The Co-development of Visual Working Memory and Executive Function in Early Childhood.

Eleanor L. Johns
100157237

School of Psychology
University of East Anglia

A thesis submitted for the degree of
Doctor of Philosophy

November 2024

Abstract

Executive function is considered fundamental to cognition and academic achievement, and executive function continues to improve with age across early childhood. However, there are major challenges in how executive function is currently conceptualised in early development. There is a clear need for more longitudinal studies that track the co-development of candidate components of executive function. These components are working memory, inhibition control, and cognitive flexibility. It is important to understand these components during infancy, however there are few measures available to assess these components. A good starting point to address this is using established measures of visual working memory that are robust from infancy. In this thesis, I aim to track the co-development of visual working memory and executive function during early development, something that no prior study has currently done. This was investigated in three steps. Chapter 2 examines the longitudinal stability of visual working memory across two tasks from 6 to 54 months of age. Findings suggest visual working memory is longitudinally stable across this period. Furthermore, findings demonstrate cross-task relationships in the first longitudinal study to examine this. Next, Chapter 3 examines the longitudinal stability of executive function beginning in the toddler period. Findings suggest executive function is longitudinally stable when assessed using the same measure from 30 to 78 months of age. Importantly, results show interactions with maternal education level and gender. To examine the co-development of these two cognitive systems, Chapter 4 examines whether early measures of visual working memory predict later executive function. Findings show that visual working memory measures from 6 months of age predict executive function performance at 30 months of age. These measures at 30 months also predict executive function four years later at 78 months of age. An additional measure of visual working memory capacity from a separate task was also found to be predictive of executive function in childhood. Findings from these three chapters all show that VWM measured in infancy and childhood robustly predicts later executive function skills. These findings are discussed in relation to dynamic systems models of visual working memory and executive function as well as how these findings may inform future assessment and intervention research.

Acknowledgements

Firstly, I would like to express my sincere gratitude to John Spencer, my supervisor, and Larissa Samuelson, my secondary supervisor, for creating the Developmental Dynamics Lab. You have both nurtured my growth as a researcher, welcoming me openly since I became a part of your lab during the second year of my undergraduate degree. Over the past seven years, the lab you created has become my home-away-from-home. You have both created a fierce research lab, and are responsible for developing highly ambitious students into successful researchers. I only hope that I can go on to be a good example of what you have created. You have somehow managed to keep this space an accepting and friendly environment, within which I have been able to create lifelong friendships. There is no overly competitive or highly pressuring environment, and I hope that other students are as lucky to be able to develop their skills in such an inclusive and welcoming space. I would also like to thank you both for the opportunity to bond with Daisy, Ghost, Summer, Katniss, and Nymeria. Animals are the biggest comfort, and I will cherish every cuddle I have had with each one of you.

To John, thank you for allowing me to undertake this journey with you. Being a first-generation university student, before working with you I never thought this level of attainment was possible. Your belief in me is what allowed me to be here, pursuing an academic career. No matter what I threw at you, you continued to support me and my education. Your excellent guidance and mentorship is the only reason I feel prepared to go off on my own in the world of research, thank you.

To Larissa, I still remember my interview for the role of voluntary research assistant in 2017. Since then, you continued to provide me with mentorship and support whenever needed. I am extremely grateful I got to learn from such a powerful and knowledgeable woman, who has consistently worked so hard to nurture and empower fellow women. Thank you for passing even a slither of your confidence and knowledge onto me. This project would not be where it is today without your guidance, and without your emotional support.

To Dr Stats himself, the great Sam Forbes, I am indebted to you and you[R] knowledge. Thank you for teaching me what a growth curve is. Thank you for educating me on open science. Thank you for encouraging me to ask the stupid-actually-not-stupid questions. Thank you for teaching me how to have a day off when I am sick. Thank you for every word of encouragement and email full of knowledge. You have helped me more than I can ever give enough credit for. I hope you enjoy my retro R code and the wonderful analyses that have stemmed from it.

I would also like to thank Jo Walther and Nicky Bridgham. Jo - there would be no brain study without you. This is as much yours as it is mine, thank you. Nicky - I am sorry for making you learn lines for data collection, thank you for keeping me company on some long drives and for always being a light on dark days. I hope the lab continues to be filled with the laughter and kindness I always experienced from both of you. I would also like to thank my fellow colleagues and friends within the lab, Prerna, Stacey, Niki, Jade. This would not exist without all of your support and the many coffee breaks. To Milena especially, thank you for going on the journey of attempting research during the pandemic with me. You have been there since the very beginning, I am so proud of how far we have come.

Lastly, I would like to thank all of the people behind the scenes. To my family, thank you for the emotional (and at times financial) support I have needed throughout my education. To Bianca Gama, thank you for always believing in me, and for supporting me for over 13 years from near to far and back to near again. You are a wonderful teacher, who educated me on much more than just Of Mice and Men. To Martin Pelling, thank you for reading over every cover letter and PhD application. To Chloe, Pete, Samara, Sarah, Colin, and everyone else I have met through our shared support of City—thank you for being with me on much needed days off, following the greatest team. To all my friends at UEA, you have given me the best memories and I cherish each and every one of you. Thank you for being there through the best and worst of times.

This thesis is dedicated to Kayson, thank you for being the brightest shining star. You bring me joy everyday. I hope I can inspire you to pursue all of your dreams, no matter how big they may seem.

And to my younger self—you did it, you made it. Don't give up.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Over the last two decades, executive function (EF) has become an increasingly 'hot topic' with research drawing attention from educators, parents, and commercial companies. EF is considered fundamental to cognition and academic achievement, and EF continues to improve with age across early childhood (Carlson et al, 2004; Carlson, 2021). Although there is no broad consensus on the definition of EF, the most accepted definition suggests that EF is composed of three factors: an inhibition factor, a switching factor, and an updating factor (Diamond, 2013; Miyake et al., 2000). These are a set of goal directed skills under which the inhibition control, cognitive flexibility, and working memory systems are engaged (Zelazo, Blair, & Willoughby, 2016). During early childhood, there are periods of striking improvement in young children's inhibitory control, cognitive flexibility, and working memory abilities (Fuhs et al., 2014; Diamond, 2013; Garon et al., 2008). Whilst there is general agreement that these three component processes are core EFs associated with improvements in reasoning, planning, problem solving, and adaptive behaviour (Collins & Koechin, 2012; Lunt et al., 2012; Zelazo et al., 2013), there is disagreement surrounding how these processes interact to enable EF. There is also disagreement about how EF should be conceptualised over development. This is important as a developmental model of EF could usefully inform early intervention practices. However, this is not achievable without first understanding the developmental trajectory of the processes involved in EF. In this context, it is important to note that much of the research conducted within the executive function sphere has been cross sectional, raising challenges for assessing the developmental trajectory of EF.

In this introductory chapter, I will begin by discussing current theories of EF. Whilst the adult literature identifies the separable components of inhibitory control, cognitive flexibility, and working memory that work together to form EF, researchers examining EF during early development are conflicted on the dissociable nature of these components. Moreover, these theories may be restricted by their statistical method. Consequently I will discuss a number of other theories of EF that do not use the same methods.

From the evaluation of theory, it becomes clear that examinations of early EF are sparse. This is likely due to the difficulty in assessing EF using tasks that do not require verbal responses. Therefore, next I discuss the challenges of examining early EF and what measures have been used in prior work.

Given these challenges of examining early EF, an arising new approach for understanding early EF suggests examining the co-development of components and how they relate to EF. Consequently, I will review the literature on the development of each candidate component, with a focus on research conducted in infancy. Finally, I will argue for the need to understand the role of multiple systems within EF, particularly in early development.

The main goal of this thesis is to track candidate predictors of later EF from infancy to determine if these predictors co-develop with EF skills longitudinally. The EF abilities of working memory and updating are strongly associated with academic achievement in English, Maths, and Science (St Claire-Thompson & Gathercole, 2006); thus, increasing understanding of how working memory co-develops with EF skills may be crucial in increasing the effectiveness of early EF and academic interventions. To assess working memory in infancy, I focus on the visual cognition system which has shown to be promising in predicting later EF from infancy (Rose et al., 2012). In particular, I track the early development of visual working memory (VWM). This is a component system of working memory that processes visual information. VWM develops within the first year of life (Reyes et al., 2020), so this system may allow us to begin to track the co-development of working memory and EF from early infancy.

1.1 Theories of Executive Function.

EF has been defined in over 32 separate ways within the literature (Goldstein & Naglieri, 2013). These different definitions stem, in part, from a prolonged theoretical debate surrounding the component structure of EF. Whilst we see periods of vast improvement in young children's working memory, cognitive flexibility, and inhibitory control abilities during early childhood (Fuhs et al., 2014; Diamond, 2013; Garon et al., 2008), there is little agreement about whether these components make up one unitary EF factor that differentiates across development, or whether EF is a unified yet diverse construct with interconnected but separable components across the entire developmental life span. The root of these disagreements stems from a reliance on the adult literature to explain EF in childhood. Therefore, in this section I will first give an overview of the main theories of EF proposed within the literature. I will then discuss how the statistical approach used as the basis for investigations of EF structure and the resulting reliance on the component structure seen in adults leads to these theoretical disagreements. Finally, I will discuss theories of EF that are not limited by the use of this specific approach.

Due to the influence of non-executive skills, such as language ability, on widely used measures of EF during early childhood, confirmatory factor analysis (CFA) has become a common analysis method for attempting to understand the processes involved in executive function (see Friedman & Miyake, 2004; Wiebe et al., 2008). CFA involves using a battery of tasks differing in stimulus and response demands, and modelling task performance scores as indicators of underlying latent variables to create a construct, in this case, called "purified" EF (Wiebe et al., 2011). Beginning in adulthood, CFA of nine commonly used EF tasks with adults indicated evidence of three subconstructs of EF: working memory, shifting (cognitive flexibility), and inhibition control. These were shown to be correlated but discrete in adulthood and can be measured separately (Miyake et al., 2000).

The unity yet diversity model of EF proposed by Miyake et al. (2000) suggests that working memory, cognitive flexibility, and inhibition control are independent components of EF working together in unity. These components are considered differentiated, meaning they are distinct and specialised. Within this model, individual differences in these proposed executive functions are unable to be measured by one sole task. Instead, multiple behavioural tasks must be used (Miyake & Friedman, 2012) from which commonality can be extracted to produce latent variables, uncovering a 'pure' measure of these executive functions and removing the influence of non-executive skills. Performance on these tasks is correlated indicating a common underlying ability, referred to as 'unity', but during extraction the three latent variables of working memory, cognitive flexibility, and inhibition control are revealed (Miyake et al., 2000). This three-factor model is supported by similar empirical studies utilizing CFA to reveal working memory, inhibition, and cognitive flexibility as EF subcomponents in adulthood (e.g., Friedman et al., 2008; Lehto et al., 2003; Vaughan & Giovanello, 2010).

Whilst many agree EF involves working memory and cognitive flexibility, there is disagreement surrounding the presence of an inhibition factor. Updated research with adults suggests there is a common EF factor made up of the overlapping variance of the three proposed executive functions, but that the separable components making up the diversity side of the model involve only updating and shifting (Miyake & Friedman, 2012), representing the components of working memory and cognitive flexibility respectively. Considerations of overlapping variance demonstrate there is no longer unique variance attributed to inhibition. Karr et al. (2018) report numerous studies struggling to identify inhibition control as a separable latent variable, with the inhibition specific tasks demonstrating weak factor loadings. Instead of this separable inhibition control factor, the updated model indicates inhibition is absorbed as a part of the common EF factor reflecting domain general executive control (Miyake & Friedman, 2012; Best & Miller, 2010). Here, common EF is posited as a construct representing the ability to bias attention towards a particular task or goal, alongside the ability to create mental representations of that specific task or goal enabling goal maintenance. The unity within this model stems from shared attentional control, with the diversity reflecting the specialised subcomponents of working memory and cognitive flexibility.

Whilst the multi-component nature of EF appears evident from around 4-years of age, this is less conclusive in younger children. It has been proposed that EF in preschool may be best characterised by a single unitary factor (Hughes et al., 2010; Wiebe et al., 2011). This unitary model posits that structural differentiation may occur across development with the unitary model during childhood evolving into the unity-diversity structure of EF in adulthood (Wiebe et al., 2011). Utilising the same CFA method as Miyake et al (2000), Wiebe et al. (2008) found that at 3 to 6 years of age, a single factor model of EF best fit the observed data in contrast to a model including separable working memory and inhibition factors. Nevertheless, Howard et al. (2015) indicate EF may go through a period of integration, not fractionation, by school age, with performance on working memory, inhibition control, and cognitive flexibility tasks at 3 years being largely unrelated to one another but performance in these tasks at 4 years becoming increasingly related. Here, a two-factor model with working memory and inhibitory control was proposed as a best fit for the data. Similarly, Miller et al. (2012) proposed that this two-factor model best fit the data with the inclusion of more tasks and when age ranges were restricted. These differences in model structure may reflect the proposed hierarchical development of EF. Garon et al. (2008) indicate that working memory emerges first, followed by inhibition. This sequence of development then allows the emergence of cognitive flexibility. As suggested by Zelazo (2015) and Devine et al. (2019), it is only when all of these processes have developed and become specialised that EF may begin to differentiate. This hierarchical view hypothesises that basic abilities are necessary before structural change in EF occurs and children can complete complex tasks.

To date, there have been no longitudinal examinations of this potential sequence to examine this unitary theory of EF empirically. Doing so would require the examination of EF subcomponents from early in development, examining longitudinal stability over time. This would require understanding the stability and development of specific components as well as the identification of suitable tasks that can be used to track each component longitudinally from infancy. This is further complicated by the demands of CFA. CFA requires tasks that provide only one outcome measure per task due to the likelihood of high correlations between multiple outcome measures. For instance, many candidate EF tasks involve measuring both reaction times and accuracy. This is problematic as one typically does not include multiple, within-task measures in a factor analysis. Thus, the specific tasks being administered and the specific dependent variables chosen to include in factor analytic studies are quite constrained and may not clearly reflect the processes thought to underlie each component (Miller et al., 2012).

Transitioning from statistical models to examinations of brain and behaviour, there is some evidence that changes in the brain reported over development, such as neural maturation, supports the view of an early unified EF that becomes more diverse during early development. Behavioural changes in EF correspond with neural maturation of the prefrontal cortex (Roberts & Pennington, 1996). So, whilst EF may begin unified, EF is considered to develop into distinct and separable components across childhood as the prefrontal cortex (PFC) matures. Developmental changes in the structure and connectivity of PFC are important for the development of higher order processing skills and are associated with improvements in EF (Diamond, 2002; Welsh & Pennington, 1988). This neural development in early infancy and childhood may constrain the age at which EF can become stable and distinct, with evidence suggesting changes in PFC development appear around age 4 (Fiske & Holmboe, 2019). These changes in PFC development support longitudinal studies that propose a discrete yet related structure that has fractionated by around 4-years of age (Brydges et al., 2014). This fractionation may be related to emerging distinctness within working memory, inhibition control, and cognitive flexibility systems. This view predicts an increasingly differentiated EF across early preschool years, although other research has demonstrated components are highly related even from early childhood (Wiebe et al., 2011). Examined as separate components, working memory, cognitive flexibility, and inhibition control are differentially related to developmental outcomes in early school age children (Brock et al., 2009). The suggestion here is that whilst these processes may begin unified, each has a unique developmental trajectory that corresponds to developmental changes occurring in the brain, leading to separation across development. However, this neural evidence has also been explained by a number of theories present alternatives to the unity and diversity view of development.

For instance, Zelazo and colleagues have proposed the cognitive complexity and control theory, which stipulates that both EF and reflective processing depend on the PFC in a similar way (Zelazo, 2004). This cognitive complexity and control theory proposes that children's development occurs from hierarchical rule-representation and conscious control. Children must be able to reflect on proposed rules and behaviours in order to make a decision and select a response; this is EF according to the theory. Zelazo (2015) also proposes that other systems such as attention are more heavily involved in supporting early EF. Zelazo (2015) states EF serves to modulate attention to fulfil a goal, and this goal may be simple or complex. By adulthood, these lower level systems may no longer contribute as significantly. From this perspective, age-related changes in EF are understood in terms of corresponding biological changes in the ability to formulate higher-level hierarchical representations throughout childhood. Zelazo (2015) acknowledges that EF may begin as a unified, and not diverse, system. However, it is proposed that across development, EF begins to show diversity as components become more specialised. Within this account, early EF is considered reactive, where behaviour is directly triggered by a stimulus and occurs from automatic processes. Over development, attention processes, such as attentional control, mature and EF becomes more deliberate. Children become able to focus and sustain attention to the task or problem at hand. As this occurs, EF becomes more differentiated. Attentional control allows the ability to suppress automatic responses, leading to reflective processing; children develop the ability to stop and think about actions before deciding how to respond to a stimulus. This ability to utilise attentional control and sustained attention from childhood is thought to underlie how the specialised subcomponents of working memory and cognitive flexibility interact based on the demands of a specific task. Although this theory has many strengths, it has been criticised for failing to explain how people form and engage rule representations in real-time and how the dynamics of activation or inhibition work within this proposed hierarchical system (Buss & Spencer, 2014). Furthermore, whilst cognitive complexity and control theory has provided descriptive accounts of the integration of both brain and behaviour, it fails to specify any neural mechanisms as the theory is a verbal / descriptive theory.

Other theories of EF development have focused on specific tasks which show dramatic change in early development such as the Dimensional Change Card Sort (DCCS; Frye et al., 1995) task. This task is commonly used from 3 years of age and upwards and involves sorting bivalent test cards according to one dimension (e.g. object colour; preswitch phase) and then according to another (e.g. object shape; postswitch phase). For instance, children might be shown two target cards: an orange monkey on the left and a green lion on the right. In the 'colour' game, orange lions would be sorted to the left while green monkeys would be sorted to the right. After repeatedly playing the 'colour' game, children are asked to switch and play the 'shape' game. Here, orange lions would be sorted to the right and green monkeys would be sorted to the left. This task is thought to tap the different components of EF as children have to maintain the correct rule in working memory, switch the rule across the preswitch and postswitch phases, and inhibit the prepotent response to continue sorting by the previous dimension.

According to the Attentional Inertia account of EF (Kirkham et al., 2003), children experience attentional inertia during the DCCS task, and are therefore unable to inhibit thinking of the objects' initially relevant attribute. That is, perseveration reflects difficulty in inhibiting a prepotent focus on the first relevant aspect of the stimulus, disengaging from that mindset, refocusing attention to the newly relevant aspect, and then being able to switch attention. Although compelling, this account has been criticised for a failure to consider all processes involved within this task. For example, this account does not consider the influence of working memory (Munakata et al., 2003).

Other researchers have argued that the ability to overcome perseveration on the DCCS task may depend on the strength of the representation of the rules (Morton & Munakata, 2002). Here, success in the DCCS task stems from the development of sufficiently strong active representations of the rules that correspond to current goals of the task. Switching on the DCCS occurs when the actively maintained rules in working memory outcompete the 'latent' representations that correspond to the prepotent response from the pre-switch rules. To demonstrate these ideas, Morton and Munakata (2002) built a connectionist model. This was the first attempt to quantitatively simulate children's behaviours in the DCCS task.

As outlined by Perone et al. (2021), cognitive complexity and control theory and the connectionist model of EF have notable similarities to a recent theory of EF using the framework of Dynamic Field Theory (DFT) proposed by Buss and Spencer (2014). This is a dynamic systems model (Spencer & Schöner, 2003) able to capture both the micro-scale moment-to-moment changes that influence performance in the DCCS task as well as macro-scale changes across development. Unlike component based models of EF, DFT does not assume that specific components are assigned unique "jobs" in specific tasks. Rather, the theory examines how an integrated set of generic cortical fields interact together and support the completion of tasks in a goal-directed manner.

In particular, DFT posits that patterns of neural activity within cortical fields underlie a diverse array of empirical findings from the DCCS task. These cortical fields consist of neurons 'tuned' to specific dimensions (e.g., space, colour, orientation). The cortical fields are organised so that neighbouring locations in a field will respond maximally to similar features, as they have similar receptive fields. 'Peaks' of activation form due to localised excitation. Excited neurons become active and stimulate neighbouring neurons. At the same time, the activation of these neurons leads to inhibition of activity for neurons coding for features distant from the activated site in the cortical field. The excitation of these local neurons and inhibition of distant neurons allows a peak to form and remain stable. This peak of activation, then, represents a neural decision that a particular feature is present. The ability to maintain this peak of activation over time reflects working memory. For example, a peak might represent a specific feature of an object to be remembered, such as an object's colour. The introduction of new inputs, such as a new feature, can lead to these peaks shifting, with new peaks arising across a network of integrated cortical fields. Importantly, inhibition plays a role in such models, reflecting the suppression of stimulation caused by actively maintained peaks which suppress other options via global inhibition. Similarly, dynamic fields can display aspects of flexibility as peaks can be updated 'on the fly' as task demands change.

DFT is able to explain performance on specific EF tasks, for example, in the widely used DCCS task. According to DFT, when a child is asked to sort by colour, a peak will form in a "colour-space field" representing the specific colours of the card presented. A peak will also form in the "spatial field" for the side the card is consistently sorted to. For instance, if a child is consistently asked to sort by colour and there is an orange monkey target card on the left sorting tray in the task space, then a presentation of an orange lion test card will build a peak at the left location in space due to the overlap between the colour orange and this feature at the left location in space. This sorting behaviour is supported by a frontal system that is biased by the instruction to 'sort by colour', boosting the activity of the colour-space field and biasing the model to sort by this dimension. When the child is then asked to switch to sorting this card by its shape (i.e., put the orange lion in the right tray to match the green lion), they must now inhibit the irrelevant colour information. According to DFT, younger children (and younger models) have weaker neural interactions and less refined connections between the frontal and posterior cortical fields. This reflects a less organised mapping between their representation of, for instance, the word 'colour' and selective activation of the colour-space field. Consequently, they are unable to robustly activate their frontal system to selectively attend to shape and they erroneously respond by colour. Older children, by contrast, have stronger neural interactions and more refined frontal-posterior connectivity. Thus, they are able to robustly activate the word 'shape', boost the shape-space field, and sort by this dimension (correctly placing the orange lion to the right tray to match the green lion).

Buss and Spencer (2014) formalised these concepts in a DF model that quantitatively simulated findings from 14 different DCCS conditions using the same model with the same parameters. In addition, Buss and Spencer (2014) tested several novel predictions, including predictions about the role of space in DCCS performance. Here, they showed how simply moving the target cards in space at the start of the first postswitch trial could improve the performance of 3-year-old children and impair the performance of 4-year-old children. Subsequent work has also tested additional novel predictions derived from this theory (e.g. Perone et al., 2019; Perone et al., 2015). Thus, the DF model of EF development has captured more data quantitively than any other theory and has established its ability to generate novel predictions. This theory has also led to novel insights into development. For instance, Lowery et al. (2022) demonstrated that early label learning is related to later dimensional attention and performance on the DCCS task, with improvements in task switching abilities being related to a child's label learning ability. This is supported by Spencer et al. (2025) who simulated children's behaviour in an expanded version of the DF model, demonstrating that dimensional label learning provides the first quantitively defined learning mechanism to explain the development of EF.

To conclude, we can see the unity yet diversity and unitary models of EF coming from the CFA studies are heavily reliant on the concept of latent factors. In these CFA models, we are defining EF by the processes we attribute to specific tasks from specific dependent variables. While this is useful, it fails to specify how these systems work together to complete the tasks examined. Moreover, developmental work from this perspective has yielded mixed findings. An alternative approach has emerged using the concepts of DFT which specifies how multiple processes are integrated to succeed on complex EF tasks, such as the DCCS task. DFT also specifies what is changing over development, which is a unique feature of this approach.

Now that I have reviewed the key theories of the development of EF, I take a closer look at how EF has been measured in early development in the next section.

1.2 Examining Early Executive Function.

There are significant changes in social and language development during infancy which effect how EF can be measured at different stages of development. As a result, there are many different tasks used to examine EF in early childhood. Within this section, I will begin by explaining the most widely used tasks of early EF, from infancy to childhood. Early EF is difficult to measure due to the requirement of language comprehension in many robust EF tasks. This constrains their use until after the development of language comprehension. By contrast, many of the tasks that do not require a language component reach ceiling levels early on, meaning we are forced to use different tasks at different ages. Therefore, I will focus on the use of early EF measures longitudinally and the resulting challenges. Whilst there are statistical methods that may allow us to examine performance on separate tasks over time, we cannot be sure we are measuring the same executive and non-executive skills across these tasks. This creates difficulty in tracking the stability of EF from infancy.

Measures of Early EF

Over the first to second year of life, infants begin to be able to search for hidden objects (Bell, 2012; Diamond, 1990). For example, during early infancy, the A-not-B task has been used to examine EF. A-not-B is a visual task where a desirable object is placed in one of two hidden (e.g. closed box) locations (location “A”). After a delay, the child is then allowed to retrieve the object. The object is repeatedly placed in location A, building a prepotent response to that location. The object is then hidden in the alternate location (location B; other closed box) and the child must inhibit the prepotent response in order to correctly retrieve the object from location B. In order to be successful in this task, the child must hold the object location within their working memory and update this information when the object location is changed. After this location change and a delay, the child must then flexibly switch their attention to the new object location, and inhibit the prepotent response to the old location, in order to correctly locate the object (Thelen et al., 2001). Therefore, this task is considered an early EF task as it requires the use of working memory, cognitive flexibility, and inhibition control using only visual and motor responses.

Once language comprehension develops, verbal instruction becomes possible. From two years of age, infants can begin to play games in which they have to understand and keep multiple rules in mind and engage cognitive flexibility. This allows us to examine 2 to 5 year old children's ability to switch between tasks (Prerner & Lang, 2002; Brooks et al., 2003), inhibit actions (Steelandt et al., 2012) and maintain and use information held in working memory (Garon et al., 2014; Cheng et al., 2020) all in the context of verbal instruction.

