using PsychoPy 1.83.01 (Peirce, 2007). They were presented on a laptop/tablet computer with a screen 256mm x 144mm. The visual working memory task was presented with the tablet upright plugged into a keyboard. For the other measures the tablet was detached from the keyboard and placed flat on a table. Participants completed two sessions. In the first session they completed one condition of the visual WM task along with forward digit recall (FDR), and backward digit recall (BDR). In the second session they completed the other condition of the primary task and the Corsi and odd-one-out tasks. Visual WM condition order was counterbalanced across participants.

Visual working memory task. All stimuli measured approximately 1.5 x 1.5 degrees of visual angle and were presented on a white background. Within each trial the 3 stimuli were presented sequentially at three (randomly selected) corners of an invisible square 4 degrees of visual angle wide around a central fixation. Stimuli were selected from a pool of 6 colours and 6 shapes. Two fixed sets of 30 trials were created with 3 stimuli presented in each trial. Use of these two sets in each condition was counterbalanced across participants. Within a set, each shape and colour was presented at each serial position 5 times and probed as the response 5 times. Each serial position was probed for response 10 times within each set. For each condition, participants completed 6 practice trials followed by 30 test trials. Within a trial, three shapes were individually presented before participants had to respond to the test probe by saying aloud what colour the shape was in the trial set (see Figure 3.1 for detailed timings). The experimenter recorded the participant's response by pressing a key on a second keyboard plugged into the laptop. There was a 1000ms inter-trial interval. In the prioritisation condition participants were instructed to try especially hard to remember the colour of the first item in the sequence (Experiments 1 & 2). They were told that either two (Experiment 1) or four (Experiment 2) (purely notional) points would be awarded for successfully recalling this item if it was probed, with one point being awarded for the other items. In the baseline condition participants were instructed to try equally hard to remember each shape in the sequence. In Experiment 3, rather than being asked to prioritise the first item, participants were asked to try especially hard to remember the final item (with 4 notional points attached to a correct answer for this item). Within each experiment, all participants completed both the baseline and the prioritisation conditions.

3.2.3 Additional measures

Four of the tasks described in Chapter 2 were used for this study: forward digit recall, backward digit recall, Corsi block recall, and odd-one-out. The tasks were selected such that we had both simple and complex verbal and visuospatial tasks. This combination of additional measures is typical in the literature (e.g. Gathercole and Alloway, 2004) allowing the role of verbal versus visuospatial memory, and simple versus complex tasks to be assessed. However, as noted in Chapter 2, these measures are highly related.

Forward digit recall (FDR; simple verbal WM). Participants completed 16 trials of serial digit recall at span lengths from 3 to 6 (see Chapter 2 for details).

Backward digit recall (BDR; complex verbal WM). Participants completed 16 trials of backward digit recall at span lengths from 2 to 5.

Corsi block recall (simple visuospatial WM). Sixteen trials of block recall

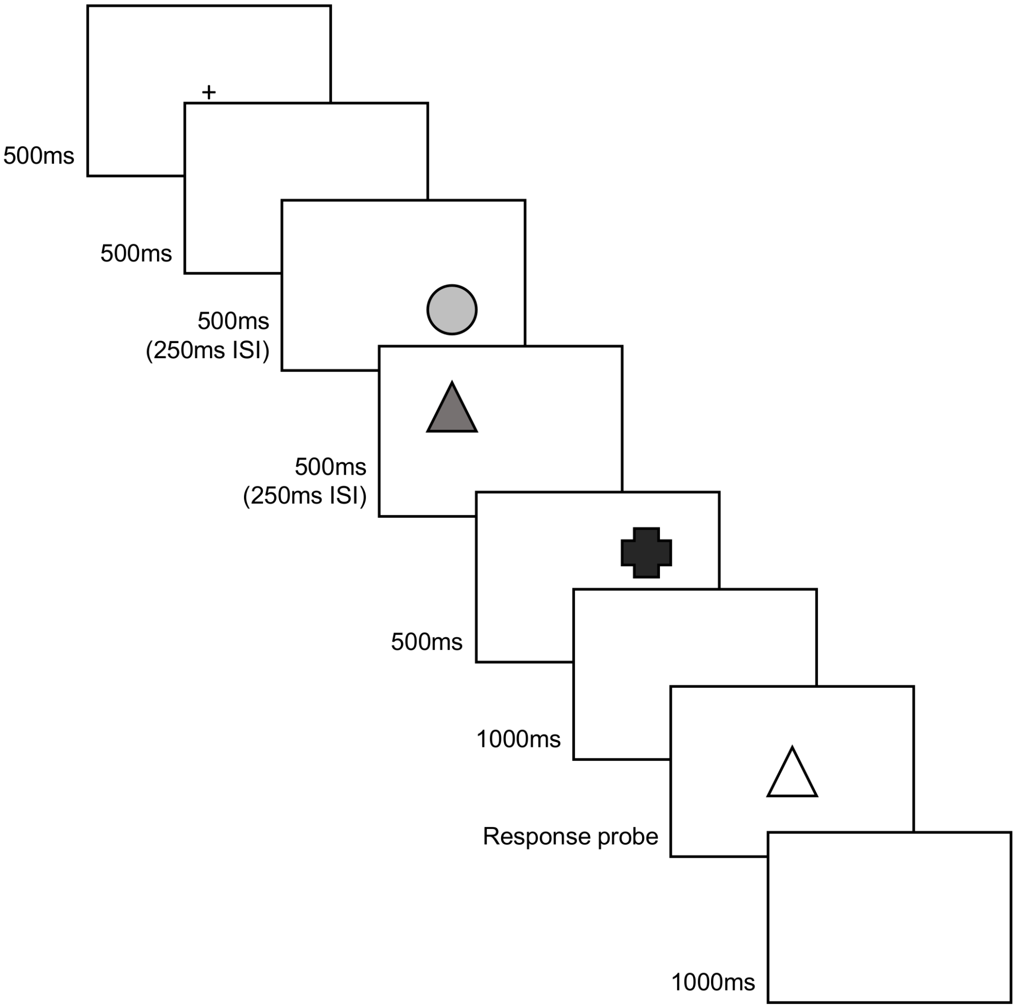


Figure 3.1: *Schematic for the visual working memory task. Figure is not to scale, and grey shades represent different colors.*

were completed by each participant with sequence lengths ranging from 3 to 6.

Odd-one-out (complex visuospatial WM). Finally, participants completed 16 trials of the odd-one-out task with sequence lengths ranging from 2 to 5.

3.3 Analysis plan

Outcome measures. For the primary task the outcome measure was proportion correct at the trial level (number of correct trials / total number of trials). For the four additional WM tasks the outcome measure was proportion correct at the item level across all 16 trials (number of items correctly recalled / total number of items presented).

Primary task analyses. For the primary task both frequentist and Bayesian analyses are reported. For all three experiments a mixed condition (prioritisation, baseline; within) x serial position (1, 2, 3; within) x age group (7, 8, 9; between) ANOVA was carried out. Where main effects were significant follow-up *t* tests are reported. These follow-up *t* tests were corrected for multiple comparisons using the Bonferroni-Holm method. Degrees of freedom and *p* values for the ANOVAs were Greenhouse-Geisser corrected for sphericity where applicable.

Bayesian ANOVAs with the same structure are also reported. These were carried out using the BayesFactor package in R (Morey et al., 2015; R Core Team, 2016). The default priors described in Rouder et al. (2012) were used. Bayesian analyses are reported to quantify evidence in favour of the null hypothesis, with a Bayes Factor describing the ratio between the likelihood of the data under the alternate model versus the null. More specifically this analysis gives the most likely set of effects given the data. Models with and without particular effects can also be compared to determine how much more (or less) likely a model with a certain effect is. ANOVAs, rather than logistic regression as in Chapter 2, were used to facilitate comparison with previous studies (e.g. Hu et al., 2014, 2016). Bayes

Factors allow us to quantify evidence in favour of the null hypothesis.

Individual differences analysis. Two sets of individual differences analyses were implemented. Firstly, regardless of whether we observe a prioritisation effect overall, it would be useful to investigate whether individual differences in WM relate to the effect of condition. It may be that some children, such as those with high WM, are able to prioritise serial positions. While this may not be reflected in the analysis of overall performance it may be revealed by predicting the prioritisation boost at the critical serial position with our additional WM measures. Secondly, the relationship between performance on the primary task and the additional individual difference measures was also explored, using performance in the baseline condition. This was achieved by looking at how individual differences relate to performance at each serial position, combining across the three experiments. The baseline condition was identical throughout, making this possible. Here we addressed how the relationship between individual differences and working memory might vary by serial position.

3.4 Experiment 1

For Experiment 1, participants were asked to either try to remember all items equally (baseline condition), or to try especially hard to remember the first item in the sequence (prioritisation condition). For the prioritisation condition, 2 notional points were awarded for correctly recalling the first item, if tested. One point was awarded for the 2nd and 3rd position as well as all three positions in the baseline condition, though these points were not linked to a reward.

3.4.1 Method

Participants. 87 (47 girls) participants took part in Experiment 1 (Mean age = 8.98, SD = 0.95, Range = 7.5 - 10.47). Of this dataset, 15 children were excluded due to having special educational needs, and an additional 3 children were excluded due to being distracted on the primary tasks. Finally, 1 child was excluded due to lacking data for both conditions of the primary task. The final sample used for primary task analysis had 68 participants (Mean age = 9.02, SD = 0.92, Range = 7.5 - 10.47). There were 21 eight year olds (Mean age = 7.94, SD = 0.27, Range = 7.5 - 8.43), 22 nine year olds (Mean age = 8.89, SD = 0.3, Range = 8.5 - 9.43), and 25 ten year olds (Mean age = 10.05, SD = 0.32, Range = 9.57 - 10.47).

Materials & Procedure. See General Methods (above) for a description of the materials, procedure and analysis plan.

3.4.2 Results

3.4.2.1 Primary task analysis

Proportion correct by condition and age group for the primary task is illustrated in Figure 3.2. A condition (prioritisation, baseline; within) x serial position (1, 2, 3; within) x age group (8, 9, 10; between) mixed ANOVA was carried out. There was no main effect of condition: $F(1, 65) = 0.65$, $p = .42$, $\eta_p^2 < .01$, $\eta_G^2 < .01$. The main effect of year was significant but small: $F(2, 65) = 3.39$, $p = .04$, $\eta_p^2 = .094$, $\eta_G^2 = .035$. Finally, there was a substantial main effect of serial position: $F(2, 130) = 46.56$, $p < .001$, $\eta_p^2 = .42$, $\eta_G^2 = .18$. None of the interactions were significant (all ps > 0.59).

Bayes Factor analysis revealed that the most likely model given the data had effects of age group and serial position (6.58 times more likely than a model with

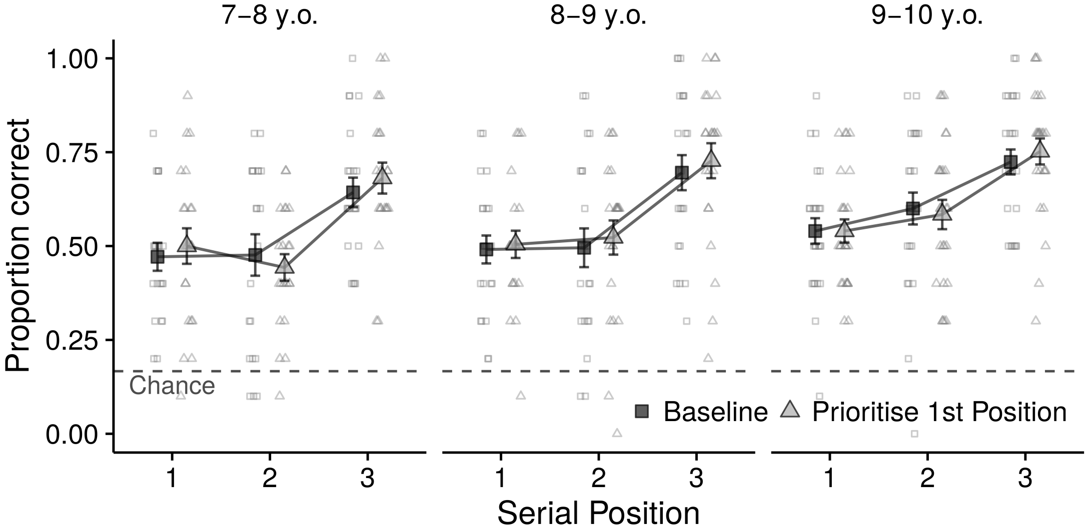


Figure 3.2: *Serial position curves for Experiment 1 by age group and condition for the primary task. Error bars show standard error. The horizontal dotted line shows chance performance. The unfilled grey shapes show the raw data (see Weissgerber et al., 2015).*

effects of age, serial position, and condition.) However, this model was only 1.03 times more likely than a model with effects of serial position only.

Planned pair wise comparisons revealed a non-significant difference between serial positions 1 and 2 ($t(67) = -0.77$, $p = .45$, BF = 0.18, $d = -0.09$). The Bayes Factor analysis showed that the null model was 5.6 times more likely than the alternate model. Positions 1 and 3 significantly differed with performance at position 3 being higher ($t(67) = -8.09$, $p < .001$, BF > 10000, $d = -0.98$). Equally, positions 2 and 3 differed significantly with higher performance at position 3 ($t(67) = -7.6$, $p < .001$, BF > 10000, $d = -0.92$).

8 year-olds and 9 year-olds did not differ significantly ($t(40.6) = -1.08$, $p = .29$, BF = 0.48, $d = -0.33$), nor did 9 year-olds and 10 year-olds ($t(44) = -1.45$, $p = .27$, BF = 0.68, $d = -0.42$). 8 year-olds and 10 year-olds, on the other hand, did differ significantly in performance with 10 year-olds performing better than 8-year-olds ($t(43.89) = -2.67$, $p = .037$, BF = 4.45, $d = -0.78$).

Table 3.1: *Standardised and unstandardised coefficients for predicting the difference in performance at position 1 between the prioritisation and baseline conditions.*

	B	SE	Beta	p-value
Intercept	0.043	0.28		0.88
FDR	0.16	0.23	0.11	0.49
BDR	-0.029	0.18	-0.03	0.87
Corsi	-0.0022	0.21	-0.0019	0.99
Odd-one-out	-0.066	0.17	-0.074	0.69
Age	-0.011	0.031	-0.051	0.73

3.4.2.2 Individual differences analysis

While no overall effect of prioritisation was observed there could be an effect for some participants that is related to individual differences in WM. This was tested with a linear regression predicting the difference in performance for the first item between the prioritisation and baseline conditions. One participant was excluded from this analysis due to having not completed the Corsi task. Table 3.1 shows the results of this analysis. Clearly none of the predictors were related to the boost at the prioritised position. The R^2_{adj} for this model was .01.

Bayes Factor analyses also supported the absence of a relationship between the additional WM measures and the prioritisation boost. The intercept-only model was 3.6 times more likely than any alternative model.

3.4.3 Discussion

The analyses of the primary visual WM task showed no effect of condition; telling children to try especially hard to remember the first item did not improve memory for that item. There was an effect of year-group driven by the difference between the youngest (7 to 8) and oldest (9 to 10) groups, however, the Bayes Factor analysis did not support including the effect of year. In addition, there were no interactions with age, such that older children were no more able to utilize the instructions than the youngest children.

The individual difference analyses further supported the suggestion that children are unable to prioritise items in visual WM. If the ability to prioritise is emerging at the ages we tested then we would expect those children with the most developed WM abilities to show a prioritisation effect. This would be revealed if our additional WM measures related to the difference in performance at position 1 between the prioritisation and baseline conditions. Instead none of the measures were related to this difference in performance. Thus, while some participant performed better at position 1 in the prioritisation condition this is likely due to random fluctuations in performance rather than meaningful developmental changes. As predicted, there was a large effect of serial position with recall from the final position relatively more accurate than from the first two positions. This recency effect is larger than those observed when an entire visual sequence is recalled (Hitch et al., 1988; Pickering et al., 1998), instead resembling the effects observed using a precision-based single item probe (Burnett Heyes et al., 2012).

These results provide the first suggestion that, unlike adults (Hu et al., 2014, 2016), 7 to 10 year-olds are unable to prioritise the first item in a sequential visual WM task. In contrast, like adults, they do clearly show improved recall of the final sequence item. These findings might indicate a developmental contrast between controlled, effortful processing on the one hand, and relatively effortless and automatic processing on the other. However, one possible alternative account of the outcomes from Experiment 1 is simply that children were not sufficiently motivated or that they forgot the prioritisation instructions. To address this concern and establish whether the Experiment 1 findings replicate, we increased the motivation to prioritise in Experiment 2 by adjusting the notional points rewarded for the prioritised item, and telling participants that if they got enough points they would be given a reward upon completion of the task. In addition, participants were shown an instruction screen every 10 trials reminding them which item in the sequence they should try especially hard to remember.

3.5 Experiment 2

Here we wanted to replicate the result of Experiment 1 while ensuring it was not the result of a simple lack of motivation or forgetting of instructions. Children aged 7 to 12-years-old have been shown to adapt their performance in response to points and small rewards (e.g., Chevalier, 2017).

3.5.1 Method

Participants. A new sample of 88 (44 girls) participants initially took part in Experiment 2 (Mean age = 9.19, SD = 0.85, Range = 7.68 - 10.62). Fifteen children were excluded due to having special educational needs, and 4 children due to lacking data for both conditions of the primary task. The final sample used for primary task analysis included 69 participants (Mean age = 9.19, SD = 0.8, Range = 7.68 - 10.62). In the final sample there were 22 eight year olds (Mean age = 8.3, SD = 0.37, Range = 7.68 - 9.55), 25 nine year olds (Mean age = 9.16, SD = 0.33, Range = 8.72 - 9.65), and 22 ten year olds (Mean age = 10.12, SD = 0.31, Range = 9.67 - 10.62).

Materials & Procedure. The materials and procedure were identical to Experiment 1 except that in the prioritisation condition participants were told that 4 points would be awarded for successfully recalling the first item, rather than the 2 points in Experiment 1. Participants were also told that they would be rewarded with a sticker if they got sufficient points (though in fact all participants were rewarded at the end of the study). The instructions for the baseline condition were identical to Experiment 1. In addition, to ensure that children remembered the priority instructions and remained motivated to follow them, a screen was displayed every 10 trials containing a reminder to try especially hard to remember the first shape (prioritisation condition) or to remember all three shapes equally (baseline condition).

3.5.2 Results

3.5.2.1 Primary task

Mean proportion correct by condition, serial position, and age group is shown in Figure 3.3 A condition (prioritisation, baseline; within) x serial position (1, 2, 3; within) x age group (8, 9, 10; between) mixed ANOVA was carried out. There was no main effect of condition: $F(1,\ 66) = 0.64$, $p = .43$, $\eta_p^2 < .01$, $\eta_G^2 < .01$), nor of year: $F(2,\ 66) = 1.06$, $p = .35$, $\eta_p^2 = .031$, $\eta_G^2 < .01$. There was a large and significant main effect of serial position: $F(1.73,\ 114.5) = 21.28$, $p < .001$, $\eta_p^2 = .24$, $\eta_G^2 = .13$.

Bayes Factor analysis revealed that the most likely model given the data had a main effect of serial position. This model was 7.7 times more likely than a model with effects of condition and serial position, and 9.36 times more likely than one with effects of year and serial position.

Planned paired wise comparisons revealed a non-significant difference between serial positions 1 and 2 ($t(68) = 0.97$, $p = .3$, BF = 0.21, $d = 0.12$). The Bayes Factor showed that the null model of no-difference was 4.8 times more likely than the alternative. Positions 1 and 3 significantly differed with performance at position 3 being higher ($t(68) = -4.44$, $p < .001$, BF = 567.6, $d = -0.53$). Equally, positions 2 and 3 differed significantly with higher performance at position 3 ($t(68) = -7.18$, $p < .001$, BF > 10000, $d = -0.86$).

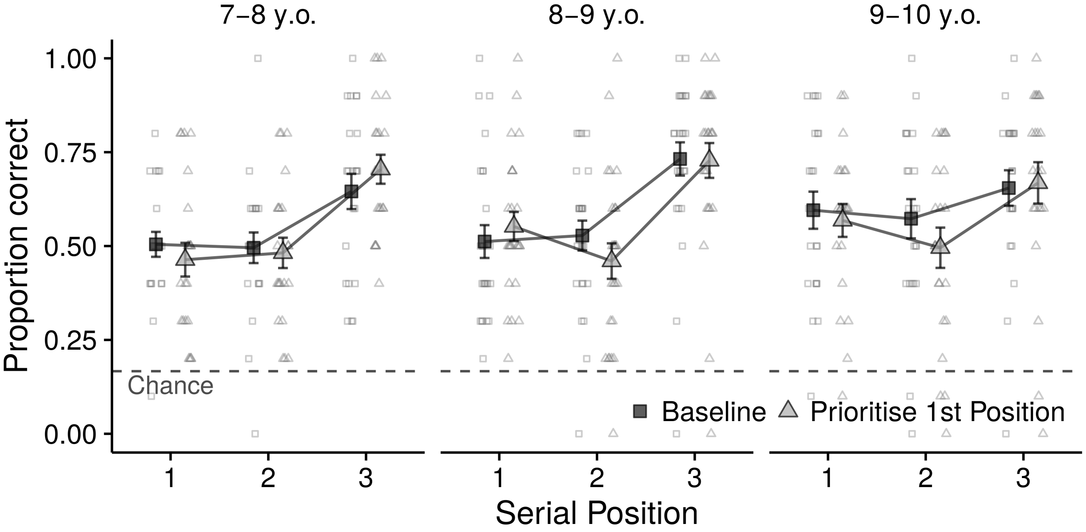


Figure 3.3: *Serial position curves for Experiment 2 by age group and condition for the primary task. Error bars show standard error. The horizontal dotted line shows chance performance. The unfilled grey shapes show the raw data (see Weissgerber et al., 2015)*

3.5.2.2 Individual differences analysis

As with Experiment 1 the relationship between individual differences in WM and the difference between the prioritisation and baseline condition at the first serial position was investigated. Table 3.2 shows the results of this analysis. Again none of the predictors were related to the boost at the prioritised position. The R^2_{adj} for this model was .01.

Bayes Factor analyses also supported the absence of a relationship between the additional WM measures and the prioritisation boost, though less clearly than Experiment 1. The intercept-only model was 2.2 times more likely than any alternative model. 2.2 is less than the conventional cut-off of 3 used to conclude evidence from a Bayes Factor (Jeffreys, 1961). Models with FDR, odd-one-out or BDR as single predictors were all between 2 and 3 times less likely than the intercept-only model. Any other combination of predictors were over 3 times less likely than the intercept-only model.

Table 3.2: *Standardised and unstandardised coefficients for predicting the difference in performance at position 1 between the prioritisation and baseline conditions.*

	B	SE	Beta	p-value
Intercept	0.27	0.37		0.47
FDR	-0.17	0.24	-0.11	0.49
BDR	-0.069	0.27	-0.041	0.8
Corsi	-0.27	0.34	-0.13	0.43
Odd-one-out	0.29	0.2	0.22	0.15
Age	-0.0085	0.04	-0.027	0.83

3.5.3 Discussion

The analyses of the primary task showed, as with Experiment 1, no effects of condition, alongside a large recency effect for the final item. Unlike Experiment 1, no effect of age group was found. This is unsurprising given the size of the effect and the lack of support from Bayes Factor analysis observed in Experiment 1.

Experiment 2 suggests that children's inability to prioritise the first item in a sequence is not the result of auxiliary factors such as motivation or instructions. In line with this, informal questioning of participants following the experiment indicated no difficulties with understanding the task or the priority instruction. Combining the data from Experiments 1 & 2, Bayes Factor analysis showed that the null hypothesis of no difference in performance at serial position 1 for the two conditions is 6.5 times more likely than the alternative. Thus, we do not simply observe an uninformative absence of a priority instruction effect. Rather, we have strong evidence for the lack of a difference. This runs counter to the consistently observed priority boost at the first sequence position in young adults on an essentially identical WM task (Hu et al., 2014, 2016), and suggests children are unable to engage in effortful, goal-directed attention to prioritise items.

Combining the data in the same way gave a Bayes Factor over 10000 in support of a difference between positions 1 and 3. This also applied when comparing positions 2 and 3 (BF > 10000). In contrast, the BF in support of no difference between positions 1 and 2 was 7.2. In addition to strong evidence for the absence of a prioritisation effect we have very strong evidence for the recency effect.

The results of the individual difference analysis were less clear than Experiment 1 providing inconclusive evidence as to whether the intercept-only model was supported more than any alternative. Nevertheless, the intercept-only model still remained the most likely model given the data. Thus, there is little reason to suggest that our additional WM measures related to the prioritisation boost. In the absence of additional evidence it is reasonable to assume that those children with superior WM abilities do not show a prioritisation effect.

So far we have only examined whether children show a priority boost at the first position in a short sequence of visual stimuli. Adults also show a prioritisation effect for the final item in a sequence, supplementing relatively automatic boosts for this recency item via controlled attention (Hu et al., 2014, 2016). It is possible that children cannot achieve a measurable boost in performance for the first item in a sequence in addition to processing subsequent items, as required by the primary task in Experiments 1 and 2. In contrast, processing the first two items in the sequence followed by prioritising the final item may represent a less demanding and complicated form of goal-directed working memory resource management. Children may therefore be able to engage in this more easily, with observable boosts to performance at the prioritised final position. Thus, in Experiment 3 we investigated this possibility by asking participants to prioritise the final item in a sequence.