For instance, one key task used in this early developmental period is the DCCS task outlined previously (Frye et al., 1995). Whilst often considered a cognitive flexibility based task, this task does require both working memory to remember the current rule to be followed and maintain the dimension to be attended to, and attention shifting skills to be able to switch stimulus response mappings (Diamond, 2013). Inhibition is also involved to inhibit the prepotent responses, similarly to the A-not-B task.

Another example of a common EF task that can start to be used from 3 to 5 years of age is the classic 'Simon Says' task (Carlson, 2005). Here participants need to respond to the instructions of the experimenter only when those instructions are preceded by the phrase "Simon says" (e.g., "Simon says, turn around"). In this task, the prepotent response would be to perform any action immediately when the experimenter gives the instruction. This task is traditionally considered an inhibitory control task, requiring the child to suppress performing a behaviour when there is no "Simon says" stated before the instruction. Working memory is also used to maintain and remember the rule of only following a "Simon says" along with remembering and updating the actions to be performed. This task has been allocated to the group of 'Go/No-Go' (Simpson & Riggs, 2006) tasks also used from 3 years. These tasks involve 'go' trials where a child is asked to respond to a stimulus or perform a specific action. 'No-go' trials are also recorded, where a child is asked to make no response to a different stimulus or not to perform a specific action. Another example of this type of task is the tower-task (Kochanska et al., 1996), where a child is asked to place a block on a tower (go), but must then wait for the experimenter to place a block on the tower before placing another (no-go). These are used as inhibitory control tasks, where responding on no-go trials reflects the inability to inhibit go responses (Simpson & Riggs, 2006).

During early childhood, Stroop-like tasks such as the Day/Night task (Gerstadt et al., 1994) are also used. This task involves asking children to respond with the name of stimuli B when presented with stimuli A, e.g. 'Night' when shown a picture of day, and to respond with the name of stimuli A when presented with Stimuli B, e.g. 'Day' when shown a picture of night. This task involves inhibition to inhibit the natural prepotent response of saying what is seen on the card, cognitive flexibility to switch response mappings, and working memory to remember which word is assigned to which card.

Here, by simply detailing the requirements of widely used tasks, we begin to see the challenges of examining early EF. The literature tends to assign each task to a specific component, ignoring task requirements that may need engagement from other EF components, such as working memory. Furthermore, there are differences in the level of reliance on specific EF components that each task requires, with many of the tasks relying more heavily on inhibition control. The level of non-executive skills required also differs across tasks. For example, the reliance on motor skill for A-not-B is replaced with a reliance on language ability for the DCCS and Simon Says tasks. Consequently, the longitudinal assumptions we can make by using these tasks repeatedly over ages is limited.

Longitudinal examinations of early EF and the current challenges faced

Few studies have examined EF longitudinally and those that have tend to focus on short intervals (e.g. Miller & Marcovitch, 2015; Johansson et al., 2016; Hughes et al., 2009; Willoughby et al., 2012). These longitudinal studies often use multiple tasks across development and often struggle to show stability from early infancy, likely due to the different requirements and demands of the tasks.

The aforementioned A-not-B task shows relationships to increased fronto-parietal activation from as young as 8 months (Bell, 2012). However, this task may not show longitudinal stability. For example, Miller and Marcovitch (2015) found no correlation in performance on the A-not-B task from 14- to 18-months of age. Moreover, whilst the A-not-B task is appropriate in early infancy, it reaches ceiling levels from around 24-months of age. After this, other tasks are required to examine EF performance longitudinally (Broomell & Bell, 2022).

This A-not-B task is predominantly designated as an inhibition control task within the literature, despite working memory and cognitive flexibility requirements. Therefore, in longitudinal examinations of stability, A-not-B is often paired with more specific inhibition control tasks, such as the go/no go tasks, later (see Broomell & Bell, 2022). Whilst the A-not-B task does involve inhibitory control, the delay results in more of a reliance on working memory to correctly identify where the object is. The go/no-go tasks it is commonly paired with, however, have been shown to be inherently examining inhibitory control (Prerner & Lang, 2002). Whilst working memory is required to remember which trials are go and which are no-go, performance on pre-potent go trials is increased over that of no-go trials, specifically in 6- to 12-year-old children (Ciesielski et al., 2004). This failure in the no-go trials is due to increased demand on the inhibitory control system. Using a button press task where children were asked to press a button on the same side as a stimulus or on the opposing side to the stimulus, Wright and Diamond (2014) indicated the order of trials on these tasks could lead to lower performance due to an inability to update to the new rule, demonstrating a working memory difficulty. However, it was shown that 6- to 10-year old children were generally able to remember task rules. For instance, when the orders were reversed and incongruent, opposite side trials were examined first, children performed in a comparable manner to when congruent, same side trials were examined first. This illustrates it is the inhibiting of the automatic response, such as that posed on no-go trials, that is the challenging element of the task (Wright & Diamond, 2014). Furthermore, the A-not-B task can be conducted without a reliance on language, but many go/no-go tasks such as simon says have a reliance on verbal skill to understand the given instructions. Measures requiring language comprehension place additional cognitive load on the child (Hughes & Graham, 2002) and these tasks requiring comprehension or verbal responses cannot be used during early infancy. This demonstrates the main issue with examining early EF: a lack of tasks that can be consistently used over ages longitudinally.

Stroop-like tasks have also been used to examine EF longitudinally. Johansson et al. (2016) paired a hide and seek task at 12-months, similar to the A-not-B task, with a Stroop-like task at 36 months. This Stroop-like task required children to identify a sound whilst viewing a matching (e.g. picture of a dog whilst hearing a woof) or mismatching (e.g. picture of a dog whilst hearing a car honking) stimulus. Therefore, the child must be able to effectively identify the sounds, tapping into their auditory recognition memory. However, there was no element of sound or requirement to match an object to a sound within the hide and seek task, which was similar to the classic A-not-B task. Johansson et al. (2016) found no correlation between 12-month EF tasks with EF at 24 months or 36 months, but 12-month hide and seek task performance negatively correlated with performance on a Stroop-like task at 36 months. Whilst both tasks may tap into inhibition and switching abilities, there are differing demands from non-executive skills and executive skills, and this may be influencing the relationships among measures.

By using different tasks with different requirements over time, the conflicting findings surrounding the structure and stability of EF may simply demonstrate a mismatch in our classification and pairing of tasks over time rather than informing us about the development of EF. There is support for this use of multiple tasks across development, with Fuhs and Day (2011) finding EF tasks are invariant among preschool aged children. An invariance suggests the tasks are examining the same construct regardless of differences in measurement as children are performing the same across tasks despite the task differences. Willoughby et al. (2012) also found

EF tasks demonstrated strong measurement invariance from 3 to 5 years of age. However, these researchers concluded that the battery of tasks used were not as reliable in capturing the EF abilities of higher performing children, and this invariance was only reliable for the low performing children. This may be related to the task impurity problem and the reliance on a single dependent variable when using confirmatory factor analysis. The single dependent variable chosen for certain tasks may not have been sensitive enough for examining the high-performing children's scores, where increased task demands, for example, may have been required to capture further variance and examine EF capabilities.

Evidence from DCCS tasks shows whilst successful at sorting via the requested dimension during a preswitch phase, most 3-year-olds continue to sort the test cards by this first dimension during a postswitch phase. Here, the ability to inhibit interference is critical as response inhibition supports children's ability to inhibit attention to irrelevant stimuli and sustain attention to relevant stimuli to encode information accurately (Roderer et al., 2010). Canonical versions of this task use bivalent test cards that match the target cards along one dimension. Adaptations of this task have found 3-year-olds are able to demonstrate successful switching when these dimensions are more separated, e.g. the shape, a star, is presented in the middle of the card as black/colourless, and the colour information comes from a coloured background, e.g. blue, behind that shape (Diamond, Calrson, & Beck, 2005). With these separated dimensions, children can describe a patch of colour as 'blue' and the outline of the shape as 'star', as opposed to attempting to apply these labels to the same object when the dimensions are not separated. By 5 years of age, children are successfully able to apply these labels to less separated dimensions, e.g. "blue" and "star" are within the same object, and switch flexibly (e.g., Dick et al., 2005; Kirkham et al., 2003; Zelazo, Müller, Frye, & Marcovitch, 2003). The DCCS task has successfully been examined longitudinally in different forms, such as on a tablet task, from as young as 2 years of age and is able to capture age-related changes in EF (Carlson, 2021). Given the robustness of this task, and the multitude of theories that have engaged with this task, a version of the DCCS will be used within the present study. However, the use of this task still leaves a large period of early development, early infancy, understudied.

Given the difficulty in examining EF early on in childhood, and the resulting conflicts across theories, the next section examines what we know about the unique developmental trajectories of the three component processes hypothesised to underlie EF with an emphasis on studies beginning in infancy.

1.3 Possible precursors to Executive Function.

Differing definitions of EF have led to the identification of 18 sub-components, however many of these are overly complex and the exact number of constructs rightfully labelled as components of EF is largely unknown (Karr et al., 2019). The most commonly emphasised components identified within the literature remain working memory or updating, cognitive flexibility or shifting, and inhibition control. Modelling approaches indicate the best way to further understand EF and the systems involved may be to track the co-development of these components involved in EF (Spencer et al., 2025). In this section, I will give an overview of the development of each of these sub-components starting in infancy.

1.3.1 Inhibitory Control

Inhibitory control relates to the ability to withhold a dominant or highly practised prepotent response, habit, or impulse to respond to a stimulus (Diamond, 2013), and often requires the ability to redirect attention. The development of inhibitory control across early infancy has been well documented. Inhibitory control emerges from around 6 to 12 months, observed from the ability to inhibit dominant responses (Bell, 2012; Cuevas et al., 2012; Holmboe et al., 2018). Holmboe et al. (2018) indicate inhibitory control begins to show stability from 6 months, although there is only a limited ability to control actions. Before 6 months, responses are driven largely by automatic processes, such as reflexes and immediate needs. Infants may briefly focus on a specific object or face demonstrating signs of early attention regulation.

The emergence of attention control, motor inhibition, and basic delayed gratification can be seen from 6 to 12 months. Infants begin to hold focus on specific objects and shift attention away from distractions. Motor inhibition also becomes observable, where infants show hesitation before reaching for forbidden objects or items out of reach (Hendry et al., 2016). As aforementioned, infants can also demonstrate basic object permanence in simple versions of the A-not-B task (Bell, 2012).

Garon et al. (2008) demonstrate that infants begin to engage in more complex tasks that require inhibition, such as the A-not-B task involving motor movement/reaching from 12 to 18 months. Object permanence, often linked to inhibitory control, becomes evident and simple rule-following also begins to emerge: toddlers show early signs of complying with rules like ”Don’t touch” or ”Stop,” though control over these actions is still inconsistent (Garon et al., 2008).

From 14 months onwards, toddlers demonstrate delayed gratification for longer periods. Kochanska et al. (2001) assessed toddler inhibitory control using tasks referred to as Do and Don’t, for example a child was asked to put toys into a specific basket (do), or to not touch specific toys on a shelf (don’t). Toddlers demonstrated increasingly better following of the multi-step instructions on ’do’ trials from 14- to 45-months, and their emotional regulation also improved, with toddlers showing the ability to inhibit immediate emotional responses like frustration by self-soothing (Kochanska et al., 2001). However, within these stages of inhibitory control development we can already see the influence of other EF factors. For example, to be able to demonstrate good inhibitory control by refraining from touching certain toys, the child must be able to utilise their working memory to remember the specific toys they have been instructed not to touch.

Inhibitory control is the most commonly studied component of EF (Silva at al., 2022). However, inhibition control tasks struggle to assess inhibition as a separable component, with many requiring elements of cognitive flexibility and working memory. As discussed previously, the contribution of inhibitory control to EF as its own isolated cognitive process is contested. This contention may also stem from differing definitions of inhibitory control, with a separation between the inhibition of distraction in order to focus attention and the inhibition of certain behaviours requiring emotion regulation. Inhibition control tasks vary from a focus on inhibiting a prepotent response to others that focus on oculomotor inhibition (Nigg, 2000). Whilst the study of inhibitory control has challenges, there is a well-described trajectory of inhibition control development that would benefit from updated measures providing clarity. This is an area currently being studied (see Broomell & Bell, 2022).

1.3.2 Cognitive Flexibility

Attention control is often used interchangeably with shifting, also referred to as cognitive flexibility. This capacity to flexibly direct attention between differing mental tasks or rules is important for being able to problem solve and switch between certain concepts. Poor cognitive flexibility can lead to getting stuck in previous ways of completing tasks and not being able to consider an alternative perspective (Miller et al., 2015; Stuss et al., 2000).

The foundation for complex cognitive flexibility stems from early attention shifting. As demonstrated by the early A-not-B task, over the first year of life infants become able to shift their attention between objects, people, and different locations (Bell 2012; Diamond, 1990). From around 2.5 years onwards, cognitive flexibility is studied using a number of tasks, for example, the Dimensional Change Card Sort task (DCCS; Frye et al., 1995). As described previously, children are required to match cards based on colour/shape and then asked to switch and match cards on the other dimension. A similar task used from 3 years onwards is the Flexible Item Selection Task (FIST; Jacques & Zelzo, 2001). In this task, children are presented with 3 cards and asked to pair two cards up based on a specific dimension, e.g., shape. The child is then asked to switch and pair two of the same three cards based on a different dimension, e.g., colour. These cognitive flexibility tasks have demonstrated that children are able to reliably switch from one rule to another from 4-years of age (Müller et al., 2006; Zelazo et al., 2003; Jacques & Zelzo, 2001). The main challenge of these rule-switching tasks is the requirement to update behaviour depending on a change in rule (Blakey & Carroll, 2018). These tasks are proposed as sole cognitive flexibility tasks, and in comparison to inhibitory control tasks, these tasks appear to posses more commonality in terms of task requirements. However, as demonstrated previously when discussing the DCCS task, these tasks often involve elements from other cognitive processes.

During early development, cognitive flexibility needs to be understood as an emerging skill supported by developments in working memory and inhibitory control (Carroll et al., 2016; Chevalier et al., 2012). It is especially difficult to isolate inhibition control and cognitive flexibility tasks, as the task possess similar requirements. Moreover, the majority of cognitive flexibility tasks require language or motor skills, which may influence performance. The emergence of cognitive flexibility is an important area of development to be considered for future research, but first updated tasks more suitable for tracking cognitive flexibility from early infancy need to be developed.

1.3.3 Working Memory

Working memory has been shown to be important for academic success (Jaroslawska et al., 2016). Here, working memory is important not only in terms of capacity but also in terms of working memory resources that can be devoted to attending to instructions and information alongside the ability to remember what is required at different times or in different subjects (Cheie et al., 2017). This requires monitoring and encoding incoming information, revising items held in working memory, and replacing irrelevant information with new more relevant information (Morris & Jones, 1990).

Historically, working memory has been proposed to act as a buffer between long term memory and EF which allows for previous experiences to be brought into consideration when making decisions or planning (Baddeley, 2000). This multi-component system reportedly involves the phonological loop, a short term storage system for verbal information; the visuo-spatial sketchpad, a short term storage system for visual and spatial information; the central executive, an attentional control mechanism directing attention between competing options to enable decision making and task completion; and the episodic buffer; a multidimensional storage system that can access information from the long term memory store (Baddeley, 2000). Distinctions have been made between these domains of working memory and the short term memory system (see Cowan, 2001). Tasks measuring short-term memory are focussed only on storage, whilst working memory tasks are considered to require additional manipulation. The adult literature indicates this additional manipulation has stronger predictive power for higher cognitive skills than short term memory measures, likely due to the more complex nature of the tasks (Unsworth & Engle, 2007).

Despite claims that this multi-component model has not been challenged by a better model over the 50 years since Baddeley and Hitch first proposed it (see Baddeley and Hitch, 1974), there are more recent models of working memory that can explain changes in working memory over development. For instance, DFT has been used to explain developmental changes in working memory capacity. DFT proposes that the maintenance and retrieval processes of working memory can be attributed to activation in specialised cortical networks. Specifically, Wijeakumar and Spencer (2020) explain that neural systems move in and out of three attractor states: resting; stabilised; and self-sustaining. It is in this self-sustaining state where we see working memory. The interactions within the system itself, and between other systems, can sustain activation after an input is removed, leaving a self-sustaining state which is maintained and can be retrieved when an input is received. Working memory is considered to be a specific state of the neuro-cognitive system (Spencer et al., 2025), and multiple types of information can be actively maintained depending on the 'tuning' of the cortical fields examined. For instance, multiple working memory fields have been examined in prior work including spatial, visual, and object-based working memory (Schutte et al., 2003; Schutte & Spencer, 2009; Simmering, 2016; Schneegans et al., 2014). Each of these fields can show the aforementioned self-sustaining activation patterns. This is considered to be one hypothesis for the neural basis of working memory within the brain (Wei et al., 2012).

Considering the consensus within the EF literature regarding the importance of the working memory system, it is surprising to note a lack of interaction between the EF and working memory literatures. A large focus of EF intervention research has been on inhibitory control and shifting; many EF tasks are based on inhibition of prepotent responses, switching paradigms, and are primarily considered cognitive flexibility tasks. Working memory seems to only be considered when attempting to examine the structure of EF. Moreover, EF measures often rely on what are considered to be short term memory measures to examine memory performance, such as the picture sequence memory test in the NIH Toolbox (Bauer & Zelazo, 2013) or span tasks (e.g. Blankenship et al., 2019; Broomell & Bell, 2022). Basic span tasks, such as forward span, are widely accepted as a measure of short term memory (Baddeley, 2020), but may not be complex enough to require working memory. When short term memory tasks were made more complex, they related more strongly to measures of general intelligence (Unsworth & Engle, 2007) and this added complexity is thought to require engagement from the working memory system. Many EF measures that add a more complex element to attempt to examine working memory generally only do so using measures of verbal span plus manipulation, such as backward span tasks (e.g. Broomell & Bell, 2022). These tasks are not suitable for use during infancy and fail to examine an integral aspect of memory: visual working memory. Visual working memory is crucial in language development which likely supports the development of EF. For instance, Gooch et al. (2016) found a bidirectional relationship between language ability and EF, with visual working memory specifically being a strong predictor of language ability.

Visual working memory can be examined as young as four months of age using preferential looking tasks (Ross-Sheehy, 2003; see Reyes et al., 2020). Between 3- and 5-years of age, children are able to complete more complex visual working memory tasks, showing increases in capacity on change detection tasks (Buss et al., 2014, Simmering, 2016). At this age, children can also engage in verbal working memory tasks, such as backward span tests, which as discussed are more commonly used within EF research. Verbal and visual-spatial working memory have consistently been demonstrated to be separable, particularly in childhood (Gathercole et al., 2004). However, this separability is often not accounted for within EF research.

Working memory is highly predictive of later success. For example, kindergarten working memory predicted third-grade mathematics and reading achievement, even after accounting for other EF skills (Nguyen & Duncan, 2019). Preschool working memory has also been shown as the sole EF component that can predict later academic achievement at 15-years of age in mathematics (Watts et al., 2014) and reading (Ahmed et al., 2019).

In order to increase our understanding of executive functions and how the systems involved within EF may co-develop, we must integrate the study of EF with the existing working memory literature. One way to do this would be to examine the co-development of EF and visual working memory as this system can be readily measured from infancy.

Are these components truly separable?

EF is widely viewed as a set of domain-general components that underlie higher-order goal-direct behaviours (Miyake et al., 2000; Diamond, 2013). However, it has been argued that this view of EF is reductive (Doebel, 2020; Perone, Simmering, & Buss, 2021).

Doebel (2020) argues for a move away from viewing EF as components that simply support other cognitive domains. It is argued that the typical correlational analyses used cannot be indicative of the role of EF within that domain in the way current literature tends to imply. Instead, Doebel (2020) argues for the consideration of an individuals' knowledge, beliefs, norms, values, and preferences, which may be activated in the presence of specific goals. Here, EF development reflects the use of certain skills to guide behaviour in a context-specific manner. For instance, instead of a separable inhibitory control system, children are able to demonstrate inhibition from utilising skills in control from previous knowledge, values, and beliefs. For example, Doebel (2020) explains that to restrain from hitting a friend, the child will use previous knowledge of what it feels like to be hit, engage the values that we should avoid hitting others, and assess the belief that hitting others leads to being scolded. This aligns with the dynamic systems account of EF that underlies DFT (Perone et al., 2021; Spencer et al., 2024).

Perone et al. (2021) explain that cognition and behaviour are part of a system involving multiple interactive states. Development reflects the ability to enter preferred states, however these are context dependent. For example, as a child begins learning to walk, they engage multiple components involving muscle control and balance capacity. Over time, walking becomes a preferred state of motion, however in certain contexts we return to the previous state of motion, crawling. Consequently, these states are considered in real-time, only existing within a specific moment, but are also historical and previous states can be recreated. In terms of EF, it has been argued that there should be less focus on the separability of components, and more focus on clarifying how this integrated system uses multiple states together to achieve different goals

(Spencer et al., 2024). Here, the proposed components including attention, inhibitory control, and working memory need to be further understood in terms of their integration with higher-level cognitive control systems to enable EF.

In many real world contexts, multiple integrated systems are used to enact behaviours. The relationship between pre-school working memory, EF, and later academic success (Ahmed, 2019) may indicate this. Underlying the ability to flexibly switch between different classroom contexts is the ability to engage a previous state of learning relevant to that particular classroom. For example, when asked a maths question, a child must engage the working memory system to call on and manipulate mathematical equations using the new information given in the question. However, when going to answer this question, the child must also act in response to their knowledge and beliefs of the classroom context, for example this question may have been asked by a teacher who requires you write down the answer and then raise your hand. To do this, the child must engage the motor state of raising their hand and an inhibition state where they know calling out will lead to being scolded and they must wait to be called on. So, whilst working memory was mapped to later academic success, the ability to engage working memory is not solely responsible for this success. Instead, throughout schooling, multiple systems have been integrated to allow specific states to be entered and result in the appropriate behaviour.

Similarly, Hendry and Scerif (2023) argue that individuals have an upper and lower limit of EF ability, and the context of the environment may influence performance in momentary situations. For example, measuring a child's EF when they are motivated by a specific goal and the environment is a good fit to their personal needs may lead to capturing their upper-limit of EF capability. Whereas examining the child when under stress and the child is not motivated by a goal may lead to capturing a lower limit of this EF capability. Consequently, the assumptions we make about EF abilities need to account for the contexts within which they are studied. As stated by Carlson and Zelazo (2022), EF skills do not act in isolation. It seems that knowledge, values, language, motor skills, and specific person-environment contexts must be considered. Specifically, Carlson (2009) indicates that understanding the role of context in terms of these social influences, but also in terms of proximal task-related variables, is imperative to advancing the study of EF.

To summarise this section, a detailed look at possible precursors to EF in infancy reveals, again, a complex literature with few longitudinal studies that begin in infancy and no studies that consistently use the same tasks over time to probe the co-development of candidate 'components' of EF with emerging EF skills. Moreover, there is emerging dissatisfaction with the component perspective, with multiple researchers arguing for a re-conceptualisation of EF skills. In this context, however, it is clear that one aspect of working memory that can be measured early in development, visual working memory, has not been tracked longitudinally along with changes in EF skills. This is a clear gap in the developmental literature. Given that working memory is consistently identified as an important component of EF in adult studies, this will be the focus of this dissertation. Thus, in the next section, I take a more detailed look at the development of visual working memory beginning in infancy.

1.4 Visual working memory in early development.

Visual working memory (VWM) is important for written and verbal communication (Daneman & Merikle, 1996), and also for creativity and innovation (Vandervert et al., 2007). For those with a poorer VWM, deficits are associated with learning difficulties such as Dyscalculia (Szucs et al., 2013). Consequently, VWM is demonstrated to have strong predictive value. As discussed, VWM is also able to be examined from early infancy (Reyes et al., 2020). Consequently, VWM is ideal for early assessment, particularly in relation to the development of EF. However, to date, there are few examinations of the development of VWM from infancy using longitudinal samples, and no longitudinal examinations of this system from infancy into childhood.

In early infancy research, eye-tracking tasks examining looking behaviours are a useful method of examining visual cognition in the absence of any need for language abilities. For example, Papageorgiou et al. (2014) revealed individual differences in infants' looking behaviour, examined by fixation duration, were positively related to parent-report measures of infants' effortful control. Similarly, Blankenship et al. (2019) examined the look duration and shift rate of five-month-old infants to examine infant attention and its relation to EF. Path analyses revealed this infant attention was predictive of EF at 10 months, and EF followed a continuous pattern of development from 10 months to 6 years.

Although several studies have examined links between visual cognition and later EF, it is not always clear which looking tasks in infancy robustly measure visual working memory specifically. One task that was designed to probe this cognitive system in infancy is the preferential looking VWM task designed by Ross-Sheehy and colleagues (2003). Later in development, researchers tend to focus on tasks that estimate working memory capacity (although see Johnson et al., 2014 and Shimi & Scerif, 2021 for an alternative focus on the resolution of VWM). A canonical task here is the change detection paradigm. In this task, children are shown a memory array of, for instance, multiple coloured squares. They must remember the array over a short delay (e.g., 1 second) and then report if a second, test array is the 'same' or 'different'. Capacity estimates are consistently demonstrated to increase with age, and as such, capacity measures appear a good candidate for examining the stability of VWM over time. However, there are differences in capacity based on whether estimates are examined utilising visual or verbal working memory (see Simmering & Perone, 2013 for a review).

More recently, Simmering (2016) examined how the preferential looking measure of VWM was related to the 'change detection' task. Simmering (2016) confirmed that performance across the preferential looking and change detection VWM tasks were related. She then used a dynamic neural field model of VWM development to understand how these tasks were related, showing that the same model of VWM could capture performance from both tasks including cross-task correlations. The present study builds on this work by tracking VWM longitudinally using both VWM tasks.