3.6 Experiment 3

Adults can prioritise the final item in sequential visual WM tasks, with the resulting boost to performance further improving the already accurate recall of this recency position (Hu et al., 2016). Here we used a procedure identical to Experiment 2 to ask whether children demonstrate the ability to prioritise the final item in a sequence. This allows us to address whether the absence of an effect in Experiments 1 & 2 results from the relatively complex task of having to prioritise an item while processing subsequent items. As with Experiments 1 and 2, if we observed a prioritisation effect for the final serial position we would also expect a drop in performance at the non-prioritised positions, as is observed with adults (Hu et al., 2014). Alternately, if the outcomes of Experiments 1 & 2 represent a more general under-development of executive resources in children that undermines the ability to prioritise any item, then the absence of a prioritisation effect would be expected to remain in Experiment 3.

3.6.1 Method

Participants. 85 participants took part in Experiment 3 (Mean age = 9.11, SD = 0.9, Range = 7.68 - 10.64). Seven children were excluded due to having special educational needs. An additional 1 child was excluded due to being distracted on the primary tasks. Five children were excluded due to lacking data for both conditions of the primary task. Finally, one participant was excluded due to having an overall accuracy of less than chance (16%) on the primary visual WM task. The final sample used for primary task analysis had 71 participants (Mean age = 9.15, SD = 0.87, Range = 7.68 - 10.64). There were 23 eight year olds (Mean age = 8.11, SD = 0.29, Range = 7.68 - 8.64), 25 nine year olds (Mean age = 9.19, SD = 0.3, Range = 8.67 - 9.6), and 23 ten year olds (Mean age = 10.13, SD = 0.36, Range = 9.67 - 10.64).

Materials & Procedure. The materials and procedure were identical to Experiment 2 except that in the prioritisation condition participants were instructed to try especially hard to remember the final item in the sequence rather than the first.

3.6.2 Results

3.6.2.1 Primary task

A condition (prioritisation, baseline; within) x serial position (1, 2, 3; within) x age group (8, 9, 10; between) mixed ANOVA was carried out. Unlike Experiments 1 and 2 there was a main effect of condition: $F(1, 68) = 6.94$, $p = .01$, $\eta_p^2 = .093$, $\eta_G^2 < .01$. No main effect of age group was observed: $F(2, 68) = 3.03$, $p = .055$, $\eta_p^2 = .082$, $\eta_G^2 = .032$. As with Experiments 1 and 2 there was a large main effect of serial position: $F(2, 136) = 58.37$, $p < .001$, $\eta_p^2 = .46$, $\eta_G^2 = .22$. No interactions were significant (all ps > 0.09).

Bayes Factor analysis revealed that the most likely model given the data had effects of condition and serial position (see Figure 3.4). However, this model was only 1.03 times more likely than a model with an effect of serial position only.

Planned pairwise comparisons revealed a non-significant difference between serial positions 1 and 2 ($t(70) = -1.46$, $p = .12$, BF = 0.4, $d = -0.17$). Positions 1 and 3 significantly differed with performance at position 3 being higher ($t(70) = -8.99$, $p < .001$, BF > 10000, $d = -1.07$). Equally, positions 2 and 3 differed significantly with higher performance at position 3 ($t(70) = -8.78$, $p < .001$, BF > 10000, $d = -1.04$).

Finally we looked at the difference between the prioritisation and baseline conditions at each serial position. There were not significant differences at serial positions 1 ($t(70) = -2.02$, $p = .095$, BF = 0.9, $d = -0.24$) or 3 ($t(70) = 0.47$, $p = .64$, BF = 0.15, $d = 0.06$). Performance in the prioritisation condition was significantly lower than the baseline condition at position 2 ($t(70) = -2.53$, $p = .041$, BF = 2.52, $d = -0.3$), Though the difference was significant for position 2 the alternate model is only 2.5 times more likely than the null model. For position 3 the null model is 6.7 times more likely that the alternative.

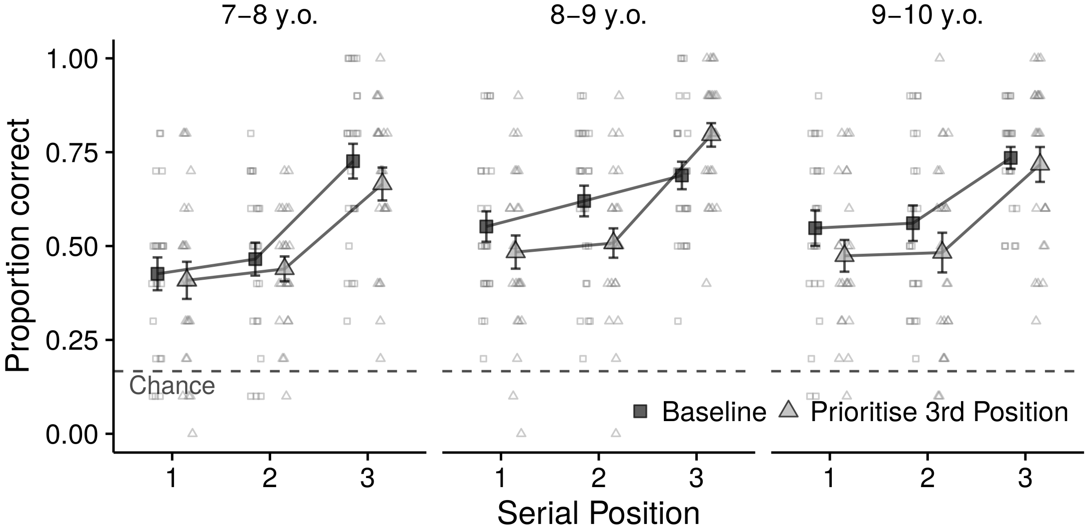


Figure 3.4: *Serial position curves for Experiment 3 by age group and condition for the primary task. Error bars show standard error. The horizontal dotted line shows chance performance. The unfilled grey shapes show the raw data (see Weissgerber et al., 2015)*

3.6.2.2 Individual differences analysis

Table 3.3 shows the results of an analysis of whether individual differences in WM were related to the difference between the prioritisation and baseline condition at the final serial position. Again, none of the measures significantly predicted the prioritisation boost. The R^2_{adj} for this model was .02.

Table 3.3: *Standardised and unstandardised coefficients for predicting the difference in performance at position 3 between the prioritisation and baseline conditions.*

	B	SE	Beta	p-value
Intercept	-0.23	0.3		0.45
FDR	0.28	0.21	0.2	0.19
BDR	0.29	0.21	0.24	0.17
Corsi	-0.21	0.25	-0.15	0.41
Odd-one-out	-0.079	0.16	-0.074	0.63
Age	0.0048	0.032	0.019	0.88

Bayes factor analyses showed that the most likely model given the data included FDR as a predictor. However, this model was only 1.5 times more likely than the intercept-only model. This provides inconclusive evidence either way as to whether WM related to the difference between prioritisation and baseline performance.

3.6.3 Discussion

In line with the outcomes from Experiments 1 and 2, we observed large effects of serial position with a substantial advantage for the final item over positions 1 and 2, across both instruction conditions. This adds further weight to the conclusion that children demonstrate relatively automatic recency benefits in visual working memory. The question posed by this experiment was whether children are then able to effectively supplement this improved performance by deliberately prioritising the final item, as observed in adults (Hu et al., 2014, 2016). Unlike the primacy-focused instruction used in Experiments 1 and 2, here we found a small effect of condition when instructing children to prioritise the final item. However, this reflected lower performance for the non-prioritised items in the prioritisation condition compared with the baseline, with no measurable concomitant increase in accuracy for the final item. Follow-up comparisons showed that the effect was driven by performance at the second position. Finally, as with Experiment 2 there was no effect of age group.

Given the effect size, the significant effect of condition should be treated with caution. Nevertheless, if genuine it could reflect an unsuccessful attempt by children to prioritise the final position, resulting in a drop in performance at the second serial position. In adults, increases in accuracy for the prioritised item are accompanied by declines in performance at non-prioritised positions, relative to a baseline condition (Hu et al., 2014, 2016). In the present study, we did not see a boost in performance for the final item, despite the possible drop in performance at position 2. The difference in performance between the two conditions speaks against the idea that children simply do not understand the prioritisation instructions and thus perform equivalently to baseline in that condition. Nevertheless, they remain unable to achieve a boost in performance at the prioritised position.

For Experiment 3 the individual difference analyses provided inconclusive evidence as to whether our additional WM measures related to the difference in performance between the two conditions. While none of the predictors showed a significant relationship, the Bayes Factor analysis had a model including FDR as a predictor as the most likely given the data. That said, this model was less than two times more likely than the intercept-only model. Thus, the most modest interpretation of this analysis across the three experiments would be that WM ability is not related to the prioritisation effect.

3.7 Summary of the prioritisation effects

Figure 3.5 provides a summary of the (absent) prioritisation effects across Experiments 1-3. The difference between performance in the prioritisation and baseline condition for each experiment, at the prioritised serial position (position one, Exp 1 & 2; or position three, Exp 3), is shown on average and for each participant.

3.8 Cross-experiment individual differences analysis

Increased accuracy for the final item in a sequence, as observed across Experiments 1-3 in the present study, may indicate it is stored in a different state to earlier items (Allen et al., 2014). Whereas earlier items require executive resources to be

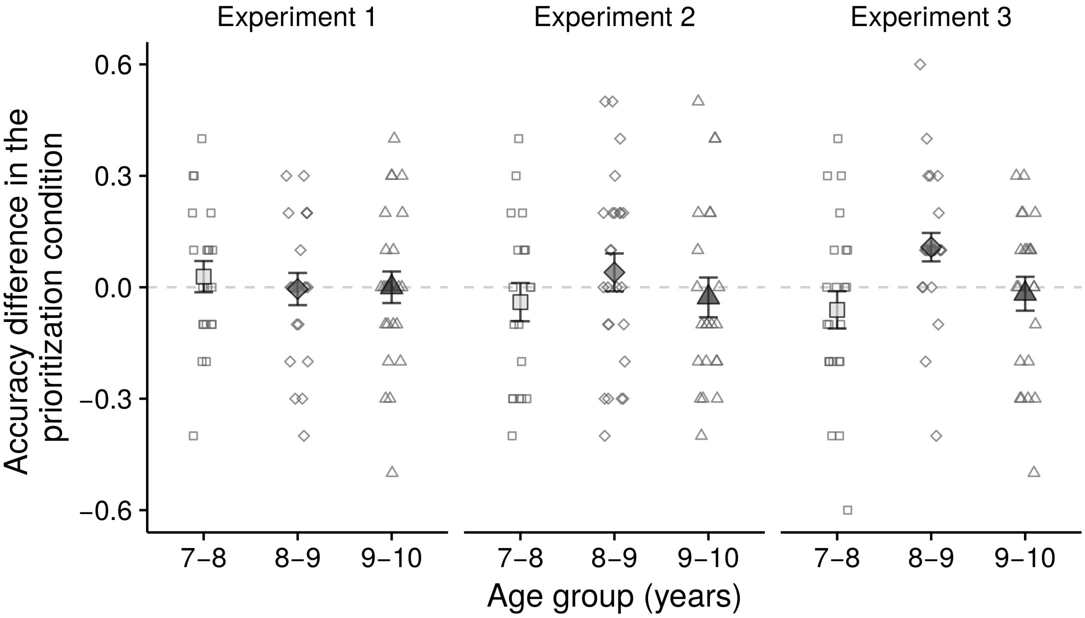


Figure 3.5: *The difference between performance in the prioritisation condition and the baseline condition at the prioritised serial position for each experiment. Unfilled grey shapes show the raw data, and error bars show standard error.*

actively maintained, the most recent item may be automatically maintained in a privileged state. This claim would lead to two predictions about the relationship between individual differences in WM and performance in the baseline condition: (i) those children with better working memory will also perform better at serial positions 1 and 2; (ii) working memory will not predict performance at the final serial position.

3.8.1 Method and Results

The data from the baseline condition of the primary task was combined for each serial position as the task and instructions for this condition were identical across the three experiments (N = 210). Tables 3.4, 3.5 and 3.6 show the coefficients for predicting performance at the first, second, and third serial positions, respectively.

For serial position 1 (Table 3.4), Corsi and odd-one-out significantly predicted

Table 3.4: *Standardised and unstandardised coefficients for predicting performance at the first serial position in the baseline condition.*

	B	SE	Beta	p-value
Intercept	-0.16	0.13		0.25
FDR	0.18	0.096	0.14	0.057
BDR	0.072	0.091	0.065	0.43
Corsi	0.29	0.11	0.22	0.01
Odd-one-out	0.17	0.077	0.18	0.026
Age	0.022	0.015	0.095	0.14

Table 3.5: *Standardised and unstandardised coefficients for predicting performance at the second serial position in the baseline condition.*

	B	SE	Beta	p-value
Intercept	-0.13	0.14		0.35
FDR	0.24	0.1	0.16	0.022
BDR	0.19	0.098	0.16	0.049
Corsi	0.33	0.12	0.22	0.007
Odd-one-out	0.17	0.082	0.16	0.038
Age	0.0061	0.016	0.024	0.7

performance whereas FDR, BDR, and age did not. The $R^2{}_{\text{adj}}$ for this model was 0.25. Bayes Factor analysis revealed that the most likely model given the data is the one that included FDR, Corsi and odd-one-out as predictors. It was 1.6 times more likely than a model that also included age as a predictor, and over 10000 times more likely than the intercept only model. The inclusion of FDR in the most likely model reflects the fact the p value was .057.

For serial position 2 (Table 3.5), FDR, BDR, Corsi and odd-one-out significantly predicted performance, with $R^2{}_{\text{adj}} = 0.3$. The most likely model given the data includes FDR, BDR, Corsi, and odd-one-out as predictors. This model was 10000 times more likely than the intercept only model but only 1.3 times more likely than a model without BDR as a predictor.

For the final serial position (Table 3.6), none of the individual difference measures significantly predicted performance ($R^2{}_{\text{adj}} < .01$). Analysis using Bayes Factors

Table 3.6: *Standardised and unstandardised coefficients for predicting performance at the third serial position in the baseline condition.*

	B	SE	Beta	p-value
Intercept	0.48	0.15		0.0014
FDR	0.12	0.11	0.092	0.28
BDR	-0.015	0.1	-0.014	0.88
Corsi	0.082	0.12	0.065	0.51
Odd-one-out	0.0065	0.084	0.0069	0.94
Age	0.0093	0.016	0.042	0.57

showed that the intercept-only model was 1.58 times more likely than any alternative.

For serial positions 1 and 2 the Bayes Factor analysis consistently supported the inclusion of Corsi, odd-one-out and FDR as predictors of performance, though the relationships were stronger for the visuospatial measures. The likelihood of the data under a model assuming a relationship between these predictors and performance at the final serial position was calculated. The intercept-only model, where none of the predictors are assumed to relate to performance at the final position, was 17 times more likely than the model that included Corsi, odd-one-out and FDR as predictors.

To further investigate the automaticity of a recency boost a residual score was created. A linear regression was carried out with performance at positions one and two as predictors and performance at the final position as the outcome. The residuals from this regression were then taken as an index of the recency boost ‘controlling’ for baseline performance. This score represents the difference between actual performance at the final position and the performance at the final position predicted from performance at the first two positions. This score was created using data from the baseline condition only, combining across the three experiments. It is important to control for performance at the first two positions because accuracy scores are bounded between 0 and 1. If a simple difference score was used then a negative relationship between WM and the recency boost might be observed

Table 3.7: *Standardised and unstandardised coefficients for predicting the recency residual score.*

	B	SE	Beta	p-value
Intercept	-0.13	0.15		0.39
FDR	0.083	0.11	0.067	0.43
BDR	-0.042	0.1	-0.04	0.68
Corsi	0.037	0.12	0.03	0.76
Odd-one-out	-0.017	0.084	-0.018	0.84
Age	0.0085	0.016	0.039	0.6

that simply reflects the fact that low WM participants do worse at positions 1 and 2 leaving more ‘room’ for a recency boost. High performing participants may already be performing at, say, over 75% for the first two positions meaning the potential recency boost is limited before they hit ceiling. As Table 3.7 shows none of the additional WM measures were significantly related to this residual score. The $R^2{}_{\text{adj}}$ for this model was $< .01$.

The Bayesian analysis also supported an absence of a relationship between the residual score and WM. The intercept-only model was 4.76 times more likely than the best alternative model, which included FDR as a predictor. Figure 3.6 shows the relationship between each WM measure and the residual score.

3.8.2 Discussion

Both predictions with respect to individual differences in WM and performance by serial position were supported, with WM measures predicting visual WM recall at serial positions 1 and 2, but not position 3. Importantly, the Bayes Factor analysis provided evidence against those WM abilities that relate to performance at the first and second positions relating to performance at the final serial position. This suggests that the recency effect for the final item is indeed automatic and does not draw on executive resources. Not only is this recency effect present in children but also, like in adults, it does not relate to executive attention. This

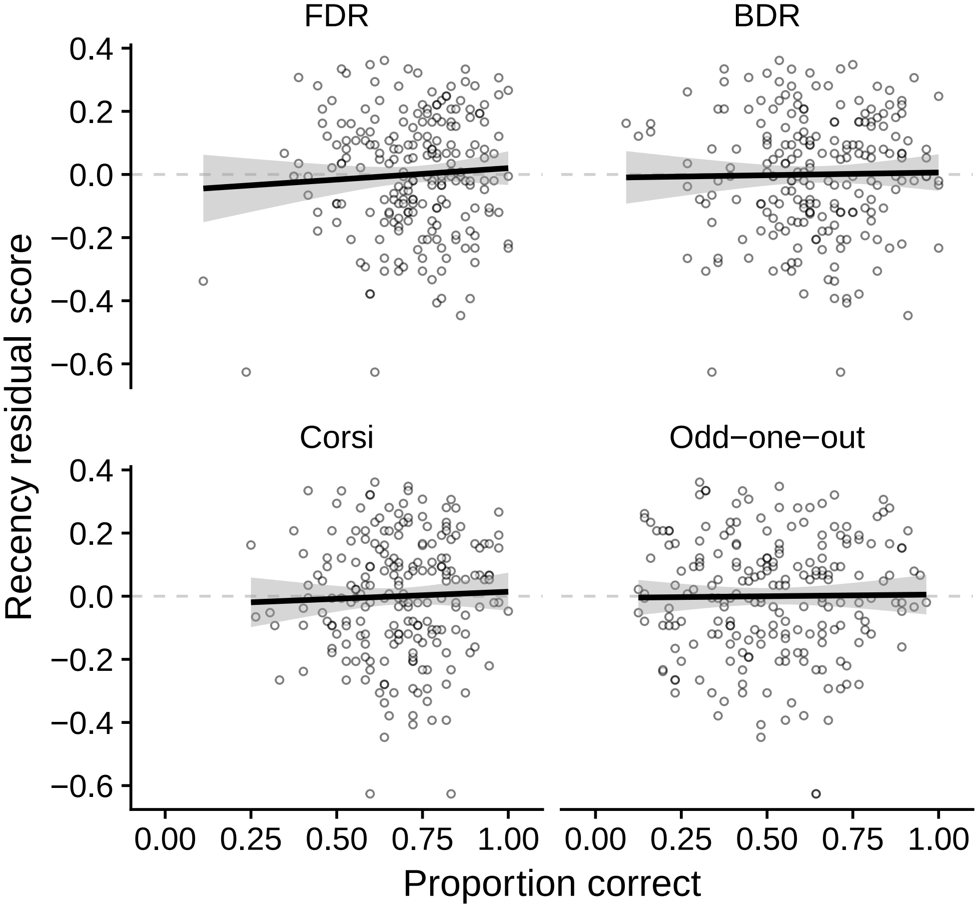


Figure 3.6: *Working memory predicting the residual performance at the final position that is not explained by performance at the first and second serial positions. Plots combine data from the baseline condition of all three experiments.*

finding therefore extends evidence from dual-task methods in adults (Allen et al., 2014; Hu et al., 2016) to an individual differences approach with children. The results presented here also bolster the idea that presentation order can be used to effect changes in the availability of information for all participants, regardless of their individual WM abilities.

3.9 General Discussion

Across three experiments, we found strong evidence that children more effectively recall items when they were encountered in the final sequence position, relative to earlier positions. In contrast, we also found strong evidence that children do not prioritise particular serial positions in visual WM when instructed to do so. In Experiments 1 and 2 children did not prioritise the first serial position; the combined data provided strong support for the absence of a priority effect for the first item. In Experiment 3 we observed a small effect overall, with children being unable to prioritise the final serial position, but showing a small drop in performance at the second position. This suggests an attempt to prioritise the final item leading to a reduction in accuracy at earlier positions, but with no concurrent boost for the final position. The claim that 7 to 10 year olds are unable to prioritise serial positions in visual WM was further supported by the individual difference analyses reported for each experiment. It was not the case that children who performed better on WM measures were able to prioritise, as would be expected if this ability was emerging for the age ranges tested.

Thus, it appears that children aged 7 to 10 years old are unable to selectively attend to a serial position. This runs counter to consistent observations of this ability in young adults using near-identical procedures (Hu et al., 2014, 2016). One possible explanation for the contrasting findings between these age groups might lie in the observation that young adults show a substantially diminished or abolished priority boost when the availability of executive control resources is reduced by concurrent performance of a more attention-demanding verbal task (Hu et al., 2016). This would suggest that the strategic prioritisation in visual WM of items within a sequence draws heavily on executive resources, which are known to develop through childhood and not reach adult-like levels until late adolescence (Jurado and Rosselli, 2007; Waszak et al., 2010). Thus, children may lack the necessary executive control abilities to be able to effectively prioritise items within a sequence; carrying out this task might be analogous to adult performance under high cognitive load. Under this account the basic task is ‘too difficult’ for children to have any resource available for prioritisation. In line with this, the effect of condition at serial position two in Experiment 3 could suggest that children are trying to prioritise the final item by diverting executive resources away from the first two items, while being unable to boost performance for this final item. Alternatively, it could be that the task is in fact ‘too easy’: children do not choose to prioritise because they incorrectly think that they can remember all 3 items (Cowan, personal communication). This possibility could be tested by using four items instead of three.

In contrast to the absence of a priority boost at either the first or final serial position, the substantial recency advantage for the final item observed in all three experiments is in line with similar effects observed in adult studies (e.g. Allen et al., 2006, 2014). This extends limited previous work with children using different visual stimuli (Burnett Heyes et al., 2012; Hitch et al., 1988; Pickering et al., 1998) to the current task and stimulus sets. In addition, and in line with our initial predictions, recall accuracy for this final item was not related to individual differences in WM, while performance at serial positions 1 and 2 did relate to these measures. These outcomes were supported by Bayesian analysis, with strong support for a model containing WM ability for positions 1 and 2, contrasting with strong support for the intercept-only model for the third serial position. The observation that our visuospatial tasks (Corsi and odd-one-out) more consistently relate to performance at the first and second serial positions is in line with the primarily visual nature of the experimental task, though the additional relationship with verbal tasks (particularly FDR) might indicate a degree of verbal recoding.

To our knowledge, this is the first demonstration of the shifting relationship between memory across different positions in a sequence and broader WM ability. This finding supports the suggestion that incoming items are automatically processed but require additional resources to be maintained once they have been displaced from an active and accessible 'privileged state' by subsequent items, with the most recently encountered item automatically remaining in this state at least for a brief period (Allen et al., 2014; Hu et al., 2014, 2016). In terms of supporting WM performance in children, these results suggest that engaging executive control is not a feasible strategy, at least for children aged 7 to 10 years. In contrast the heightened availability of the most recently presented item can be leveraged for all participants, whatever their WM ability. This last point is crucial when considering how children with poor WM might be supported. Such children would still benefit from the most important information in a sequence being presented last. However, this study involved probed recall where participants were not required to recall the whole sequence. Serial position curves differ, though recency effects remain, for whole-sequence visual recall (Pickering et al., 1998).

Why do no priority effects emerge in the present study, when previous work has identified evidence that children are able to direct attention in certain contexts? To understand the limitations of children's executive abilities it is useful to consider the differences between this paradigm and others that have been used. With simultaneous presentation, children as young as 7 can allocate attention in response to visual cues (Shimi et al., 2014) and probe frequency in visual WM (Cowan et al., 2010). However, it is possible that children are unable to actively prioritise visual items within a temporal sequence. One reason for this could be the inability to resist interference from subsequent items. Perhaps children are trying to prioritise the first item but find it difficult to resist interference caused by encoding and maintaining the following items in the sequence. This interference may be particularly pronounced in our task where each item is equally likely to be tested and must, therefore, be maintained in memory until the recall probe is presented. In adults prioritised items are particularly vulnerable to suffix interference from a to-be-ignored item presented after the test sequence, suggesting a cost to the heightened accessibility that results from prioritising an item (Hu et al., 2014, 2016).