1.5 Summary.

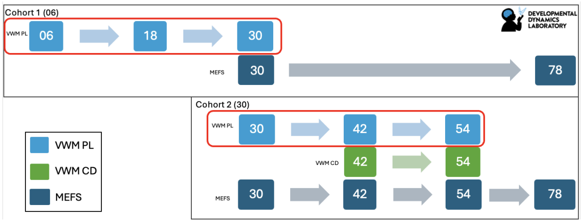
To summarise, there are major challenges in how EF is conceptualised in early development as well as major challenges in assessing EF in early infancy. One clear need is for more longitudinal studies that track candidate components of EF along with the emergence of EF skills. Here, I identified a clear gap in the literature with no prior work relating the early development of VWM to EF development. Thus, the goal of this thesis is to track the co-development of VWM and EF in a longitudinal study.

To bring together the research areas of VWM and EF, we must first understand the development and longitudinal stability of both VWM and EF. In the next chapter, I examine the stability of individual differences in VWM using a preferential looking task in a longitudinal sample. As seen in Figure 1.1a, this will involve investigating two cohorts: the first cohort will start the study at 6 months and completed the preferential looking task at 6, 18, and 30 months of age, and a second cohort will start the study at 30 months and complete the preferential looking task at 30, 42, and 54 months of age. Performance on this preferential looking task will be assessed on a year-by-year basis to determine whether measures from this task are longitudinally stable over this developmental period.

To further enrich our understanding of VWM development, I will also measure performance in the change detection task once children reach the age of 42 months (the youngest age studied by Simmering, 2016). As outlined in Figure 1.1b, children from cohort two will be examined in this change detection task at 42 and 54 months of age. Previous research has related task performance across these two VWM measures in cross-sectional samples. However, no previous research has examined cross-task relationships within the same participants longitudinally, nor has any study examined the stability of visual working memory from early infancy through childhood. Thus, in Chapter 2, after exploring the longitudinal stability of each measure, I will investigate cross-task relationships among measures from these tasks. As seen in Figure 1.2, this will include an exploration of performance on the preferential looking task at 30 months as a predictor for performance on the change detection task at 42 and 54 months of age.

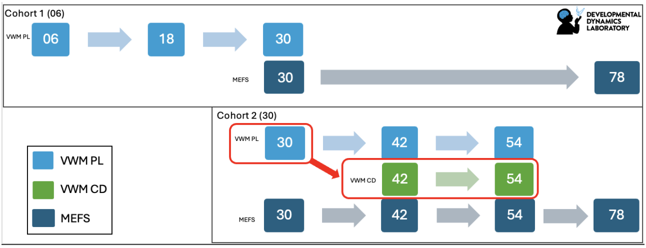
Figure 1.1
Overview of Chapter 2

(a) The longitudinal trajectory of VWM on a preferential looking task (VWM_{PL})



(b) The longitudinal trajectory of VWM on a change detection task (VWM_{CD})

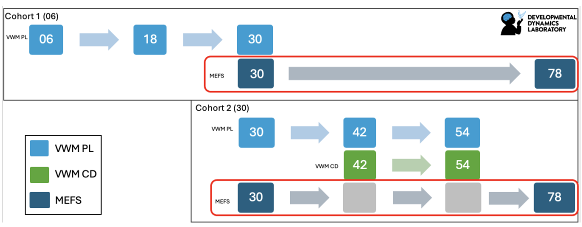
Figure 1.2
Overview of Chapter 2
Does VWM_{PL} predict performance on VWM_{CD}?



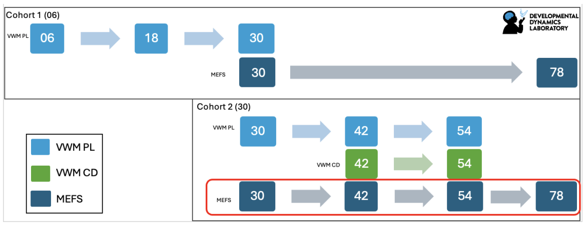
Chapter 3 uses the same longitudinal sample to examine the longitudinal stability of executive function. I explore performance on an executive function task from 30 to 78 months, guided by previous literature that highlights the effects of SES (Hackman et al., 2015) and gender (Palomino & Brudvig, 2022; Yamamoto & Imai-Matsumura, 2019) on EF development. Specifically, as outlined in Figure 1.3, I will examine the trajectory of EF from 30 to 78 months of age across both cohorts. Next, a more fine-grained, step-by-step examination of EF will ensue using data from the second cohort. Here, EF performance will be assessed at 30, 42, 54, and 78 months of age.

Figure 1.3
Overview of Chapter 3

(a) The longitudinal trajectory of EF.



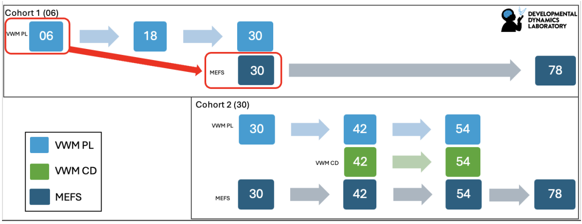
(b) The longitudinal of EF, cohort two



Where Chapter 2 and Chapter 3 will provide insights into the developmental trajectory of VWM and EF separately, Chapter 4 synthesises these trajectories. The literature suggests we must track the co-development of systems. Thus, in Chapter 4, I examine how VWM and EF are related across development. No previous research has examined this issue, particularly using measures obtained in infancy and used consistently over time. As seen in Figure 1.4a, I will investigate performance on the preferential looking VWM task in early infancy at 6 and 18 months of age as separate predictors of EF at 30 months of age. Secondly, as outlined in Figure 1.5, I will investigate performance on the preferential looking task at 30 months of age as a predictor of EF at 78 months of age in both cohorts.

Figure 1.4
Overview of Chapter 4

(a) Does infant VWM at 6 months predict toddler EF?



(b) Does infant VWM at 18 months predict toddler EF?

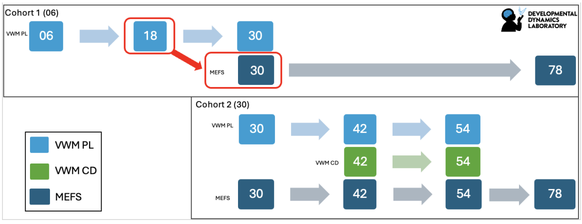
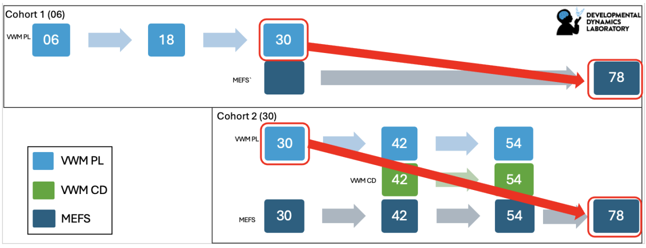


Figure 1.5
Overview of Chapter 4
Does toddler VWM predict child EF?



Lastly, as seen in Figure 1.6, I will investigate performance in the change detection task to examine whether VWM capacity at 42 months of age predicts EF at 54 months of age. This analysis will be extended to examine VWM capacity at 42 months of age and EF at 78 months of age. Next, as outlined by Figure 1.7, I will then examine whether VWM capacity at 54 months of age predicts EF at 78 months of age.

Figure 1.6
Overview of Chapter 4
Does child VWM capacity predict child EF?

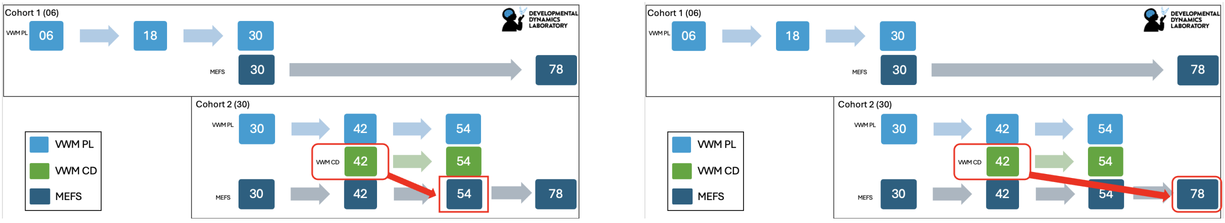
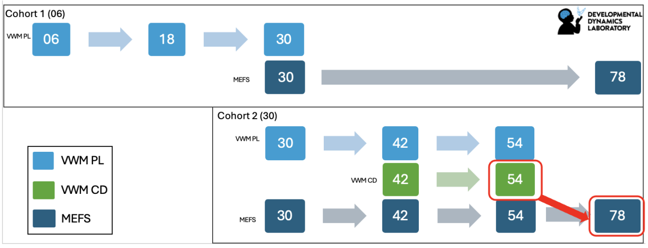


Figure 1.7
Overview of Chapter 4
Does child VWM capacity predict child EF?



In Chapter 5, I combine the findings from this research and discuss their contributions to the literature. I also discuss the real-world implications of our findings and consider how expanding our understanding of infant cognition can contribute to EF interventions.

Chapter 2

The Developmental Trajectory of Visual Working Memory during Early Childhood.

2.1 Introduction

Working memory is a cognitive system deemed to explain one-third to one-half of variance in fluid intelligence (Conway et al., 2003). Moreover, working memory has been established as a key predictor of academic attainment (Jaroslawska et al., 2016; Alloway & Alloway, 2010). Specifically, working memory has been shown as a more powerful predictor of academic achievement than IQ (Alloway & Alloway, 2010). Preschool working memory has similarly been related to later cognitive outcomes, being predictive of math and reading abilities at 15 years of age (Ahemed et al., 2019). Working memory is generally considered to contain a highly limited capacity. Deficits in this working memory capacity are associated with reduced cognitive function, with children with a reduced working memory demonstrating symptoms of inattentiveness, distractibility, and difficulty in problem-solving and monitoring quality-control in academic settings (Gathercole et al., 2008).

Working memory is typically divided into separate verbal and visual-spatial subsystems (Baddeley, 2000), each with a limited capacity (see Cowan, 2001). Within children, individual differences in visual-spatial memory abilities have been shown to predict future academic achievement. Bull et al. (2008) conducted a longitudinal study examining cognition and school-related learning capacity in primary school children. Using the Corsi-blocks task, Bull et al. (2008) established that children's visual-spatial memory ability across the first three years of primary school was a significant predictor of performance in mathematics. This chapter focuses on the early development of a central type of working memory that is part of this visual-spatial subsystem: visual working memory (VWM).

Early measures of VWM show infants begin to be able to maintain mental representations over a short delay from five to eight months of age (Bell, 2012; Pelphrey et al., 2004), and VWM has been analysed in infants as young as four months using change preference tasks (Ross-Sheehy et al., 2003; Wijeakumar et al., 2019). Therefore, VWM is ideal for early assessment. Whilst VWM has been successfully captured in early infancy, the developmental trajectory of VWM from this early infancy period is currently unclear.

There have been few longitudinal studies examining VWM, and those that have been conducted have focused on the mid-to-late childhood period (Heyes et al., 2016; Darki & Klingberg, 2015). Examinations of VWM beginning in infancy are sparse, with only one study examining VWM longitudinally from early infancy (Forbes et al., in prep). It was found that VWM, measured using a preferential looking task (VWM_{PL}), was longitudinally predictive across two separate samples from the UK and India. This task involves children exploring a display with colourful squares on the left and right side of the screen that blink on and off. On one side, one square changes colour after each blink. The ability to detect this change utilises working memory capabilities as children must be able to hold the colours in mind during each blink and update their working memory when changes occur. VWM in year 1 (at 6 and 9 months of age) predicted VWM in year 2 (at 18 and 21 months of age). However, this study only focussed on tracking VWM over a single year period. Thus, an open question is how VWM changes from infancy into early childhood.

Other examinations of the VWM_{PL} task have indicated frontal cortex activation and the inability to suppress distraction may also be related to VWM performance. For example, children show suppression in localised areas of the frontal cortex on more difficult levels of the VWM_{PL} task in a manner that suggests increases in demand require frontal cortex suppression to support distraction suppression (Wijeakumar et al, 2019; McKay et al., 2021). As demand increased, activity in the frontal cortex was suppressed allowing children to sustain attention to the changing side and suppress the non-changing side from capturing their attention. With higher memory loads where more squares are presented on each display, children with a reduced amount of looking to the changing side appear unable to suppress distracting information, shifting back and forth between displays. At the neural level, these children failed to suppress key parts of the frontal cortex. Importantly, Wijeakumar et al. (2019) indicate children from lower educated mothers and lower income backgrounds were more likely to perform poorly on the VWM_{PL} task and more likely to demonstrate this inability to suppress distraction. Given this preferential looking task is able to capture differences in VWM ability in a manner that is sensitive to socioeconomic factors, captures individual differences in VWM from as young as four months (Reyes et al., 2020), and has been shown to be longitudinally predictive (Forbes et al., in Prep), VWM_{PL} is a good candidate for examining VWM from early infancy into childhood. In line with this previous research, maternal education will also be included.

Cross-sectional research later in development has also shed light on the development of VWM using change detection tasks, shown to be effective in capturing VWM capacity levels in both childhood and adulthood (Simmering, 2016). This task involves verbally responding whether a set of stimuli in a test array, such as colourful squares or differing shapes, is the same or different to a memory array presented one second prior to the test array. Buss, Fox, et al. (2014) used a shape-based change detection task and found that three-year-olds had a capacity of 1.2 items, rising to 1.8 items by four-years of age. Simmering (2016) used a colour change detection task to examine similar age-related changes in VWM capacity. Here, three-year-olds showed a VWM capacity of 1.5 to 2 items, raising to 2 to 3 items by five-years of age. Capacity estimates were found to be higher in the same individuals on the VWM_{PL} task relative to capacity estimates from change detection, but performance across both tasks was correlated.

There appears to be a positive relationship between VWM captured by preferential looking in infancy and VWM captured using the widely used change detection task. Simmering (2016) found that a higher VWM demonstrated through a higher level of looking to the changing side on the VWM_{PL} task was positively related to VWM capacity on the change detection task, specifically when accounting for the number of switches between the changing displays on the VWM_{PL} task. Thus, the VWM_{PL} and change detection tasks may be utilised together to successfully track the longitudinal stability of VWM from early infancy onwards.

In order to understand how visual working memory changes over time and the robustness of individual differences in VWM performance, VWM must be examined in a longitudinal sample, starting in infancy. The goal of the present chapter is to conduct the first longitudinal study of VWM across a large longitudinal range, looking at VWM from 6 to 54 months of age across two cohorts of children.

Firstly, the stability of VWM in the VWM_{PL} task will be assessed to answer the first research question: is VWM, examined through preferential looking, longitudinally stable from early infancy? I expect to find stability in individual differences, with VWM performance predicting itself at each year and age-related increases in performance being demonstrated in this task. Next, the stability of VWM in the change detection VWM task will be assessed to answer the second research question: is VWM capacity, examined through change detection, longitudinally stable from early childhood? I expect to replicate capacity estimates from Simmering (2016) and to demonstrate a similar age-related increase in the first longitudinal examination of performance in this task. The first longitudinal cross-task examination of a preferential looking and change detection task will then be completed using the same children from 42 to 54 months of age in both tasks. This will assess the third research question: is VWM performance captured in a preferential looking task related to VWM performance captured in a change detection task? I expect to replicate the positive relationship found by Simmering (2016) between measures of performance on VWM_{PL} and the change detection task. Given the longitudinal nature of this study, the developmental predictability of the VWM_{PL} task will then be examined to answer the fourth research question: do measures of VWM examined through preferential looking predict performance on a change detection task up to two years later? Given the relatedness of the

VWM_{PL} and change detection task concurrently in within-subjects examinations (Simmering, 2016), I expect that VWM_{PL} will be predictive of later change detection, with a higher level of looking to the changing side on VWM_{PL} predicting a higher VWM capacity in the change detection task.

2.2 Methods

2.2.1 Participants

179 children completed the visual working memory tasks. There were two cohorts. Cohort one began the study at six months and were tested at 6, 18, and 30 months-of-age. Cohort two began the study at 30 months and were tested at 30, 42, and 54 months-of-age. Demographics are shown in Table 2.1. Families of 5 children did not complete the demographics questionnaire. Average maternal education level was a Bachelor's Degree, and mean income was £40,404.94 (SD = 11970.63). Participants had normal or corrected-to-normal vision. Colour vision was examined through family history of colour blindness risk; at-risk children were excluded. All participants were full-term infants.

This project was reviewed and approved by the Ethics Committee at NHS England (IRAS ID 196063). Parents signed an informed consent form on behalf of the child. Children received a toy and a t-shirt for participating at each lab visit. Parents were given £20 for each visit to the lab. The data reported here are a subset of a larger study examining the neural basis of visual working memory and attention in early development.

Data counts revealed 3 children did not complete the visual working memory preferential looking eye-tracking task (VWM_{PL}) in the first year of participation, referred to as year 1. There was 1 child in this first year who did complete VWM_{PL} but did not fixate on an area of interest fast enough to calculate the key behavioural measures of interest, and thus this child was removed. There were 38 children who did not complete the VWM_{PL} task in the second year, year 2, and 50 who did not complete VWM_{PL} in the third year, year 3 (see Table 2.2). The higher missing data in years 2 and 3 were due to the Covid-19 pandemic affecting our ability to collect data in these years.

A visual working memory change detection task (VWM_{CD}) was administered to cohort two only. Of the 85 children in cohort two, data counts revealed 18 children who did not complete the VWM_{CD} task in any year. Of the 67 children who did complete this task, 9 only completed the VWM_{CD} in year 2 and not year 3, and a different 9 only completed the task in year 3, not in year 2 (see Table 2.3). Data counts revealed that every child who completed the VWM_{CD} task at 42 and 54 months of age also completed the VWM_{PL} task at each year of participation, from 30 to 54 months of age (see Table 2.4).

Table 2.1
Demographic Characteristics

Variable	Cohort One N = 89	Cohort Two N = 85
Gender		
Boys	46 (52%)	41 (48%)
Girls	43 (48%)	44 (52%)
Maternal Education level		
Left School before 16	2 (2.2%)	1 (1.2%)
GCSE/O Levels or equivalent	10 (11%)	14 (16%)
A Levels or equivalent	10 (11%)	11 (13%)
Trade Apprenticeship	1 (1.1%)	5 (5.9%)
Some University	6 (6.7%)	9 (11%)
Bachelor's Degree	41 (46%)	30 (35%)
Master's Degree	13 (15%)	11 (13%)
Doctorate or Professional Degree	6 (6.7%)	4 (4.7%)
Ethnicity		
White British	75 (84%)	76 (89%)
Asian	1 (1.1%)	0 (0%)
Black African	0 (0%)	1 (1.2%)
South African	2 (2.2%)	0 (0%)
White British and Arabic	1 (1.1%)	0 (0%)
White British and South American	2 (2.2%)	0 (0%)
White British and Asian	4 (4.5%)	2 (2.4%)
White European and Asian	1 (1.1%)	0 (0%)
White British and Black African	0 (0%)	2 (2.4%)
White British and Black Caribbean	0 (0%)	2 (2.4%)
White British and Other European	3 (3.3%)	2 (2.4%)

Table 2.2
$\mathbf{VWM_{PL}}$ data counts and descriptive statistics.

	Year 1			Year 2			Year 3		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
CP_C	175	0.65	0.12	141	0.65	0.12	129	0.63	0.10
CP_{NC}	175	0.43	0.14	141	0.43	0.12	129	0.42	0.13
CP_{10}	176	0.53	0.05	141	0.53	0.05	129	0.52	0.05
Total Participants	**176**			**141**			**129**		
Total N	179								

Table 2.3
$\mathbf{VWM_{CD}}$ data counts and descriptive statistics.

		Year 2			Year 3		
Variable		**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
K_{MAX}		58	1.38	0.67	58	2.08	0.68
A prime	SS1	55	0.85	0.18	58	0.96	0.07
	SS2	53	0.70	0.25	58	0.91	0.09
	SS3	46	0.55	0.25	57	0.78	0.18
Total Participants		**58**			**58**		
Total N	67						

Table 2.4
$\mathbf{VWM_{PL}}$ and $\mathbf{VWM_{CD}}$ data counts and descriptive statistics, cohort two.

	Year 1			Year 2			Year 3		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
K_{MAX}				58	0.72	0.72	58	2.07	0.68
CP_C	67	0.66	0.10	58	0.64	0.18	58	0.63	0.16
CP_{NC}	67	0.42	0.12	58	0.44	0.19	58	0.41	0.19
Total Participants	**67**			**58**			**58**		
Total N	67								

2.2.2 Procedure

Cohort one completed only a preferential looking visual working memory task. Cohort two completed both a preferential looking and a change detection visual working memory task.

Participants were seated and asked whether they would like to watch a clip from Peppa Pig, Dinosaur Train, or Paw Patrol. The selected clip was shown while the experimenter placed an fNIRS cap over the participant's head and a small target sticker on the participant's forehead. Note that the fNIRS data will not be analysed within this thesis. The eye-tracking camera was then adjusted so that the pupil was in focus. Before all eye-tracking tasks, a calibration procedure took place once eye-tracking adjustments had been completed. During calibration, participants were shown a black and white geometric shape looming in five locations across the screen - middle, top, bottom, left, and right. This mapped the raw eye position data to the camera image data allowing the mapping of gaze position to stimulus presentation. Once a successful calibration had been recorded, the experiment started.

Participants attended the lab twice each year for three years. All participants completed a preferential looking visual working memory task (VWM_{PL}) at each visit to the lab. This task involved watching a screen with colourful squares blinking on and off on left and right displays.

One display was randomly selected as the "change display", and on this side the squares changed colour after each blink. On the no change display, the colour of the squares remained constant for the entire trial. Three load levels were presented. At six months these were: low load, one square on each side; medium load, two squares on each side; high load; three squares on each side. From 18 months these were: low load, two squares on each side; medium load, four squares on each side; high load, six squares on each side. These changes in load level are age appropriate and allow us to detect visual working memory changes across development. The colour of the squares was randomly selected from nine colours: green, black, violet, brown, cyan, yellow, blue, red, and white. The colours were always different from one another on a single display, but colours could be repeated to appear on both displays simultaneously (i.e. the same colour could appear on both the change and no change side). The squares appeared for 500ms and disappeared for 250ms. Within this task, participants are expected to engage their working memory to hold the colours of squares in mind, enabling the participant to detect that one side is changing and focus on this changing display due to its novelty. Figure 2.1 shows a schematic of the visual working memory trials.

At 42 and 54 months, cohort two also completed a change detection task (VWM_{CD}) during one of their visits to the lab. First, children were familiarised with the task using flashcards. A flash card showing a memory array of colourful squares was presented. There were three set sizes, SS1 showed one square, SS2 showed two squares, and SS3 showed three squares. The children were asked to "Remember the colours". A flashcard showing a test array was then presented, in which either all of the squares matched the memory array (no-change trials), or the colour of one square changed (change trials). The squares were presented in the same spatial location on the memory and test array, despite a colour change. If a change were to occur, only one square changed regardless of set size. The children were asked if the test array card matched the memory array presented prior. They were instructed to verbally respond with "same" or "different". The experimenter either confirmed the correct response e.g. "Yes, the cards are different, this square changed from red to green", or explained an incorrect response e.g. "These squares were different, see the red square here changed to green". Familiarisation flashcards were presented in a specific order: SS1, no change; SS1, change; SS2, no change; SS2, change;

SS3, no change; SS3, change. Once familiarised, the children watched a screen presenting the memory array for 2000ms. After a 900ms delay, the test array was presented until a response was received (see Figure 2.2). Memory arrays were manipulated so that the memory and test array were presented on the left on one trial. Then, on the next trial, the memory and test arrays were presented on the right side of the screen. This helped children identify which arrays to compare. A gaze contingent fixation cross was presented on the side the arrays would appear at the start of each trial to ensure children were focusing on the correct side. The experimenter recorded the child's verbal response on a keyboard by pressing one key for same and one key for different.

Figure 2.1
Schematic demonstrating the VWM_{PL} task.

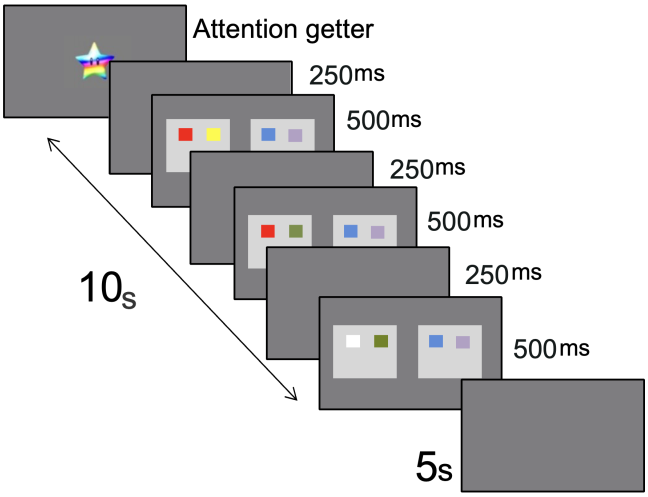
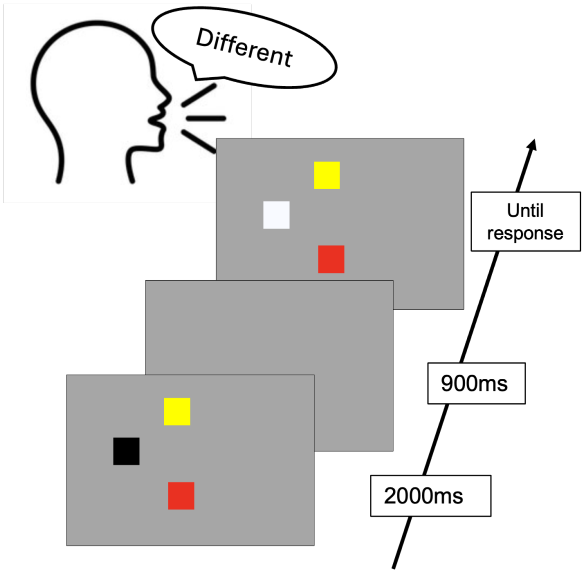


Figure 2.2
Schematic demonstrating the VWM_{CD} task.