Sequential presentation, in general, also infers costs compared to simultaneous presentation (Gorgoraptis et al., 2011), particularly at earlier positions in a sequence. This appears to be driven by the fact that sequentially presented items are more vulnerable to interference from subsequent items (Allen et al., 2006; Gorgoraptis et al., 2011). Support for an interference account of our findings comes from Shimi et al. (2014)'s suggestion that resisting interference drives the ability to use a retro-cue to selectively attend to an item in an array; interference from the other items in an array must be managed to direct attention onto the cued item. However, such an explanation struggles with the absence of a prioritisation effect for the final item in Experiment 3, as there are no subsequent to-be-remembered items that might cause retroactive interference. A response could be that the prioritisation boost for the final item is too small in children to be detected by this paradigm, over and above the substantial recency effects that are observed. Indeed, it is noteworthy that the prioritisation effect for the final item is slightly smaller in adults than for the first item in a sequence (Hu et al., 2014, 2016). This could be tested in future work using a paradigm that allowed for continuous responses (Burnett Heyes et al., 2012; Sarigiannidis et al., 2016); Gorgoraptis et al. (2011) observed improvements in the visual memory precision of adult participants for more frequently cued items at all positions in a sequence, including the final one.

Another possibility for future work to examine concerns the validity of priority instructions for the purposes of optimising task performance. Shimi et al. (2014) found that the benefit for retro-cues was removed when cue validity was reduced from 100 to 50%. This suggests that children are able to selectively ignore a cue when it becomes less informative. In the present study each serial position was probed an equal number (33%) of times. The instruction to prioritise a serial position may be insufficiently predictive of a probe for children to follow. However, this does not explain why the prioritisation effect was not observed in children when it is consistently seen in adults (Hu et al., 2014, 2016). Shimi et al. (2014) found that the cuing benefit for both children and adults was removed by reducing cue validity. Thus, any account of our finding that turns on the question of probe frequency would have to explain why this absence of an effect is limited to children with sequential but not simultaneous presentation. Cowan et al. (2010) also showed that young children are able to adjust the amount of attention they allocate to an item dependent on probe frequency. However, this ability broke down for the youngest group (7 to 8 year-olds) under high cognitive load. It could be that sequential presentation has a difficulty analogous to simultaneous presentation under high load, interfering with children's ability to selectively attend. Moreover, with prioritisation participants are asked to share their attention among a set of items while allocating more attention to one item. This sort of division involves a more complicated allocation of attention than simply focusing on a single item.

General developmental changes in availability of executive control resources provide one strong candidate for accounting for the absence of prioritisation effects in the present study. The developmental mechanisms underpinning children's performance on visual WM tasks could also be framed in terms of a distinction between reactive and proactive control strategies (Braver, 2012; Chevalier et al., 2014; Chevalier, 2015b; Chevalier et al., 2015; Morey et al., 2017). Proactive control involves planning prospectively for future responses. Reactive control, on the other hand, is simply a response to currently presented information. Under this approach, a developmental shift can be observed in which younger children primarily demonstrate reactive control strategies, responding to the stimulus at hand and only planning ahead when the nature of the task makes reactive control less useful. In contrast, older children and adults show proactive control, planning their recall of items to optimise performance (Chevalier et al., 2014; Chevalier, 2015a). Chevalier et al. (2014), for example, showed the shift from reactive to proactive control strategies over development with a working memory task. They compared the response times for recalling the first item in a sequence with additional items for children aged 3 to 10-years-old, as well young adults. Children under 6-years-old responded reactively by quickly recalling the first item but were then slower for subsequent items. Older children and adults were slower to recall the first item, but relatively faster for later items in the sequence. This suggests that they proactively planned recall for multiple items, reflecting in faster response times following the first item (Chevalier et al., 2014).

Chevalier et al. (2015) contrasted reactive and proactive control by exploring the effect of task demands on strategy choice. They presented children with a task where images had to be sorted by either colour or shape, manipulating when cues to the relevant feature were presented. For a range of outcome measures 10-year-olds showed evidence of using cue information to proactively plan their responses whenever it was available. However, 5-year-olds only engaged proactive control when a reactive strategy was made less useful by only presenting a cue to the relevant feature prior to the item to be sorted. This suggests that in addition to a developmental increase in executive control abilities, children also show an increased metacognitive ability to flexibly engage appropriate control strategies with age (Chevalier, 2015b; Chevalier et al., 2015). These 'meta-control' abilities are crucial for children to appropriately select from a range of abilities in response to specific task dynamics. Importantly, control strategies will vary in their associated cognitive costs, meaning that choice of strategy is likely to be influenced by a child's available executive resources (Chevalier et al., 2015).

While proactive abilities have previously been observed in the age-range used in the current study, they continue to develop over a prolonged period (Chevalier, 2015b). Speculatively, our sequential task might encourage a reactive control style whereby participants attempt to remember all the items before making a retrospective search of memory following the probe. Crucially, while proactive control strategies might emerge for some tasks at 7-years-old we should not expect development to involve monolithic shifts that immediately apply to all tasks (Siegler, 1994, 2007). Rather, proactive control should develop over time, perhaps being applied to tasks such as ours, where the cues to the optimal strategy are less salient, later in development (Chevalier et al., 2014; Chevalier, 2015b).

As noted in Chapter 2, the majority of children tested were from low socioeconomic (SES) neighbourhoods. This factor could be relevant here considering that prioritisation is thought to draw on executive resources (Hu et al., 2016). While evidence for a relationship between SES and WM storage is mixed, it is more consistently found to be related to executive abilities (Alloway et al., 2014; Arán-Filippetti, 2013; Engel et al., 2008; Farah et al., 2006; Hackman and Farah, 2009; Hackman et al., 2014; Noble et al., 2005, 2007). When comparing across other studies, it is important to be mindful of the possible role of extraneous factors such as SES on task performance, whether the sample is predominately low or high SES. Indeed, future research could investigate potential influences of socio-economic factors on tasks involving executive control.

3.9.1 Conclusions

The work presented in this Chapter suggests that children are unable to engage executive control to prioritise a particular position within a visual sequence, when instructed to do so. This suggests that, for children 7 to 10 years old, simply ‘paying more attention’ to a subset of a sequence is not a fruitful way to support performance. A number of alternate possibilities could be considered to attempt to elicit prioritisation effects in children (see General Discussion). For example, the frequency with which a prioritised item is probed could be manipulated to investigate whether this interacts with performance. Alternatively, more meaningful stimuli could be used (e.g. Shimi and Scerif, 2015), with the idea that children may be more willing or able to prioritise stimuli with more intrinsic meaning. However, such a change means sacrificing experimental control and methodological purity; long term memory may contribute to performance as well as verbal recoding. Here the aim was to reproduce work with adults as closely as possibly to allow direct comparisons. The recency effects observed here demonstrate a way that performance for some items is supported relatively automatically as a result of the dynamics of a task. Chapter 4 extends this idea by investigating the influence that structure in the task environment has on WM performance.

This study represents the first attempt to delineate the boundaries of children's ability to selectively attend to items in visual WM by using sequential presentation and explicit prioritisation instructions. Across three experiments, we observed no evidence that children aged 7-10 years can selectively prioritise an item within a sequence. This runs counter to repeated observations using the highly similar methodology in young adults (Hu et al., 2014, 2016), suggesting the ability (or proclivity) to do so emerges after 10 years of age. In contrast, we provide convergent evidence for the automaticity of the recency effect in visual WM, robustly demonstrating this boost for the final item in a different population to the previous work with adults (Allen et al., 2014; Hu et al., 2014, 2016) and providing evidence for the absence of a relationship with cognitive control using individual difference rather than dual-task methodology. Thus, children do show the same overall profile as adults concerning the relative effortful and automatic processing of earlier versus final sequence items. However, unlike adults, they appear unable to selectively allocate more attentional resources to particular items in a sequence.

Chapter 4

Using the environment to support working memory

4.1 Introduction

Chapter 3 described an investigation into the ability of children to use executive attention to increase the accessibility of a subset of a visual sequence. Ultimately this is something that children aged 7 to 10 years seem unable to do. In light of this result a different approach to supporting WM in children was taken here. Rather than looking to internal attentional resources, 'resources' available outside of an individual were investigated. This investigation was informed by embodied approaches to cognition. Accounts of embodied cognition vary greatly (see Shapiro, 2010; Wilson, 2002) but all emphasise the role of the body and environment in shaping cognition. Embodied cognition represents a useful starting point for supporting WM with its emphasis on real-world performance and ecologically valid tasks. If one is ultimately concerned with supporting WM in the real world it seems appropriate to take a perspective informed by real-world performance.

The ‘extended mind’ theory argues that the cognitive system is, at times, constituted by elements of the brain, body, and environment (Clark and Chalmers, 1998; Clark, 2008, 2016). When we are faced with complex problems we solve them with a ‘soft assembly’ of resources that span brain, body, and environment. The phrase ‘soft assembly’ captures the dynamism and flexibility — common to many theories of embodied cognition – at the heart of this approach. It is not the case that the environment is always a constituent of the thinking system. Instead, resources in the environment enter into this system for as long as they are used. For example, someone might use a scrap of paper to make notes while working through a problem. While this dynamic movement of information between brain, body, and the environment continues there is a ‘soft assembled’ cognitive system including the scrap of paper. However, a piece of note paper may well be discarded after it has been used. Similarly, information relevant to a task may be represented in WM until it is no longer needed and ‘discarded’ from memory. Thus, tools in the environment are used in support of cognition in much the same way, functionally speaking, as internal resources. Understanding WM performance in real-life situations will ultimately involve understanding the range of ways information can be represented over short periods of time in support of complex cognition. The most parsimonious account of behaviour may involve describing a system that spans brain, body, and environment. A perspective that emphasises factors outside of the brain seems all the more important when we think in terms of children in the classroom – that is, children embedded in a certain sort of environment. An analysis of performance need not stop at interactions between brain and body; the classroom environment is replete with exactly the sorts of tool that we routinely use to augment our cognitive capacities.

The work in this chapter involves an attempt to look at the way task-relevant structure in the environment influences WM performance. The aim was to carry out some of the experimental groundwork necessary to understanding this idea, rather than immediately exploring applications. Here an original ‘proof of concept’ for the relevance of structure in the environment is presented. The experimental manipulation is intentionally subtle in order to provide the groundwork for manipulating the influence the environment has on cognition within the context of WM performance.

Making use of our environment to support cognition is ubiquitous, from writing notes (Clark, 2008; Hertel, 1993; Intons-Peterson and Fournier, 1986) to arranging a kitchen (Kirsh, 1995). The importance of spatial organization to performance has also been demonstrated experimentally with computerized (Hess et al., 1999; Solman and Kingstone, 2017) and virtual reality tasks (Ragan et al., 2012). The organization of material is not only relevant to performance, but also to metacognitive judgements of one’s knowledge, such that individual who organize their environment (e.g. office space) are more confident in their knowledge (Hamilton et al., 2016; Hertel, 1988). Additionally, domain knowledge likely interacts with the benefit observed for expected versus unusual spatial arrangements, such as chess players’ familiar with certain arrangements of pieces on a board (Bonny and Castaneda, 2016; Chase and Simon, 1973; Connors et al., 2011; Holding and Reynolds, 1982).

Here we consider the role of task-relevant groupings of objects in the environment in the context of children completing a WM task. Task-relevant arrangements of items in the environment can be seen as an ‘intelligent use of space’ (Kirsh, 1995). Kirsh (1995) outlines the various ways in which space (i.e. the environment) can be used to support and constrain cognition. The influence of space on cognition can be classified into, (i) “spatial arrangements that simplify choice” (ii) “spatial arrangements that simplify perception”, and (iii) “spatial dynamics that simplify internal computation” (Kirsh, 1995, p. 35). In most real-world examples these functions of space will combine. For example, arranging the ingredients needed to cook some meal on a counter simplifies choice by removing the need to choose the relevant ingredients stored among irrelevant ingredients. This spatial arrangement also simplifies internal computation as the ingredients will act as cues to memory in the process of following a recipe. Perception could also be simplified in this case by grouping together ingredients used at the same point in a recipe. This would obviate the need to visually search the counter-top for the ingredients needed at a given point.

Making use of the environment in real life situations have been investigated in a range of occupational settings such as bartenders mixing drinks (Beach, 1993), waitresses taking orders (Stevens, 1993), and workers at dairy plants (Scribner, 1986). Beach (1993) notes that the different glasses used for different drinks act as a constraint on cognition for bartenders (see also, Clark, 2008). Each glass only affords a small number of possible drinks, simplifying the search of memory required to decide what drink to make for a given glass. The temporal dynamics of the bartender's actions further illustrate the ways we make use of space, in that the immediate arrangement of glasses for a list of drinks capitalises on the representation of the list before it has undergone substantial decay or interference. Later in the task, when this representation is weaker, the glasses serve as cues to recall. Stevens (1993) observed how waitresses reorganize orders from multiple customers to map onto locations within a restaurant. Both these examples show the way in which people use spatial configurations in their environment to support ongoing cognition (see also, Scribner, 1986).

The role of arranging objects in the environment could also be characterized by the idea of offloading (Dunn and Risko, 2015; Gilbert, 2015; Risko and Gilbert, 2016). A (sub)task is offloaded onto the environment when the environment is organized to substitute a functional role previously carried out internally. Clark (2008) gives the example of the dynamic flow of information between internal and external representations via a notepad when working on some problem. This offloading is functionally useful in reducing the load placed on capacity-limited cognitive systems such as working memory. This would then free-up cognitive resources to allow for more complex and sophisticated forms of processing or behaviour.

Participants have been shown to offload aspects of a task by organizing stimuli to reflect task goals (Gilbert, 2015). Participants were particularly likely to offload task instructions when the load of the task was increased either by having more complicated instructions or by interleaving a processing task (Gilbert, 2015). Participants also show an increased tendency to offload as a task becomes more difficult in the context of STM (Risko and Dunn, 2015). Risko and Dunn (2015) presented participants with a sequence of letters for serial recall of varying lengths. In one condition participants were given the opportunity to write down the sequence of letters. Unsurprisingly, participants' performance was better when they chose to write down items from a sequence versus cases where they either did not or could not. Participants were also more likely to offload the task as sequence lengths increased. In a further experiment participants were asked to make metacognitive judgements about a task where they would either have to rely on memory to recall a sequence of letters or would be required to write them down. Participants who imagined having to write down sequences predicted that their recall would be more accurate and less effortful than participants who imagined having to rely on internal memory resources. Thus, the availability of external memory resources affects both actual and imagined performance (Risko and Dunn, 2015).

Given the pervasive difficulties experienced by children with poor working memory abilities, researchers have investigated ways in which to support such children. St Clair-Thompson et al. (2010) identify two possible routes to supporting children with WM difficulties. One is to adapt tasks and teach strategies to provide environmental support, while the other tries to improve working memory directly through training. The failure of training (Melby-Lervåg and Hulme, 2013) to provide convincing beneficial outcomes leads us to ask whether monitoring, strategy use, and task adaptation offer a more fruitful approach to supporting children with WM difficulties (Elliott et al., 2010;

Gathercole et al., 2008; St Clair-Thompson, 2011; St Clair-Thompson et al., 2010). Grounded or embodied accounts of cognition that emphasize the role of the body and environment (Clark, 1999, 2008, 2016; Shapiro, 2010; Smith and Sheya, 2010; Smith and Thelen, 2003; Smith and Gasser, 2005; Thelen and Smith, 1996; Wilson, 2002; Yu and Smith, 2017) offer a way to develop an approach based on strategies and task adaptation. If WM training does not transfer to broader cognition we might, nevertheless, be able to adapt the environment and affordances of a task to support performance.

The effects of enactment in the following instructions literature (Allen and Waterman, 2015; Gathercole et al., 2008; Jaroslawska et al., 2016; Waterman et al., 2017) could be characterized as a case of making use of the body to support working memory performance. The enactment effect describes the benefit that performing a sequence of actions has on recall, over simple verbal repetition. This was initially an unexpected observation (Gathercole et al., 2008) that has since been investigated more systematically (Allen and Waterman, 2015; Gathercole et al., 2008; Jaroslawska et al., 2016; Waterman et al., 2017). Allen and Waterman (2015) investigated the effect of enactment at encoding on either verbal or enacted recall. Enacted recall resulted in more accurate performance overall without any additional benefit from enactment at encoding. In contrast, enactment at encoding boosted verbal recall. This absence of a benefit for enactment at encoding for subsequent enacted recall has been challenged in work using different stimuli (Jaroslawska et al., 2016). Waterman et al. (2017) investigated these differences showing that the complexity of the required actions was crucial. Where a large set of objects and action is used (Allen and Waterman, 2015) enactment at encoding does not benefit recall. If the stimulus set is simpler (Jaroslawska et al., 2016) then enactment at encoding does boost subsequent enacted recall (Waterman et al., 2017). Demonstration of a sequence of instructions has also been found to benefit recall (Waterman et al., 2017; Yang et al., 2015b). The benefit for enacted recall is also not vulnerable to cognitive load (Yang et al., 2014).

Taken together this work suggests a role for the body in supporting working memory performance for a task originally designed to be analogous to classroom activities (Gathercole et al., 2008). The automaticity of the enactment effect (Yang et al., 2014) is also important in suggesting it provides a ‘cost-free’ way to boost performance, like recency effects (Chapter 3). That said, the complexity of the stimuli used is clearly an important factor to consider (Jaroslawska et al., 2016; Waterman et al., 2017). Waterman et al. (2017) found that demonstration benefited recall, even for more complicated stimulus sets. One natural extension of this work would be to investigate the influence of structure in the environment on performance. Such investigations would extend the scope of influence on performance from brain and body into the environment (Clark, 2008).

Embodied cognition has been explicitly applied to theory in WM. Macken et al. (2015) argue that performance on a given task results from the dynamic interplay of three factors, (i) the *task* that must be completed, (ii) the *material* the task must be completed on (words, shapes, etc.), and (iii) the *repertoire* of competencies and skills available to the participant (see also, Macken et al., 2016; Hughes et al., 2016). For example, in a verbal serial recall task the materials might be a list of letters, the requisite task would be to reproduce the sequence in the same serial order, and the repertoire available to a participant might include processes to represent order and perceptual-motor systems that allow letters to be articulated. Importantly, this approach highlights that training an individual’s competencies is not the only route to improving performance. The materials for a task could, instead, be adapted to be more congruent with task goals (Macken et al., 2015). The enactment effect for following instructions is a case of a change in the task resulting in a different repertoire of skills being recruited that improve performance. The importance of stimulus sets in this literature (Waterman et al., 2017) offers an example of the materials of a task influencing emergent performance.

In summary, making use of the environment to support cognition provides a potential avenue to supporting children with low WM. Classrooms are complex environments where successful performance might emerge from complex interactions between classically cognitive constructs (memory, attention), the body, and features of the environment. The current research considers whether task-relevant groupings of objects in the environment can help support the ability to remember sequences of instructions. Here the 'materials' that participants must use to recall verbal sequences are manipulated to be more, or less, task relevant. This task-relevant grouping of objects in the environment is a type of offloading (Dunn and Risko, 2015; Gilbert, 2015; Risko and Gilbert, 2016), which reduces the load placed on limited capacity cognitive systems, or increases the perceptual-motor congruence between the materials and required task (Hughes et al., 2016; Macken et al., 2015, 2016).

4.2 Experiment 1

Participants were asked to encode short verbal sequences of colours and recall them by picking up physical coloured bricks. These bricks were either arranged randomly, or grouped by colour. This latter, ordered, arrangement could be described as offloading visual search and action-planning onto the environment (Dunn and Risko, 2015; Gilbert, 2015; Risko and Gilbert, 2016). Equally, it is an example of using space to simplify perception and choice (Kirsh, 1995). Perception is simplified compared to a random arrangement by salient colour groupings reducing the need to search for a particular brick. Choice was also simplified as participants only had to choose between four groups of bricks to select a particular colour.

We, therefore, expected that participants would perform better when the bricks were grouped by colour. We also took additional WM measures to explore the influence of individual differences in WM on performance on the primary task.

Children with poor WM show deficits even on simple storage tasks (Alloway et al., 2009a; Gathercole et al., 2016; Holmes et al., 2014), such as recalling the series of colours presented here. If grouping objects in the environment allows participants to 'offload' some processing demands, we might expect participants with low WM to particularly benefit. Children with better WM, on the other hand, may not find even a pseudorandom arrangement particularly demanding, and so would show reduced benefits of object grouping.

Analyses are presented both for the whole sample and comparing 'extreme groups' from the top and bottom 15% of participants. This extreme groups analysis answers the same basic question of the overall analysis, attempting to provide convergent evidence with an alternate commonly used method. A cut-off of 15% was used as it has been common in previous studies (e.g. Gathercole et al., 2016; Holmes et al., 2009). In addition, a 15th percentile cut off is often used to identify children with motor difficulties (Lingam et al., 2012; Pieters et al., 2012). Finally, 15% was found to provide reasonably equal numbers of participants for the low and high groups, considering the small number of scores afforded by span measures.

4.2.1 Method

Participants. 88 participants took part in Experiment 1. Four were excluded due to having special educational needs, leaving 84 children (Mean age = 10.34, SD = 0.28, Range = 9.89 – 10.85). Participants were recruited from a predominately low SES Pakistani British area of Bradford, UK. Ethical approval was obtained from the School of Psychology Ethics Committee, University of Leeds, UK.

4.2.1.1 Materials

Primary task. For the primary task participants were presented with 12 coloured bricks measuring 3 x 3 x 2cm. The bricks were placed in a line with a gap of 3cm between each brick. The experimenter read out a sequence of colours at a rate of approximately one per second in a neutral voice. At the end of the sequence the participant was required to pick up a single coloured brick for each colour in the sequence and place it in a response area in front of them. At the end of a trial the bricks were returned to their original location. Four colours were used (red, blue, yellow, and green). A span procedure was used starting at sequences of length three and increasing to a maximum of eight. An item was scored as correct only if it was recalled in the correct location. Participants completed three trials at each span length, with span being increased by one if the participant got two or more trials correct. This procedure was continued until the participant got fewer than two trials correct at a given length. A participant's span score was then the longest length at which they got two or more trials correct. An additional 0.33 was added on to the score if they got one trial correct at the next sequence length. For example, if a participant got 2/3 trials at span five and 1/3 trials at span six, then they would have a span score of 5.33. Participants completed two conditions for this task. One where the bricks were in a pseudorandom arrangement and another where the bricks were arranged by colour. A schematic of the task is depicted in Figure 4.1. The order of the two arrangement conditions was counterbalanced.

Forward digit recall (FDR). The experimenter read out digits at a rate of one per second and the participant was required to repeat them back in the same order. The same span procedure as the primary task was used with sequence lengths starting at three and increasing to a maximum of eight. Performance was scored in the same way as the primary task.

Backward digit recall (BDR). This task was identical to forward digit recall except that participants were required to repeat the letters in backwards order. The task began with sequences of length two increasing to a maximum of six.

Corsi block recall. A board with nine pseudorandomly arranged blocks was placed in front of the participant. The experimenter touched a sequence of blocks

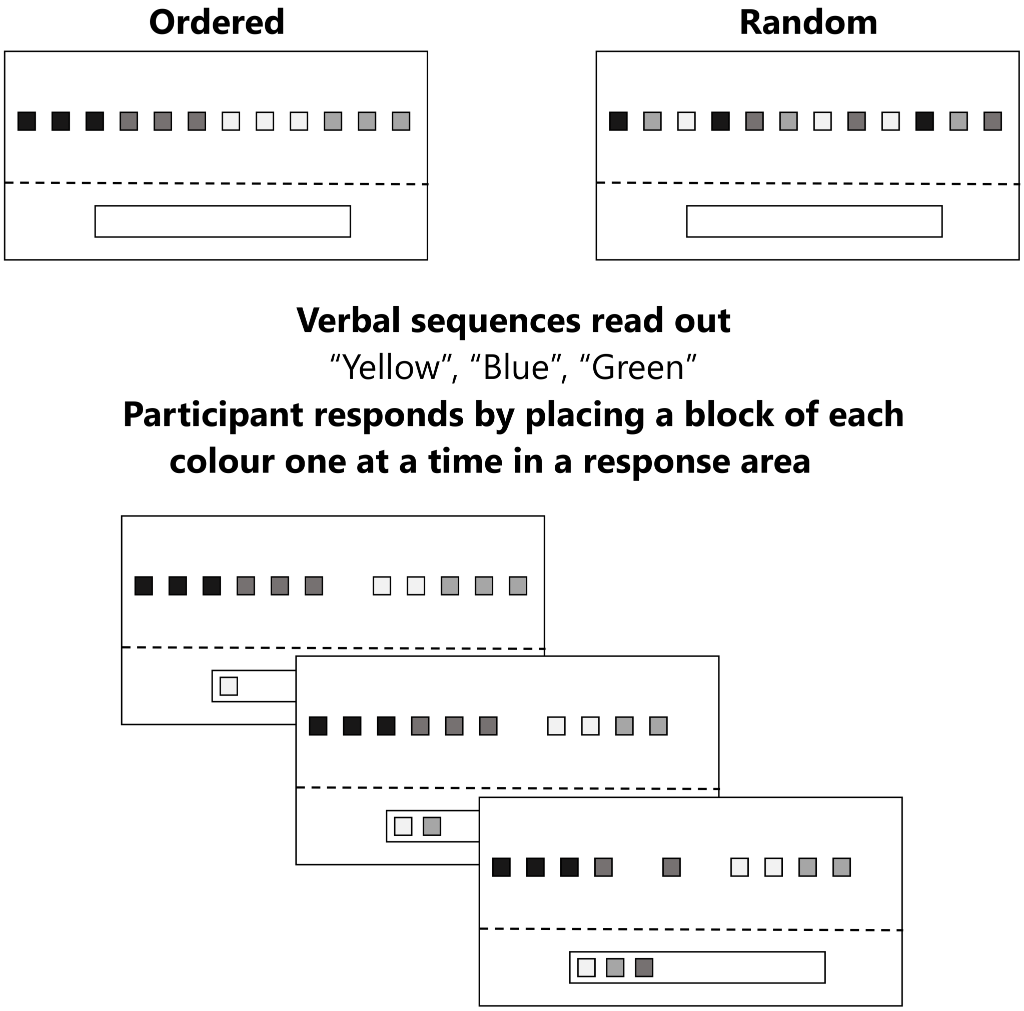


Figure 4.1: *A schematic of the procedure for the primary task. Shades of grey represent different colours. (Not to scale)*

at a rate of approximately one per second. The participant was then required to touch the blocks in the same order. This task used the same span procedure as previous tasks beginning at sequences of three up to a maximum of eight.

4.2.2 Results

4.2.2.1 Primary task

Model details. The data were analysed using a Bayesian linear mixed-effects model in R (R Core Team, 2017) using rstanarm (Gabry and Goodrich, 2017) with random intercepts for subjects. As there were only two observations per participant, this was the 'maximal' (Barr et al., 2013) random effect structure afforded by the data. Such a model differs from a standard regression in that rather than having a single intercept term for all participants each participant has their own intercept term. This can be represented as:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i,\ \epsilon_i \sim N(0, \sigma_y^2)$$

Where α_j is the intercept associated with the j^{th} participant. Importantly, these participant-level intercepts are themselves drawn from a common distribution.

$$\alpha_j = a + \eta_j,\ \eta_j \sim N(0, \sigma_\alpha^2)$$

This analysis was implemented using Bayesian estimation via rstanarm in R (see Chapter 2 for further details on this method). The model was run for 20,000 iterations, split across 4 chains. An additional 20,000 iterations were used for warm-up prior to the 20,000 draws from the posterior used to make the inferences below. The default, very weak, priors in rstanarm (Gabry and Goodrich, 2017) were used for this model. These are shown in Figure 4.2. For example, the

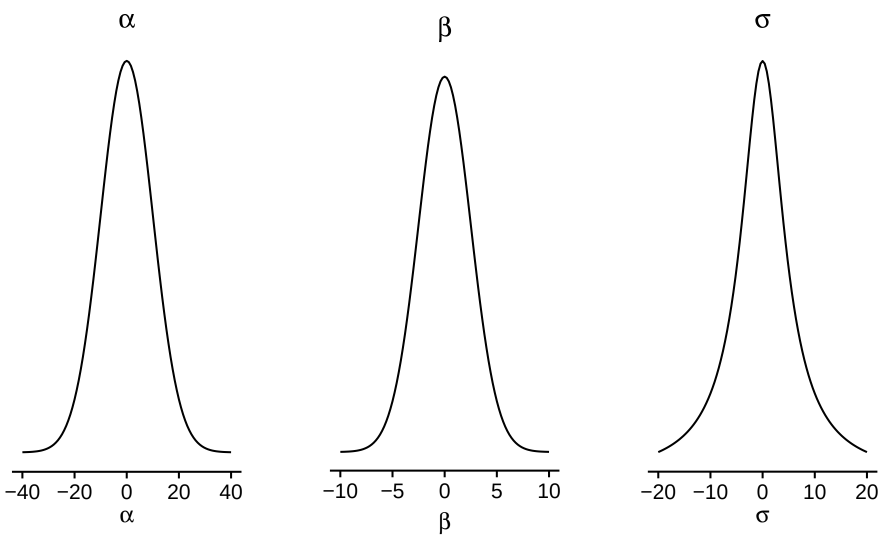


Figure 4.2: *Default priors for rstanarm Version 2.15.3 for the models used here.*

Table 4.1: *Descriptive statistics for the variables in Experiment 1.*

Variable	Mean (SD)	Range
Age	10.34 (0.28)	9.89 – 10.85
Backward digit recall	3.36 (0.85)	1.33 – 6
Forward digit recall	5.19 (0.88)	3.33 – 8
Corsi	4.55 (0.89)	3 – 6.33
Ordered	4.85 (0.54)	4 – 6.33
Random	4.66 (0.77)	3.33 – 6.33
Ordered advantage	0.19 (0.58)	-1.33 – 1

prior for the coefficients in the model was a normal distribution with mean 0 and standard deviation 2.5. Priors this weak will have essentially no influence on the final estimates, particularly given the sample sizes used.

Results. The model included condition, FDR, BDR, and Corsi as predictors as well as interactions for each of FDR, BDR, and Corsi with condition. Descriptive statistics for the different variables are provided in Table 4.1.

Posterior odds were used to compare the different effects in the model. The posterior odds describe the ratio between posterior samples where an estimate is positive versus negative. For example, posterior odds of 50:1 in favour of a positive

Table 4.2: *Coefficients and model diagnostics for the main model in Experiment 1. See main text for a description of the diagnostics. Except for the intercept parameter estimates are standardised coefficients*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	4.752	4.648, 4.854	20000:1	9133	1.000
Condition	0.095	0.037, 0.154	1052:1	20000	1.000
FDR	0.291	0.177, 0.404	20000:1	8364	1.000
BDR	0.061	-0.058, 0.18	6:1	7881	1.001
Corsi	0.099	-0.013, 0.212	23:1	8586	1.000
Condition x FDR	-0.094	-0.158, -0.029	487:1	20000	1.000
Condition x BDR	0.004	-0.064, 0.071	1:1	20000	1.000
Condition x Corsi	-0.062	-0.127, 0.002	35:1	20000	1.000

[1] Mean refers to posterior distribution

[2] 95% CI is the 95% Bayesian credible interval

[3] Odds are the posterior odds

[4] ESS = effective sample size

coefficient would indicate that, for the 20,000 samples taken from the posterior, 50 estimates were positive for every 1 that was negative for the effect of interest. Table 4.2 provides the coefficients the main model in Experiment 1.

For the effect of condition the odds were 1052:1, with performance being superior in the ordered condition. The odds were 20000:1 in favour of a positive coefficient for FDR predicting performance on the primary task, and were 23:1 in favour of a positive coefficient for Corsi predicting task performance. However, for BDR the odds were only 6:1, and therefore did not support BDR predicting performance on the primary task. The interaction between FDR and condition was favoured 487:1, as was the interaction between Corsi and condition (odds of 35:1). The interaction between BDR and condition was not supported (odds of 1:1). Table 4.2 provides the parameter estimates as diagnostic for this model.

As the overall effect of condition interacted with FDR and Corsi, it was evaluated at different values of both predictors. Figure 4.3 shows the distribution of posterior estimates from the model where the values of FDR and Corsi were set at the sample minimum, mean, or maximum (see Table 4.1). This allows us to show

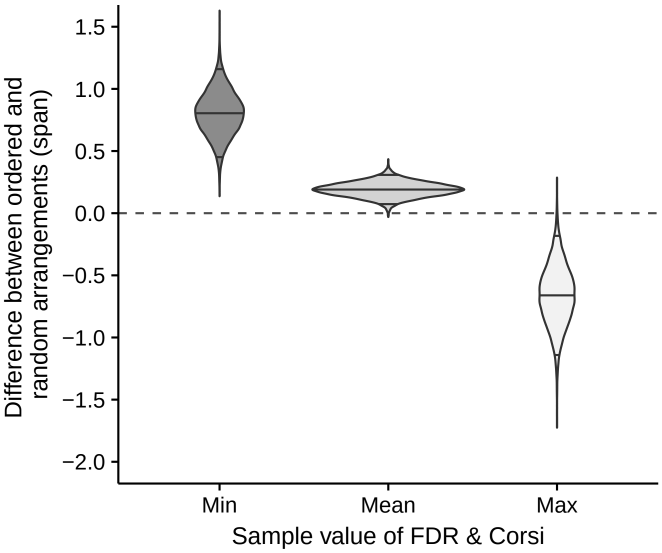


Figure 4.3: *The distribution of predicted differences between the ordered and random conditions at the sample minimum, mean and maximum for forward digit recall and Corsi.*

how the predicted performance in the two conditions interacts with performance on our additional WM measures. In other words, Figure 4.3 shows the predicted difference in performance between the ordered and random conditions from the coefficients in our model for an individual performing at, for example, the sample minimum of FDR and Corsi.

At the sample mean of FDR and Corsi the posterior estimate for the difference between the ordered (posterior median = 4.85, 95% credible interval = 4.73 – 4.97) and random 4.66, 95% credible interval = 4.54 – 4.78) conditions was 0.19 (95% credible interval = 0.07 – 0.31). This suggests a slight boost in performance in the ordered condition for the average participant. As Figure 4.3 shows, the estimated boost in performance in the ordered condition was more pronounced for participants with poor working memory.

Model diagnostics. The performance of MCMC sampling can be summarised in a number of ways (Kass et al., 1998; Kruschke, 2014). For the model all parameters had an effective sample size greater than 5951 and $\hat{R}$ values < 1.001, with the majority of parameters having effective samples sizes of 20,000 (see

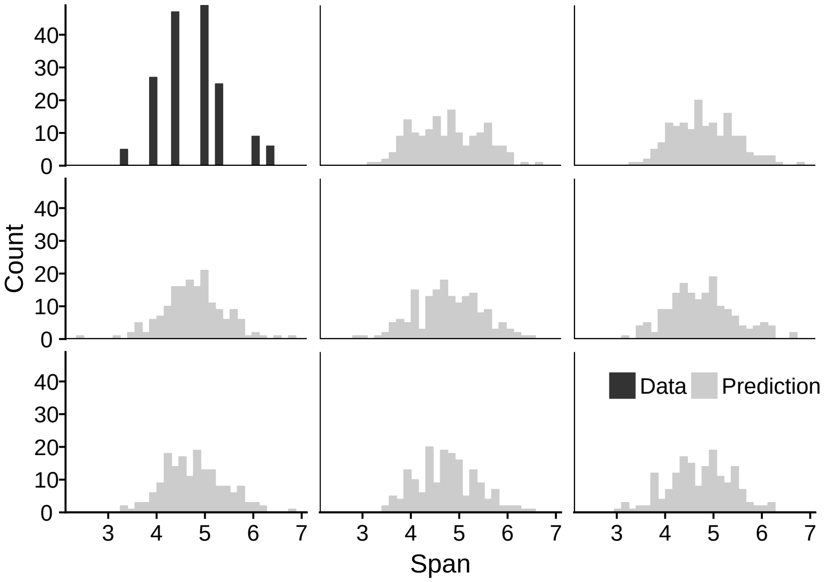


Figure 4.4: *Posterior predictive check for the main model in Experiment 1 showing the raw data and sets of fitted values from the model. The upper left panel shows the raw data with the other panels showing sets of predicted values from the model. Figure generated using the pp_check function from rstanarm (Gabry and Goodrich, 2017).*

Table 4.2). The effective sample size is a measure of the number of samples from the posterior taking into account autocorrelation between samples. This can be thought of as indexing the amount of unique information in the posterior with respect to a particular parameter (Kruschke, 2014). $\hat{R}$ (R-hat in Table 4.2) is a measure of the variance between chains versus the combination of all the chains, indexing convergence. Trace-plots were also visually inspected using ShinyStan (Gabry, 2017). Figure 4.4 below shows a posterior predictive check for the main model. The unusual distribution for y (the data) is a result of span measures only being able to take a small number of values. y_{rep} shows the distribution of sets of predicted values from the model. Posterior predictive checks can be used to assess the fit of a model by observing how closely the predicted values of the model resemble the raw data, showing whether estimates generated from the model approximate the underlying data. If a model can produce estimates that resemble the data used to create it then the model could be thought to capture the key features of the process that generated the raw data.

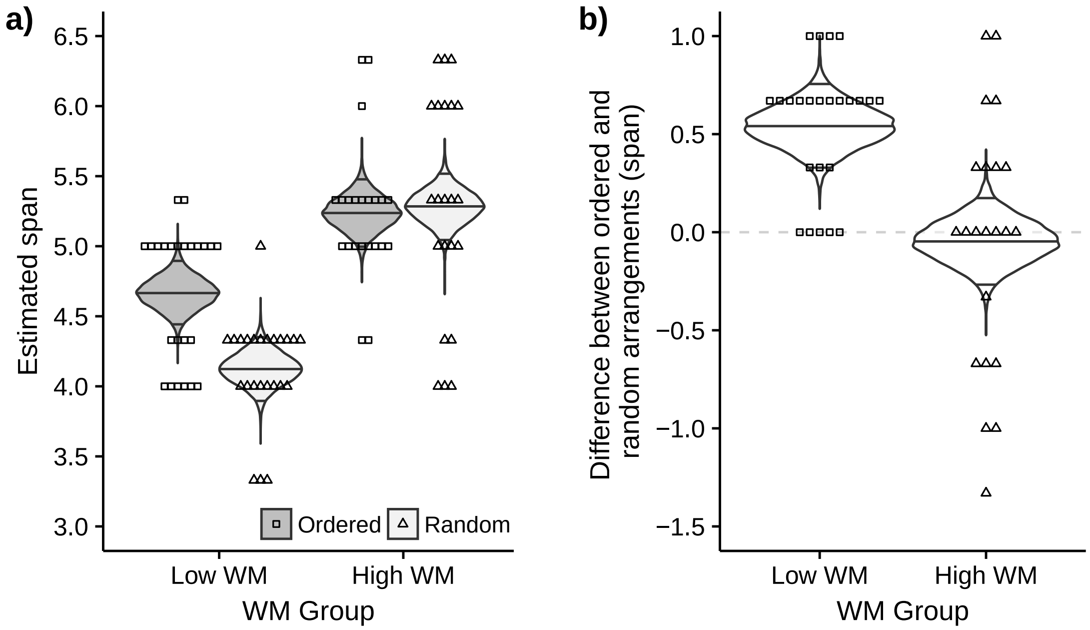


Figure 4.5: *a) Estimated performance in the ordered and random conditions for the low and high WM groups. (b) The estimated difference between the ordered and random conditions for the low and high WM groups. Unfilled shapes show the raw data.*

4.2.2.2 Extreme groups

Results. The interaction between condition and our additional measures was further explored using an extreme-groups approach. Low (N = 24) and High WM (N = 22) groups were selected by taking participant below the 15th percentile (Low WM) and above the 85th percentile (High WM) on the forward digit recall task (e.g. Gathercole et al., 2016). FDR was used to select the groups due to showing the largest interaction with condition. As with the primary task analyses a linear mixed-effect model was fitted to these subgroups, with random intercepts for subjects. The model included condition, WM group, and a condition by WM group interaction. Figure 4.5 shows the estimated difference between the ordered and random condition for the two groups, as well as the raw data.

Posterior odds were again used to gauge support for different effects in the model. The effect of condition was favoured 1999:1. The effect of WM group was favoured

Table 4.3: *Parameter estimates for the extreme group model in Experiment 1. Except for the intercept parameter estimates are standardised coefficients.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	4.829	4.686, 4.972	20000:1	8328	1.000
Condition	0.124	0.048, 0.201	1999:1	20000	1.000
WM Group	0.433	0.289, 0.576	20000:1	8481	1.001
Condition x WM Group	-0.147	-0.224, -0.07	6666:1	20000	1.000

[1] Mean refers to posterior distribution
[2] 95% CI is the 95% Bayesian credible interval
[3] Odds are the posterior odds
[4] ESS = effective sample size

by a factor of 20000:1. Finally, the interaction between condition and WM group was favoured 6666:1. The posterior odds did not support a difference between conditions in the high WM group (odds of 2:1 in favour of superior performance in the random condition). In contrast for the low WM group the odds were 20000:1 in favour of better performance in the ordered condition. Table 4.3 provides the parameter estimates and diagnostics for this model. Finally, the predictive performance of the model was assessed using posterior predictive checks, as with the main model (Figure 4.6).

4.2.3 Discussion

The results of Experiment 1 suggest that, on average, participants' spans were higher when objects had a task-relevant versus a pseudorandom arrangement. This effect was much more pronounced for participants with lower WM. Why do children with poor working memory particularly benefit from these sorts of stimulus arrangements?

One possibility is that children with low WM find the relatively complex visual search in the random condition particularly taxing. With the ordered arrangement participants can 'offload' (Dunn and Risko, 2015; Gilbert, 2015;

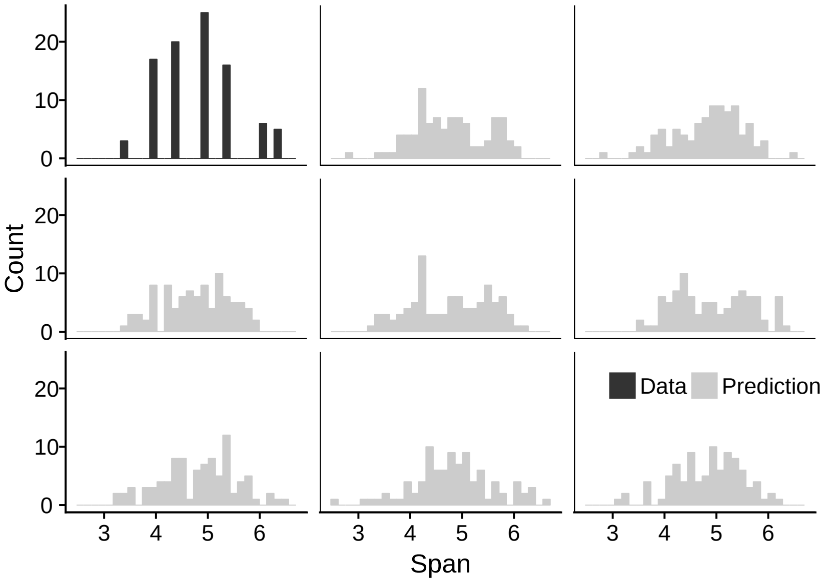


Figure 4.6: *Posterior predictive check for the extreme group model in Experiment 1 showing the raw data and sets of fitted values from the model. The upper left panel shows the raw data with the other panels showing sets of predicted values from the model. Figure generated using the pp_check function from rstanarm (Gabry and Goodrich, 2017).*

Risko and Gilbert, 2016) some of the visual search elements of the task onto the environment (Kirsh, 1995). In line with this, previous work has found that low WM children are less accurate (although not slower) on measures of visuospatial attention (Holmes et al., 2014; St Clair-Thompson, 2011). Another task element that may be offloaded onto the environment in the ordered condition is action planning. In the ordered condition there are only four possible actions, one for each colour groupings. This could represent a simpler action space than the random condition, reducing the load placed on internal resources. When investigating action-planning abilities in low WM children, St Clair-Thompson (2011) found that low WM children required more attempts to complete a Tower of London action planning task than participants with average WM. Whilst Holmes et al. (2014) did not replicate the deficit in overall planning performance on the Tower of London task, children with low WM did make more rule violations when completing the task.

A complementary account of our finding relates to the approach of Macken and colleagues (Macken et al., 2015, 2016) where performance on a given task is thought to result from the interplay between the task itself, the task materials, and the competencies of the individual. In the current experiment, grouping the bricks by colour increases the perceptual-motor congruence between the materials and the required task (Macken et al., 2015), which may reduce the need to recruit attentional resources. This reduction in attentional demands would then clearly benefit children with low WM.

In summary, this experiment showed that task-relevant grouping of objects in the environment improved performance on a task, and was particularly beneficial for children with low working memory ability. One interesting question is whether children are aware of this difference between ordered and random arrangements, and, further, whether children will spontaneously make use of the environment to structure a task to optimize performance. These issues relate to the concept of metacognition; a range of abilities that involve knowledge about cognition (Kuhn, 2000; Schneider, 2008). Research has shown that metacognition develops throughout childhood (Flavell, 1979; Flavell et al., 1999; Siegler, 1994), with children's meta-strategic knowledge (awareness of task difficulty, task performance, and appropriate strategies to employ) improving with age (Kuhn, 2000). Indeed, Schneider (2008) argues that meta-strategic knowledge continues to develop into adolescence. In addition, children with low WM show poorer metacognition; they are less effective at selecting higher-level strategies, and are less able to transfer taught strategies across domains or tasks (Alloway et al., 2009a).

Experiment 2 therefore sought to replicate the findings from Experiment 1, but also to assess children's awareness of the potential benefits of structuring the environment. In addition, while the current experiment used three WM measures — simple verbal, simple visuospatial, and complex verbal WM – Experiment 2 included a complex visuospatial task. Finally, one limitation of the current experiment was that span scoring tends to result in a small number of possible scores, thus limiting the ability to discriminate between participants. Therefore, in Experiment 2 we used a method where children are given a fixed number of sequences at lengths three to six (see Waterman et al., 2017), and performance was scored at the item-level, giving a greater range of possible values (for a review of scoring methods see Conway et al., 2005).

4.3 Experiment 2

Experiment 2 attempted to replicate and expand on Experiment 1. Firstly, we expected both the main effect of condition and interaction with WM (observed in Experiment 1) to replicate. Secondly, we investigated whether children were aware that an ordered arrangement can boost performance, and whether they would spontaneously choose such an arrangement. To address this, participants rated the difficulty of the ordered and random conditions, and were asked to judge which condition was easier. In addition, a 'free arrangement' condition was added where participants were asked to arrange the bricks for themselves before completing a set of trials.

If a participant chooses an ordered arrangement in our 'free arrangement' condition, they would need to have recognized both that the ordered arrangement made the task easier, and translated this knowledge in selecting an ordered arrangement (Kuhn, 2000; Schneider, 2008). Recent work using a WM task has shown that 7 to 12-year-old children are sensitive to the cognitive effort required to complete a task and will avoid a more difficult task because it requires greater cognitive effort to complete (Chevalier, 2017). Therefore, children in this age range can show awareness of task difficulty (see also, Efklides et al., 2006). As such, we expected that children in the current experiment would recognize that an ordered arrangement facilitates better performance.

The ability to anticipate task difficulty, and plan an effective strategy to improve performance (proactive control), is less well developed in younger children (Chevalier, 2015a), and is likely influenced by concurrent developments in WM (Chevalier, 2015a). It, therefore, could be that children may not choose an ordered arrangement in the free arrangement condition, even if they can retrospectively judge the ordered condition as easier. In addition, children with poor WM may be less able to plan effective strategies due to difficulties holding and manipulating possible strategies in mind.

4.3.1 Method

4.3.1.1 Participants

100 participants took part in Experiment 2. Fifteen participants were excluded due to having special educational needs. A further 2 participants were excluded due to distraction, and one participant due to missing data on one of the additional WM measures, leaving 82 children (Mean age = 9.62, SD = 0.62, Range = 8.51 – 10.49). Participants were recruited from a predominately low SES Pakistani British area of Bradford, UK. Ethical approval was obtained from the School of Psychology Ethics Committee, University of Leeds, UK.

4.3.1.2 Materials

Primary task. For Experiment 2 all participants completed the same set of trials, rather than using a span procedure as per Experiment 1. For each condition participants completed a block of four trials at sequence lengths three, four, five and six. Performance was then scored as the total number of colours correctly recalled in the correct position. An additional colour (orange) was added to reduce the chance of participants getting items correct by guessing. This meant there were 15 objects, three of each colour. The order of the ordered and random conditions was counterbalanced. Participants then completed an additional block of trials where they were instructed to arrange the bricks in whatever order they wanted so long as they remained equally spaced in a row, as in the ordered and random conditions. After all three conditions had been completed, participants were asked to rate whether the task was easier when the bricks were ‘mixed up’ or ‘grouped by colour’. They then rated the difficulty of each condition on a 7-point scale.

For Experiment 2, computerised versions of the additional measures from Experiment 1 were used as well as an additional complex visuospatial WM measure. Using computerised measures removes potential differences in task presentation that come from different researchers presenting stimuli. Full descriptions of these tasks can be found in Chapter 2.

Forward digit recall (simple verbal WM). Participants listened to sequences of digits over headphone and recalled them in ordered using response boxes on-screen. Sixteen trials were completed with sequence lengths ranging from 3 to 6.

Backward digit recall (complex verbal WM). The same as forward digit recall except that sequences were recalled in reverse order. Sixteen trials were completed with sequences from 2 to 5 items in length.

Corsi block task (simple visuospatial WM). A sequence of nine possible boxes lit up and participants recalled the sequence by touching the boxes in the same order that they lit up. Sixteen trials were completed with sequence lengths ranging from 3 to 6.

Odd-one-out (complex visuospatial WM). Processing trials were interleaved with the recall of spatial locations to provide a complex measure of visuospatial WM. Sequence lengths ranged from 2 to 5 across 16 trials.