2.2.3 Materials

For all eye-tracking tasks, a 42-inch LCD television was connected to a PC running SR Research Experiment Builder to display stimuli to participants. Participants were seated on a caregiver's lap or in a high-chair approximately 100cm from the television screen. An Eye-Link 1000 plus eye-tracker (SR Research, Ontario, Canada) was placed on a small stand approximately 60 to 70cm from the participant. To track the participant's head and eye position despite participant movement, a small target sticker was placed on each participant's forehead. The eye-tracker used monocular recording, tracking the gaze position of a single eye using pupil and corneal reflections of an infrared light source, at a sampling rate of 500Hz. An additional camera was located to the left of the eye-tracker on the same stand. This recorded the participants face. A further camera was located behind participants' heads in the ceiling at the back of the room. This recorded the experiment as displayed on the television monitor.

2.2.4 Methods of Analysis

Eyetracking data were exported using SR research DataViewer on a frame-by-frame basis. Following Spencer et al. (2023), for any cases with no recorded eye tracking data, data was hand coded. All looking data were processed in R using the eyetrackingR package (Dink and Ferguson, 2021). Trials were excluded in cases where more than 75% of the data were classed as not looking to the screen.

The VWM_{PL} task is usually analysed using a standard change preference score, CP_{10}, at each load level (see Ross-Sheehy et al., 2003). This is the proportion of looking to the changing side divided by the total look duration to both the change and no change side across the full 10 second time window. However, Forbes et al. (in prep) reported that this measure is not effective in capturing individual differences in visual working memory longitudinally. Instead, in-depth analyses revealed two new measures of performance which will be used in the present study. These new measures consider the location of the first fixation the child makes when the first change array is presented, as different demands are placed on the visual working memory system depending on the side the child first looks to. In line with Forbes et al. (in prep) and Spencer et al. (2023), infants' looking data were sorted into the two types of trials based on where they were looking at the onset of the first change (at 1000 ms). As over 13 per cent of trials had missing data within the time window from 1000-1100ms, the 'first- look' classification in Spencer et al. (2023) was determined based on the first frame of non-missing eye-tracking data up to 2500ms. This reduced missing data. In addition, these two prior studies also trimmed data from the last few seconds of each trial, focusing on the time window from 1750ms to 6750ms. These same timing parameters were used here. The two change preference scores were then calculated similarly to CP_{10} by dividing the proportion looking to the change side by the total look duration within this time window. The two 'first-look' based measures will be referred to as CP_{NC} and CP_C.

CP_{NC} is the proportion of looking to the changing side when the child's first look is to the non changing side. This measure takes into account the additional demands placed on the visual working memory system when starting on the non changing side. The child must first detect 'no change' by recognising the colours are the same, that is, the child must hold the colours in mind, compare the colours presented on the next flash, and detecting this 'sameness'. The child should then release fixation due to a lack of novelty, and switch to the changing side where they must again utilise their working memory to detect that a change is occurring.

CP_C is the proportion of looking to the changing side when the child's first look is to the changing side. This measure captures the child's ability to sustain attention to novelty. Here, the child must hold the colours displayed in their working memory, compare these to the colours presented on the next flash, and detect the colour change. Due to children's novelty preference, the children should find this changing side more engaging and sustain attention to it.

The present study will also examine total looking time (TLT), as this has previously been demonstrated as an important visual exploration measure within this task (Forbes et al., in prep; Wijeakumar et al., 2023). This is the length of time the child looked to either display across each 10 s trial. Switch rate will be examined in order to replicate analyses conducted by Simmering (2016). This is the number of shifts in fixation from one side of the screen to the other, divided by the total looking time in seconds, to provide a number of switches per second.

All VWM_{PL} measures will be aggregated over load level. Forbes et al. (in prep) found load level was only related to CP_{NC}, with no load effects found for the CP_C measure. I aggregate over load here to reduce missing data given the focus is to understand longitudinal changes in visual working memory. Note that preliminary analyses showed similar results with versus without aggregation over loads.

To analyse the VWM_{CD} task, two scores are examined. Firstly, I calculate A' (A prime) which aggregates over correct and incorrect responses across change and no change trials (see Simmering, 2016). Change trial responses are classified into hits and misses. No-change trial responses are classified into correct rejections and false alarms. A prime was calculated using the updated formula from Aaronson and Watts (1987):

If H ≥ FA:

$$A' = 1/2 + [(H - FA) * (1 + H - FA)]/[4 * H * (1 - FA)]$$

If H < FA:

$$A' = 1/2 - [(FA - H) * (1 + FA - H)]/[4 * FA * (1 - H)]$$

This measure only uses the hits (correctly detected changes) and false alarms (incorrectly detected changes i.e. the child reported a change when no change occurred). As such, the A prime score provides us with an aggregate score measuring sensitivity to change whilst allowing for instances where false alarms exceed number of hits.

The second score is a capacity measure using the Pashler's K formula (Pashler, 1988). I first calculated capacity for each set size:

$$K = SS * (H - FA)/(1 - FA)$$

This formula is appropriate for our variant of the change detection paradigm using a whole-array test as opposed to presenting a single item at test (Rouder, Morey, Morey, & Cowan, 2011). The highest estimate across set sizes is used at the maximum capacity estimate (Simmering, 2016). This is the K_{MAX} score.

All participants were included in initial analyses. At each stage, participants were removed for missing data in the variable being analysed at the year being predicted. The stability of visual working memory was first examined using the CP_{10}, CP_C, CP_{NC}, and TLT measures from the VWM_{PL} task across both cohorts. The stability of visual working memory will also be examined within the canonical VWM_{CD} task from 42 to 54 months. Concurrent relationships between VWM_{PL} and VWM_{CD} will then be examined before the 30 month VWM_{PL} task data from cohort two will be used to examine whether individual differences in the VWM_{PL} task at 30 months of age predict individual differences in performance at 42 and 54 months of age in the VWM_{CD} task.

Where maternal education level was included in models, it was entered as a scaled numerical variable. Here, a maternal education of "left school at or before 16" was entered as 1, "GCSE/O levels or equivalent" as 2, "A Levels or equivalent" as 3, "Trade Apprenticeship" as 4, "Some University" as 5, "Bachelor's Degree" as 6, "Master's Degree" as 7, and "Doctorate or Professional Degree" as 8. Gender was also scaled to create a numerical variable, with boys being entered as - 0.5 and girls as 0.5 in all models.

Where models assessed one outcome variable per participant (i.e., non-nested scores and no random effects), linear models were run using the lm function from the R package (R. C. Team, 2021). In these models, scores were split by year before being entered into the model. For example, performance in year 1 was added as a predictor of performance in year 2. Consequently, the outcome variable only contained one score per participant and year was not needed as a predictor within the model. Maternal education level, gender, and cohort were added to these models as fixed predictors. The summary function from the R package (R. C. Team, 2017) was used to provide regression coefficients. The variance inflation factor (VIF) calculated from the car package in R (Fox & Weisberg, 2019) is also reported to evaluate possible multicollinearity. For significant predictors, the estimated magnitude and direction of the effect are reported.

Where outcome variables contained nested scores, such as where scores on the outcome variable over time were predicted by year within the model, linear mixed effect models specifying participant ID as a random effect were run using the lmer function from the Lme4 package in R (Bates et al., 2015). Here, the outcome variable contained multiple scores per participant, as scores at each year or set size were not separated as unique predictors. These variables were entered as levelled factors. Within R, levelling the data provides the model with a reference level at which any change in the outcome variable will be examined relative to. For instance, year was entered as a predictor with 3 levels (year 1, year 2, and year 3) with year 3 as the reference. Here year 3 was selected as the reference level due to this being the level with the expected highest performance. Participant ID was always the grouping variable. To indicate significance, p values were calculated using Satterthwaite's method from the R package LmerTest (Kuznetsova et al., 2017). A type III Wald Chi-squared test from the car package in R (Fox & Weisberg, 2019) was used to assess the contribution of each parameter in reducing residual deviance of the model. Where additional fixed effects were explored and model comparison needed, an ANOVA and Akaike's Information Criterion (AIC; Wagenmakers & Farrell, 2004) were used to compare models.

Normality was assessed by examining residuals from the DHARMa R package (Hartig, 2024) producing Q-Q plots and DHARMa residuals. In all cases, final models had well-distributed residuals. This suggests normality assumptions were not violated.

2.3 Results

2.3.1 The longitudinal stability of visual working memory examined through preferential looking ($\mathbf{VWM_{PL}}$)

To answer the first research question of whether there is longitudinal stability in VWM assessed through preferential looking, measures from the VWM_{PL} task were modelled across all ages. The following analyses were conducted with both cohorts one (enrolled at 6 months) and two (enrolled at 30 months). As analyses were conducted cross-cohort, year refers to the year of participation in the study. Cohort was included in all models as a covariate, with ages at each year as follows: for cohort one, 6 months of age in year 1, 18 months in year 2 and 30 months in year 3; for cohort two, 30 months of age in year 1, 42 months in year 2, and 54 months in year 3.

$\mathbf{CP_{10}}$ measure

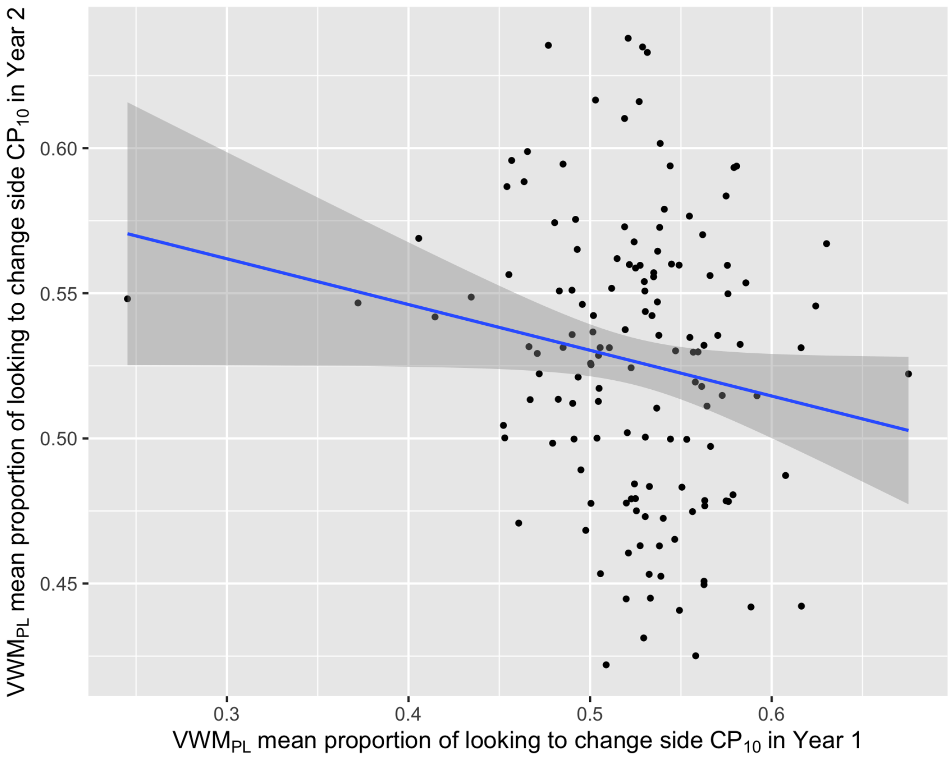
The first question was whether there was longitudinal stability in the canonical measure of the VWM_{PL} task, the CP_{10} measure, over time. This question was assessed in two stages. Firstly, I examined whether individual differences in year 1 predicted year 2. CP_{10} was averaged over loads to produce a single aggregate score for each child. A linear model predicting CP_{10} in year 2 as a function of CP_{10} in year 1, maternal education level and gender was run. Cohort was entered as a main effect covariate to control for possible overall differences related to age cohort. All predictors were added as fixed effects. 37 participants were excluded for not having a CP_{10} in both year 1 and year 2. 138 participants were included in the model, run as:

$$lm(CP_{10}[year2] \sim CP_{10}[year1] \times Gender \times Maternal\,Education\,level + Cohort)$$

The overall regression model was not statistically significant (R^2 = [0.06], F(8,130) = 1.01, p = .43), explaining only 5.8% of the variance (adjusted R^2=0.00, RSE[residual standard error] = 0.05).

Despite the non-significant overall model, mean CP_{10} in year 1 (M = 0.52, SD = 0.05) significantly predicted mean CP_{10} in year 2 (M = 0.53, SD = 0.05) , β = - 0.21, 95% CI [- 0.39, - 0.03], p=.0246 (see Table 1). As seen in Figure 2.3, a higher mean CP_{10} in year 1 predicted a lower mean CP_{10} in year 2. This effect demonstrated a small to medium effect size, d = 0.37. There was no significant main effect of age, and no significant main or interaction effects of maternal education level or gender.

Figure 2.3
Graph showing mean CP_{10} in year 1 and year 2 across both cohorts.



Note: The figure indicates two possible outliers with a mean CP_{10} in year 1 below 0.3. Results remained the same upon removing CP_{10} scores below 0.3. As results and assumptions were not influenced, these data points were not deemed as needing removal.

Secondly, I examined whether individual differences in CP_{10} in year 2 predicted individual differences in CP_{10} in year 3 using a linear model predicting CP_{10} in year 3 as a function of CP_{10} in year 2, maternal education level and gender. Cohort was entered as a main effect covariate to control for possible overall differences related to age cohort. 51 participants were excluded for not having a CP_{NC} in both year 2 and year 3. 124 participants were included in the model, run as:

$$lm(CP_{10}[year3] \sim CP_{10}[year2] \times Gender \times Maternal\,Education\,level + Cohort)$$

The overall regression was not statistically significant (R^2 = [0.02], F(8,115) = 0.36, p = .94), explaining only 2.4% of the variance (adjusted R^2= - 0.04, RSE 0.05). As seen in Table 2, there were no significant main or interaction effects of mean CP_{10} in year 2. Furthermore, there were no significant main effects of cohort, and no significant main or interaction effects of maternal education level or gender.

First-look based measures

Next, I examined whether the 'first-look' based measures, CP_{NC} and CP_C, demonstrated longitudinal stability. This question was addressed in two models, firstly whether individual differences in year 1 predicted individual differences in year 2, and secondly whether individual differences in year 2 predicted individual differences in year 3.

CP$_{\mathbf{NC}}$

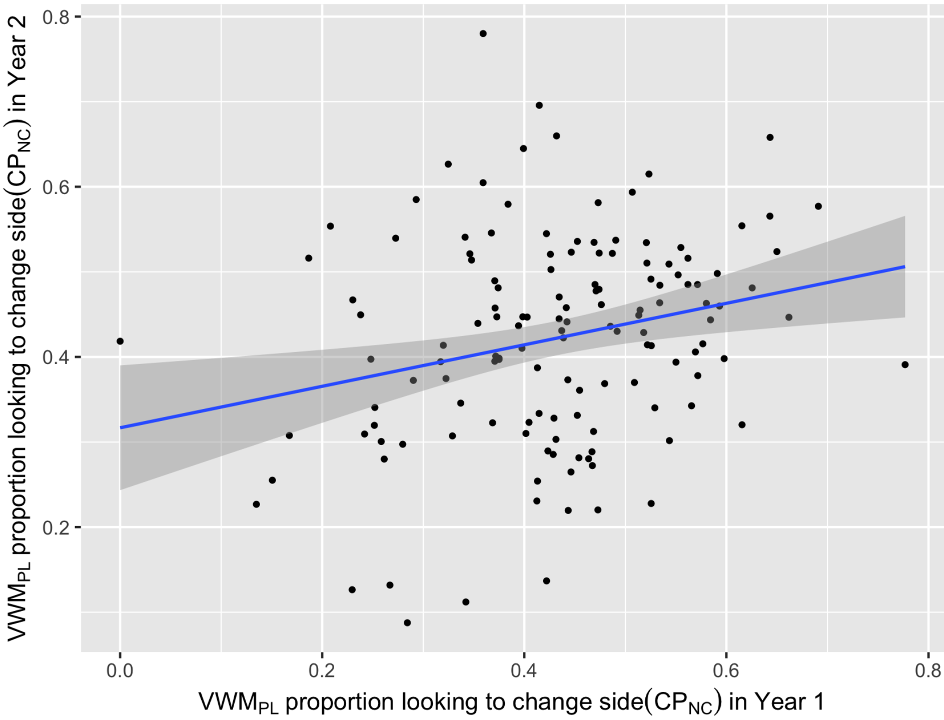
Mean CP_{NC} averaged over loads was used to produce a single aggregate score in each year for each child. A linear model predicting CP_{NC} in year 2 as a function of CP_{NC} in year 1, maternal education level and gender was run. Cohort was entered as a main effect covariate to control for possible overall differences related to age cohort. 37 participants were excluded for not having a CP_{NC} in both year 1 and year 2. 138 participants were included in the model, run as:

$$lm(CP_{NC}[year2] \sim CP_{NC}[year1] \times Gender \times Maternal\,Education\,level + Cohort)$$

The overall regression was not statistically significant (R^2 = [0.07], F(8,129) = 1.28, p = .26), explaining only 7.4% of the variance (adjusted R^2= 0.02, RSE = 0.122).

Although the overall model was not significant, mean CP_{NC} in year 1 (M = 0.43, SD = 0.12) significantly predicted mean CP_{NC} in year 2 (M = 0.43, SD = 0.12), β = 0.24, 95% CI [0.07, 0.42], p=.00663 (see Table 3). As seen in Figure 2.4, a higher mean CP_{NC} in year 1 predicted a higher mean CP_{NC} in year 2. A moderate effect size indicated this to be meaningful despite the non-significant overall model, d = 0.51. There was no significant main effect of age, and no significant main or interaction effects of maternal education level or gender.

Figure 2.4
Graph showing mean CP_{NC} in year 1 and year 2 across both cohorts.



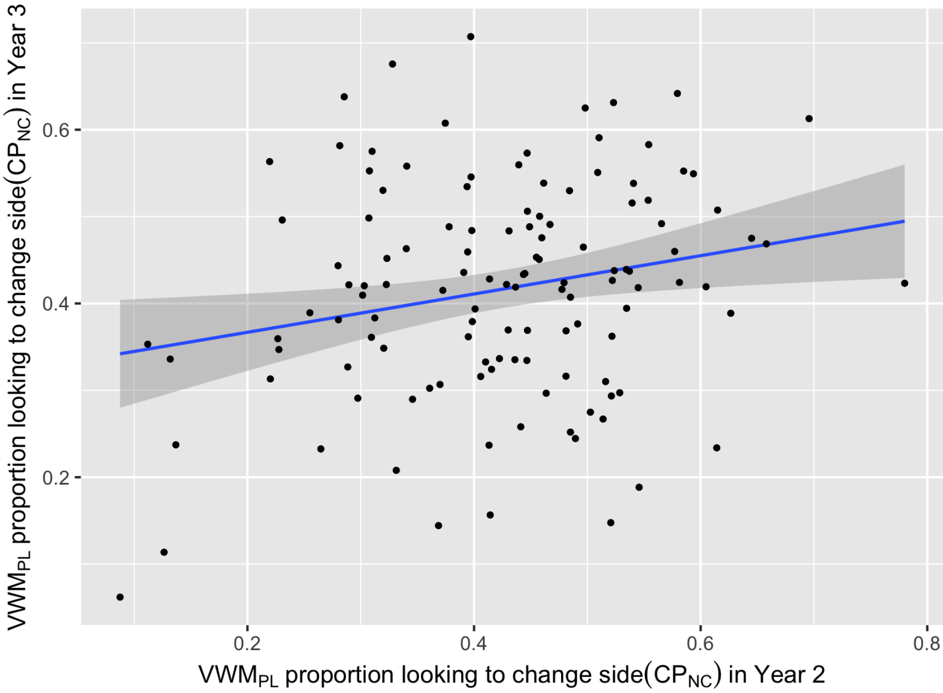
Next, I examined whether individual differences in CP_{NC} in year 2 predicted individual differences in CP_{NC} in year 3. A linear model predicting CP_{NC} in year 3 as a function of CP_{NC} in year 2, maternal education level and gender was run. Cohort was entered as a main effect covariate to control for possible overall differences related to age cohort. 51 participants were excluded for not having a CP_{NC} in both year 2 and year 3. 124 participants were included in the model, run as:

$$lm(CP_{NC}[year3] \sim CP_{NC}[year2] \times Gender \times Maternal\,Education\,level + Cohort)$$

The overall regression was not statistically significant (R^2 = [0.08], F(8,115) = 1.31, p = .25), explaining only 8.3% of the variance (adjusted R^2= 0.02, RSE = 0.123).

Even with the non-significant overall model, mean CP_{NC} in year 2 (M = 0.42, SD = 0.13) significantly predicted mean CP_{NC} in year 3 (M = 0.42, SD = 0.13), β = 0.23, 95% CI [0.05, 0.40], p=.0123 (see Table 4). As seen in Figure 2.5, a higher mean CP_{NC} in year 2 predicted a higher mean CP_{NC} in year 3. A moderate effect size indicated this to be meaningful despite the non-significant overall model, d = 0.47. There was no significant main effect of age, and no significant main or interaction effects of maternal education level or gender.

Figure 2.5
Graph showing mean CP_{NC} in year 2 and year 3 across both cohorts.



Overall, CP_{NC} showed robust longitudinal individual differences. Individual differences in mean CP_{NC} in year 1 predicted individual differences in mean CP_{NC} in year 2, and individual differences in mean CP_{NC} in year 2 predicted individual differences in mean CP_{NC} in year 3.

$\mathbf{CP_C}$

CP_C was assessed for longitudinal stability in a similar manner. Mean CP_C averaged over loads was used to produce a single aggregate score in each year for each child. A linear model predicting CP_C in year 2 as a function of CP_C in year 1, maternal education level and gender was run. Cohort was entered as a main effect covariate. 37 participants were excluded for not having a CP_C in both year 1 and year 2. 138 participants were included in the model, run as:

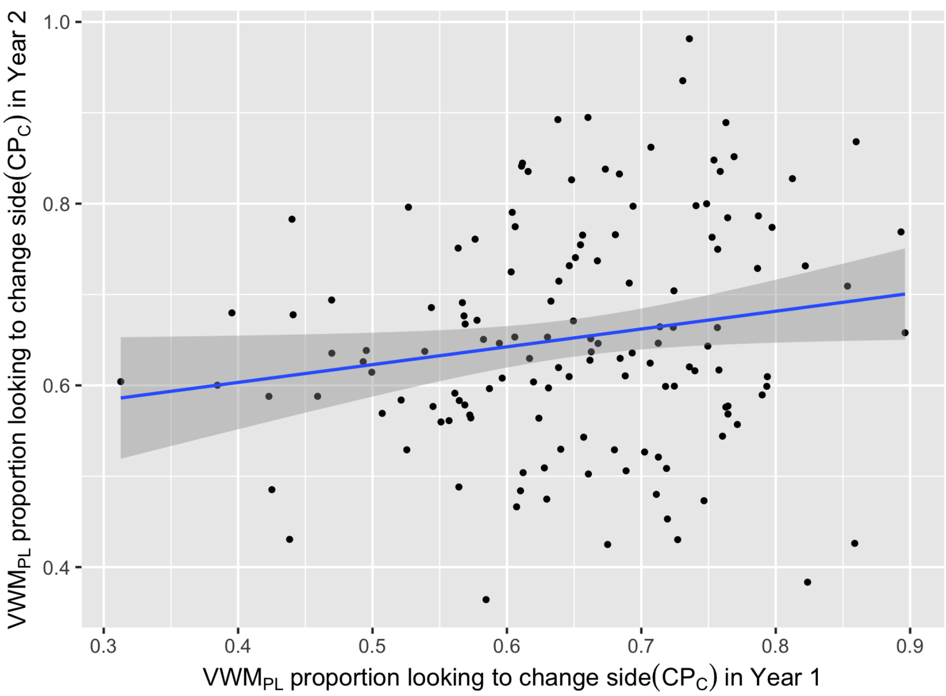
$$lm(CP_C[year2] \sim CP_C[year1] \times Gender \times Maternal\,Education\,level + Cohort)$$

The overall regression was not statistically significant (R^2 = [0.05], F(8,129) = 0.88, p = .53), explaining only 5.1% of the variance (adjusted R^2= - 0.01, RSE = 0.124).

Even with the non-significant overall model, mean CP_C in year 1 (M = 0.65, SD = 0.11) significantly predicted mean CP_C in year 2 (M = 0.65, SD = 0.12), β = 0.22, 95% CI [0.02, 0.42], p=.0292 (see Table 5). As seen in Figure 2.6, a higher mean CP_C in year 1 predicted a higher mean CP_C in year 2. A moderate effect size indicated this to be meaningful despite the non-significant overall model, d = 0.45. There was no significant main effect of age, and no significant main or interaction effects of maternal education level or gender.

Figure 2.6

Graph showing mean CP_C in year 1 and year 2 across both cohorts.



CP_C was then examined for stability from year 2 to year 3. A linear model was conducted predicting CP_C in year 3 as a function of CP_C in year 2, maternal education level and gender. Cohort was entered as a main effect covariate. 51 participants were excluded for not having a CP_C in both year 2 and year 3. 124 participants were included in the model, run as:

$$lm(CP_C[year3] \sim CP_C[year2] \times Gender \times Maternal\,Education\,level + Cohort)$$

The overall regression was not statistically significant (R^2 = [0.09], F(8,115) = 1.59, p = .14), explaining only 9.9% of the variance (adjusted R^2= 0.04, RSE = 0.096). As seen in Table 6, there were no significant main or interaction effects of mean CP_C in year 2. There were also no significant main effects of age, and no significant main or interaction effects of maternal education level or gender.