Table 4.4: *Descriptive statistics for the variables in Experiment 2*

Variable	Mean (SD)	Range
Age	9.62 (0.62)	8.51 – 10.49
Backward digit recall	0.7 (0.18)	0.14 – 1
Forward digit recall	0.78 (0.12)	0.44 – 1
Corsi	0.74 (0.15)	0.36 – 0.97
Odd-one-out	0.59 (0.21)	0.16 – 1
Free	0.76 (0.11)	0.47 – 0.99
Ordered	0.75 (0.1)	0.53 – 0.94
Random	0.75 (0.11)	0.46 – 0.93
Ordered advantage	0.01 (0.08)	-0.17 – 0.24

4.3.2 Results

4.3.2.1 Primary task

Model details. For Experiment 2 the outcome measure was accuracy (rather than span, as in Experiment 1). To respect the fact accuracy in not normally distributed the data were analysed using a binomial logistic regression. This approach prevents the spurious effect that can emerge from averaging over accuracy data and using analyses such as ANOVAs (Jaeger, 2008). Random slopes for the effect of condition for participants were included, in addition to random intercepts. This, again, was the maximal random effect structure afforded by the data. Including random slopes is preferable to random intercept-only models, which have substantially inflated error rates (Barr et al., 2013). The models were run for the same number of iterations as with Experiment 1. The same (default) priors for the intercept and coefficients as Experiment 1 were used; there is no σ parameter with logistic regression. However, these priors have a different interpretation here as the model is parametrised in log odds. Zero in log-odds translates to a probability of 0.5 when transformed.

Results. Descriptive statistics are provided in Table 4.4, with the caveat that accuracy data is not normally distributed (Jaeger, 2008).

Table 4.5: *Parameter estimates and diagnostics for the primary model in Experiment 2. Parameters are expressed in log-odd units. For continuous predictors the coefficients represent the change in log-odds associated with a 1 SD change in the predictor.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	1.160	1.076, 1.248	20000:1	8782	1
Condition	0.021	-0.03, 0.074	4:1	20000	1
FDR	0.262	0.141, 0.382	20000:1	7668	1
BDR	0.085	-0.033, 0.202	12:1	8283	1
Corsi	0.029	-0.078, 0.135	2:1	8611	1
Odd-one-out	0.020	-0.092, 0.132	2:1	8562	1
Condition x FDR	-0.065	-0.135, 0.005	29:1	20000	1
Condition x BDR	-0.010	-0.079, 0.059	2:1	20000	1
Condition x Corsi	0.027	-0.037, 0.092	4:1	20000	1
Condition x Odd-one-out	0.015	-0.052, 0.082	2:1	20000	1

[1] Mean refers to posterior distribution
[2] 95% CI is the 95% Bayesian credible interval
[3] Odds are the posterior odds
[4] ESS = effective sample size

The model included condition, FDR, BDR, Corsi, and odd-one-out as predictors[1]. Interactions between each predictor and condition were also included. For this primary analysis, the free arrangement condition was not included in the model as it had a number of differences with the random and ordered conditions. Excluding the free condition from this analysis also allows for more direct comparisons with Experiment 1.

In contrast to Experiment 1, the effect of condition was not supported by the model (odds of 4:1 in favour of superior performance in the ordered condition). The odds were 20000:1 in favour of a positive coefficient for FDR predicting performance on the primary task. However, for BDR the odds were only 12:1 in favour of a positive coefficient, and for Corsi and odd-one-out the odds were only 2:1 in favour of positive relationships with performance on the primary task. The

---

[1] The model was run with age group as a predictor. While there was a small main effect of age group it did not interact with condition. Thus, the simplified model is reported to aid comparison with Experiment 1.

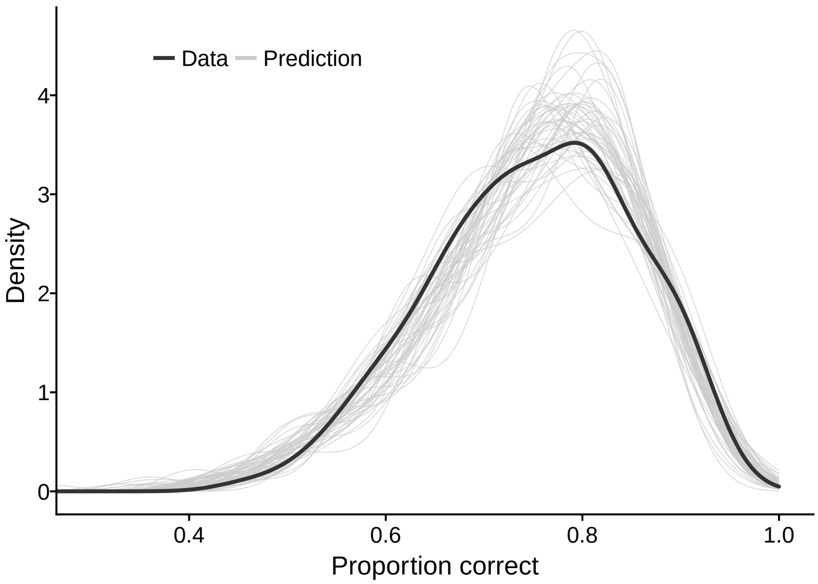


Figure 4.7: *Predicted distributions for the outcome variable in Experiment 2 with the raw data overlaid for the main model. Figure generated using the pp_check function from rstanarm (Gabry and Goodrich, 2017).*

interaction between FDR and condition was favoured 29:1. On the other hand, the interactions between BDR and condition, and odd-one-out and condition were not supported (both with odds of 2:1); neither was the interaction between Corsi and condition (odds of 4:1). The results of a posterior predictive check are shown in Figure 4.7.

As with Experiment 1 the predicted difference between the ordered and random condition from the model was estimated for different values. Here the value of FDR only was varied as an interaction between Corsi and condition was not supported, unlike Experiment 1. Figure 4.8 shows the distribution of predicted differences between the two conditions for different values of FDR.

To summarise, the data provide strong evidence for the claim that FDR predicts performance on our primary task, and moderate evidence that this interacts with condition. The evidence is minimal for the effects of condition in isolation, or the other WM measures.

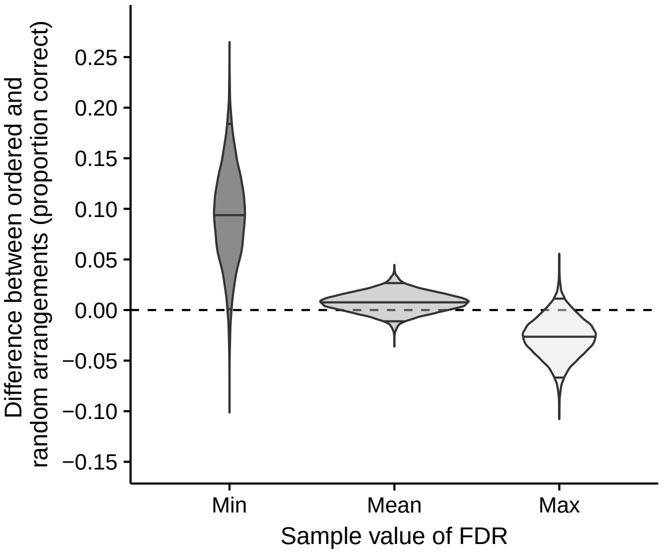


Figure 4.8: *The distribution of predicted differences between the ordered and random conditions at the sample minimum, mean and maximum for forward digit recall.*

4.3.2.2 Extreme groups

As with Experiment 1 an extreme groups analysis was also carried out. Again, participants were selected by being above the 85th (high WM) or below the 15th (low WM) percentiles for FDR. The model included condition and WM group as predictors as well as their interaction. Table 4.6 shows the parameter estimates and diagnostics for this model.

The effect of condition was not supported (odds of 3:1 in favour of better performance in the ordered condition). The effect of WM group was strongly favoured by a factor of 20000:1. The interaction between condition and WM group was supported with odds of 99:1. When looking at the WM groups separately, the effect of condition was not strongly supported for the high WM group (odds of 8:1 for superior performance in the random condition). In contrast for the low WM group the odds were 43:1 in favour of superior performance in the ordered condition.

On average, the low WM group (N = 13) were 5.2% more likely to recall an item

Table 4.6: *Parameter estimates and diagnostics for the extreme group model in Experiment 2. Parameters are expressed in log-odd units.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	1.059	0.907, 1.213	20000:1	4705	1
Condition	0.024	-0.056, 0.101	3:1	15805	1
WM Group	0.473	0.324, 0.625	20000:1	5324	1
Condition x WM Group	-0.089	-0.165, -0.015	99:1	15865	1

[1] Mean refers to posterior distribution

[2] 95% CI is the 95% Bayesian credible interval

[3] Odds are the posterior odds

[4] ESS = effective sample size

in the ordered condition. The high WM group (N = 20) were 1.9% more likely to recall an item in the random condition. However, in both cases the 95% credible interval overlapped with zero (see Figure 4.9). The posterior predictive check for this model is shown in Figure 4.10.

4.3.2.3 Free arrangement condition and metacognitive measures

The majority (78%) of participants rated the ordered condition as easier than the random condition. The modal response on a 7-point scale for the ordered condition was 7 (i.e. “very easy”). For the random condition the modal response was 4 (“neither easy or difficult”). Figure 4.11 shows the distribution of difficulty ratings for the two conditions. Table 4.7 shows the difference in performance between the ordered and random conditions grouped by the difference in difficulty rating between the ordered and random conditions.

For the low WM participants 57.1% rated the ordered condition as easier compared to 84.6% for the high WM group. This was despite the low WM participants performing better in the ordered condition. In addition, only 64.3% of the low WM group rated the ordered condition as very easy versus 84.6% for the high WM group. The modal rating of the random condition was neither easy nor difficult for both groups (50.0% vs 53.8%)

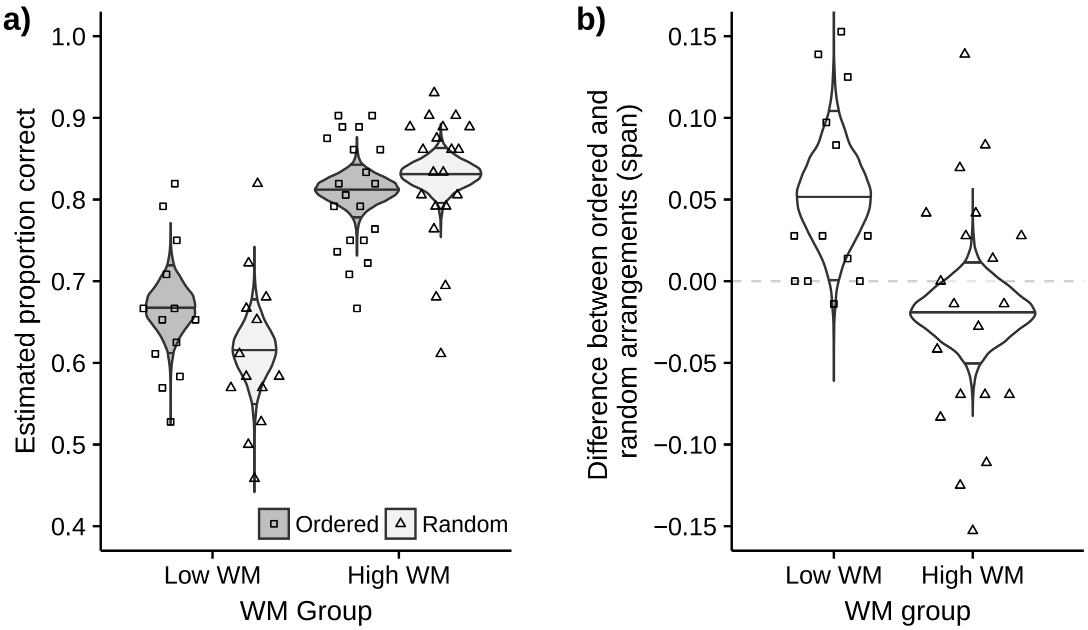


Figure 4.9: *a) Estimated performance in the ordered and random conditions for the low and high WM groups. (b) The estimated difference between the ordered and random conditions for the low and high WM groups.*

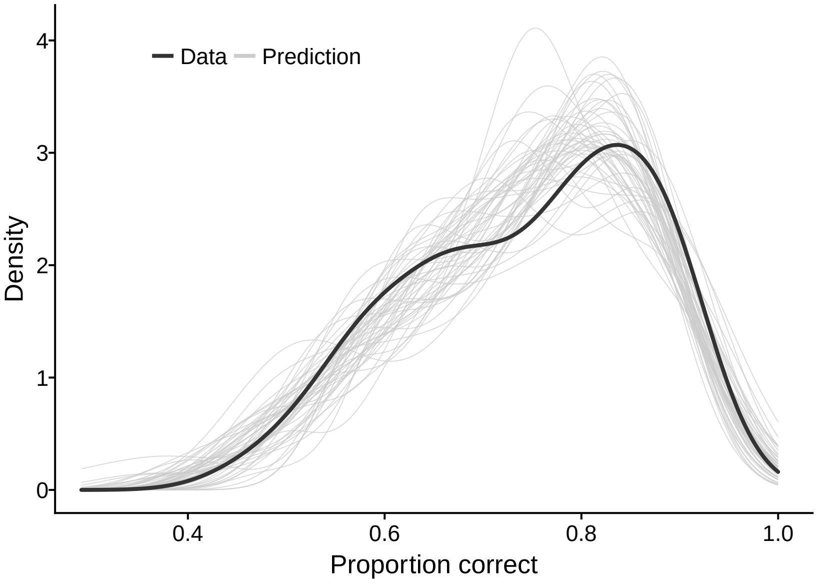


Figure 4.10: *Predicted distributions for the outcome variable in Experiment 2 with the raw data overlaid for the extreme groups model. Figure generated using the pp_check function from rstanarm (Gabry and Goodrich, 2017).*

Table 4.7: *The difference between performance in the ordered and random conditions grouped the difference in difficulty rating given for the two conditions*

Ordered preference	Mean (SD)	N
-3	0.01 (0.03)	3
-2	0.06 (0.03)	3
-1	0.11 (0.02)	3
0	-0.02 (0.07)	11
1	-0.02 (0.1)	17
2	0.04 (0.08)	18
3	-0.01 (0.08)	24
4	-0.06 (0.13)	2
6	0.08	1

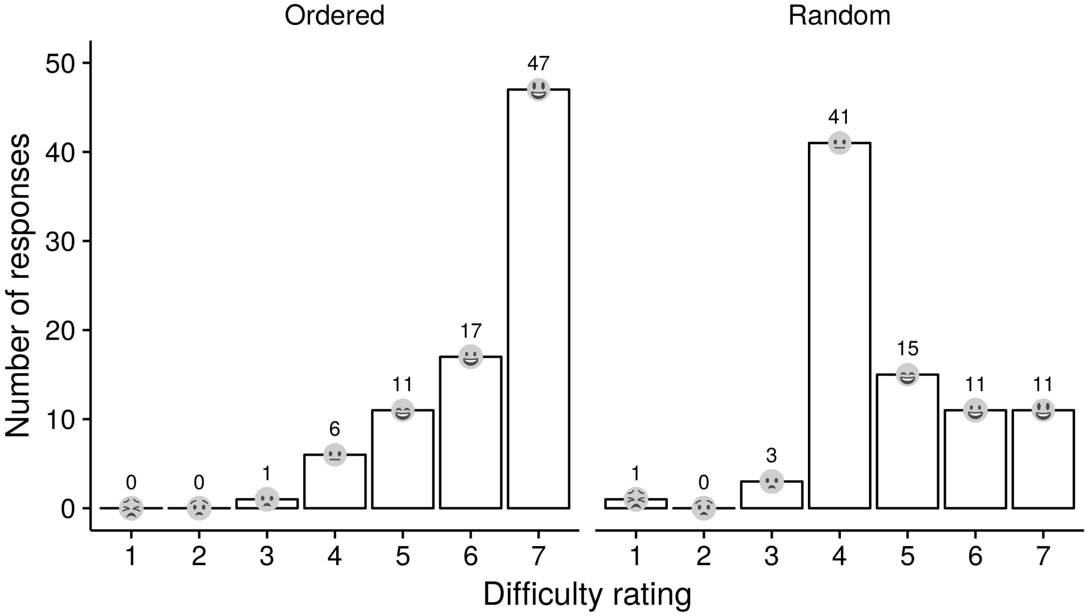


Figure 4.11: *The distribution of difficult ratings for the two condition. The emoji atop the bars were placed on a scale for participants to make the difficulty ratings. The emoji are Copyright 2016 Twitter, Inc and other contributors, CC-BY 4.0. The graph created with help from emoGG (github.com/dill/emoGG)*

To evaluate the randomness of each participant's chosen arrangement, a 'randomness metric' was created. The 15 locations in which the bricks could be placed were numbered 1 to 15. The absolute distance between each pair of bricks of the same colour was calculated and squared. The sum of these squared distances was then a metric of the randomness of the chosen arrangement. This metric was used to capture the assumption that a ordered arrangement would involve placing bricks of the same colour close together, whereas for a random arrangement the bricks would, on average, be further apart.

The distribution of the randomness metric can be seen in Figure 4.12, where lower values indicate a less random arrangement (the ordered arrangement used for the primary task would have a score of 30 on this metric). As Figure 4.12 shows, the majority of participants chose a relatively random arrangement. In addition, the mean randomness of the chosen arrangements for the low (Mean = 504.43, SD = 278.58) and high (Mean = 568, SD = 194.78) WM groups was very similar, so this comparison was not analysed further.

With regards to overall performance in the free condition (M = 0.76, SD = 0.11), this was very similar to performance in the ordered (M = 0.75, SD = 0.10) and random (M = 0.75, SD = 0.11) conditions. When considering the impact of participants' chosen arrangement on task performance, the randomness metric did not relate to performance in the free arrangement condition (odds of 10:1). There was some support for a negative relationship between randomness metric and overall performance in the ordered and random conditions, averaging across condition (odds of 54:1) such that participants who performed better overall in the ordered and random conditions chose a less random arrangement in the free arrangement task. However, the predicted change of accuracy for a 1 SD increase in the randomness metric was only -0.017. The randomness metric also did not interact with condition (posterior odds = 18:1).

Finally, the relationship between the randomness metric and performance was

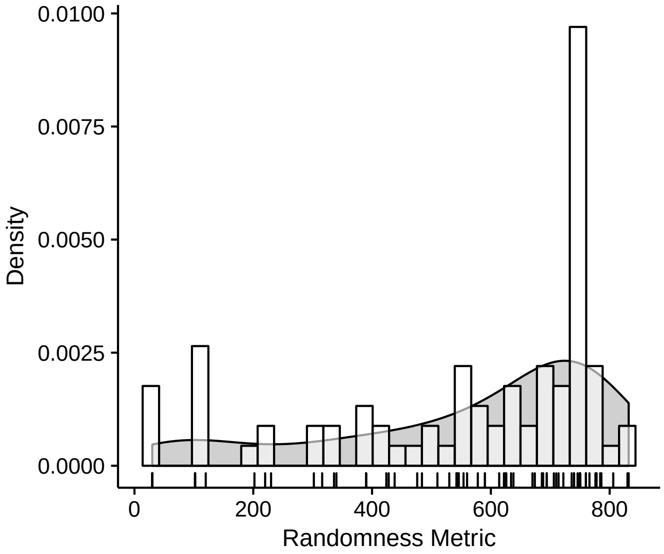


Figure 4.12: *The distribution of scores on the randomness metric with overlaid density estimate.*

investigated with a model where the metric was the outcome and performance in each condition, as well as the additional WM measures, were included as predictors. A negative relationship between performance in the ordered condition and the randomness metric was supported by odd of 34:1 (β = -0.29, 95% credible interval = -0.55 – -0.04). All other predictors were supported by odds of 18:1 or less, as Table 4.8 shows.

4.3.3 Discussion

We replicated the interaction with WM observed in Experiment 1 whereby children with low WM children performed worse in the random condition. This reinforces the idea that task-relevant structure in the environment (Kirsh, 1995) and the affordances of task materials (Macken et al., 2015) can be utilized to support performance. By 'off-loading' some of the processing requirements of the task onto the environment, participants with a limited WM capacity have more resources available to support storage. This result may be useful to inform how tasks are adapted to reduce the load placed on children with poor WM (Alloway et al.,

Table 4.8: *Parameter estimates for a model predicting the randomness metric in Experiment 2. All parameters except the intercept are standardised coefficients.*

Parameter	Mean	95% CI	Odds	ESS	R-hat
Intercept	0.000	-0.213, 0.213	1:1	20000	1
Free	-0.185	-0.509, 0.133	7:1	20000	1
Ordered	-0.294	-0.597, 0.01	34:1	20000	1
Random	0.131	-0.25, 0.516	3:1	20000	1
FDR	0.310	-0.062, 0.681	18:1	20000	1
BDR	-0.088	-0.397, 0.221	2:1	20000	1
Corsi	-0.107	-0.388, 0.172	3:1	20000	1
Odd-one-out	0.071	-0.21, 0.351	2:1	20000	1

[1] Mean refers to posterior distribution
[2] 95% CI is the 95% Bayesian credible interval
[3] Odds are the posterior odds
[4] ESS = effective sample size

2008; St Clair-Thompson et al., 2010). In addition to changing how material is presented, support could also be provided through changing the structure of how material must be recalled. This also has implications for a more integrated (Amso and Scerif, 2015) or ecological (Moreau and Conway, 2014) approach to training WM. This experiment did not replicate the overall effect of condition for the average participant. This overall effect was small in Experiment 1, so, taken together, these experiments suggest that task-relevant organization specifically benefits children with low WM.

The majority of participants rated the ordered condition as easier, but did not choose an ordered arrangement in the free-arrangement condition. This supports the idea that children show awareness of task difficulty before they can effectively engage in strategies to facilitate task performance (Chevalier, 2015a). However, it is important to note that children with average or high WM did not actually perform better in the ordered condition. One possible explanation is simply that the random condition is not particularly difficult for these children, making it harder to find performance differences between the two conditions. An alternative explanation is children may have been able to respond more quickly in the ordered condition, resulting in an experience of lower cognitive effort that is not captured by an accuracy performance metric. Future work might therefore examine the impacts of organization on both response accuracy and latency.

In contrast, children with low WM did perform better in the ordered condition, and so should clearly benefit from choosing an ordered arrangement in the ‘free-arrangement’ condition. However, as with the majority of the participants, children with low WM did not chose to structure the bricks in an ordered fashion before completing the task. This conclusion is, of course, contingent on the randomness metric used. It could be that children ordered the bricks in some sense, just not in the same way as our ‘ordered’ condition. Interpretation is also complicated by the fact we asked children to arrange the brick ‘how ever they like’ rather than to ensure maximal performance. Future work could emphasise performance in a free arrangement condition and ask children to explain their chosen arrangement. When asked to rate which condition was easier, low WM children did not discriminate between the conditions, with just over half rating the ordered condition as easier. This supports the idea that children with poor WM have less well developed metacognitive skills, leading to poorer judgements of task difficulty, limited understanding of facilitative strategies, and lack of ability to utilize such strategies effectively (Alloway et al., 2009a; Chevalier, 2015a).

4.4 General Discussion

This study investigated the role of task-relevant perceptual organization in WM, and found that, in two experiments, task-relevant arrangements improved performance for children with low WM. In addition, participants in Experiment 2 rated the difficulty of ordered versus random arrangements. Despite performing better in the ordered condition, children with low WM did not consistently rate this condition as easier, nor did they choose an ordered arrangement when allowed to arrange the bricks themselves before completing the task. There was some weak indication that overall performance in the ordered condition predicted the randomness of a participants chosen arrangement.

The consistent advantage of an ordered arrangement for children with low WM affords a number of explanations. One possibility is that this is driven by difficulty with visual search and action planning (Holmes et al., 2014; St Clair-Thompson, 2011). Additionally, grouping the environment in a task-relevant manner may allow participants to offload processing functions that would otherwise need to be performed (Dunn and Risko, 2015; Gilbert, 2015; Risko and Gilbert, 2016), which might particularly benefit children with fewer available cognitive resources (e.g. low WM children). If performance on a given tasks emerges from the interaction of task-specific factors and the capabilities of an individual (Macken et al., 2015) then adapting task factors may be particularly important for participants with a more limited 'repertoire' of skills. Risko and Gilbert (2016) also discuss the importance of metacognition for successful offloading. Appropriate strategy choice involves awareness of the internal resources available and how external aids can be successfully employed to improve performance.

In both experiments, high WM children performed similarly in the ordered and random conditions. A speculative explanation for the absence of an arrangement effect is that they are able to focus in on a subset of the bricks, perhaps reducing response times. For example, the first six bricks from the left for the random arrangement in Experiment 1 (Figure 4.1) include all the colours needed to response to sequence lengths up to four. If some participants were responding in this way it might suggest a role of individual difference in the formation of the objects of WM (Macken et al., 2015). The skills and competencies that contribute to WM vary by individual such that the objects of memory vary by individual. Here the 'objects of memory' would result from the interaction between the presented sequences, the bricks used for recall, and the capabilities of the individual. For some participants focusing on a subset of the bricks could mean that a simpler set of responses can be planned using that subset. Under such an account different individuals are, in some important sense, not remembering the same thing when presented with the same materials, due to individual differences in the approach they take to the task. Some participants may simply be remembering a sequence of colours names, while others might be remembering a sequence of pre-planned actions on a subset of the bricks.