In summary, the longitudinal stability of CP_C varied. Individual differences in mean CP_C in year 1 predicted individual differences in mean CP_C in year 2; however, this was not the case moving from year 2 to year 3.

TLT measure

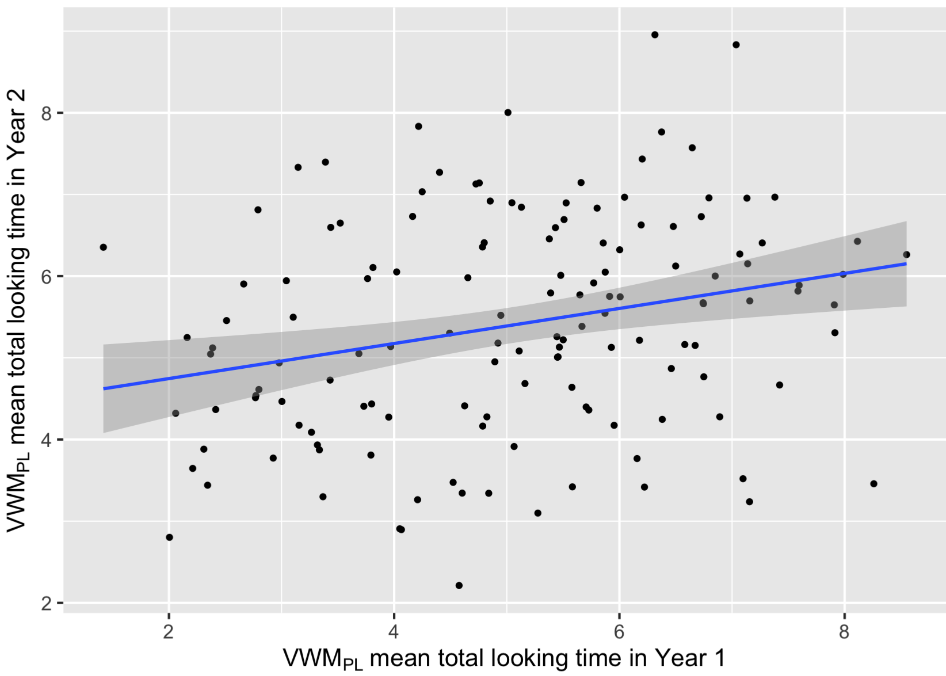
The third question was whether there was longitudinal stability in visual exploration. This was assessed using the total looking time (TLT) measure. Mean TLT averaged over loads was used to produce a single aggregate score in each year for each child. A linear model predicting TLT in year 2 as a function of TLT in year 1, maternal education level and gender was run. Cohort was entered as a main effect covariate. 36 participants were excluded for not having a TLT in both year 1 and year 2. 139 participants were included in the model, run as:

$$lm(TLT[year2] \sim TLT[year1] \times Gender \times Maternal\,Education\,level + Cohort)$$

The overall regression was partially significant (R^2 = [0.10], F(8,130) = 1.59, p = .0689), explaining 10% of the variance (adjusted R^2= 0.05). The residual standard error was 1.32.

Mean TLT in year 1 (M = 5.06, SD = 1.62) significantly predicted mean TLT in year 2 (M = 5.41, SD = 1.34), β = 0.25, 95% CI [0.10, 0.41], p=.00173 (see Table 7). As seen in Figure 2.7, higher mean TLT in year 1 predicted a higher mean TLT in year 2. This effect was found to be of a moderate size indicating a meaningful effect even with the overall regression being only partially significant, d = 0.54. There was no significant main effect of age, and no significant main or interaction effects of maternal education level or gender.

Figure 2.7
Graph showing mean TLT in year 1 and year 2 across both cohorts.



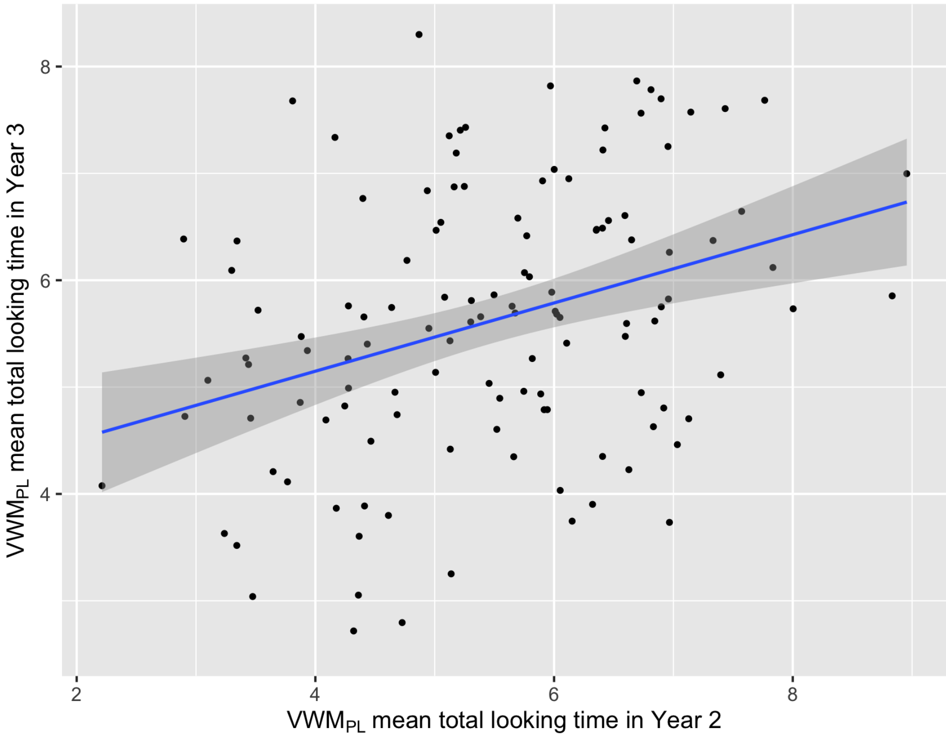
Next, I conducted a linear model predicting TLT in year 3 as a function of TLT in year 2, maternal education level and gender. Cohort was entered as a main effect covariate. 51 participants were excluded for not having a TLT in both year 2 and year 3. 124 participants were included in the model, run as:

$$lm(TLT[year3] \sim TLT[year2] \times Gender \times Maternal\,Education\,level + Cohort)$$

The overall regression was significant (R^2 = [0.17], F(8,115) = 1.59, p = .004156), explaining 17% of the variance (adjusted R^2= 0.12). The residual standard error was 1.18.

Mean TLT in year 2 (M = 5.47, SD = 1.33) significantly predicted mean TLT in year 3 (M = 5.60, SD = 1.29), β = 0.28, 95% CI [0.12, 0.45], p=.00101 (see Table 8). As seen in Figure 2.8, a higher mean TLT in year 2 significantly predicted a higher mean TLT in year 3. This effect demonstrated a moderate effect size, d = 0.62. There were no significant main effects of age and no significant main or interaction effects of maternal education level or gender

Figure 2.8
Graph showing mean TLT in year 2 and year 3 across both cohorts.



Overall, then, TLT showed robust longitudinal stability. Mean TLT in year 1 predicted mean TLT in year 2, and mean TLT in year 2 predicted mean TLT in year 3.

2.3.2 The longitudinal stability of visual working memory examined through change detection ($\mathbf{VWM_{CD}}$)

The VWM_{CD} task is a commonly used measure of VWM, providing two measures of performance. The first is A prime, a sensitivity to change measure examined across set sizes. The second is K_{MAX}, a maximum capacity measure (over set sizes). Critically, examinations of this task across different ages have all been cross-sectional in early development. Therefore, here I asked if these key measures of performance from this task show robust individual differences in early development.

The following analyses contain data from cohort two only. Here, year 1 refers to 30 months of age, year 2 refers to 42 months, and year 3 refers to 54 months.

What influences VWM$_{\text{CD}}$ performance within a longitudinal sample?

Following Simmering (2016), I first examined A prime. Previous results from this cross-sectional study showed an interaction between age and set size, with lower accuracy in set size 3 than set sizes 1 and 2, and a larger decrease from set size 1 to set size 3 for 4-year-olds than 3- and 5-year-olds. A prime scores also increased overall from 3 to 4 years of age. Next, I examined K_{MAX}. Previous estimates of K_{MAX} indicate an age related increase, with 3-year-olds' VWM capacity to be 1.90 items and 4-year-olds' VWM capacity to be 2.20 items (Simmering, 2016).

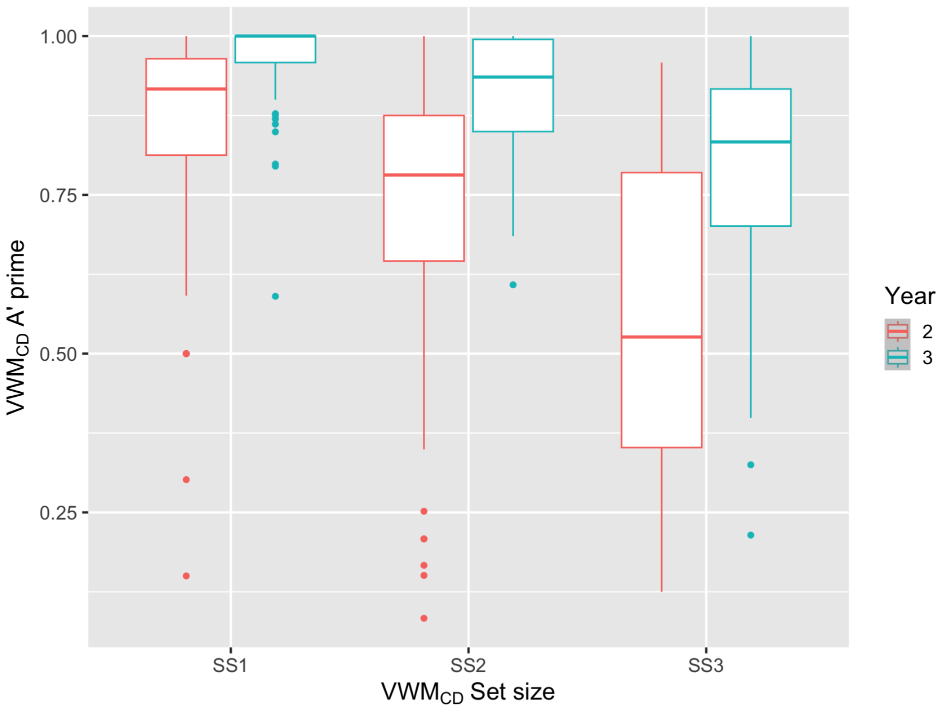
A prime

To answer the question of whether performance on VWM_{CD} is similar within our longitudinal sample relative to Simmering (2016), A prime was examined in a linear mixed effects model containing year and set size. This model included year as a predictor to replicate the inclusion of age in cross-sectional models. Models also included a random effect for participant. Year and set size were included as fixed effects, and both were levelled as factors. Within R, levelling the data provides the model with a reference level at which any change in the outcome variable will be examined as relative to. Here, the reference level was performance in year 3 in set size 3. Here year 3 set size 3 was selected as the reference level due to this being the level with the expected highest performance within the hardest set size. 67 participants were included in the model, run as:

$lmer(A'\,prime \sim Year \times Set\,Size + (1|Participant\,code))$

A Wald Chi-square test indicated a significant interaction between year and set size ($X^2(2)$ = 8.69, p = .01295). As seen in Figure 2.9, accuracy decreased across set sizes, with a larger decrease across set sizes occurring in year 2, at 42 months of age, than in year 3, at 54 months of age. Individually, both year ($X^2(1) = 137.47$, $p < .001$) and set size ($X^2(2) = 26.71$, $p < .001$) significantly explained variance in A prime score (see Table 9 for regression coefficients and effect sizes). These results replicate the age by set size interaction demonstrated in Simmering (2016).

Figure 2.9
Graph showing A prime score across set sizes in year 2 and year 3.



Next, I ran a second linear mixed effects model with the addition of maternal education level and gender as fixed effects. Due to the increased number of predictors, year and set sizes were scaled to be included as numeric variables. This reduced the overall number of interaction comparisons as year and set size were entered into each interaction once, not for each level. 67 participants were included in the model, run as:

$$lmer(A'\,prime\ \sim\ Year \times Set\,Size \times\ Maternal\,Education\,level \times\ Gender + (1|Participant\,code))$$

A Wald Chi-square test indicated the previous interaction between year and set size remained ($X^2(1) = 6.73$, $p = .009461$). Individually, year ($X^2(1) = 132.65$, $p < .001$), set size ($X^2(1) = 167.86$, $p < .001$), and maternal education ($X^2(1) = 7.64$, $p = .005704$) significantly explained variance in A prime scores (see Table 10 for regression coefficients and effect sizes). There was no main effect of gender.

A Wald Chi-square test indicated the main effect of set size and maternal education level was superseded by a significant interaction between maternal education level and set size ($X^2(1) = 4.62$, $p = .03161$). As seen in Figure 2.10, whilst performance decreased over set size for all children, children with a less educated mother showed greater decrease in A prime scores across set sizes.

A Wald Chi-square test also indicated an interaction between set size and gender was significant ($X^2(1) = 5.23$, $p = .022204$). As seen in Figure 2.11, performance decreased over set size, with girls showing increased performance to boys in set size 2, and boys showing increased performance to girls in set size 3.

Figure 2.10
Graph showing A prime score across set sizes and Maternal Education level.

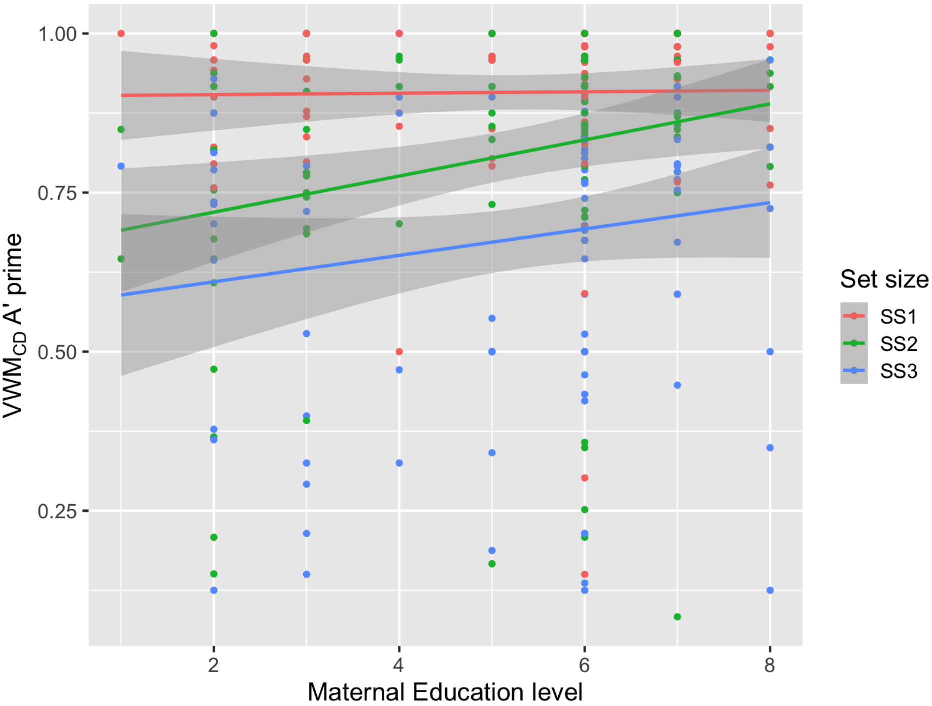
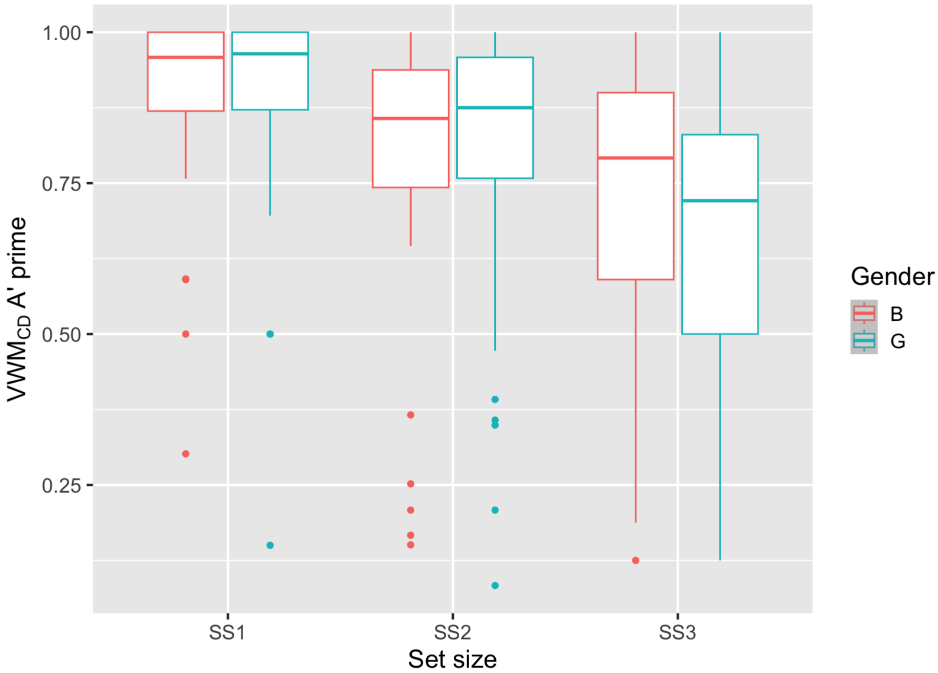


Figure 2.11
Graph showing A prime score across set sizes and gender.



The Q-Q plots of both A prime models indicated significant deviation of the observed values from the expected values. This suggests normality assumptions were violated. DHARMa residuals revealed quantile deviations from the predicted model were also detected. This indicates poor goodness-to-fit. In part due to this poor fit, I opted to focus subsequent analyses on the summary measure of change detection, K_{MAX}.

$\mathbf{K_{MAX}}$

To examine whether K_{MAX} demonstrated a similar age-related increase in capacity as reported by Simmering (2016), I conducted a linear mixed effects model predicting K_{MAX} as a function of year, maternal education level, and gender. Participant ID was included as a random effect. Year, maternal education level, and gender were included as fixed effects. Year was entered as a levelled variable, with year 3 being the reference. 67 participants were included in the model, run as:

$$lmer(K_{MAX} \sim Year \times Maternal\, Education\, level \times Gender + (1|Participant\, code))$$

A Wald Chi-square test indicated year significantly explained variance in K_{MAX} score ($X^2(1) = 58.42$, $p < .001$). As seen in Figure 2.12, K_{MAX} increased from year 2, at 42 months of age, to year 3, at 54 months of age. Mean K_{MAX} was found to be 1.28 (*SD* = 0.72) for 42 month-olds and 2.07 (*SD* = 0.68) for 54month olds. Overall, these are lower capacity estimates than those found previously for 3- and 4-year olds, but there was a greater increase in capacity from 42 months of age to 54 months of age than the increase demonstrated by Simmering (2016).

A Wald Chi-square test indicated maternal education level also significantly explained variance in K_{MAX} score ($X^2(1) = 12.80$, $p <.001$). Table 11 provides regression coefficients and effect sizes. As seen in Figure 2.13, children with a higher educated mother demonstrated a higher K_{MAX} score. Mean K_{MAX} was found to be 1.82 (*SD* = 0.79) for children with a more highly educated mother and 1.45 (*SD* = 0.79) for children with a less educated mother.

No interactions between terms were found. There was no dependence between year and maternal education level in their influence on K_{MAX} ($X^2(1) = 0.01$, $p = .93$). There was also no main effect or interaction effects with gender.

Figure 2.12
Graph showing K_{MAX} across year of participation.

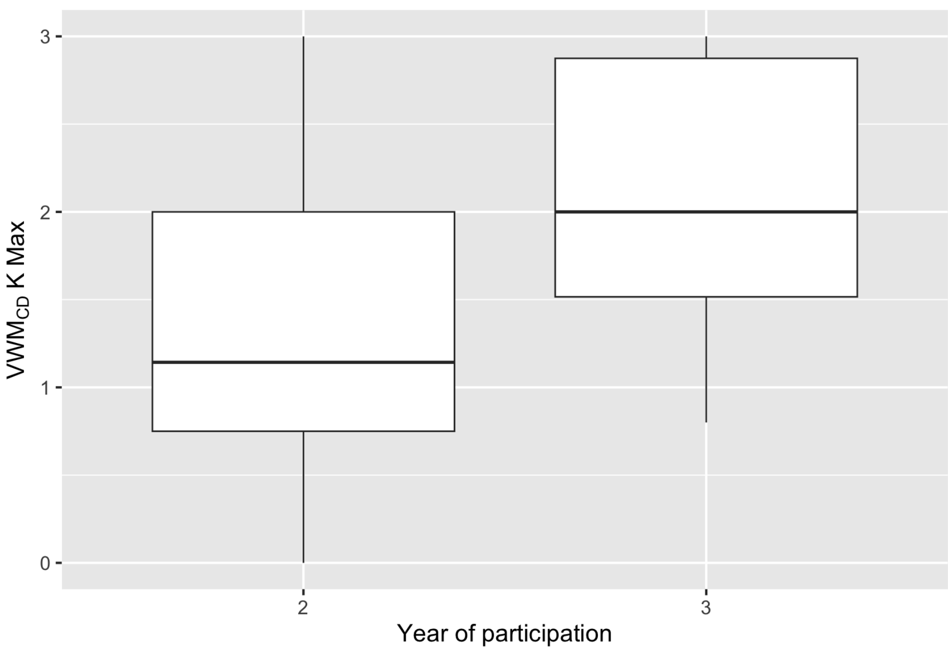
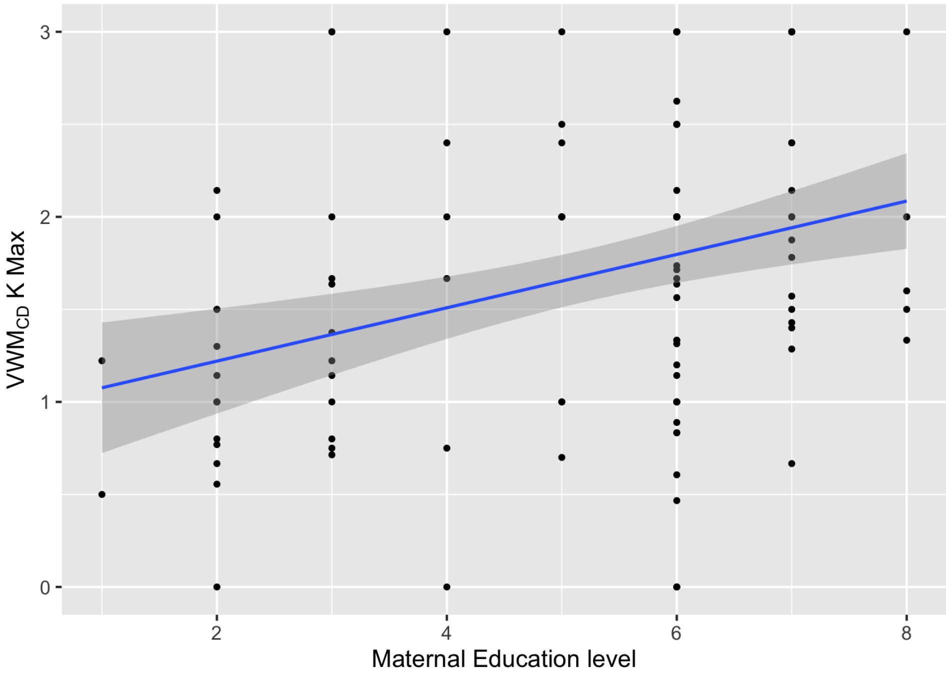


Figure 2.13
Graph showing K_{MAX} by maternal education level.



2.3.3 Cross-task performance

How does $\mathbf{VWM_{PL}}$ relate to $\mathbf{VWM_{CD}}$?

To examine the question of whether cross-sectional findings of cross-task performance on VWM_{PL} and VWM_{CD} were replicated in our longitudinal sample, performance was first examined following Simmering (2016) using a hierarchical regression with three models. The first model examined only year as a predictor of K_{MAX}. The second model included the CP_{10} measure from VWM_{PL}, examining both year and CP_{10} as predictors of K_{MAX}. The third model examined year and CP_{10} alongside switch rate. Simmering (2016) found the third model to be the best fitting model. This model demonstrated three predictors that explained significant variance in K_{MAX} scores of 3- to 5-year-olds: age; CP_{10} in the VWM_{PL} low load level (2 items on each display); and the switch rate in the VWM_{PL} high load level (6 items on each display).

Within all models, year was entered as year 2, representing 42 months of age, and year 3, representing 54 months of age. As the data from the present study were longitudinal, participant

ID was included as a random effect. 67 participants were included in the models.

As stated above, Model 1 included only year as a predictor of K_{MAX}. As seen in Figure 2.12 above, K_{MAX} increased from year 2 to year 3. This increase was found to be significant (see Table 12).

Model 2 included year and CP_{10} at each load level: low load containing 2 squares (SS2); medium, 4 squares (SS4), and high, 6 squares (SS6). Beyond the variance explained by year, CP_{10} in SS6 was found to account for a significant additional proportion of the variance in K_{MAX} score (see Table 12). As seen in Figure 2.14, children with a lower level of looking to the changing side in SS6 demonstrated a higher K_{MAX} score. These results differ to that of Simmering (2016). Firstly, the present study found it was CP_{10} in the high load that was important for K_{MAX}, whereas Simmering (2016) found it to be CP_{10} within the low load. Furthermore, the present study found CP_{10} was negatively associated with K_{MAX} in the low and high loads, with a lower CP_{10} in the high load significantly predicting a higher K_{MAX}. Simmering (2016) indicated CP_{10} was positively correlated with K_{MAX} in the low and high loads.

Model 3 included year, CP_{10} at each load level, and switch rate at each load level. As seen in figure 2.15, children with a higher switch rate in SS2 demonstrated a higher K_{MAX} score. In this model, switch rate significantly explained an additional proportion of variance in K_{MAX}, beyond that of year and CP_{10} in SS6 (see Table 12). These results are similar to that of Simmering (2016), however the present study found that switch rate in the low load was positively associated with K_{MAX}, whereas Simmering (2016) found positive associations with switch rate in the high load.

Figure 2.14
Graph showing K_{MAX} and CP_{10} in SS6 at 42 and 54 months of age.

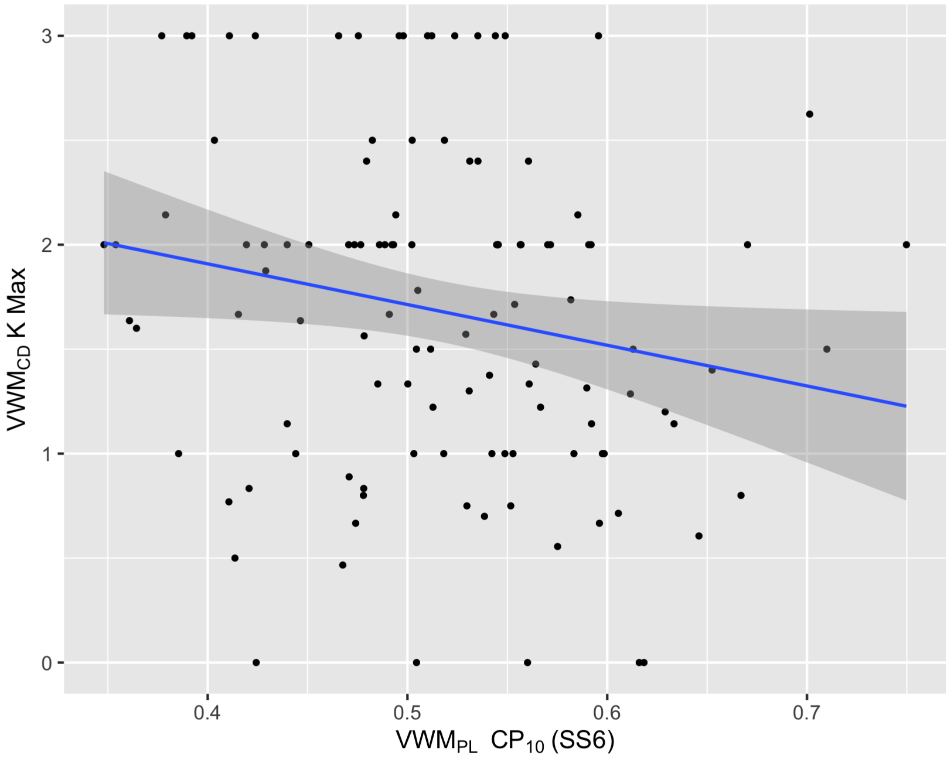
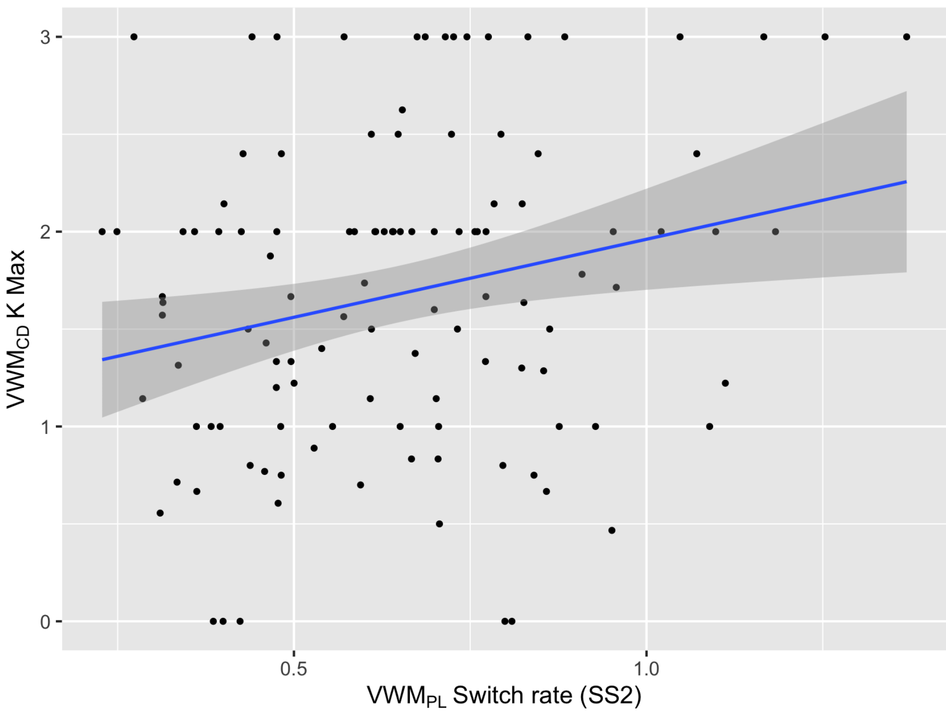


Figure 2.15
Graph showing K_{MAX} and switch rate in SS2 at 42 and 54 months of age.

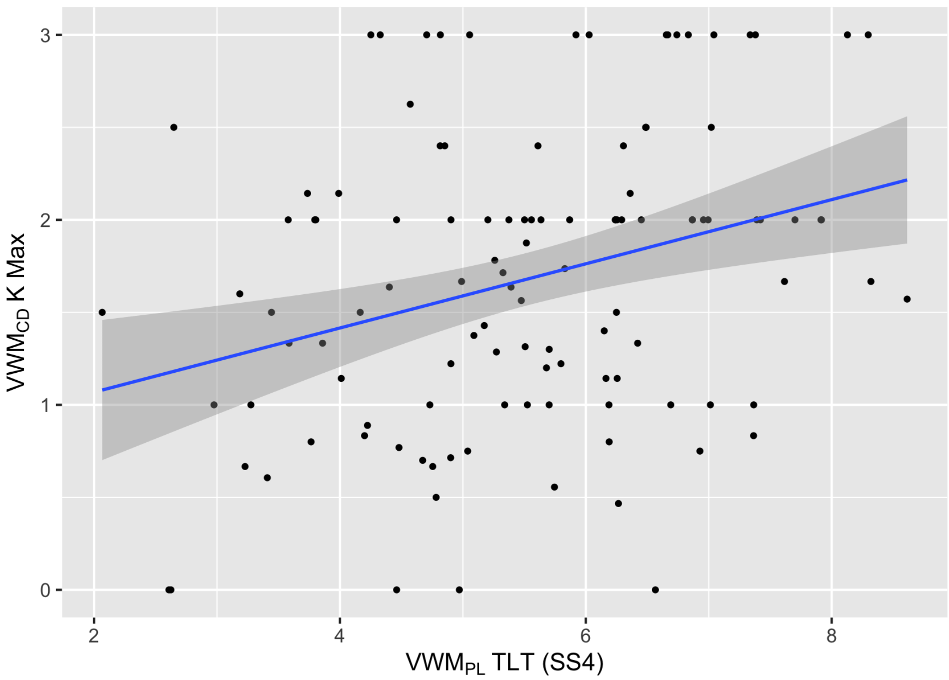


Model comparisons were carried out on the three models. Comparisons demonstrated model 2 (AIC = 240.69) had a lower AIC than model 1 (AIC = 242.05) suggesting better model fit. As data were longitudinal and therefore had to be examined in linear mixed effects models, a Chi-Square likelihood ratio test was performed in place of an ANOVA. This was only partially significant (X^2= 7.36, p= .06138). Model 3 (AIC = 240.22) demonstrated the lowest AIC, indicating the best fitting model. This reduction in AIC was very small, and a likelihood ratio test showed model 3 was not a statistically significant better fitting model (X^2= 6.47, p=.09072). This indicates Model 3 appears to be adding unnecessary complexity.

Previous longitudinal examinations of the VWM_{PL} task indicate TLT is an important variable to consider when assessing preferential looking (Forbes et al., in Prep; Wijeakumar et al., 2023). Consequently, a new model was run to assess the contribution of TLT in each load within the present data. TLT was added to the above best fitting model, model 2. This was assessed as model 4. Model 4 included year, CP_{10} in each load, and TLT in each load, as predictors of K_{MAX}.

As seen in Figure 2.16, children with a higher TLT in SS4 demonstrated a higher K_{MAX} score. In this model, TLT in SS4 explained an additional proportion of variance K_{MAX}, but this was only partially significant (see Table 13). The year and CP_{10} effects remained.

Figure 2.16
Graph showing K_{MAX} and TLT in SS4 at 42 and 54 months of age.

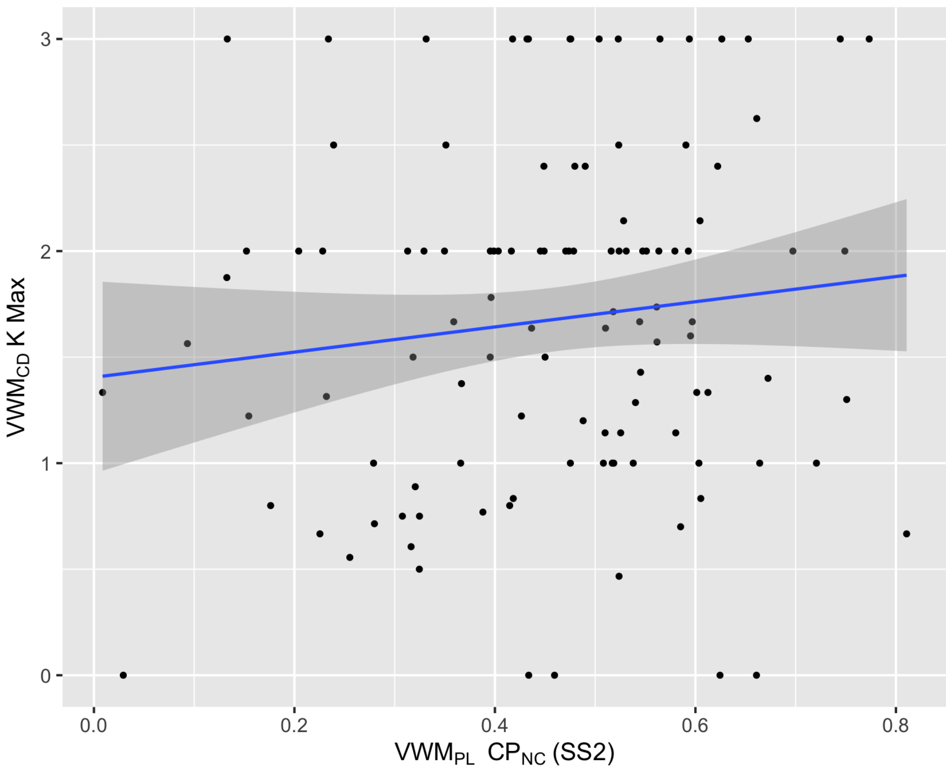


Model comparisons revealed model 4 (AIC = 236.85) had the lowest AIC out of all models. This difference in AIC was lower than model 2, the previous best fitting model, by more than 2 points, a common threshold for indication of a best fitting model (Sutherland et al., 2023). A likelihood ratio test also showed model 4 was a better fitting model than model 2 (X^2= 9.84, p= .01997).

As CP_{10} was not longitudinally predictive within earlier analyses, an additional model was run replacing the CP_{10} measure with the longitudinally predictive CP_{NC} measure. Alongside being longitudinally stable, this measure has previously been found to capture variance in VWM across load levels within the VWM_{PL} task (Forbes et al., in prep). Consequently, a new model was run to assess the contribution of CP_{NC} in each load within the present data. This was assessed as model 5. Model 5 included year, CP_{NC} in each load, and TLT in each load, as predictors of K_{MAX}. Two participants were removed for having no CP_{NC}.

Within this model, the previous effect of TLT in SS4, as seen in Figure 2.16, was significant. Additional variance in K_{MAX} was significantly explained by CP_{NC} in SS2 (see Table 14). As seen in Figure 2.17, a higher CP_{NC} in SS2 predicted a higher K_{MAX}.

Figure 2.17
Graph showing K_{MAX} and CP_{NC} in SS2 at 42 and 54 months of age.



Model comparisons revealed this model 5 (AIC = 235.04), containing TLT and CP_{NC}, was better fitting than the original best fitting CP_{10} model, model 2 (AIC 237.44). A likelihood ratio test also showed model 5 was a better fit than model 2 (X^2= 8.39, p= .03854). Comparing model 4 and 5, there was no addition to the model to make a likelihood ratio test viable; model 5 replaced CP_{10} with CP_{NC} directly, so there was no additional parameter to be assessed for improving model fit. Using only AIC, model 4 (AIC = 233.39) and 5 (AIC = 235.04) did not differ by more than 2 points, and thus the threshold for determining a better fitting model was not reached. As there was no significant difference in model fit, and the parameters within model 5 are heavily motivated by recent literature, model 5 was selected as the most appropriate model.

Overall, the present analyses partially replicates the findings of Simmering (2016) regarding the positive shift rate effect; however, analyses did not replicate the positive relationship between CP_{10} and K_{MAX}. The CP_{NC} measure did, however, show a positive relationship with K_{MAX}.

Does $\mathbf{VWM_{PL}}$ predict later $\mathbf{VWM_{CD}}$?

To explore the novel question of whether cross-task performance can be examined in a longitudinally predictive manner, VWM_{PL} from 30 months of age was explored as a predictor of 42 and 54 month VWM_{CD} capacity. To examine this relationship, the new VWM_{PL} measures explored previously within this chapter will be examined. As before, all measures were examined as aggregate measures over load. I first examined whether VWM_{PL} measures at 30 months of age predicted K_{MAX} at 42 months of age. Then, I examined whether VWM_{PL} measures at 30 months of age predicted K_{MAX} at 54 months of age. Maternal education level and gender were included in all models along with TLT as this showed robust effects in the concurrent models above.

A linear model examining mean CP_C in year 1, at 30 months of age, mean TLT in year 1, maternal education level, and gender as predictors of 42 month K_{MAX} was run. 58 participants were included in the model, run as:

$$lm(K_{MAX}\,[year2] \sim CP_C\,[year1] \times TLT\,[year1] \times Maternal\,Education\,level \times Gender)$$

The overall regression was not significant (R^2 = [0.29], $F(15,\, 42)$ = 1.14, p = .35), although it explained 29% of the variance in K_{MAX} (adjusted R^2= 0.04). The residual standard error was 0.71. There was the main effect of maternal education, as seen in previous models of K_{MAX}, but there were no main or interaction effects of mean CP_C, mean TLT, or gender (see Table 15).

Secondly, a linear model examining mean CP_{NC} in year 1, mean TLT in year 1, maternal education level, and gender as predictors of 42 month K_{MAX} was run. 58 participants were included in the model, run as:

$$lm(K_{MAX}[year2] \sim CP_{NC}\,[year1] \times TLT\,[year1] \times Maternal\,Education\,level \times Gender)$$

The overall regression was significant (R^2 = [0.44], $F(15,\, 42)$ = 2.19, p = .02307), explaining 44% of the variance in K_{MAX} (adjusted R^2= 0.24, RSE = 0.63).

There was a main effect of maternal education, as seen in previous models of K_{MAX}, but there were no main effects of mean CP_{NC}, mean TLT, or gender (see Table 16). There was a significant interaction between CP_{NC} and gender, β = 3.54, 95% CI [0.10, 0.41], p=.0426 (see Table 16). For girls, a higher mean CP_{NC} at 30 months of age predicted a higher K_{MAX} at 42 months of age (see Table 17 for means). As seen in Figure 2.18, for boys, a lower mean CP_{NC} at 30 months of age demonstrated a higher K_{MAX} at 42 months of age. Boys with a lower CP_{NC} at 30 months of age demonstrated the highest K_{MAX} at 42 months of age.

There was also an interaction between mean CP_{NC} and mean TLT: mean CP_{NC} at 30 months of age significantly predicted K_{MAX} at 42 months of age dependent on mean total looking time at 30 months of age, β = 1.55, 95% CI [0.07, 3.03], p=.0409 (see Table 16). As seen in Figure 2.19, for children with a higher mean TLT, a higher mean CP_{NC} at 30 months of age predicted a higher K_{MAX} at 42 months of age. For children with a lower mean TLT, a lower mean CP_{NC} at 30 months of age predicted a higher K_{MAX} at 42 months of age. Here, children with a lower mean TLT and a lower mean CP_{NC} at 30 months of age demonstrated the highest K_{MAX} scores at 42 months of age (see Table 18 for means). There were no interaction effects with maternal education level.

Figure 2.18
Graph showing K_{MAX} at 42 months and mean CP_{NC} at 30 months by gender.

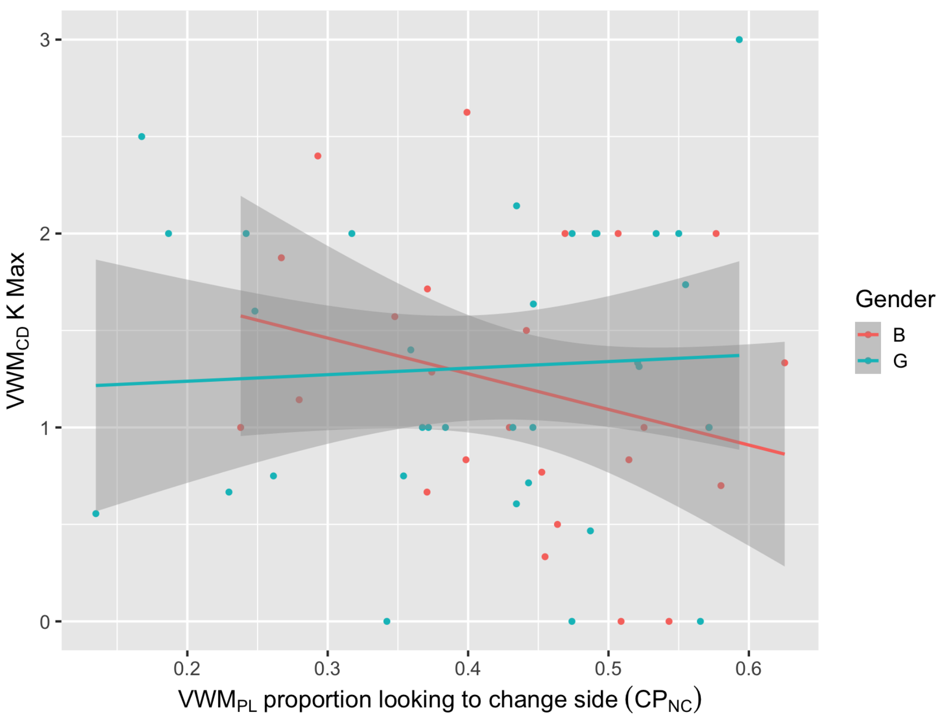
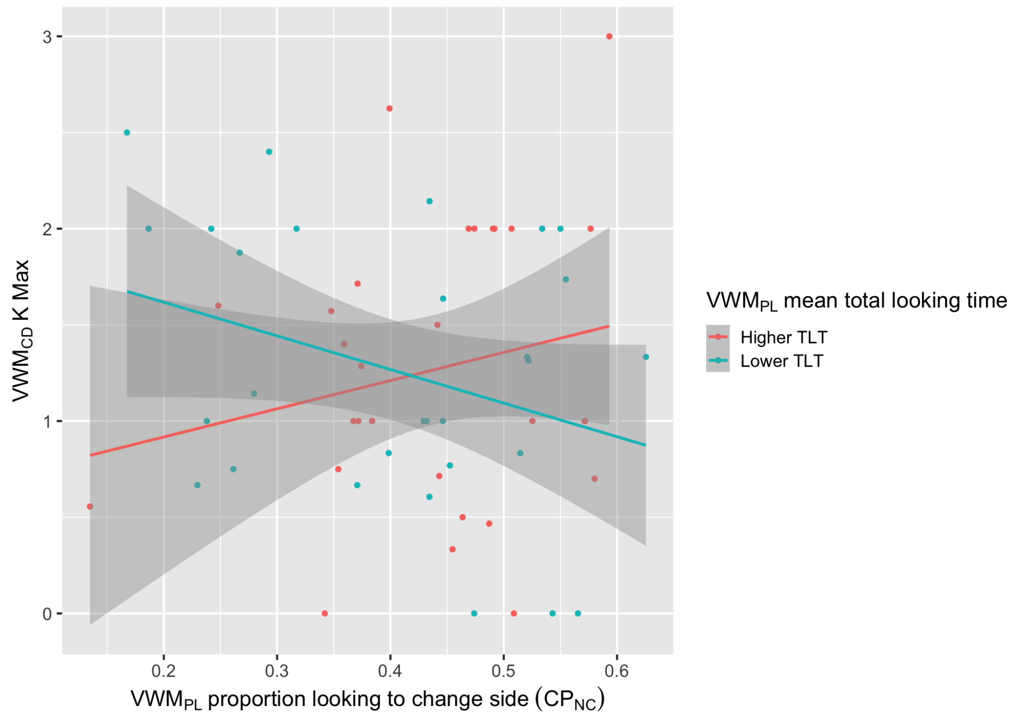


Figure 2.19
Graph showing K_{MAX} at 42 months and mean CP_{NC} at 30 months by mean TLT at 30 months.



Note: mean TLT level is demonstrated here using a median split to categorise low and high levels, but TLT was entered as continuous in all models.

This model including the interaction terms showed a large effect size, f^2=0.51, indicating that the two interactions between mean CP_{NC} at 30 months of age and gender, and mean CP_{NC} and mean TLT at 30 months of age, explained a substantial proportion of the variance in K_{MAX} at 42 months of age.

Overall, VWM_{PL} is predictive of VWM_{CD} a year later, but this predictive relationship is only found for the mean CP_{NC} measure, and must take into account gender and TLT.

Next, I examined whether VWM_{PL} predicted VWM_{CD} performance over a longer two year period. Mean CP_C in year 1, at 30 months of age, mean TLT in year 1, maternal education level, and gender were examined as predictors of year 3 K_{MAX}, at 54 months of age. 6 children were excluded for not having both a VWM_{PL} at 30 months of age and a K_{MAX} at 54 months of age. 52 participants were included in the model, run as:

$$lm(K_{MAX}[year3] \sim CP_C\,[year1] \times TLT\,[year1] \times Maternal\,Education\,level \times Gender)$$

The overall regression was not significant (R^2 = [0.38], F(15, 42) = 1.69, p = .09), although it explained 38% of the variance in K_{MAX} (adjusted R^2= 0.15). The residual standard error was 0.63. The main effect of maternal education seen in the K_{MAX} base model remained, but there were no main or interaction effects of mean CP_C, mean TLT, or gender (see Table 19).

Next, this question was explored with the CP_{NC} measure. Mean CP_{NC} in year 1 and mean TLT in year 1, at 30 months of age; maternal education level; and gender were then examined as predictors of year 3 K_{MAX} at 54 months of age. The same 52 participants were included in the model, run as:

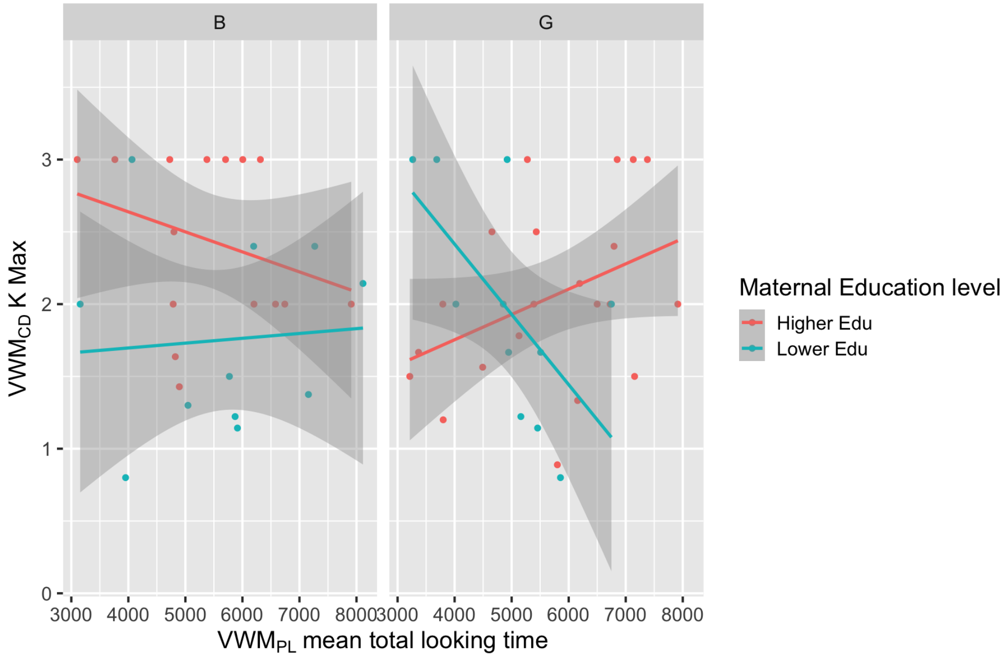
$$lm(K_{MAX}[year3] \sim CP_{NC}\,[year1] \times TLT\,[year1] \times Maternal\,Education\,level \times Gender)$$

The overall regression was not significant (R^2 = [0.37], F(15, 42) = 1.63, p = .11), although it explained 37% of the variance in K_{MAX} (adjusted R^2= 0.14). The residual standard error was 0.63.