One outstanding question concerns why participants (with average or high WM) judged the ordered condition to be easier, despite performing similarly to the random condition. It could be that a difference in response time afforded by the two conditions underpins this judgement. Retrospective judgements of difficulty are not only influenced by performance but also prior judgements of the effort or time required to complete a task (Efklides et al., 2006). If an ordered arrangement affords faster responses then this could underpin the observed difficulty judgements in a way that would not be reflected in an accuracy performance measure. This possibility could be tested by presenting the task on a tablet computer or by using motion capture technology. Presenting the task on a tablet screen might introduce concerns about the ecological validity of using virtual objects. However, it would bring benefits in terms of more fine-grained control over stimulus presentation.

4.4.1 Relationship to previous work

While some research has investigated the influence of spatial configurations on WM performance in adults (Morey et al., 2015; Woodman et al., 2003), the focus has been on the spatial properties of visual arrays. The work presented here differs by using a task that combines verbal storage with visuospatial recall where the organisation is applied to the response method not the memoranda. Organising the memoranda themselves would involve contrasting sequences of, for example, “red, red, yellow, blue, blue, green” with sequences of “red, blue, yellow, blue, red, green". That is, the grouping by colour would be applied to the presented to-be-remembered sequences themselves. In our study the same material is, arguably (see below), being encoded into WM in both conditions, but it is being recalled differently. The subtlety of this manipulation relative to previous work might explain why the observed effects of condition (ordered vs. random) were relatively small. Nevertheless, consistent evidence for a beneficial effect of organisation was observed for low WM children, highlighting possible avenues to supporting children with low WM. This is particularly important given the broad ranging negative consequences of WM difficulties (Alloway et al., 2009a; Holmes et al., 2014) and the failure of WM training to achieve far transfer (Melby-Lervåg and Hulme, 2013; Sala et al., 2017).

Despite the verbal sequences being presented in this task, organization may still play a role in how the sequences are represented in WM as the response bricks were visible at encoding. As the sequence is read out participants may begin planning a series of actions in preparation to recall the sequence, which are represented in WM. In the random condition, this would involve selecting individual bricks to pick up. In the ordered condition, it could involve simply encoding regions of space where the bricks of the required colour were arranged. Extending this work by manipulating the presence of the response-bricks at encoding could interrogate this possibility.

Some support for the role of automatic action planning in WM comes from previous work on following instructions (Allen and Waterman, 2015; Waterman et al., 2017; see also Jaroslawska et al., 2016). Allen and Waterman (2015) found that there was no additional benefit from acting out sequences at encoding when participants were preparing for enacted recall. Where participants were required to verbally recall sequences enactment at encoding did boost performance. This suggests that motoric codes are recruited when planning for enacted recall resulting in no additional benefit for explicitly acting out a sequence at encoding. In more recent work a benefit for enactment at encoding on enacted recall has been demonstrated with less complex stimuli (Jaroslawska et al., 2016; Waterman et al., 2017). This finding could contribute to an explanation of why we did not observe an overall effect of arrangement in Experiment 2, where five colours were used. The complexity of having five possible groupings of bricks to act on in the ordered condition may have removed the benefit for an ordered arrangement.

The approach explored here could complement the task-analysis approach to supporting children with poor WM (Gathercole et al., 2008; St Clair-Thompson et al., 2010), though it would need adapting to have practical applications. Presenting verbal sequences of actions is analogous to a number of classroom activities (Gathercole et al., 2008). Thus, thinking about the way that objects are organized in space offers avenues to supporting children with poor WM in the classroom. However, as the results of our metacognitive questions demonstrate, it would not be sufficient to ensure that the task environment has been adapted. Children would also need the metacognitive awareness of this change and the strategies it affords (Schneider, 2008; Chevalier, 2015a). Embedding task analysis, environmental adaptation, and strategy development within training programs might also be key to achieving far transfer (Amso and Scerif, 2015; Dunning and Holmes, 2014) and supporting children with low WM.

4.4.2 Extending and amplifying the effect

Initially the aim of this work was to investigate the role of structure in the environment in a controlled setting where the numerous factors that contribute to real-world performance were removed. By considering related experimental work and real-world performance several possibilities for extending and, critically, increasing the magnitude of the effect present themselves. One possibility would be to use more distinct spatial groupings as objects are not typically neatly arranged in rows in real-world situations. This would more closely reflect how people group objects in real-world settings (e.g. Beach, 1993). Figure 4.13 shows

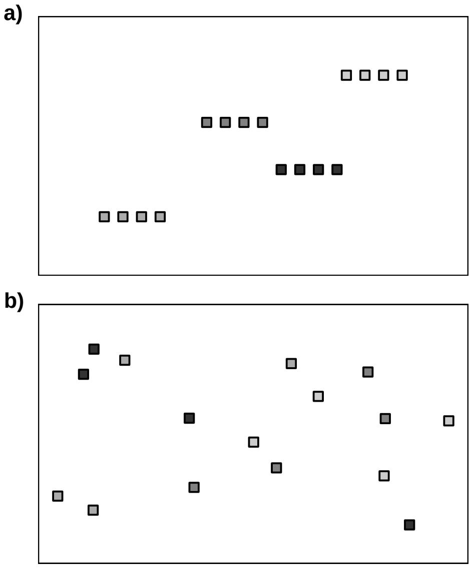


Figure 4.13: *Two possible arrangements for the ordered (a) and random (b) conditions with more distinct spatial groupings. Shades of grey represent different colours.*

an example of the more distinction spatial groupings that could be used.

One complication for the groupings in Figure 4.13 might be the cost in terms of reaching time for picking up a brick from the far top grouping. An alternative adaptation of the arrangements could be made by introducing an element of interference in the random condition. Objects with multiple features could be used such that irrelevant distracting features are introduced. In Figure 4.14a the ordered arrangement uses objects where shape and colour are always congruent – all the squares are one colour, all the triangles are another colour. The random arrangement in Figure 4.14b includes coloured objects that vary in shape. This irrelevant shape information may make searching for the appropriate object in the random condition more difficult, accentuating differences between the conditions.

Aside from adjusting the particular spatial arrangements used, it would be useful to apply this manipulation to a more naturalistic setting using everyday objects such as stationary items. The choice between naturalistic and abstract materials has been shown to be important for the emergence of beneficial effects of enactment in the following instructions literature (Jaroslawska et al., 2016;

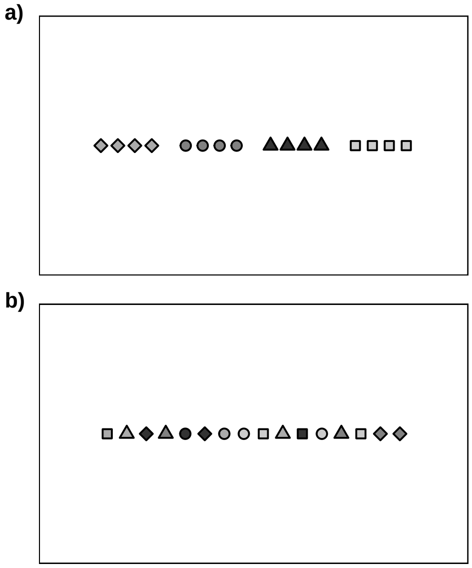


Figure 4.14: *Two possible arrangements for the ordered (a) and random (b) conditions with additional shape features. Shades of grey represent different colours.*

Waterman et al., 2017). Benefits for particular strategies are only observed when more naturalistic objects and actions are used.

In addition to the meaningfulness of the materials, using less abstract groupings might prove useful to extending the effect observed here. That said, grouping objects by colour is certainly meaningful, particularly for children. When adults make use of spatial grouping in naturalistic settings, objects are grouped to reflect their immediate relevant to the task at hand and shared meaning (Beach, 1993). Beach observed that bartenders grouped glasses of the same type together as this often signals they share ingredients within an order. This grouping behaviour was present even when identical black glasses were used; subjects still grouped glasses as if they were different, leaving space between groupings (Beach, 1993). Children also demonstrate a sensitivity to using space to represent category from before 2 years of age, reflected in grouping objects (Namy et al., 1997; Smith and Sheya, 2010) and touching behaviour (Mandler et al., 1991; Rakison and Butterworth, 1998). For example, Namy et al. (1997) observed the spontaneous grouping behaviour of infants longitudinally between 16 and 21 months when presented with set of toys from two clearly distinct categories (e.g. brightly coloured people vs. green plates). The frequency of spatial groupings increased over the period with all participants forming complete groupings of one kind of object by 19 months old. In extending the current paradigm, more familiar objects such as pens and rulers could be combined with stationary organizers children use in the classroom (pen pots etc.) to create groupings that are more familiar for participants (Yang et al., 2014, 2015a,b).

One way to understand the role of familiar objects and groupings is that they offer a route to leveraging LTM in support of WM. Such links between LTM and WM have been investigated in the context of expertise. For example, chess skill is related to recall of chess positions so long as the positions used are plausible rather than random (Chase and Simon, 1973; Connors et al., 2011; but see, Holding and Reynolds, 1982). In both cases the objects (chess pieces) are equally familiar but the way they are arranged varies in familiarity. The importance of plausible arrangements has also been demonstrated for the maps used in complex strategic video games (Bonny and Castaneda, 2016). 'Bootstrapping' effects in WM also capture the interactions between LTM and WM performance (Calia et al., 2015; Darling and Havelka, 2010; Darling et al., 2012; Darling et al., 2014; for a review, Darling et al., 2017). The core effect involves the finding that short-term recall is supported by the presentation of digits within a familiar keypad display. The effect has been demonstrated in children as young as 9 years-old (Darling et al., 2014) and suggests a route to 'leveraging' long-term memory in support of short-term memory recall that may be applicable to using spatial arrangements to support WM.

One possible limitation of the present paradigm was that grouping objects was not made sufficiently motivating. Adults have been shown to form task-relevant groupings in space when motivated to maximize performance by linking it to a food reward (Solman and Kingstone, 2017). Adults were presented with a task where goals were completed by moving and selecting objects within a virtual space. Participants were free to arrange these objects across the course of a session. Over time participants organized objects within this space based on their goal relevance. In a second experiment, where the given goal was more structured such that participants could choose to focus on a subset of the task objects , spatial organization was particularly pronounced, with high performing participants structuring the task environment more than lower performing participants (Solman and Kingstone, 2017). It may be that the effect of spatial organisation on WM performance would be accentuated in a situation where participants are more strongly motivated to group the objects. For example, rather than telling participants to group bricks ‘how they like’, the importance of maximising performance could be emphasized at this stage.

4.4.3 Conclusions

We observed consistent benefit to structuring the environment for children with low WM. This finding highlights possibilities for supporting children with poor WM in the classroom. A number of related lines of research within and outside of WM suggest ways to bolster this basic effect to maximize potential real-world benefits. The task could also be adapted to allow a more detailed investigation of the mechanisms underlying the effect.

Despite task-relevant arrangements in the environment supporting performance for low WM children, they were less likely to rate the ordered condition as easier than the random condition and, like the majority of participants, did not choose to organise the bricks when given the opportunity to do so. This disconnect between objective measures of performance and metacognitive evaluations of difficulty is critical to consider when aiming to support low WM children. Teaching strategies may not benefit those children most in need (i.e. low WM children) if they lack the relevant insight into task difficulty.

Chapter 5

General Discussion

This chapter will begin with a recap of the key findings of the previous chapters. While the work in the previous chapters was not primarily aimed at testing the predictions of different models of WM, a number of general implications for development and WM in children will be drawn out here. Finally, a range of future possibilities for supporting WM in children are discussed, in light of the results presented in this thesis.

5.1 Thesis overview

The aims of this thesis were to develop a set of measures that could be used to rapidly assess working memory in primary school aged children, and to investigate different ways that WM performance can be supported in children.

5.1.1 Chapter 2

Chapter 2 described the development of a computerised set of WM measures. This work was driven by our involvement with the *Born in Bradford* cohort study

(Wright et al., 2013), a longitudinal study following thousands of children from birth onwards in Bradford, UK. For this project a set of cognitive tasks were required that were quick to administer while remaining good quality measures of the constructs of interest. Four measures were presented in Chapter 2 that capture the core features of working memory: (i) forward digit recall, a simple verbal measure; (ii) backward digit recall, a measure of complex verbal WM; (iii) Corsi block recall, a simple visuospatial WM measure; and (iv) odd-one-out, a complex visuospatial WM task. Except for some floor effects with young children on odd-one-out, all the measures were set at the appropriate difficulty for the target ages. The validity of the measures was tested in Chapter 2 using a range of criteria, all of which were met. Firstly, the expected effects of serial position and sequence length were observed; performance decreased for longer sequences and for later positions within a sequence. Performance on the measures also improved with age, particularly for the complex measures. Additionally, the expected relationships between the tasks were observed, with stronger relationships within than between modalities. Finally, the measures predicted academic attainment in line with predictions drawn from the literature. Complex measures were more strongly related to maths ability, whereas verbal measures predicted reading performance.

The modality effects in Chapter 2 – where relationship within modalities were larger than between – would be predicted by a multiple component model of WM, which argues that storage is served by specialised modality-specific stores (Baddeley, 1986, 2007; Logie, 2011). That said, the relationships could also be predicted by an account that conceptualises WM in terms of domain-general storage (Cowan, 2001, 2005). The stronger relationships within modalities would be predicted due to the overlap in peripheral processes for, say, forward and backward digit recall. The set of processes that two verbal tasks tap into have greater overlap than the set of processes involves in completing a verbal task and a visuospatial task. This approach to understanding modality effects within the embedded processes model is exemplified by Gray et al. (2016), who include a component for ‘specific verbal processes’ in their structural modelling.

One of the lessons from Chapter 2 was that computerising measures offers a way to carry out rapid group assessment. While this point may seem obvious, a number of expensive commercial batteries still use paper booklets and manual scoring (Wechsler, 2003). The use of on-screen response boxes, rather than verbal recall, in Chapter 2 further facilitated large-scale group testing. Importantly, the verbal tasks in Chapter 2 showed the same response profile as measures using verbal recall, suggesting that using on-screen boxes to collect non-verbal responses does not undermine the verbal nature of the tasks. Chapter 2 also shows the benefit of item-level scoring, something that has been previously demonstrated (Conway et al., 2005), but is not often implemented. When using a span procedure, even with some partial credit, a task only has ~10 possible scores. In contrast, by scoring at the item-level a task can have approximately 50 to 70 possible scores. The increase in between-subject variability this affords is critical for developing individual difference measures for use in projects such as *Born in Bradford* (Wright et al., 2013).

5.1.2 Chapter 3

Chapter 3 described the first investigation of children’s ability to prioritise serial positions in visual WM. Across three experiments children did not show the ability to prioritise a particular item in a sequence, when instructed to do so. The question still remains whether this is something children are unable to do, or simply are not inclined to do within our particular paradigm. Assuming children are unable to prioritise, limitations in executive control would seem to offer a plausible explanation of this. While primary school aged children show the ability to flexibly allocate attention in some settings, performance remains limited (Cowan et al., 2010; Shimi et al., 2014). The prolonged development of the frontal lobes, linked to executive control (Smith and Jonides, 1999), means that adult-like performance on all tasks should not be expected until late adolescence (Jurado and Rosselli, 2007). By identifying the limitations of executive control in children, the work in Chapter 3 highlights the fact that simply 'paying attention' is not achievable for children in all contexts. This is important to consider in developing recommendations for strategies children might be able to employ to support WM.

One useful way to frame the work in Chapter 3 is in terms of proactive versus reactive control (Braver, 2012; Chevalier et al., 2014; Chevalier, 2015a). Proactive control describes an approach to a task where participants prospectively plan and prepare future responses. Reactive control, in contrast, involves merely responding the whatever is currently presented when completing a task. The prioritisation paradigm in Chapter 3 requires participants to take a proactive approach to the task: the process of prioritising the item must be done, at least in part, at encoding. Indeed, even prioritising the item during a maintenance interval involves proactive planning beyond merely reacting to what is currently presented. It would not be possible to prioritise a serial position with a reactive strategy where all items are encoded and maintained equivalently, with only the recall probe driving responses. While children aged 7 years do demonstrate proactive control on some tasks (Chevalier et al., 2014), these abilities are sensitive to task factors and undergo prolonged development (Chevalier, 2015b; Chevalier et al., 2015). Previous tasks demonstrating a degree of attentional flexibility in children (Cowan et al., 2010; Shimi et al., 2014) arguably required less complicated proactive control strategies than prioritisation. For example, Shimi et al. (2014) showed that 7 year-olds are able to boost performance for a cued item, as well as ignoring an uninformative cue. In that study, the cues were 100% valid, meaning participants could allocate all their attention to a cued item. In contrast, with prioritisation participants must divide attention between the three items such that more, but not all, of it is allocated to the prioritised position. This more complicated allocation of resources between multiple items likely represents a more difficult task.

Chevalier (2015a) argue that metacognitive factors are also important to the application of different attentional control strategies in children. It is not the case that a monolithic shift to proactive control occurs which immediately transfers to all tasks (Chevalier, 2015b; Siegler, 1994, 2007). Rather, children must also develop an understanding of the set of strategies available to them, as well as the tasks to which they can be applied (Chevalier, 2015a,b). This approach allows for a more nuanced treatment of the concern that the participants in Chapter 3 were not sufficiently motivated or ‘inclined’ to prioritise. Even if this was the case, it would represent an important difference between adults and children in that adults choose to prioritise under near-identical conditions. At times humans are required to engage in cognitively effortful tasks where the appropriate strategy is not always transparent. Thus, developing these metacognitive abilities alongside attentional control is central to the optimisation of behaviour with development.

In addition to highlighting the limits of visual WM in children, Chapter 3 presented a novel individual difference analysis investigating the automaticity of recency effects in WM. Recency effects are commonly observed in visual WM (Allen et al., 2006; Burnett Heyes et al., 2012; Hu et al., 2014, 2016; Walker et al., 1994) and are thought to be partially automatic, representing a boost to the most recently presented material that does not draw on executive resources (Allen et al., 2014). In Chapter 3 this suggestion was investigated by estimating the relationship between a set of WM measures and performance at the first, second, and third serial positions of the primary visual task. In line with the proposed automaticity of recency effects, the additional WM measures were related to performance at the first two serial positions, but not the final position. This analysis extends previous work with adults to an individual difference setting with children, providing strong convergent evidence. Thus, Chapter 3 provided evidence for a large and possibly automatic recency effect, alongside the absence of a more deliberate and effortful prioritisation effect.

5.1.3 Chapter 4

In light of the absence of prioritisation effects in Chapter 3, an alternate approach was taken for Chapter 4. Rather than focusing on the way that cognitive faculties can be leveraged to support performance, Chapter 4 investigated how the task environment might support performance. A novel paradigm was developed in which participants were required to recall verbal sequences of colours by picking up coloured blocks that were either arranged pseudorandomly or grouped by colour. It was found that structured colour groupings in the environment consistently boosted performance for low WM children. This finding highlights the way in which the spatial features of the task environment can be utilised to support performance in some children. Low WM children, who may find even simple verbal storage taxing, benefited from 'offloading' (Dunn and Risko, 2015; Risko and Gilbert, 2016) some of the processing requirements of the task on the environment.

In the second experiment of Chapter 4 the role of metacognition in the effect of task-relevant groupings was investigated, identifying further possibilities and complications in supporting WM in children. Overall children rated the ordered arrangement as easier, though few chose to arrange the blocks in this way, when given the opportunity to do so. In addition, low WM children did not consistently rate an ordered arrangement as easier than a random arrangement, despite being the group that consistently benefited from structure in the task environment. These findings highlight the importance of metacognitive factors in any attempt to support WM in children. It is insufficient for some strategy or manipulation to benefit children if they do not recognise this and spontaneously employ the strategy. This complication could be addressed either by building some strategy into the materials and task, or by attempting to train metacognitive awareness. Given the metacognitive limitations children exhibit (Schneider, 2008), designing a task or set of materials to automatically elicit some strategy might be more tractable. The importance of metacognition to understanding the results of Chapter 4 reflects the importance it is given in a number of interventions for children with developmental disorder. For example, the Cognitive Orientation to Daily Occupational Performance intervention involves teaching children general cognitive strategies that allow them to identify and monitor task-specific strategies (Polatajko et al., 2001; Rodger and Brandenburg, 2009).

The work in Chapter 4 was inspired by embodied theories of cognition (Clark, 2008, 2016; Shapiro, 2010; Wilson, 2002). All embodied theories share the view that the importance of the body and environment to cognition should be emphasised. Rather than test specific predictions of one theory, Chapter 4 aimed to show how this broad perspective could be usefully applied to supporting WM in children. Inasmuch as Chapter 4 showed the importance of the organisation of the task environment to performance, the results cohere with an embodied account of cognition. However, the predictions made in Chapter 4 would likely also be made by other approaches – cognitivist accounts of cognition could also say that the random condition should be more difficult due processing load of searching for a block. The results in Chapter 4 also extend recent work investigating the influence of action processing on WM performance when following instructions (Allen and Waterman, 2015; Jaroslawska et al., 2016; Waterman et al., 2017; Yang et al., 2014). Chapter 4 shows that action processes can also influence performance where the complexity of the action participants are required to complete is manipulated. For the ordered arrangement in Chapter 4 the set of possible actions was smaller than for a random arrangement, potentially underpinning the effect.

Chapter 4 also captures the idea of offloading (Dunn and Risko, 2015; Risko and Gilbert, 2016) in a WM task with children. When the blocks are grouped by colour participants are able to 'offload' visual search and action planning on the environment meaning more resources are left for storage. This offloading effect was isolated to low WM children, perhaps because the task was not sufficiently tasking for other children to benefit from offloading processing. If so, then an offloading account of performance would predict that all participants should benefit from an ordered arrangement if the demand of the task was increased.

5.2 Future directions for supporting working memory

5.2.1 Extending investigations into prioritisation effects

From Chapter 3 it seems that the possibility of engaging executive attention to focus on the most important information in a situation is limited as a strategy for children. However, it may be that with the appropriate adaptations of the task children would be able to prioritise as adults can. Perhaps children cannot prioritise positions in a visual sequence because the basic task is too attentionally demanding for any resources to be ‘left over’ for prioritisation. Support for this possibility comes from the finding that sequential visual WM is more demanding in adults (Gorgoraptis et al., 2011), and from the attentional flexibility children show with simultaneous displays (Cowan et al., 2010; Shimi et al., 2014). Future developments of this task could, therefore, consider how the basic task can be made easier such that children will exhibit an ability to prioritise. The aim of such investigations would be more about delineating the boundaries of executive control in children than developing practical tools to supporting WM. One potential way to reduce the load of the primary task (and therefore make it easier) would be to use more familiar objects (e.g. Shimi et al., 2013). This may result in performance being supported by structures within LTM. In addition, children may be more inclined to prioritise stimuli that are meaningful, such as real-world objects or cartoon characters. The task could also be made more motivating by gamifying it, and linking performance to a reward.

The absence of prioritisation effects in Chapter 3 could be the result of the fact that the prioritisation position was only probed on 1/3 trials. Children may feel that prioritisation is ‘not worth it’ when the relevant item is probed this infrequently. This possibility could be explored by manipulating probe frequency for the prioritised serial position. The role of probe frequency is particularly important to consider as children of similar ages to those tested here have been shown to alter their allocation of attention in response to manipulations of probe frequency (Cowan et al., 2010; Shimi et al., 2014). Shimi et al. (2014) found that participants ignored a retro-cue if it was only 50% predictive of which item would be probed. One possible result of manipulating probe frequency for the paradigm in Chapter 3 is that it will have an effect independent of prioritisation. Children may respond to probe frequency such that recall is improved, while prioritisation effects remain absent. Another possibility is that higher probe frequency would elicit prioritisation effects; participants would feel, under high probe frequency, that prioritisation was 'worth it'.

The prioritisation task described in Chapter 3 could also be fruitfully extended by using simultaneous presentation. If sequential presentation is more tasking then children may be more inclined or able to prioritise with a simultaneous array. Contrasting sequential and simultaneous visual presentation would also have real-world implications for how information is presented.

5.2.2 Expanding on effects of task-relevant groupings

As discussed in Section 4.4, there are a number of possibilities for amplifying the effect observed in Chapter 4. Some of the effects in the following instructions literature (Allen and Waterman, 2015; Jaroslawska et al., 2016; Waterman et al., 2017; Yang et al., 2015b) would be useful to investigate in combination with manipulations to the organisation of objects in the environment. The required recall method could be manipulated to explore the role of action processes in effects of object organisation. If the effect results from the simplified action space afforded by grouping objects, then this would likely interact with whether participants are required to verbally recall the sequence, or by moving blocks. If the mechanism for the organisation effect is limited to perceptual-motor processes at recall then it should disappear with verbal recall.

Acting out a sequence of instructions at encoding has been shown to benefit subsequent recall, for at least some sets of materials (Jaroslawska et al., 2016; Waterman et al., 2017). This factor could be investigated with the task described in Chapter 4 by requiring participants to pick up a block for each colour as the sequence is encoded. One possible result of such a manipulation would be to remove the difference between the ordered and random conditions: participants would select a series of blocks to pick up in the random condition at the encoding stage, removing the need to search among the blocks and select a set of actions at recall. That said, enactment at encoding might be more demanding in the random condition, interfering with performance. If this prediction was confirmed it would demonstrate how pre-planning actions offers a way to reduce the load of some task, where organisation in the environment is not an option.