The significant main effect of maternal education seen in previous K_{MAX} models remained, and a significant main effect of gender was also found (see Table 20). These main effects were superseded by an interaction between maternal education level, gender, and TLT. Despite the overall non-significant model, mean TLT at 30 months significantly predicted K_{MAX} at 54 months dependent on gender and maternal education level, β = 0.59, 95% CI [0.14, 1.04], p=.0108 (see Table 20). As seen in Figure 2.20, girls with a more highly educated mother and a higher mean TLT at 30 months of age demonstrated higher K_{MAX} scores at 54 months of age. For girls with a less educated mother, this appears inverse. Girls with a less educated mother and a lower mean TLT at 30 months of age demonstrated higher K_{MAX} scores at 54 months of age. Overall, differences in boys scores were less dependent on TLT, with larger differences between K_{MAX} at 42 months of age appearing more dependent on maternal education level (see Table 21 for means). Boys with a more highly educated mother demonstrated higher K_{MAX} scores than boys with a less educated mother.

There were no main or interaction effects of mean CP_{NC} (see Table 20).

Figure 2.20
Graph showing K_{MAX} at 54 months and mean TLT at 30 months by Maternal Education level and gender.



Note: Maternal Education level is demonstrated here using a median split to categorise low and high levels, but Maternal Education level was entered as continuous in all models. Here, a lower education represents below University level.

The three-way interaction between maternal education level, gender, and TLT in year 1 was found to be meaningful. The model including the interaction terms showed a moderately large effect size, f^2=0.319, indicating that interactions explained a moderate proportion of the variance in K_{MAX}.

Overall, VWM_{PL} performance at 30 months is predictive of VWM_{CD} performance at 54 months and models were qualified by interactions with maternal education level, gender, and TLT.

2.4 Discussion

The present chapter has focussed on resolving a key issue within the working memory literaturea lack of longitudinal research examining the development of working memory components from early infancy. This study focussed on the specific component of visual working memory, as this has previously been reliably assessed in the early infancy period.

First, VWM was examined using the VWM_{PL} task from 6 months of age. Previous examinations of VWM in childhood have used the same VWM_{PL} task (Ross-Sheehy et al., 2003; Simmering, 2016; Wijeakumar et al., 2019; Delgado-Reyes et al, 2020; Forbes et al., in Prep). These studies tend to focus on examining VWM using a change preference score referred to in the present study as CP_{10}. Results of the present study show this CP_{10} measure demonstrates inverted longitudinal stability from year 1 to year 2, where a higher CP_{10} in year 1 predicted a lower CP_{10} in year 2. No robust longitudinal stability from year 2 to year 3 was found. This lack of stability from year 2 to year 3 is consistent with the lack of longitudinal stability in Forbes et al. (in Prep). It is not clear why CP_{10} shows inverted stability from year 1 to year 2. It is possible this is related to the increase in set size from 6 to 18 months of age (switching from 1, 2, and 3 squares on each display to 2, 4, and 6 squares on each display); however, the set size was consistent for the second cohort of children from 30 to 42 months of age, so this increase does not fully explain the inverted result. Whilst the reasons for these results are not yet fully understood, these findings add to existing evidence that this CP_{10} measure is not a suitable measure for examining VWM longitudinally.

Results from the CP_C measure showed that CP_C in year 1 of participation predicted CP_C in year 2, but CP_C in year 2 did not predict CP_C in year 3. This CP_C measure is capturing the child's ability to detect the changing side on the VWM_{PL} task when the child's first look is to the changing side. This arguably places less demand on the visual working memory system than the CP_{NC} measure, which captures child's ability to detect the changing side when starting on the non-changing side. That is, because the child starts on the changing side, change is easier to detect which should help sustain looking to the novelty in this display. It is possible that the lack of longitudinal stability in this measure from year 2 to year 3 indicates that it was easy for older children to reliably stay on the changing side in this case.

Results show that the CP_{NC} measure was longitudinally stable across all age ranges examined: CP_{NC} in year 1 predicted CP_{NC} in year 2, and CP_{NC} in year 2 predicted CP_{NC} in year 3. Previous research suggests children and adults are able to distribute the limited resources of their working memory dependent on set size demands (Scerif & Shimi, 2021). When the child starts on the non-changing side in the VWM_{PL} task, it is likely they will need to recruit more resources from the working memory system to detect there is no change, release fixation and switch to the changing side, where they must restart the process of detecting the colourful squares in order to detect the change and sustain fixation to this changing side. Thus, CP_{NC} might be the most sensitive measure as it consistently places VWM in a resource-demanding context.

Beyond these change preference measures, results showed that TLT on the VWM_{PL} task was longitudinally stable, with TLT in year 1 predicting year 2, and TLT in year 2 predicting year 3. This TLT measure was found to be especially important for cross-task performance, where TLT on the VWM_{PL} task was often positively associated with later VWM capacity on the change detection task. TLT provides an index of the child's ability to sustain attention to the display, with a higher TLT indicating better sustained attention.

Interestingly, by obtaining multiple measures from the same task, we see that different aspects of performance are robust longitudinally. There are also complex interactions among these measures which I discuss further below. Finally, it is useful to note that within the VWM_{PL} task there were no effects of maternal education level or gender. This indicates that VWM_{PL} performance predicts itself even when controlling for these covariates.

Next, the present chapter focused on VWM capacity measured using the change detection task. This study is the first longitudinal study examining change detection in early childhood. The first goal was to examine if our longitudinal data replicated the findings of previous cross-sectional research. Results showed that changes in A' replicated cross-sectional findings from Simmering (2016). In particular, the age-related increase in overall A' scores was replicated, with accuracy in each set size increasing from 42 to 54 months of age similarly to the increase in accuracy from 3 to 4 years of age in Simmering (2016). Simmering (2016) also found lower accuracy scores in set size 3 than set sizes 1 and 2, and a larger decrease between set size 1 to set size 3 for 4-year-olds than 3- and 5-year-olds. This interaction was replicated within the present study, however the larger decrease between set sizes 1 and 3 was seen for children who were 42 months of age (3.5-years-old) rather than children who were 54 months of age (4.5-years-old). It is important to note that whilst Simmering (2016) examined 3-, 4-, and 5-year-olds, the mean ages for 3 and 4-year-olds within the study were closer to the 3.5 and 4.5-year-old ages used within the present study.

Simmering (2016) also examined maximum capacity, K_{MAX} from the change detection task. Firstly, Simmering (2016) reported 3-year-old maximum capacity to be 1.90 items, and 4-year-old maximum capacity to be 2.20 items. Whilst capacity estimates from the present study demonstrate a similar age-related increase in maximum capacity, with 3.5-year-old mean maximum capacity being 1.28 items and 4.5-year old mean maximum capacity being 2.07 items, these maximum capacity estimates were found to be lower overall than those in Simmering (2016). These capacity estimates are more in line with the 3-year-old capacity of 1.2 items and 4-year-old capacity of 1.8 items found by Buss, Fox, et al (2014) on a more challenging shape-based detection task within this age group. Due to the timing of our study, many children completed the VWM_{CD} task during the Covid-19 pandemic, and there is a possibility this may have negatively influenced capacity estimates as a result of a wide range of health and social related changes. Investigations of this influence of the Covid-19 pandemic are currently in progress to examine if there are differences between children examined pre- vs post-pandemic.

It is important to note that this model of K_{MAX} also demonstrated a consistent maternal education level effect, where children with a more highly educated mother demonstrated higher K_{MAX} scores regardless of age. The present study made an attempt to obtain a relatively representative sample, and a concerted effort was made to recruit children from lower socio-economic backgrounds. It is possible that this sample contained more children with a less educated mother than the sample used by Simmering (2016), and this could also explain the reductions in overall capacity estimates at each age.

Simmering (2016) found three predictors that explained significant variance in K_{MAX} scores of 3- to 5-year-olds: age; CP_{10} on the VWM_{PL} task in a set size of 2; and switch rate on the VWM_{PL} task in set size 6. These relationships were all positive, with an increased age, higher CP_{10} on the VWM_{PL} task, and higher switch rate on the VWM_{PL} task, predicting a higher K_{MAX} score on the VWM_{CD} task. Whilst the present study does replicate the age-related increases in K_{MAX}, the same predictive relationships with the CP_{10} and switch rate measures were not found within this longitudinal sample. Switch rate was found to have a positive relationship with K_{MAX}, where a higher switch rate on VWM_{PL} indicated a higher K_{MAX} on VWM_{CD}. However, this was only predictive in set size 2 as opposed to set size 6 as found by Simmering (2016). Moreover, a model without switch rate was found to better fit the longitudinal data.

In further contrast to Simmering (2016), the present study found an inverse relationship between CP_{10} on the VWM_{PL} task and K_{MAX} on the VWM_{CD} task. Here, a lower CP_{10} in set size 6 was associated with a higher K_{MAX} score. By contrast, Simmering (2016) indicated a higher CP_{10} in set size 2 was associated with a higher K_{MAX}. These differences were not solely due to the difference in the set size found to significantly explain variance in K_{MAX}. Whilst not significant, the present study also found this inverse relationship was present in set size 2. Given the finding that CP_{10} is not a longitudinally stable measure, these results provide further support that the CP_{10} measure does not effectively capture VWM within a longitudinal sample, suggesting it is important to move away from this change preference measure to more detailed measures that consider the differential demands of the task dependent on the context of the child's looking behaviours.

Using the same hierarchical regression structure as Simmering (2016), the switch rate and CP_{10} measures were replaced with the two longitudinally stable measures, TLT and CP_{NC}. Within this model, both CP_{NC} at SS2 and TLT at SS4 significantly explained variance in K_{MAX}. This adds to our previous finding that CP_{NC} is a more robust measure of VWM, particularly within longitudinal examinations, than the CP_{10} measure used previously.

The final question examined in this chapter was whether performance on the VWM_{CD} task could be predicted by performance on the VWM_{PL} task up to two years earlier. This novel question revealed complex interactions. Firstly, only the CP_{NC} and TLT measures from the VWM_{PL} task were found to relate to performance on the VWM_{CD} task later. Once again, these measures were the only measures to demonstrate longitudinal stability, and thus may be the measures most sensitive to changes in VWM across development.

Looking at the predictive nature of VWM_{PL} over a one year period, CP_{NC} on the VWM_{PL} task at 30 months of age was strongly related to K_{MAX} on the VWM_{CD} task at 42 months of age. Results show that for children who had a higher TLT at 30 months of age on the VWM_{PL} task, a higher CP_{NC} on VWM_{PL} at 30 months of age predicted a higher K_{MAX} at 42 months of age. These higher looking children appear able to successfully release fixation from the non-changing side, detect the change on the changing side, and hold fixation to this changing side for longer periods of time. Whilst being able to release fixation, switch, and maintain fixation to the changing side is important for later VWM capacity, results indicate that K_{MAX} scores for these children were only slightly higher than for the children with a higher TLT but lower CP_{NC} at 30 months of age. Consequently, it appears the overall ability to sustain attention to the task is the stronger predictor of later VWM capacity.

Nevertheless, it was found that children with a lower mean TLT and a lower mean CP_{NC} on VWM_{PL} at 30 months of age demonstrated the highest K_{MAX} scores at 42 months of age. These children were less able to release fixation from the non-changing side and detect the change on the changing side, demonstrated through a lower level of looking to the changing side when starting on the non-changing side. For these children, it appears that being able to simply hold fixation to the side they started on, even for a reduced period of time, is important for later VWM capacity.

The children with a lower TLT who attempted to release fixation and attend to the changing side, showing a higher CP_{NC}, may be demonstrating an inability to detect the change. These children may release fixation from the non-changing side to attend to the changing side, but become overwhelmed by the demands of the changing side, releasing fixation from the task early. This results in a lower TLT but a higher CP_{NC}. These children go on to demonstrate the lowest VWM capacity at 42 months of age, indicating that those who may show a poorer VWM, through a struggle to detect the change at 30 months of age, also show a poorer VWM capacity of around 1.07 items at 42 months of age.

It is important to note that gender interactions were also found. For girls, a higher mean CP_{NC}

at 30 months of age predicted a higher K_{MAX} at 42 months of age but for boys, a lower mean CP_{NC} at 30 months of age predicted a higher K_{MAX} at 42 months of age. Consequently, it appears that for girls being able to complete the more challenging demands of releasing fixation, switching, and then detecting and maintaining fixation to the changing side is important for later VWM capacity. For boys, the ability to simply hold fixation to the side they started on, demonstrating sustained attention, is most important for later VWM capacity.

Looking at the predictive nature of VWM_{PL} over a two year period, TLT on the VWM_{PL} task at 30 months of age was strongly related to K_{MAX} on the VWM_{CD} task at 54 months of age. This relationship was dependent on gender and maternal education level. It was found that for girls with a more highly educated mother, a higher TLT at 30 months of age predicted a higher K_{MAX} at 54 months of age. This aligns with previous findings that the ability to sustain attention to the task is important for later VWM capacity. However, for girls with a less educated mother, a lower TLT at 30 months of age predicted a higher K_{MAX} at 54 months of age. This may be related to the low TLT and low CP_{NC} effect seen at 42 months of age, where the ability to sustain fixation to the same side was important for later VWM capacity, even with a reduced TLT. However, this relationship is yet to be fully understood. Deeper investigations, such as using cluster analysis, should be conducted to determine categories of children across each measure and how they relate to each other. Cluster analysis is a method which identifies groups of participants that display more similar behaviour to each other across a number of measures, but less similar to behaviours to those in different groups (Mooi & Sarstedt, 2011). This would allow us to see whether these girls are within the same group of children who demonstrated the low TLT and low CP_{NC} effect on capacity at 42 months of age.

For the boys, whilst there were slight differences in K_{MAX} at 54 months of age dependent on TLT at 30 months of age, differences in K_{MAX} were strongly dependent on maternal education level. Boys with a more highly educated mother demonstrated higher K_{MAX} scores than boys with a less educated mother, regardless of TLT at 30 months of age. This reflects the consistent effect of maternal education level seen in previous models of K_{MAX}. This effect remained robust within every model of K_{MAX}, indicating maternal education level has a strong influence on VWM capacity. This effect may be related to previous findings that children with a less educated mother show more difficulty in suppressing distractions on VWM tasks (Wijeakumar et al., 2019), and thus these children may be struggling to successfully attend to and detect a change when presented with more competing items, leading to a lower capacity estimate.

In conclusion, VWM may be best examined longitudinally using measures that place the VWM system in a resource-demanding context. Using such measures, in this case CP_{NC}, the present study illustrates that VWM is longitudinally stable from 6 to 54 months of age. Moreover, this measure was related to performance on the widely used change detection task. However, this relationship also involved the longitudinally stable TLT measure. Modelling these measures together revealed the involvement of sustained attention in the completion of VWM tasks. Consequently, I provide support for previous research suggesting multiple factors should be considered when examining developmental changes in VWM (Shimi & Scerif, 2021; Forbes et al., in prep). By examining multiple measures within the same task, and when examining cross-task performance, the multiple systems that may be involved in the completion of a task can be revealed.

Chapter 3

The Developmental Trajectory of Executive Function during Early Childhood.

3.1 Introduction

Early EF improves from the first year through early childhood, with dramatic developmental changes occurring from three to five-years of age (Carlson, 2005; Diamond, 2013, Garon et al., 2008). As outlined previously, there are disagreements surrounding the stability of EF and its components during early childhood. There are, however, robust measures of EF that capture and are sensitive to age-related improvements in EF. For example, in a recent cross-sectional examination of over 51,000 participants from different studies in the United States, performance on the Minnesota Executive Function Scale (MEFS; Carlson & Zelazo, 2014) tablet task improved rapidly across early childhood, with more gradual improvements through adolescence (Carlson, 2021). Whilst mean level differences were found across different demographic populations, this trend of dramatic improvement during early childhood remained.

Whilst there are many longitudinal studies examining EF as a predictor of later academic achievement (E.g. Duncan et al., 2007; McClelland et al., 2014), there are few that examine the longitudinal stability of EF itself. Interestingly, parenting studies have contributed to this examination of EF longitudinally. Helm et al. (2019) found that children's performance on the DCCS at four-years of age was positively correlated with performance on the DCCS at six-years of age. It was concluded that four-year-olds with higher EF continued to demonstrate this elevated level of EF by 6-years of age, suggesting there is stability in EF in early childhood. However, within this study a four-year-old EF measure was comprised of performance over supposedly separate inhibition (Pig/Bull task), cognitive flexibility (DCCS), and working memory (forward digit span) tasks. A similar measure was comprised for six-year-old EF, however the only task that remained the same was the DCCS, with a stroop-like task being used for inhibitory control and a backward span task being used for working memory by this age. As discussed within Chapter 1, the use of different tasks over time may lead to differences resulting from task demands. Furthermore, this study began at school age. What do we know about longitudinal stability earlier in development?

Early EF has been probed for longitudinal stability from 18 to 26 months of age in a study that used different tasks at different assessment points (Bernier, Carlson, & Whipple, 2010). For example, at 18 months of age, Bernier et al. (2010) used a hide-the-pots task (an A-not-B style search task) and a DCCS style categorisation task. These were compared to multiple tasks introduced at 26-months of age, for example, a delay-of-gratification and two Stroop-like tasks. The hide-the-pots and DCCS style categorisation tasks were found to have a low correlation at 18 months of age. Despite this, both tasks were found to be positively related to the measure of 'conflict EF', which consisted of performance on the two Stroop-like and a spin-the-pots version of the A-not-B style task at 26 months of age. Moreover, hide-the-pots was associated with later impulse control on the delay of gratification task. Once again, it is difficult to make strong conclusions about the stability of performance over these tasks because the tasks place different demands on children and tap different components of EF. This is especially difficult when performance across tasks has been merged to create a specific EF factor. Consequently, we cannot establish how well each task captured individual differences, and do not know if correlations are driven by performance from one specific measure to another, or if this is consistent across each measure included within the 'conflict EF' factor.

In summary, there are currently no longer-term longitudinal studies that examine EF from the pre-school period using the same task over time. Due to aforementioned concerns with using different EF measures over time, the MEFS task (Carlson & Zelazo, 2014) will be used within the present study. This task is able to be used from 2-years of age through to adulthood.

Longitudinal assessments of EF are required to further understand the contribution of individual differences to EF development during early childhood. This is especially important given that EF skills are predictive of important outcomes later in development, including school success, health, and wellbeing (Moffitt et al., 2011). Moreover, examinations of childhood EF have revealed positive correlations between EF and mathematics, specifically in low-income communities at-risk of poverty (Blair et al., 2015). Here, stronger EF skill may be acting as a protective factor, supporting later school success. Furthermore, homeless children with stronger EF skills were more likely to succeed in school despite being homeless (Masten et al., 2012). EF was found to possess unique predictive value beyond that of general IQ, once again demonstrating better EF may protect against risk-factors, specifically demographic and home influences.

Given the significance of EF for future outcomes, specifically for at-risk children, it is important we understand the influence of demographic factors on the development of EF during early childhood. Variance in measures of latent EF have been related to socio-economic status in two- to five-year-old children (Hughes et al., 2010; Wiebe et al., 2011), and studies have found socio-economic status may be uniquely associated with the specific aspects of accuracy and reaction time when examining EF skills (John et al., 2019). However, other studies have reported no relationship between socio-economic status and EF (e.g. Duncan et al., 2017). Whilst there is mixed evidence surrounding the relationship between EF and socio-economic status, a recent meta-analysis including 8760 children found the correlation between EF and socio-economic status was statistically significant across all studies (Lawson et al., 2018). Moreover, Hackman et al. (2015) found effects of early socio-economic status examined through income-to-needs and maternal education level on EF remained from early to middle childhood. Here, lower socio-economic status predicted lower performance on EF tasks. Specifically, maternal education level predicted planning abilities on the Tower of Hanoi task in six- to seven-year old first graders, a skill often denoted as an executive functioning skill. Maternal education level continued to predict this planning ability by fifth grade. Consequently, we see the inclusion of demographic variables in examination of longitudinal EF development is necessary. No studies have examined whether these socio-economic measures remain impactful longitudinally, particularly starting from early childhood.

A current debate within cognitive research surrounds the inclusion of a multitude of socio-economic factors. Duncan and Magnuson (2012) argued that socio-economic status is too multifaceted to be captured by one single measure. For example, education and income measures lead to the presence of different beneficial resources that assist development in unique ways (Duncan & Magnuson, 2012). Thus, applying one measure as an overall indicator of socio-economic status may lead to broad assumptions about the impact of socio-economic status that are in fact related to only the specific measure used. A large number of studies have begun to assess the unique contributions of specific socio-economic factors in shaping child development and have found parental educational level has arisen as a strong predictor of children's cognitive outcomes (see Waters et al., 2021 for discussion). Furthermore, Waters et al. (2021) indicate that only parental education level, and not income-to-needs, was associated with all EF and academic achievement domains, including maths and reading ability. Whilst income-to-needs reflects a family's ability to provide educational resources, it appears that the parents own educational background is more important for nurturing EF and academic skills. Consequently, maternal education level will be examined within the present study to provide an examination of the longitudinal impact of a highly predictive socio-economic measure of EF.

Gender differences in EF have been demonstrated across Western and East Asian samples (Schirmbeck et al., 2020; Palomino & Brudvig, 2022; Yamamoto & Imai-Matsumura, 2019), with the majority of research finding girls outperform boys, except in Iran and Tanzania where this appears reversed (Schirmbeck et al., 2020). As such, when examining individual differences in EF longitudinally, it is important we consider gender differences, specifically when examining a western sample where gender differences have been shown to be prevalent during childhood.

By including factors such as gender and measures of socio-economic status, we can further understand the stability of EF over the early childhood period, whilst holding constant relevant covariates that could influence the interpretations of longitudinal associations.

The goal of the present chapter is to shed light on the longitudinal stability of EF across early childhood, utilising the same executive function task over time. I expect to find EF measured using the MEFS task will be predictive of itself over a multi-year period, from 30 to 78 months of age, demonstrating similar age-related improvements observed in cross-sectional research. Based on prior work showing that socio-economic status measured via maternal education influences EF development, I expect to find children with lower educated mothers will demonstrate lower EF ability. Given that this research was conducted in the UK, the sample was primarily from Western backgrounds. Consequently, I also expect to see effects of gender on EF performance, with girls outperforming boys.

3.2 Methods

3.2.1 Participants

139 children (71 girls) from the same study discussed in Chapter 2 completed the executive function tasks. There were two cohorts. Cohort one began the study at six months and were tested for EF skill at 30 and 78 months of age only. Cohort two began the study at 30 months of age and were tested for EF skills at 30, 42, 54 and 78 months of age. Demographics are shown in Table 3.1. Average maternal education level was a Bachelor's Degree, and mean income was £40645.69 ($SD = 11720.63$). Participants had normal or corrected-to-normal vision. Colour vision was examined through family history of colour blindness risk; at-risk children were excluded. All participants were full-term infants.

This project was reviewed and approved by the Ethics Committee at NHS England. Parents signed an informed consent form on behalf of the child. Children received a toy and a t-shirt for participating at each lab visit. Parents were given £20 for each visit to the lab and £5 for home visits. At the home visit, the child also received 3 toys totalling £5. The data reported here are a subset of a larger study examining the neural basis of visual working memory and attention in early development.

Data counts revealed that of the 139 children who completed the MEFS task at 30 months of age, two did not complete the follow up MEFS task at 78 months of age (see Table 3.2). Cohort two completed the MEFS task every year of participation at 30, 42, 54, and 78 months of age. Data counts revealed that of the 84 children from Cohort two who completed the MEFS task in at least one year of participation, 9 did not complete the MEFS task in year 1, 26 did not complete the MEFS task in year 2, 18 did not complete the MEFS task in year 3, and 12 did not complete the MEFS task in year 4. The higher missing data in years 2 and 3 were due to the Covid-19 pandemic effecting our ability to collect data in these years.

Table 3.1
Demographic Characteristics

Variable	Cohort One N = 64	Cohort Two N = 75
Gender		
Boys	33 (52%)	35 (47%)
Girls	31 (48%)	40 (53%)
Maternal Education Level		
Left School before 16	1 (1.6%)	0 (0%)
GCSE/O Levels or equivalent	4 (6.3%)	10 (13%)
A Levels or equivalent	6 (9.4%)	10 (13%)
Trade Apprenticeship	0 (0%)	4 (5.3%)
Some University	5 (7.8%)	6 (8.0%)
Bachelor's Degree	30 (47%)	30 (40%)
Master's Degree	12 (19%)	11 (15%)
Doctorate or Professional Degree	6 (9.4%)	4 (5.3%)
Ethnicity		
White British	54 (84%)	67 (89%)
Asian	1 (1.6%)	0 (0%)
Black African	0 (0%)	1 (1.3%)
South African	2 (3.1%)	0 (0%)
White British and South American	2 (3.1%)	0 (0%)
White British and Asian	1 (1.6%)	2 (2.6%)
White European and Asian	1 (1.6%)	0 (0%)
White British and Black African	0 (0%)	2 (2.7%)
White British and Black Caribbean	0 (0%)	2 (2.7%)
White British and Other European	3 (4.7%)	1 (1.3%)

Table 3.2
MEFS data counts and descriptive statistics.

	30 months			78 months		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
Total Score	139	15.79	5.80	137	73.27	9.56
Total N	139					

Table 3.3
Longitudinal MEFS data counts and descriptive statistics, Cohort Two.

Variable		**N**	**Mean**	**SD**
MEFS Total Score				
	Year 1	75	15.84	5.96
	Year 2	58	40.78	16.83
	Year 3	66	60.33	12.29
	Year 4	72	75.46	9.03
Total N	75			

3.2.2 Procedure

The MEFS task was completed in the lab from 30 to 54 months of age. The child was asked if they wanted to play a new game on a tablet. The tablet was placed on a table in front of the child. Parents were in a separate room. If the child requested the parent to be in the room, they were sat behind the child so as to not influence the child's responses. At 78 months of age, the MEFS task followed a similar procedure but was conducted inside the child's home. The child was asked to sit on a chair at their dining table, where the tablet was located in front of them. If no such table was present in the home, they were asked to sit on a sofa next to the experimenter and the tablet was held on the experimenter's knee facing the child. The remaining procedure was identical.