Demonstration has also been shown to benefit future recall of a sequence of instructions (Yang et al., 2015b). There are multiple ways in which demonstration could be applied to the task in Chapter 4. Firstly, the experimenter could pick up the blocks required to recall a sequence at encoding. When this manipulation is used with following instructions, performance is improved compared to hearing or reading a sequence of instructions (Yang et al., 2015b). Demonstration could also be investigated in a situation where participants choose their own arrangements for the blocks. Here the question would be whether an experimenter demonstrating using an ordered arrangement would influence future choice by participants.

As noted in Chapter 4, the objects used to complete the task could also be manipulated. A more naturalistic setting could also be used, to better resemble the classroom environment. For example, items of stationary could be used alongside various containers commonly found in classroom (e.g. Yang et al., 2015b).

5.2.3 Alternative routes to supporting working memory

Bootstrapping. Bootstrapping effects in WM (Calia et al., 2015; Darling et al., 2012, 2014, 2017) provide an example of leveraging LTM structures to support performance. When participants are presented with visual sequences of digits embedded within familiar keypad displays, performance is superior to seeing isolated digits. This bootstrapping effect has been demonstrated in 9 year-old children (Darling et al., 2014) but could be further investigated in the context of supporting WM. One question could be whether the bootstrapping effect relates to individual differences in working memory. Research could also investigate whether a bootstrapping effect can be 'trained' over time – would repeated exposure to a particular novel way of presenting information result in a bootstrapping effect over time? For example, if the some novel spatial configuration akin to a keypad was consistently used to present information, would it, over time, benefit performance compared to presenting stimuli in isolation? This could have implications for the importance of consistency in presenting information of the same type.

External aids. While making use of external aids is recommended to support low WM children (Gathercole and Alloway, 2008), there has been limited research systematically investigating this topic. In adults, writing down sequences has, unsurprisingly, been shown to benefit future recall (Risko and Dunn, 2015). Risko and Dunn (2015) also showed that imagining having the opportunity to write information down resulted in higher estimates of performance from participants. In one of Risko and Dunn (2015)'s experiments participants were given the option of whether or not to write down sequences of digits. This aspect could be particularly interesting to investigate in children given the disconnect between metacognitive judgements and objective performance observed in Chapter 4. It could be that children would not choose to write down sequences as often as adults, due to making inaccurate estimates of their future memory performance. Additionally, it may be that low WM children are particularly unlikely to spontaneously employ a strategy that would be beneficial. Risko and Dunn

(2015) also included a condition where participants were *required* to write down sequences, rather than being given the choice. In children such an instruction would allow for the investigation of whether external writing aids particularly benefit some children.

Teaching strategies. Exploring the possibility of teaching WM strategies could represent a fruitful line of future research into supporting WM skills in children. Strategy instruction has been shown to be beneficial for adults, particularly those with low WM (Turley-Ames and Whitfield, 2003). Working memory training programs appear to encourage the generation of domain-specific strategies that do not spontaneously transfer to other tasks (Dunning and Holmes, 2014; Moreau and Conway, 2014). In contrast, explicitly training more general strategies may be critical to supporting WM in children, and potentially complimenting training (Amso and Scerif, 2015; Dunning and Holmes, 2014). Elliott et al. (2010) found limited evidence for the idea that training teachers to encourage strategy use and adapt tasks improved academic attainment in a school setting. It may be that children themselves need to be taught relevant strategies, rather than strategies being taught to teachers.

Presentation. The WM literature is replete with examples of how manipulating the way in which information is presented either impairs or supports performance. Some of these experimental effects could be explored in the context of supporting WM performance in children. With verbal WM a range of perceptual properties of stimuli and sequences affect performance (Jones et al., 1997). Macken et al. (2015) argue that if sequences have certain properties then they are remembered as unified perceptual objects, rather than collections of items. This means that participants are able to recall more information than is possible with other sequences. For example, Macken et al. (2015) discuss the importance of spatial cues in verbal STM, where sounds from the same source are grouped together into a unified perceptual object. Thus, spatial cues can be employed to encourage particular segmentations of a stream of verbal stimuli (Hughes et al., 2016; Jones et al., 1997). A similar way to frame this idea is that verbal information can often be presented in a way that encourages chunking of items. If the range of factors that contribute to children's likelihood of chunking verbal sequences could be identified, then this would offer an avenue to supporting WM performance.

Within the visual domain, a range of spatial characteristics of stimuli seem to affect performance. For example, the classic Gestalt grouping principles include proximity, colour, and similarity in size, among others (Palmer and Rock, 1994). Woodman et al. (2003) investigated the influence of grouping and lines connecting stimuli ('connectedness') on performance in a visual WM task. Accuracy was higher for the recall of objects within the same spatial grouping as a cued item, compared to objects outside of the grouping at the same distance from the cued item. In a second experiment a similar effect was observed for groupings defined by connectedness. Indeed, the effect of connectedness overrode the effect of spatial groupings. Brunetti et al. (2016) showed that linking items together within a visual sequence, using spatial cues, boosted performance on a Corsi block recall task. Participants were presented with two versions of a Corsi task, one where blocks lit up as standard, and another where each item was followed by a display with a line linking that location to the next location that would light up. Connecting locations within a sequence with these trajectories resulted in better performance (Brunetti et al., 2016).

Taken together previous research demonstrates the importance of grouping in perception and memory for performance. Various factors clearly influence the formation of the 'objects of memory' (Macken et al., 2015), but, in general, information that will be remembered together should be grouped together, as much as possible. Work would need to be done first to establish these effects in children, before extending them to more practical applications. By manipulating the affordances of task materials, it may be possible to automatically elicit particular beneficial strategies by presenting information in a given way.

5.3 Conclusions

This thesis has shown how computerised measures of WM can be developed that allow for effective group testing in a school setting. Once children's WM has been effectively assessed one can turn to the question of how children with poor WM can be supported. In Chapter 3 the possibility of engaging executive attention to focus on a subset of a sequence was investigated. Ultimately, this was not something that children demonstrated an ability to do, possibly due to the under-development of executive functions. For Chapter 4, the potential for task-relevant structure in the environment to support performance was investigated. Across two experiments, structure in the environment was consistently beneficial for children with low verbal WM. Despite the consistent benefit of this manipulation for low WM children, they were less likely to recognise that structuring the task environment made the task easier, compared to other children. The importance of the task environment itself, in addition to the materials and task that have to be completed, represents a novel future avenue to supporting WM in children.

Appendix: Software used for this thesis

A range of software has been used to create this thesis. This appendix details this software and what it was used for.

Analysis. All the analysis was carried out in R (R Core Team, 2017). The Bayes Factor analysis in Chapter 3 was carried out using the `BayesFactor` package (Morey and Rouder, 2015). The ANOVAs in Chapter 3 using the `ez` package (Lawrence, 2016), and `effsize` was used to calculate some effect sizes in that chapter (Torchiano, 2017). The `broom` package was used to assist with manipulating the outputs of the analyses in Chapter 3 (Robinson, 2017).

The Bayesian analyses in Chapters 2 and 4 were carried out using `rstanarm` (Gabry and Goodrich, 2017). The functionality of `rstanarm` is supported by the `Rcpp` and `Matrix` packages (Bates and Maechler, 2017; Eddelbuettel et al., 2017).

Data manipulation. The data manipulation was carried out mainly using the `tidyverse` (Wickham, 2017c), a set of packages for cleaning and manipulating data in R (`dplyr`, Wickham et al., 2017a; `readr`, Wickham et al., 2017b; `tibble`, Müller and Wickham, 2017; `tidyr`, Wickham and Henry, 2017), as well as functional programming (`purrr`, Henry and Wickham, 2017), and visualisation (see below). Other packages were also used to assist with manipulating strings (`stringr`, Wickham, 2017b) and factors (`forcats`, Wickham, 2017a). Finally, the `magrittr` package deserves a special mention for introducing the 'pipe' operator (`%>%`) into R (Bache and Wickham, 2014).

Document writing. This thesis was written using the `bookdown` package in R (Xie, 2017a). This packages expands the functionality of `rmarkdown` (Allaire et al., 2017), which is turn, like `bookdown`, makes use of the `knitr` package (Xie, 2017b). The tables in this thesis were created using functions from the `knitr` and `kableExtra` packages (Xie, 2017b; Zhu, 2017).

Task design. All the computerised tasks in this thesis were created using PsychoPy (Peirce, 2007).

Visualisation. The core packages used to create the visualisations in this thesis was `ggplot2` (Wickham and Chang, 2016). In addition, `cowplot` and `ggthemes` were used to control the appearance and arrangement of the plots (Wilke, 2017a; Arnold, 2017). Figures 4.5 and 4.9 in Chapter 4 were created with the help of `ggbeeswarm` (Clarke and Sherrill-Mix, 2017). The density plots in Chapter 2 (e.g. Figure 2.8) were created using `ggjoy` (Wilke, 2017b).

Bibliography

Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., Wickham, H., Atkins, A., Hyndman, R., and Arslan, R. (2017). *rmarkdown: Dynamic Documents for R*. R package version 1.6.

Allen, R. J., Baddeley, A. D., and Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General*, 135(2):298–313.

Allen, R. J., Baddeley, A. D., and Hitch, G. J. (2014). Evidence for two attentional components in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6):1499–1509.

Allen, R. J. and Waterman, A. H. (2015). How does enactment affect the ability to follow instructions in working memory? *Memory & Cognition*, 43(3):555–561.

Alloway, T. P. (2007). *Automated Working Memory Assessment*. Pearson Publishing, London: Harcourt Assessment.

Alloway, T. P. and Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, 106(1):20–29.

Alloway, T. P., Alloway, R. G., and Wootan, S. (2014). Home sweet home: Does where you live matter to working memory and other cognitive skills? *Journal of Experimental Child Psychology*, 124(1):124–131.

Alloway, T. P., Gathercole, S. E., Adams, A.-M., Willis, C., Eaglen, R., and Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology*, 23(3):417–426.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., and Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, 28(7):725–734.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., and Elliott, J. (2009a). The Cognitive and Behavioral Characteristics of Children With Low Working Memory. *Child Development*, 80(2):606–621.

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

Alloway, T. P., Gathercole, S. E., Willis, C., and Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, 87(2):85–106.

Alloway, T. P., Rajendran, G., and Archibald, L. M. D. (2009b). Working Memory in Children With Developmental Disorders. *Journal of Learning Disabilities*, 42(4):372–382.

Amso, D. and Scerif, G. (2015). The attentive brain : insights from developmental cognitive neuroscience. *Nature Reviews Neuroscience*, 16(10):606–619.

Arán-Filippetti, V. (2013). Structure and Invariance of Executive Functioning Tasks across Socioeconomic Status: Evidence from Spanish-Speaking Children. *The Spanish Journal of Psychology*, 16(101):1–15.

Archibald, L. M. D. and Gathercole, S. E. (2007). Nonword repetition and serial recall: Equivalent measures of verbal short-term memory? *Applied Psycholinguistics*, 28(04):587–606.

Arnold, J. B. (2017). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. R package version 3.4.0.

Au, J., Buschkuehl, M., Duncan, G. J., and Jaeggi, S. M. (2015). There is no convincing evidence that working memory training is NOT effective: A reply to Melby-Lervåg and Hulme (2015). *Psychonomic Bulletin & Review*.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., and Jaeggi, S. M. (2014). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review*, 22(2):366–377.

Bache, S. M. and Wickham, H. (2014). *magrittr: A Forward-Pipe Operator for R*. R package version 1.5.

Baddeley, A. D. (1986). *Working Memory*. Oxford psychology series. Clarendon Press.

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

Baddeley, A. D. (2007). *Working Memory, Thought, and Action*. Oxford Psychology Series. OUP Oxford.

Baddeley, A. D. (2012). Working Memory : Theories, Models, and Controversies. *Annual Review of Psychology*, 63.

Baddeley, A. D. and Hitch, G. J. (1974). Working memory. *Psychology of learning and motivation*, 8:47–89.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

Barrouillet, P., Bernardin, S., and Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1):83–100.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., and Camos, V. (2009). Working memory span development: a time-based resource-sharing model account. *Developmental Psychology*, 45(2):477–490.

Bates, D. and Maechler, M. (2017). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-11.

Bayliss, D. M., Jarrold, C., Gunn, D. M., and Baddeley, A. D. (2003). The complexities of complex span: explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, 132(1):71–92.

Bays, P. M. (2015). Spikes not slots: Noise in neural populations limits working memory. *Trends in Cognitive Sciences*, 19(8):431–438.

Bays, P. M., Catalao, R. F. G., and Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10):7.1–11.

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., and Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, 11(10):6–6.

Bays, P. M. and Husain, M. (2008). Dynamic Shifts of Limited Working. *Science*, 321(August):2006–2009.

Beach, K. (1993). Becoming a Bartender: The Role of External Memory Cues in a Work Directed Educational Activity. *Applied Cognitive Psychology*, 7:191–204.

Blakey, E., Visser, I., and Carroll, D. J. (2016). Different executive functions support different kinds of cognitive flexibility: Evidence from two-year-olds, three-year-olds and four-year-olds. *Child Development*, 87(2):513–526.

Bonny, J. W. and Castaneda, L. M. (2016). Impact of the Arrangement of Game Information on Recall Performance of Multiplayer Online Battle Arena Players. *Applied Cognitive Psychology*, 30(5):664–671.

Bourke, L., Davies, S. J., Sumner, E., and Green, C. (2013). Individual differences in the development of early writing skills: testing the unique contribution of visuo-spatial working memory. *Reading and Writing*, 27(2):1–21.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2):106–113.

Brown, M. R. (1974). Effects of recall order, cue placement, and retention interval on short-term memory of normal and retarded children. *Perceptual and Motor Skills*, 39:167–178.

Brunetti, R., Del Gatto, C., Cavallina, C., Farina, B., and Delogu, F. (2016). Did I see your hand moving? The effect of movement-related information on the Corsi block tapping task. *Psychological Research*.

Burnett Heyes, S., Zokaei, N., van der Staaij, I., Bays, P. M., and Husain, M. (2012). Development of visual working memory precision in childhood. *Developmental Science*, 15(4):528–539.

Calia, C., Darling, S., Allen, R. J., and Havelka, J. (2015). Visuospatial Bootstrapping: Aging and the Facilitation of Verbal Memory by Spatial Displays. *Archives of Scientific Psychology*, 3:74–81.

Camos, V. and Barrouillet, P. (2011). Developmental change in working memory strategies: From passive maintenance to active refreshing. *Developmental Psychology*, 47(3):898–904.

Case, R. (1987). Neo-Piagetian Theory: Retrospect and Prospect. *International Journal of Psychology*, 22(5-6):773–791.

Case, R., Kurland, D., and Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33:386–404.

Chase, W. G. and Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1):55–81.

Chevalier, N. (2015a). Executive Function Development: Making Sense of the Environment to Behave Adaptively. *Current Directions in Psychological Science*, 24(5):363 – 368.

Chevalier, N. (2015b). The Development of Executive Function: Toward More Optimal Coordination of Control With Age. *Child Development Perspectives*, 9(4):239–244.

Chevalier, N. (2017). Willing to Think Hard? The Subjective Value of Cognitive Effort in Children. *Child Development*, 00(0):1–13.

Chevalier, N., James, T. D., Wiebe, S., Nelson, J. M., and Espy, K. A. (2014). Contribution of Reactive and Proactive Control to Children's Working Memory Performance: Insight From Item Recall Durations in Response Sequence Planning. *Developmental Psychology*, 50(7):1999–2008.

Chevalier, N., Martis, S. B., Curran, T., and Munakata, Y. (2015). Metacognitive Processes in Executive Control Development: The Case of Reactive and Proactive Control. *Journal of Cognitive Neuroscience*, 27(6):1125–1136.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9):345–351.

Clark, A. (2008). *Supersizing the Mind : Embodiment, Action, and Cognitive Extension: Embodiment, Action, and Cognitive Extension*. Oxford University Press, USA.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(03):181–204.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press USA.

Clark, A. and Chalmers, D. J. (1998). The extended mind. *Analysis*, 58(1):7–19.

Clarke, E. and Sherrill-Mix, S. (2017). *ggbeeswarm: Categorical Scatter (Violin Point) Plots*. R package version 0.6.0.

Connors, M. H., Burns, B. D., and Campitelli, G. (2011). Expertise in complex decision making: The role of search in chess 70 years after de Groot. *Cognitive Science*, 35(8):1567–1579.

Conway, A. R., Cowan, N., and Bunting, M. F. (2001). The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychonomic Bulletin & Review*, 8(2):331–335.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5):769–786.

Cowan, N. (1995). *Attention and Memory: An Integrated Framework*. Oxford psychology series. Oxford University Press.

Cowan, N. (2001). The magical number 4 in short term memory. A reconsideration of storage capacity. *Behavioral and Brain Sciences*, 24(4):87–186.

Cowan, N. (2005). *Working memory capacity.* Psychology Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? Nelson. *Progress in Brain Research*, 169:323–338.

Cowan, N. (2013). Working Memory Underpins Cognitive Development, Learning, and Education. *Educational Psychology Review*, 26(2):1–27.

Cowan, N. (2016a). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, pages 1–13.

Cowan, N. (2016b). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science*, 11(2):239–264.

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., and Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, 34(8):1754–1768.

Cowan, N. and Morey, C. C. (2006). Visual working memory depends on attentional ltering. *Trends in Cognitive Sciences*, 10(xx):1–3.

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., and Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science*, 13(1):120–133.

Cragg, L. and Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, 3:63–68.

Daneman, M. and Carpenter, P. A. (1980). Individual differences in working memory during reading. *Journal Of Verbal Learning And Verbal Behavior*, 19(4):450–466.

Darling, S., Allen, R. J., and Havelka, J. (2017). Visuospatial BootstrappingL When Visuospatial and Verbal Memory Work Together. *Current Directions in Psychological Science*, 26(1):3–9.

Darling, S., Allen, R. J., Havelka, J., Campbell, A., and Rattray, E. (2012). Visuospatial bootstrapping: long-term memory representations are necessary for implicit binding of verbal and visuospatial working memory. *Psychonomic bulletin & review*, 19(2):258–63.

Darling, S. and Havelka, J. (2010). Visuospatial bootstrapping: Evidence for binding of verbal and spatial information in working memory. *The Quarterly Journal of Experimental Psychology*, 63(2):239–245.

Darling, S., Parker, M. J., Goodall, K. E., Havelka, J., and Allen, R. J. (2014). Visuospatial bootstrapping: Implicit binding of verbal working memory to visuospatial representations in children and adults. *Journal of Experimental Child Psychology*, 119:112–119.

Davidson, M. C., Amso, D., Anderson, L. C., and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44(11):2037–2078.

Della Sala, S., Gray, C., and Baddeley, A. D. (1997). *Visual Patterns Test*. Thames Valley Test Company.

der Ven, S. H. G. V., Kroesbergen, E. H., Boom, J., and Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relation- ship. *British Journal of Educational Psychology*, 82:100–119.

Dunn, T. L. and Risko, E. F. (2015). Toward a Metacognitive Account of Cognitive Offloading. *Cognitive Science*, 40.

Dunning, D. L. and Holmes, J. (2014). Does working memory training promote the use of strategies on untrained working memory tasks? *Memory & Cognition*, 42:854–862.

Eddelbuettel, D., Francois, R., Allaire, J., Ushey, K., Kou, Q., Russell, N., Bates, D., and Chambers, J. (2017). *Rcpp: Seamless R and C++ Integration*. R package version 0.12.12.

Efklides, A., Kourkoulou, A., Mitsiou, F., and Ziliaskopoulou, D. (2006). Metacognitive knowledge of effort, personality factors, and mood state: Their relationships with effort-related metacognitive experiences. *Metacognition and Learning*, 1(1):33–49.

Elliott, E. M., Hughes, R. W., Briganti, A., Joseph, T. N., Marsh, J. E., and Macken, B. (2016). Distraction in verbal short-term memory: Insights from developmental differences. *Journal of Memory and Language*, 88:39–50.

Elliott, J. G., Gathercole, S. E., Alloway, T. P., Holmes, J., and Kirkwood, H. (2010). An evaluation of a classroom-based intervention to help overcome working memory difficulties and improve long-term academic achievement. *Journal of Cognitive Education and Psychology*, 9(3):227–250.

Emrich, S. M., Lockhart, H. A., and Al-Aidroos, N. (2017). Attention Mediates the Flexible Allocation of Visual Working Memory Resources. *Journal of Experimental Psychology: Human Perception and Performance*, Advance online publication.

Engel, P. M. J., Santos, F. H., and Gathercole, S. E. (2008). Are working memory measures free of socioeconomic influence? *Journal of Speech, Language, and Hearing Research*, 51(6):1580–1587.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1):19–23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3):309–331.

Ericsson, K. A. and Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In *Models of working memory: Mechanisms of active maintenance and executive control.*, number 5, pages 257–297. Cambridge University Press.

Ericsson, K. A. and Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102(2):211–245.

Fallon, S. J., Zokaei, N., and Husain, M. (2016). Causes and consequences of limitations in visual working memory. *Annals of the New York Academy of Sciences*.

Farah, M. J., Shera, D. M., Savage, J. H., Betancourt, L., Giannetta, J. M., Brodsky, N. L., Malmud, E. K., and Hurt, H. (2006). Childhood poverty: Specific associations with neurocognitive development. *Brain Research*, 1110(1):166–174.

Farrand, P. and Jones, D. (1996). Direction of report in spatial and verbal serial short-term memory. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, 49(1):140–58.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10):906–911.

Flavell, J. H., Beach, D. H., and Chinsky, J. M. (1966). Sponteneous verbal rehersal in a memory task as a function of age. *Child Development*, 37(2):283–299.

Flavell, J. H., Green, F. L., Flavell, E. R., and Lin, N. T. (1999). Development of children's knowledge about unconsciousness. *Child Development*, 70(2):396–412.

Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., and Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science*, 18(10):893–900.

Friedman, N. P., Robinson, J. L., and Hewitt, J. K. (2012). Developmental trajectories in toddler's self-restraint predict individual differences in executive functions 14 years later: A behavioural genetic analysis. *Developmental Psychology*, 47(5):1410–1430.

Friedman, Naomi, P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., and Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17(2):172–9.

Fuhs, M. W. and Day, J. D. (2011). Verbal ability and exec- utive functioning development in preschoolers at head start. *Developmental Psychology*, 47:404–416.

Gabry, J. (2017). *shinystan: Interactive Visual and Numerical Diagnostics and Posterior Analysis for Bayesian Models*. R package version 2.4.0.

Gabry, J. and Goodrich, B. (2017). *rstanarm: Bayesian Applied Regression Modeling via Stan*. R package version 2.15.3.

Galera, C. and Souza, A. L. P. D. (2010). Memória visuoespacial e cinestésica de curto prazo em crianças de 7 a 10 anos. *Estudos de Psicologia (Natal)*, 15(2):137–143.

García-Madruga, J., Vila, J., Gómez-Veiga, I., Duque, G., and Elosúa, M. (2014). Executive processes, reading comprehension and academic achievement in 3th grade primary students. *Learning and Individual Differences*, 35:41–48.

Gathercole, S. and Alloway, T. P. (2008). *Working Memory and Learning: A Practical Guide for Teachers*. SAGE Publications.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3(11):410–419.

Gathercole, S. E. (2014). Commentary: Working memory training and ADHD - Where does its potential lie? Reflections on Chacko et al. (2014). *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 55(3):256–257.

Gathercole, S. E. and Alloway, T. P. (2004). Working memory and classroom learning. *Dyslexia Review*, 17:1–41.

Gathercole, S. E., Alloway, T. P., Willis, C., and Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, 93(3):265–281.

Gathercole, S. E. and Baddeley, A. D. (1993). *Working Memory and Language*. Essays in cognitive psychology. Psychology Press.

Gathercole, S. E., Dunning, D. L., and Holmes, J. (2012). Cogmed training: Let's be realistic about intervention research. *Journal of Applied Research in Memory and Cognition*, 1(3):201–203.

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., and Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology*, 22(8):1019–1037.

Gathercole, S. E., Pickering, S. J., Ambridge, B., and Wearing, H. (2004a). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2):177–190.

Gathercole, S. E., Pickering, S. J., Knight, C., and Stegmann, Z. (2004b). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18(1):1–16.

Gathercole, S. E., Woolgar, F., Kievit, R. A., Astle, D., Manly, T., and Holmes, J. (2016). How common are WM deficits in children with difficulties in reading and mathematics? *Journal of Applied Research in Memory and Cognition*, 5(4):384–394.

Gilbert, S. J. (2015). Strategic offloading of delayed intentions into the external environment. *The Quarterly Journal of Experimental Psychology*, 68(5):971–92.

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., and Husain, M. (2011). Dynamic Updating of Working Memory Resources for Visual Objects. *Journal of Neuroscience*, 31(23):8502–8511.

Gottwald, J. M., Achermann, S., Marciszko, C., Lindskog, M., and Gredebäck, G. (2016). An Embodied Account of Early Executive-Function Development: Prospective Motor Control in Infancy Is Related to Inhibition and Working Memory. *Psychological Science*, 27(12):1600–1610.

Gray, S., Green, S., Alt, M., Hogan, T., Kuo, T., Brinkley, S., and Cowan, N. (2016). The Structure of Working Memory in Young Children and Its Relation to Intelligence. *Journal of Memory and Language*, 92:1–63.

Hackman, D. A., Betancourt, L. M., Gallop, R., Romer, D., Brodsky, N. L., Hurt, H., and Farah, M. J. (2014). Mapping the trajectory of socioeconomic disparity in working memory: Parental and neighborhood factors. *Child Development*, 85(4):1433–1445.

Hackman, D. a. and Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, 13(January):65–73.

Hall, D., Jarrold, C., Towse, J. N., and Zarandi, A. L. (2015). The Developmental Influence of Primary Memory Capacity on Working Memory and Academic Achievement. *Developmental Psychology*, 51(8):1131–1147.

Hamilton, K. A., McIntyre, K. P., and Hertel, P. T. (2016). Judging Knowledge in the Digital Age: The Role of External-Memory Organization. *Applied Cognitive Psychology*, 30(6):1080–1087.

Henry, L. and Wickham, H. (2017). *purrr: Functional Programming Tools*. R package version 0.2.3.

Hertel, P. T. (1988). Monitoring extenral memory. In Gruneberg, M. M. and Morris, P. E., editors, *Practical aspects of memory: Current research and issues.* New York: Wiley.

Hertel, P. T. (1993). Implications of External Memory for Investigations of Mind. *Applied Cognitive Psychology*, 7:665–674.

Hess, S. M., Detweiler, M. C., and Ellis, R. D. (1999). The Utility of Display Space in Keeping Track of Rapidly Changing Information. *Human Factors*, 41(2):257–281.

Hitch, G. J., Halliday, M. S., Dodd, A., and Little, J. E. (1989). Development of rehearsal in short-term memory: Differences between pictorial and spoken stimuli. *British Journal of Developmental Psychology*, 7:347–363.

Hitch, G. J., Halliday, S., Schaafstal, A. M., and Schraagen, J. M. C. (1988). Visual working memory in young children. *Memory & Cognition*, 16(2):120–132.

Holding, D. H. and Reynolds, R. I. (1982). Recall or evaluation of chess positions as determinants of chess skill. *Memory & Cognition*, 10(3):237–242.

Holmes, J., Gathercole, S. E., and Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, 12:9–15.

Holmes, J., Hilton, K. A., Place, M., Alloway, T. P., Elliott, J. G., and Gathercole, S. E. (2014). Children with low working memory and children with ADHD: same or different? *Frontiers in human neuroscience*, 8:976.

Hu, Y., Allen, R. J., Baddeley, A. D., and Hitch, G. J. (2016). Executive control of stimulus-driven and goal-directed attention in visual working memory. *Attention, Perception, & Psychophysics.*

Hu, Y., Hitch, G. J., Baddeley, A. D., Zhang, M., and Allen, R. J. (2014). Executive and perceptual attention play different roles in visual working memory: evidence from suffix and strategy effects. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4):1665–1678.

Hughes, R. W. (2014). Auditory distraction: A duplex-mechanism account. *PsyCh Journal*, 3(1):30–41.

Hughes, R. W., Chamberland, C., Tremblay, S., and Jones, D. M. (2016). Perceptual-motor determinants of auditory-verbal serial short-term memory. *Journal of Memory and Language*, 90:126–146.

Hurlstone, M. J. and Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2):295–324.

Hurlstone, M. J., Hitch, G. J., and Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, 140(2):339–73.

Intons-Peterson, M. J. and Fournier, J. (1986). External and internal memory aids: When and how often do we use them? *Journal of Experimental Psychology: General*, 115(3):267–280.

Isaacs, E. B. and Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology*, 7(4):377–380.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59(4):434–446.

Janczyk, M. and Berryhill, M. E. (2014). Orienting attention in visual working memory requires central capacity: decreased retro-cue effects under dual-task conditions. *Attention, perception & psychophysics*, 76(3):715–24.

Jaroslawska, A. J., Gathercole, S. E., Allen, R. J., and Holmes, J. (2016). Following instructions from working memory: Why does action at encoding and recall help? *Memory & Cognition*, 44(8):1183–1191.

Jarrold, C. and Hall, D. (2013). The development of rehearsal in verbal short-term memory. *Child Development Perspectives*, 7(3):182–186.

Jarrold, C., Hewes, A. K., and Baddeley, A. D. (2000). Tow separate speech measures constrain verbal short-term memory in children. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26:1626–1637.

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press, 3rd edition.

Jones, D. M., M, W. J., and H, C. (1997). Disruption of Short-term Recognition Memory for Tones: Streaming or Interference? *The Quarterly Journal of Experimental Psychology*, 50A(2):337–358.

Jones, G. and Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, 144:1–13.

Jurado, M. B. and Rosselli, M. (2007). The Elusive Nature of Executive Functions: A Review of our Current Understanding. *Neuropsychology Review*, 17(3):213–233.

Kahneman, D. and Treisman, A. (1984). Changing views of attention and automaticity. In Parasuraman, R. and Davis, D. R., editors, *Varieties of attention*, pages 29–61. Academic Press.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., and Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130(2):169–183.

Kane, M. J. and Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132(1):47–70.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., and Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, 133(2):189–217.

Karbach, J. and Verhaeghen, P. (2014). Making Working Memory Work: A Meta-Analysis of Executive-Control and Working Memory Training in Older Adults. *Psychological Science*, 25(11):2027–2037.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. *Developmental Psychology*, 45(1):56–63.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov Chain Monte Carlo in Practice: A Roundtable Discussion. *The American Statistician*, 52(2):93–100.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, 73(1–2):31–68.

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Elsevier Science.

Kuhn, D. (2000). Metacognitive Development. *Current Directions in Psychological Science*, 9(5):178–181.

Lawrence, M. A. (2016). *ez: Easy Analysis and Visualization of Factorial Experiments*. R package version 4.4-0.

Lee, K., Bull, R., and Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development*, 84(6):1933–1953.

Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., and Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology*, 82:82–99.

LeFevre, J.-A. A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S.-L. L., and Smith-Chant, B. L. (2013). The role of executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *Journal of Experimental Child Psychology*, 114(2):243–261.

Lingam, R., Jongmans, M. J., Ellis, M., Hunt, L. P., Golding, J., and Emond, a. (2012). Mental Health Difficulties in Children With Developmental Coordination Disorder. *Pediatrics*, 129:e882–e891.

Logie, R. (1995). *Visuo-spatial Working Memory*. Essays in cognitive psychology. Psychology Press.

Logie, R. (1996). The seven ages of working memory. In *Working Memory and Human Cognition*, pages 31–59. Oxford University Press.

Logie, R. H. (2011). The Functional Organization and Capacity Limits of Working Memory. *Current Directions in Psychological Science*, 20(4):240–245.

Logie, R. H. (2016). Retiring the Central Executive. *The Quarterly Journal of Experimental Psychology*, 69(10):2093–2109.

Logie, R. H. and Cowan, N. (2015). Perspectives on working memory: introduction to the special issue. *Memory & Cognition*, 43:315–324.

Luck, S. J. and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–281.

Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3):347–56.

Macken, B., Taylor, J., and Jones, D. (2015). Limitless capacity: A dynamic object-oriented approach to short-term memory. *Frontiers in Psychology*, 6(293).

Macken, B., Taylor, J. C., Kozlov, M. D., Hughes, R. W., and Jones, D. M. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition*, 155:113–124.

Mandler, J. M., Bauer, P. J., and McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, 23(2):263–298.

Martin, L., Tapper, A., Gonzalez, D. A., Leclerc, M., and Niechwiej-Szwedo, E. (2017). The effects of task-relevant saccadic eye movements performed during the encoding of a serial sequence on visuospatial memory performance. *Experimental Brain Research*, 235(5):1519–1529.

Melby-Lervåg, M. and Hulme, C. (2013). Is Working Memory Training Effective? A Meta-Analytic Review. *Developmental Psychology*, 49(2):270–291.

Melby-Lervåg, M. and Hulme, C. (2015). There is no convincing evidence that working memory training is effective: A reply to Au et al. (2014) and Karbach and Verhaeghen (2014). *Psychonomic Bulletin & Review*.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British medical bulletin*, 27(3):272–277.

Miyake, A. and Friedman, N. P. (2012). The Nature and Organisation of Individual Differences in Executive Functions : Four General Conclusions. *Current Directions in Psychological Science*, 21(1):8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex ”Frontal Lobe” tasks: a latent variable analysis. *Cognitive psychology*, 41(1):49–100.

Monette, S., Bigras, M., and Guay, M. C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology*, 109(2):158–173.

Moreau, D. and Conway, A. R. A. (2014). The case for an ecological approach to cognitive training. *Trends in Cognitive Sciences*, 18(7):334–336.

Morey, C. and Miron, M. (2016). Spatial sequences, but not verbal sequences, are vulnerable to general interference during retention in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2):99–127.

Morey, C. C. and Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic Bulletin & Review*, 20(1):163–70.

Morey, C. C., Cong, Y., Zheng, Y., Price, M., and Morey, R. D. (2015). The color-sharing bonus: Roles of perceptual organization and attentive processes in visual working memory. *Archives of Scientific Psychology*, 3(1):18–29.

Morey, C. C. and Cowan, N. (2004). When visual and verbal memories compete: evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, 11(2):296–301.

Morey, C. C. and Cowan, N. (2005). When Do Visual and Verbal Memories Conflict? The Importance of Working-Memory Load and Retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4):703–713.

Morey, C. C. and Mall, J. T. (2012). Cross-domain interference costs during concurrent verbal and spatial serial memory tasks are asymmetric. *The Quarterly Journal of Experimental Psychology*, 65(9):1777–1797.

Morey, C. C., Mareva, S., Lelonkiewicz, J. R., and Chevalier, N. (2017). Gaze-based rehearsal in children under 7 : a developmental investigation of eye movements during a serial spatial memory task. *Developmental Science*.

Morey, C. C., Morey, R. D., van der Reijden, M., and Holweg, M. (2013). Asymmetric cross-domain interference between two working memory tasks: Implications for models of working memory. *Journal of Memory and Language*, 69(3):324–348.

Morey, R. D. and Rouder, J. N. (2015). *BayesFactor: Computation of Bayes Factors for Common Designs*. R package version 0.9.12-2.

Myers, N. E., Chekroud, S. R., Stokes, M. G., and Anna, C. (2017a). Benefits of flexible prioritization in working memory can arise without costs. *Journal of Experimental Psychology: Human Perception and Performance*.

Myers, N. E., Stokes, M. G., and Nobre, A. C. (2017b). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences*, 21(6):449–461.

Müller, K. and Wickham, H. (2017). *tibble: Simple Data Frames.* R package version 1.3.4.

Namy, L. L., Smith, L. B., and Gershkoff-Stowe, L. (1997). Young children's discovery of spatial classification. *Cognitive Development*, 12(2):163–184.

Noble, K. G., McCandliss, B. D., and Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, 10(4):464–480.

Noble, K. G., Norman, M. F., and Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental Science*, 8(1):74–87.

Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3):411–421.

Oberauer, K. (2013). The focus of attention in working memory-from metaphors to mechanisms. *Frontiers in human neuroscience*, 7(October):673.

Oberauer, K., Farrell, S., Jarrold, C., and Lewandowsky, S. (2016). What Limits Working Memory Capacity? *Psychological Bulletin*.

Oudgenoeg-Paz, O., Leseman, P. P. M., and Volman, M. C. J. M. (2014). Can infant self-locomotion and spatial exploration predict spatial memory at school age? *European Journal of Developmental Psychology*, 11(1):36–48.

Palmer, S. and Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, 1(1):29–55.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2):8–13.

Peng, P. and Fuchs, D. (2017). A randomized control trial of working memory training with and without strategy instruction. *Journal of Learning Disabilities*, 50(1):62–80.

Peng, P., Namkung, J., Barnes, M., and Sun, C. (2015). A Meta-Analysis of Mathematics and Working Memory: Moderating Effects of Working Memory Domain, Type of Mathematics Skill, and Sample Characteristics. *Journal of Educational Psychology*.

Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory*, 9(4-6):423–432.

Pickering, S. J. and Gathercole, S. E. (2001). *Working Memory Test Battery for Children.* Psychological Corporation.

Pickering, S. J., Gathercole, S. E., Hall, M., and Lloyd, S. A. (2001). Development of memory for pattern and path: Further evidence for the fractionation of visuo-spatial memory. *The Quarterly Journal of Experimental Psychology Section A*, 54(2):397–420.

Pickering, S. J., Gathercole, S. E., and Peaker, S. M. (1998). Verbal and visuospatial short-term memory in children: evidence for common and distinct mechanisms. *Memory & Cognition*, 26(6):1117–1130.

Pieters, S., Desoete, A., Van Waelvelde, H., Vanderswalmen, R., and Roeyers, H. (2012). Mathematical problems in children with developmental coordination disorder. *Research in Developmental Disabilities*, 33(4):1128–1135.

Polatajko, H. J., Mandich, A. D., Miller, L. T., and Macnab, J. J. (2001). Cognitive Orientation to Daily Occupational Performance (CO-OP): Part II The Evidence. *Physical and Occupational Therapy in Pediatrics*, 20(42038):83–106.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Ragan, E. D., Bowman, D. A., and Huber, K. J. (2012). Supporting cognitive processing with spatial information presentations in virtual environments. *Virtual Reality*, 16(4):301–314.

Raghubar, K. P., Barnes, M. a., and Hecht, S. a. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, 20(2):110–122.

Rakison, D. H. and Butterworth, G. E. (1998). Infants' attention to object structure in early categorization. *Developmental Psychology*, 34(6):1310–1325.

Raynor, P., Duley, L., Small, N., Tuffnell, D., Wild, C., Wright, J., Airey, R., Akhtar, S., Brown, H., Crompton, J., Davies, J., Emsley, S., Farrar, D., Hepworth, S., Jankowicz, D., Johnson, W., Lodge, V., McKinney, T., Oddie, S., Pearson, S. A., Seal, S., Waiblinger, D., West, J., Adamson, K., Barratt, A., Clover, C., Lawrence, R., Riaz, M., Ali, S., Bartle, J., Esgate, N., Galbraith, J., Kausir, R., Mahmood, S., Meer, S., Gibson, N., Wall, R., Law, G., Mijal, R., Moghul, S., Sarwar, S., France, N., Smith, E., Sykes, M., Clarke, B., Jaja-Wachuku, C., Tolson, M., Shah, P., West, R., Ali, S., Dickson, P., Ingleson, S., Maden, R., Nguyen, S., Pearson, S. A., Seal, C., Seymour, C., Sims, A., Turner, M., Woffendin, C., Cameron, N., Bhopal, R., Davey-Smith, G., House, A., Lawlor, D., Leon, D., Ness, A., and McKinney, P. (2008). Born in Bradford, a cohort study of babies born in Bradford, and their parents: Protocol for the recruitment phase. *BMC Public Health*, 8:1–13.

Risko, E. F. and Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition*, 36:61–74.

Risko, E. F. and Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, 20(9):676–688.

Robinson, D. (2017). *broom: Convert Statistical Analysis Objects into Tidy Data Frames*. R package version 0.4.2.

Rodger, S. and Brandenburg, J. (2009). Cognitive Orientation to (daily) Occupational Performance (CO-OP) with children with Asperger's syndrome who have motor-based occupational performance goals. *Australian Occupational Therapy Journal*, 56(1):41–50.

Rose, S. A., Feldman, J. F., and Jankowski, J. J. (2011). Modeling a cascade of effects: The role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science*, 14:1161–1175.

Ross-Sheehy, S., Oakes, L. M., and Luck, S. J. (2003). The Development of Visual Short-Term Memory Capacity in Infants. *Child Development*, 74(6):1807–1822.

Rouder, J. N., Morey, R. D., Morey, C. C., and Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, 18(2):324–330.

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5):356–374.

Russell, J., Jarrold, C., and Henry, L. (1996). Working memory in children with autism and with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 37(6):673–686.

Sala, G., Gobet, F., Sala, G., and Gobet, F. (2017). Developmental Psychology Working Memory Training in Typically Developing Children : A Meta-Analysis of the Available Evidence Working Memory Training in Typically Developing Children : A Meta-Analysis of the Available Evidence. *Developmental Psychology*, Advanced online publication.

Sarigiannidis, I., Crickmore, G., and Astle, D. E. (2016). Developmental and individual differences in the precision of visuospatial memory. *Cognitive Development*, 39:1–12.

Schneider, W. (2008). The Development of Metacognitive Knowledge in Children and Adolescents: Major Trends and Implications for Education. *Mind, Brain, and Education*, 2(3):114–121.

Scribner, S. (1986). *Practical Intelligence: Nature and Origins of Competence in the Everyday World*. Cambridge University Press.

Sense, F., Morey, C. C., Morey, R. D., Prince, M., and Heathcote, A. (2017). Opportunity for verbalization does not improve visual change detection performance : A state-trace analysis. *Behavior Research Methods*.

Shah, P. and Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of experimental psychology. General*, 125(1):4–27.

Shapiro, L. (2010). *Embodied Cognition*. New Problems of Philosophy. Taylor & Francis.

Shimi, A., Kuo, B.-C., Astle, D. E., Nobre, A. C., and Scerif, G. (2013). Age Group and Individual Differences in Attentional Orienting Dissociate Neural Mechanisms of Encoding and Maintenance in Visual STM. *Journal of Cognitive Neuroscience*, 26(4):864–877.

Shimi, A., Nobre, A. C., Astle, D., and Scerif, G. (2014). Orienting Attention Within Visual Short-Term Memory: Development and Mechanisms. *Child Development*, 85(2):578–592.

Shimi, A. and Scerif, G. (2015). The interplay of spatial attentional biases and mental codes in VSTM: Developmentally informed hypotheses. *Developmental Psychology*, 51(6):731–743.

Shipstead, Z., Hicks, K. L., and Engle, R. W. (2012). Cogmed working memory training: Does the evidence support the claims? *Journal of Applied Research in Memory and Cognition*, 1(3):185–193.

Siegler, R. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, 3(1):1–5.

Siegler, R. S. (2007). Cognitive variability. *Developmental Science*, 10(1):104–109.

Simmering, V. R. (2012). The development of visual working memory capacity during early childhood. *Journal of Experimental Child Psychology*, 111(4):695–707.

Simmering, V. R. (2016). Ii. the Cognitive Dynamics Theory of Visual Working Memory. *Monographs of the Society for Research in Child Development*, 81(3):25–55.

Smith, E. E. and Jonides, J. (1999). Storage and Executive Processes in the Frontal Lobes. *Science*, 283:1657 – 1661.

Smith, L. and Gasser, M. (2005). The development of embodied cognition: six lessons from babies. *Artificial life*, 11(1-2):13–29.

Smith, L. B. and Sheya, A. (2010). Is Cognition Enough to Explain Cognitive Development? *Topics in Cognitive Science*, 2(4):725–735.

Smith, L. B. and Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, 7(8):343–348.

Solman, G. J. F. and Kingstone, A. (2017). Arranging Objects in Space: Measuring Task-Relevant Organizational Behaviors During Goal Pursuit. *Cognitive Science*, 41:1042–1070.

St Clair-Thompson, H., Stevens, R., Hunt, A., and Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, 30(2):203–219.

St Clair-Thompson, H. and Sykes, S. (2010). Scoring methods and the predictive ability of working memory tasks. *Behavior Research Methods*, 42(4):969–975.

St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology*, 22(2):286–296.

St Clair-Thompson, H. L. (2011). Executive functions and working memory behaviours in children with a poor working memory. *Learning and Individual Differences*, 21(4):409–414.

St Clair-Thompson, H. L. and Allen, R. J. (2013). Are forward and backward recall the same? A dual-task study of digit recall. *Memory & cognition*, 41(4):519–32.

Stevens, J. (1993). An Observational Study of Skilled Memory in Waitress. *Applied Cognitive Psychology*, 7:205–217.

Stone, J. M. and Towse, J. N. (2015). A Working Memory Test Battery: Java-Based Collection of Seven Working Memory Tasks. *Journal of Open Research Software*, 3:0–9.

Swanson, H. L. (2017). Verbal and Visual-Spatial Working Memory : What Develops Over a Life Span ? *Developmental Psychology*, 53(5):971–995.

Swanson, H. L. and Berninger, V. W. (1996). Individual differences in children's writing: A function of working memory or reading or both processes? *Reading and Writing*, 8(4):357–383.

Swanson, H. L., Zheng, X., and Jerman, O. (2009). Working Memory, Short-Term Memory, and Reading Disabilities. *Journal of Learning Disabilities*, 42(3):260–287.

Tam, H., Jarrold, C., Baddeley, A. D., and Sabatos-DeVito, M. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. *Journal of Experimental Child Psychology*, 107(3):306–324.

Thelen, E. and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press.

Thelen, E. and Smith, L. B. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action*. A Bradford book. MIT Press.

Torchiano, M. (2017). *effsize: Efficient Effect Size Computation*. R package version 0.7.1.

Turley-Ames, K. J. and Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, 49(4):446–468.

Turner, M. L. and Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28(2):127–154.

Van de Weijer-Bergsma, E., Kroesbergen, E. H., and Van Luit, J. E. H. (2014). Verbal and visual-spatial working memory and mathematical ability in different domains throughout primary school. *Memory & Cognition*, 43(3):367–378.

Walker, P., Hitch, G., Doyle, A., and Porter, T. (1994). The development of short-term visual memory in young children. *International Journal of Behavioral Development*, 17(1):73–89.

Wang, B., Cao, X., Theeuwes, J., and Olivers, C. N. L. (2016). Location-based effects underlie feature conjunction benefits in visual working memory. *Journal of Vision*, 16(11):1–13.

Waszak, F., Li, S.-C., and Hommel, B. (2010). The development of attentional networks: cross-sectional findings from a life span sample. *Developmental Psychology*, 46(2):337–349.

Waterman, A. H., Atkinson, A. L., Aslam, S. S., Holmes, J., Jaroslawska, A., and Allen, R. J. (2017). Do actions speak louder than words? Examining children's ability to follow instructions. *Memory & Cognition*.

Wechsler, D. (1997). *Wechsler memory scale (WMS-III).* Psychological corporation San Antonio, TX.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV).* The Psychological Corporation.

Wechsler, D. (2008). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV).* The Psychological Corporation.

Weissgerber, T. L., Milic, N. M., Winham, S. J., and Garovic, V. D. (2015). Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. *PLoS Biology*, 13(4):1–10.

Wickham, H. (2017a). *forcats: Tools for Working with Categorical Variables (Factors).* R package version 0.2.0.

Wickham, H. (2017b). *stringr: Simple, Consistent Wrappers for Common String Operations.* R package version 1.2.0.

Wickham, H. (2017c). *tidyverse: Easily Install and Load 'Tidyverse' Packages.* R package version 1.1.1.

Wickham, H. and Chang, W. (2016). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.* R package version 2.2.1.

Wickham, H., Francois, R., Henry, L., and Müller, K. (2017a). *dplyr: A Grammar of Data Manipulation.* R package version 0.7.2.

Wickham, H. and Henry, L. (2017). *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions.* R package version 0.7.1.

Wickham, H., Hester, J., and Francois, R. (2017b). *readr: Read Rectangular Text Data.* R package version 1.1.1.

Wiebe, S. A., Espy, K. A., and Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. latent structure. *Developmental Psychology.*

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., and Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, 108:436–452.

Wilke, C. O. (2017a). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'.* R package version 0.8.0.

Wilke, C. O. (2017b). *ggjoy: Joyplots in 'ggplot2'.* R package version 0.3.0.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & Review*, 9(4):625–636.

Woodman, G. F., Vecera, S. P., and Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, 10(1):80–87.

Wright, J., Small, N., Raynor, P., Tuffnell, D., Bhopal, R., Cameron, N., Fairley, L., A Lawlor, D., Parslow, R., Petherick, E. S., Pickett, K. E., Waiblinger, D., and West, J. (2013). Cohort profile: The born in bradford multi-ethnic family cohort study. *International Journal of Epidemiology*, 42(4):978–991.

Wu, K. K., Chan, S. K., Leung, P. W., Liu, W. S., Leung, F. L., and Ng, R. (2011). Components and developmental differences of executive functioning for school-aged children. *Developmental Neuropsychology*, 36:319–337.

Xie, Y. (2017a). *bookdown: Authoring Books and Technical Documents with R Markdown.* R package version 0.5.

Xie, Y. (2017b). *knitr: A General-Purpose Package for Dynamic Report Generation in R.* R package version 1.17.

Yang, T., Gathercole, S. E., and Allen, R. J. (2014). Benefit of enactment over oral repetition of verbal instruction does not require additional working memory during encoding. *Psychonomic Bulletin & Review*, 21(1):186–192.

Yang, T.-x., Allen, R. J., and Gathercole, S. E. (2015a). Examining the role of working memory resources in following spoken instructions. *Journal of Cognitive Psychology*, 5911(31400873):1–13.

Yang, T.-x., Allen, R. J., Yu, Q.-j., and Chan, R. C. K. (2015b). The influence of input and output modality on following instructions in working memory. *Scientific Reports*, 5:17657.

Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. *Attention and performance*, 18:73–103.

Yantis, S. and Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1):121–134.

Yu, C. and Smith, L. B. (2017). Hand-Eye Coordination Predicts Joint Attention. *Child Development*, 00(0):1–19.

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., and Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): measuring executive function and attention. *Monographs of the Society for Research in Child Development*, 78(4):16–33.

Zhang, W. and Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192):233–U13.

Zhu, H. (2017). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 0.4.0.

Zokaei, N., Heider, M., and Husain, M. (2014). Attention is required for maintenance of feature binding in visual working memory. *The Quarterly Journal of Experimental Psychology*, 67(6):1191–213.