All participants completed the Minnesota Executive Function Scale (MEFS; Carlson & Zelazo, 2014). MEFS is a tablet task taking from two to six minutes (four minutes average test duration), based on the Dimensional Change Card Sort tasks (DCCS; Zelazo, 2006). This task involves increasingly difficult levels requiring a child to sort cards into a virtual box according to dimensions such as size, shape, or colour (see Carlson, 2021 for examples of the MEFS task). Before each level, the child receives a demonstration and rule checks to ensure understanding of the rules for that level, e.g. "I have these boxes here. This one has a small elephant on it and this one has a big elephant on it. In this game the small elephants go in the small elephant box and the big elephants go in the big elephant box. See, here is a small elephant. It goes in the small elephant box (experimenter drag). Can you put this elephant where it goes? (Child drag)". The child is given a rule to follow based on one of the aforementioned dimensions. The rule is restated on the first two trials e.g. "If it's a small elephant, it goes in the small elephant box" and the relevant dimension was emphasised e.g. "Here is a small elephant".

In the next five trials, a prompt was given to ensure the child was ready, e.g. "Get ready!". After five trials, the experimenter announced a 'new game' in which the rule was switched. The child is asked to follow this new rule, e.g. "the small elephant goes in the small elephant box" switches to "the small elephant goes in the big elephant box". To complete this task, children are required to utilise the different components of EF. The children must be able to focus on the task and pay attention, remember and update the rule/s, inhibit the prepotent response from the previous rule, and engage flexibility to switch rules. The MEFS task has been shown to be valid with more than 5,000 children and is predictive of school readiness and achievement (Carlson, Zelazo, & Faja, 2013). An age-appropriate starting level is selected automatically by the app based on test norms. Testing continued with a criterion score of 80 percent at each level. If the criterion score was not met at the starting level, the app automatically regressed levels until a lower level was passed, setting the basal level.

There are seven levels within the MEFS task. The first involves simply sorting the cards by animal category. For example, in level 1, a horse may be displayed on the left and a duck displayed on the right. The child is asked to sort the cards by the animal displayed. First the duck game is played: ducks go into the duck box. Next, the child must switch to the horse game: horses go into the horse box.

Level 2 focuses on size. For example, an elephant may be presented in the same colour and shape in both boxes. On the left is a large elephant; on the right is a small elephant. First, the child is asked to sort by size: large elephants go in the large elephant box, small elephants go in the small elephant box. The child is then asked to switch the rule: small elephants go in the large elephant box, and large elephants go in the small elephant box.

Level 3 involves sorting by colour and shape. Within level 3, the background of the card is coloured and the shape is presented in black. For example, a frog shape with no coloured features is presented on a red background in the box on the left. On the right, a butterfly shape with no coloured features is presented on a blue background in the box on the right. First, the child is asked to sort by shape (frogs go in the frog box, butterflies go in the butterfly box), and later by colour (red background goes in red background box, blue background goes in blue background box). Cards always have opposing features, for example, the card on the box will show a frog with a red background, but the child will be presented a card with a frog with a blue background.

Level 4 also involves sorting by shape and colour, however, the background is plain and it is the colour of the shape that changes. For example, the child may be presented with a monkey that is coloured orange in the box on the left. On the right box, a lion that is coloured green is presented. The shapes are presented as flat 2-dimensional shapes, where only the outline of the animal is used. Here, the child is first asked to sort by shape: monkeys go in the orange-monkey box, lions go in the green-lion box. The child is then asked to sort by colour: orange shapes go in the orange-monkey box, green shapes go in the green-lion box. As above, cards are presented with opposing features.

Level 5 involves the same stimuli, however the difficulty is increased. The child is now asked to switch between the rules every few trials. For example in a continuous flow of sorting the cards, the child may be asked to play the colour game twice in a row, and then asked to play the shape game. Here, the child must listen and adapt to the rule stated by the experimenter before sorting the card. For instance, when a card is presented the experimenter will say "Play the colour/shape game".

Level 6 also involves the same stimuli, however there is the introduction of a border rule. When a card with a black border is presented, the child must sort the card by shape, but when a card with no border is presented, the child must sort by colour. In level 7, these rules are reversed. The child is asked to switch the border rule so that now, a black border means the child must sort by colour, and no border means to sort by shape.

3.2.3 Materials

For all tablet tasks, a 1st generation iPad Pro (12.9 inch) was used.

3.2.4 Methods of Analysis

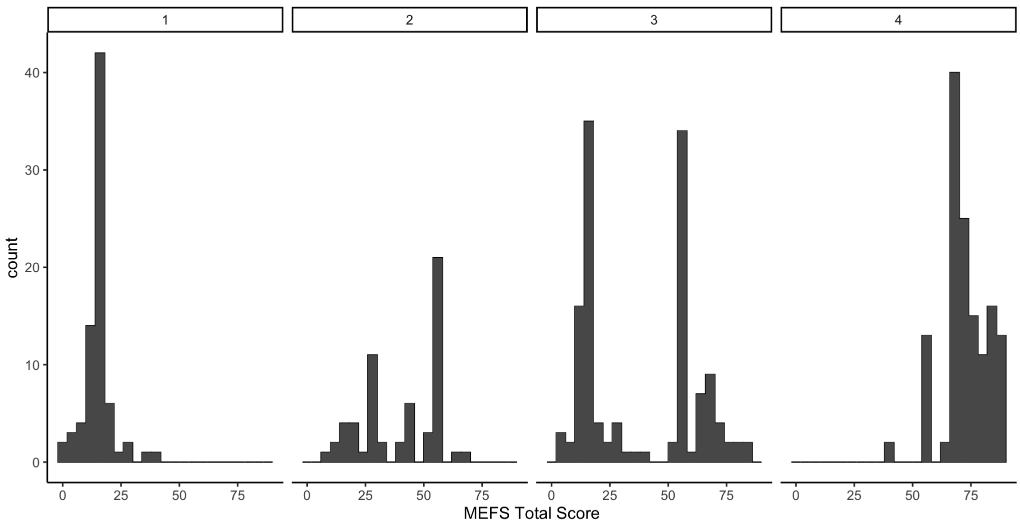
Scoring was automatically calculated by the MEFS app (MEFS App™; Carlson & Zelazo, 2014). A total score was calculated using a proprietary algorithm accounting for accuracy and reaction time. Additional scoring measures include highest level passed (0-7) and standardised score. Standardised score was calculated based on US norms; as our study took place in the UK, these were not included within our analyses. Consequently, total score was selected as the most appropriate measure of EF.

Where maternal education level was included in models, it was entered as a scaled numerical variable. Here, a maternal education of "left school at or before 16" was entered as 1, "GCSE/O levels or equivalent" as 2, "A Levels or equivalent" as 3, "Trade Apprenticeship" as 4, "Some University" as 5, "Bachelor's Degree" as 6, "Master's Degree" as 7, and "Doctorate or Professional Degree" as 8. Gender was also scaled to create a numerical variable, with boys being entered as - 0.5 and girls as 0.5 in all models. Considering the mixed findings on the importance of gender (Schirmbeck et al., 2020) and socio-economic status (Lawson et al., 2018), any partially significant findings will be explored further by removing non-contributing predictors.

Preliminary analyses showed that the distribution of total score had long tails in years 1, 2, and

4 (see Figure 3.1); thus, a Student's t-distribution was used in all models. This approximates the data distribution more robustly, leading to more normally distributed residuals. As total score provided one score per participant per year, multiple regression models were run to examine the stability of EF within individuals over time. Longitudinal models were run using the glmmTMB R package (Brooks et al., 2017). This allowed us to capture the within-subject nature of the data while also using the t family distribution. The summary function from the R package (R. C. Team, 2021) was used to provide regression coefficients. For significant predictors, the estimated magnitude and direction of the effect are reported. For models with a random intercept, a type III Wald Chi-squared test from the car package in R (Fox & Weisberg, 2019) was used to assess the contribution of each parameter in reducing residual deviance of the model. As total score had to be scaled to ensure a comparable scale for future cross-task comparison analyses presented in Chapter 4, total score was scaled within all models presented here to keep consistency in the treatment of all variables. At each stage, participants were removed for missing data. Normality was assessed by examining residuals from the DHARMa R package (Hartig, 2024) producing Q-Q plots and DHARMa residuals.

Figure 3.1
Histogram showing distribution of MEFS total score across years 1, 2, 3, and 4.



3.3 Results

3.3.1 Executive Function at 30 months

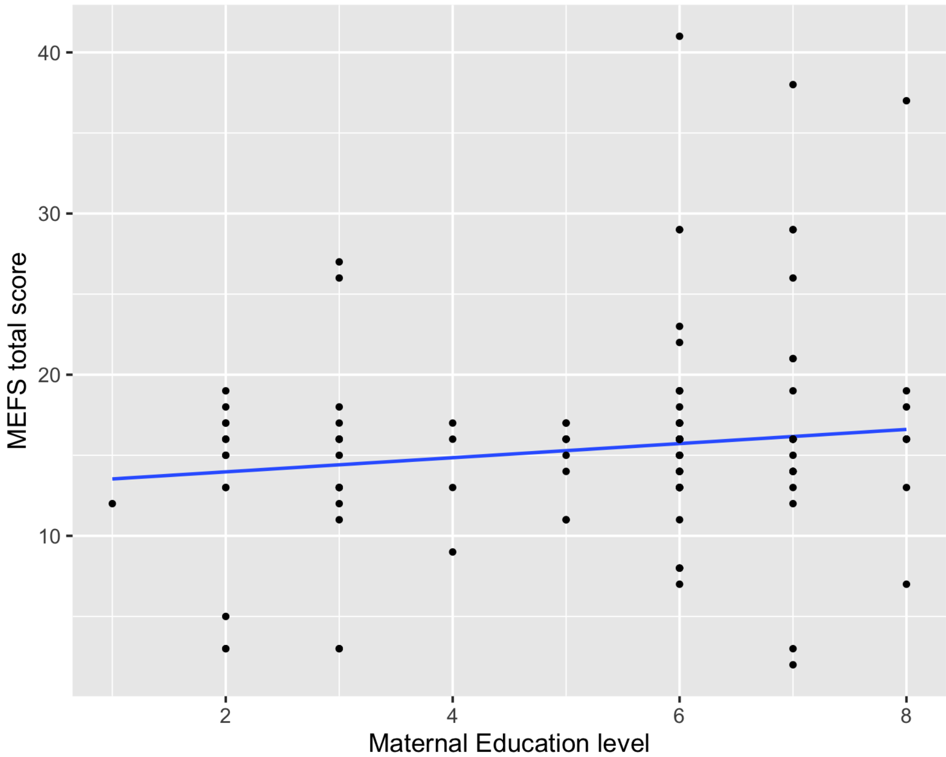
The first question I examined was whether socio-economic factors and gender were important for executive function performance at 30 months of age, the age at which the participants first completed the MEFS task. As there was data for both cohorts at this age, the data can be combined to provide a larger sample.

MEFS total score was assessed in a GLMM including maternal education level and gender as predictors of executive function performance at 30 months of age. As the outcome variable contained only one score per participant, no random intercept was added. Scale parameters were fixed to control for overdispersion and stabilise model convergence. Maternal education level and gender were added as fixed effects. All 139 participants were included in the model, run as:

$$glmmTMB(Total\,Score\,\sim\,Gender \times\,Maternal\,Education\,level)$$

There was a partially significant effect of maternal education level on total score, $\beta = 0.01$, $z = 1.90$, p =.058. There was no effect of gender and no interaction between gender and maternal education level (see Table 22). As seen in Figure 3.2, children with a more highly educated mother demonstrated a higher MEFS total score at 30 months of age.

Figure 3.2
Graph showing MEFS total score at 30 months of age by Maternal Education level



To explore this partially significant effect of maternal education level on MEFS total score, a second model was run removing gender:

$glmmTMB(Total\, Score \sim Maternal\, Education\, level)$

Within this model (model 2), maternal education level was found to significantly predict MEFS total score at 30 months, $\beta = 0.01$, $z = 1.99$, $p = .0468$. Model comparisons indicate this second model without gender provided a better model fit as the AIC was lower by more than 2, model 1 AIC = - 424.86, model 2 AIC = - 427.46. A likelihood ratio test suggested the addition of gender did not significantly improve model fit ($X^2(2) = 1.39$, $p = .50$). Therefore, model 2 was selected as the better fitting model showing an impact of maternal eduction on EF performance at 30 months of age.

3.3.2 The longitudinal stability of Executive Function

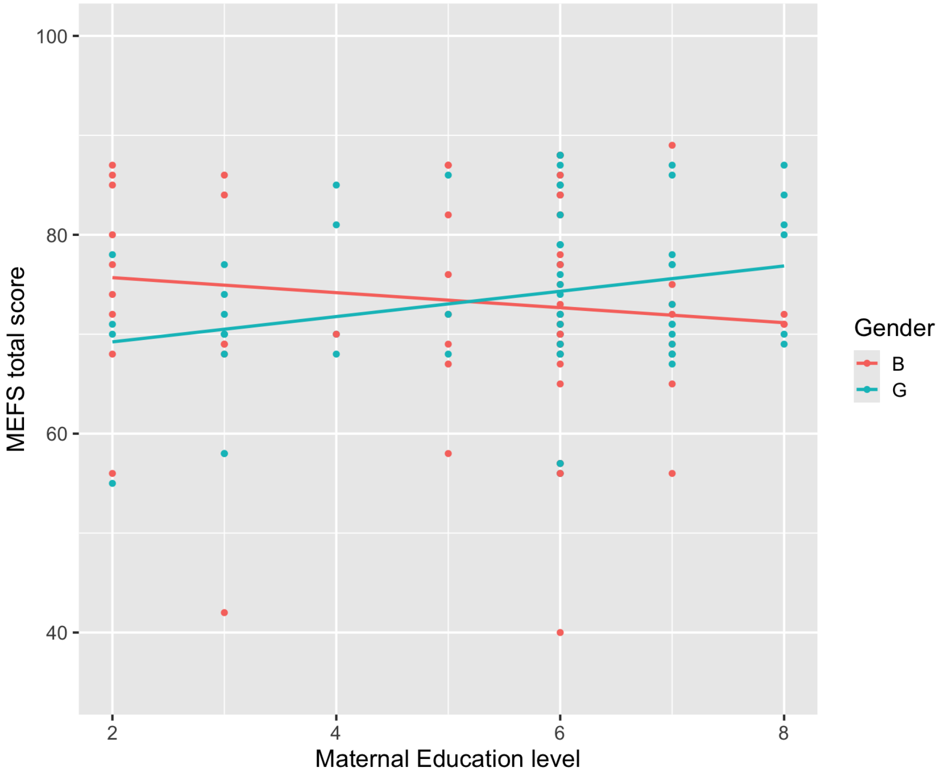
Executive Function from 30 to 78 months

To examine whether EF is longitudinally stable in early childhood, the data were first assessed across both cohorts from 30 to 78 months of age. In an initial 'baseline' model, MEFS total score at 78 months of age was assessed in a GLMM to examine the effect of maternal education level and gender on executive function performance. Gender was included in the model as evidence suggests gender is important from around 5-years of age (Palomino & Brudvig, 2022). As the outcome variable contained only one score per participant, no random intercept was added. Scale parameters were fixed to control for overdispersion and stabilise model convergence. 137 participants were included in this model, run as:

$$glmmTMB(Total\,Score \sim Gender \times Maternal\,Education\,level)$$

There was no main effect of gender or maternal education level (see Table 23). However, there was a significant interaction between gender and maternal education level, $\beta = 0.04$, $z = 2.21$, $p = .0272$. As seen in Figure 3.3, girls with a more highly educated mother demonstrated a higher MEFS total score (M = 74.90, SD = 8.31) than girls with a less educated mother (M = 71.05, SD = 8.09). This was the inverse for boys. Boys with a more highly educated mother demonstrated a lower MEFS total score (M = 72.05, SD = 10.44) than boys with a less educated mother (M = 73.79, SD = 11.33).

Figure 3.3
Graph showing MEFS total score at 78 months of age by Maternal Education level and gender.



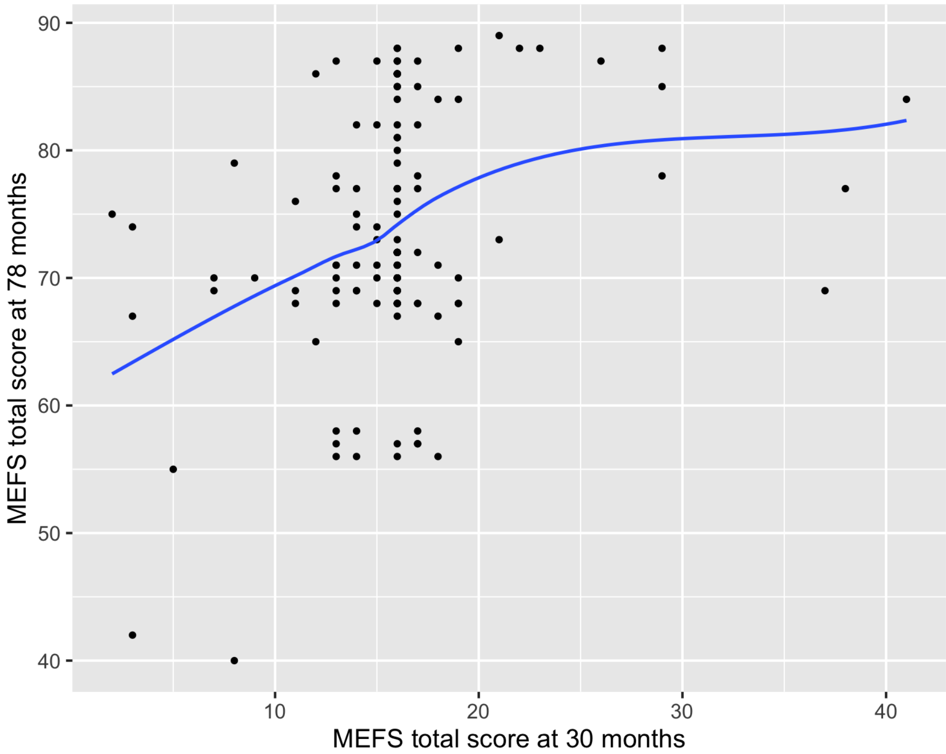
Next, 30 month MEFS total score was added as a predictor of 78 month MEFS total score in a GLMM model including maternal education level and gender as predictors. 16 children were excluded for not having a MEFS total score at both 30 months of age and 78 months of age. 121 participants were included in the model, run as:

$$glmmTMB(Total\,Score[78] \sim Total\,Score[30] \times Gender \times Maternal\,Education\,level)$$

MEFS total score at 30 months of age was found to significantly predict MEFS total score at 78 months of age, β = 0.88, z = 3.98, p <.001. As seen in Figure 3.4, a higher MEFS total score at 30 months of age predicted a higher MEFS total score at 78 months of age. Mean MEFS total score improved from 15.79 (SD = 5.80) at 30 months of age to 73.27 (SD = 9.56).

There was no main effect of gender or maternal education level (see Table 24). There was a marginal interaction between gender and maternal education level, $\beta = 0.03$, $z = 1.87$, p= .061.

Figure 3.4
Graph showing MEFS total score at 78 months of age by MEFS total score at 30 months of age.



Executive Function from 30, 42, 54, to 78 months

The previous results within this chapter demonstrate robust individual difference in EF over time but also a complicated relationship between the contributions of maternal education level and gender on EF development during early childhood. To explore this further, EF performance was examined in a more detailed model over age groups. Cohort two completed the MEFS task across all four years of test. Consequently, MEFS total score was examined in cohort two only at 30, 42, 54, and 78 months of age.

MEFS total score was examined in a GLMM including year, maternal education level, and gender as predictors. As data contained multiple scores per participant, participant code was added as a random intercept. 75 participants were included in the model, run as:

$$glmmTMB(Total\,Score\ \sim\ Year \times Gender \times Maternal\,Education\,level + (1|ParticipantCode))$$

A Wald Chi-square demonstrated both year ($X^2(3) = 3434.48$, $p < .001$) and gender ($X^2(1) =$

11.19, $p<.001$) significantly explained variance in MEFS total score (see Table 25 for regression coefficients). These effects were superseded by a significant interaction of year and gender on MEFS total score, $X^2(3) = 36.71$, $p < .001$. As seen in Figure 3.5, MEFS total score increased across years. This increase was influenced by gender, with girls demonstrating a higher increase in MEFS total score in year 2 than boys. By years 3 and 4, both genders were comparable (see Table 26 for means).

A Wald Chi-square test also demonstrated a significant effect of maternal education level on MEFS total score, ($X^2(3) = 3434.48$, $p <.001$) . As seen in Figure 3.6, children with a more highly educated mother ($M = 50.62$, $SD = 25.43$) demonstrated higher overall mean total scores on the MEFS task than children with a less educated mother ($M = 44.11$, $SD = 25.66$).

Figure 3.5
Graph showing MEFS total score over year from 30 to 78 months of age by gender.

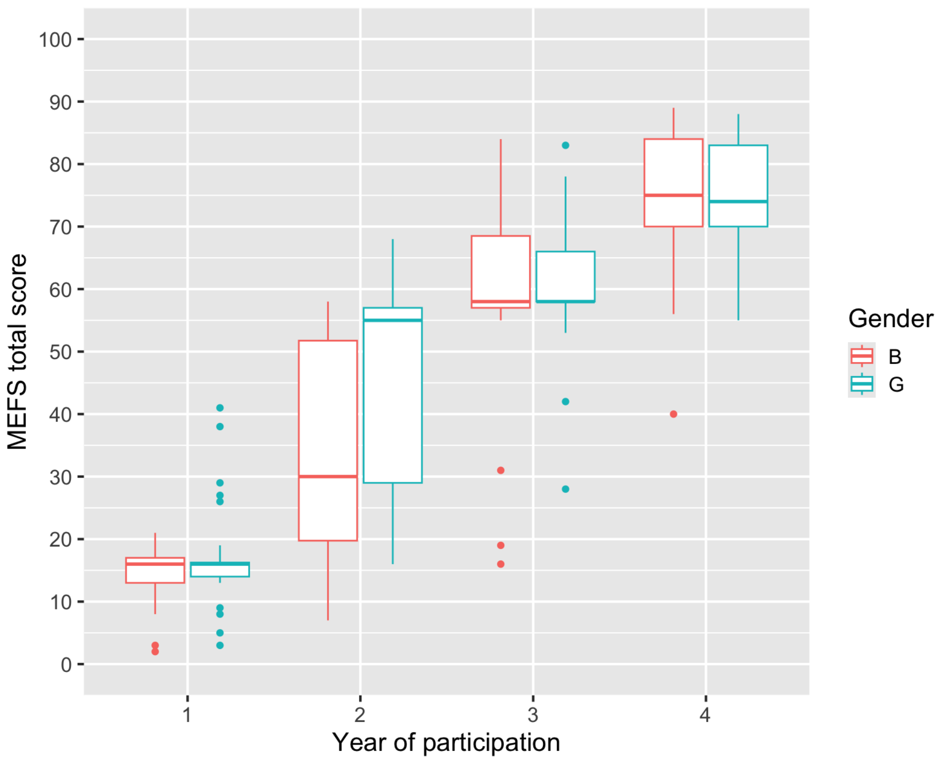
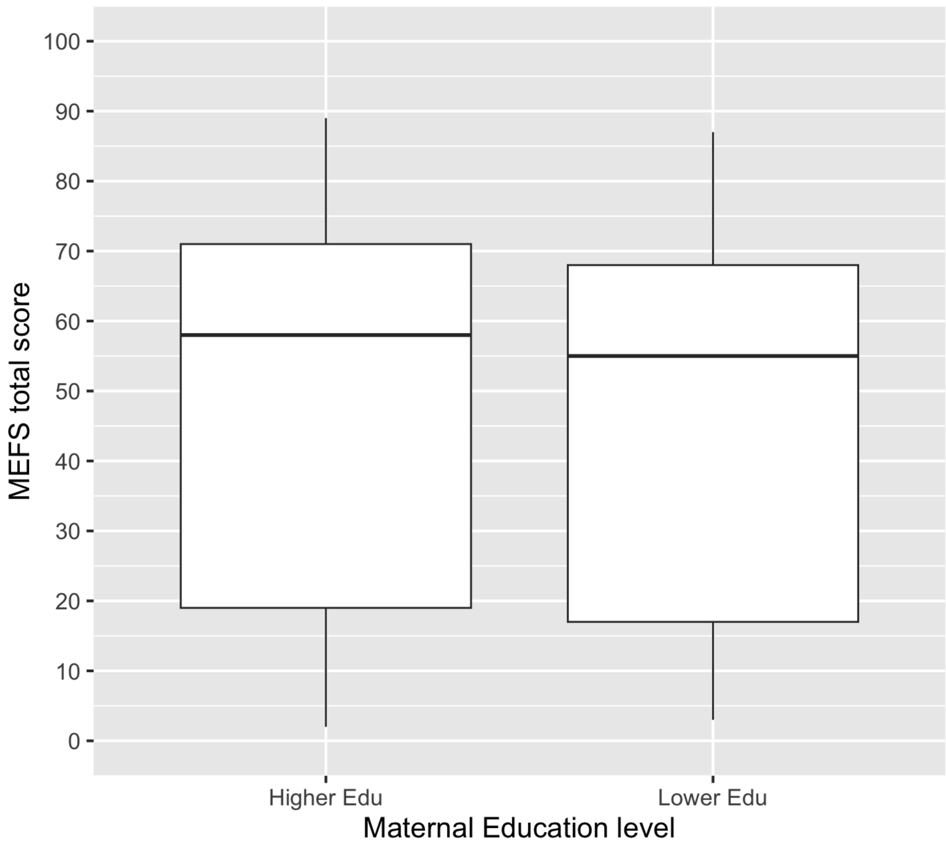


Figure 3.6
Graph showing MEFS total score by Maternal Education level category.



Within this model, the random intercept for participant code had a variance of 0.0012, suggesting very low variability in MEFS total score across participants. This suggests fixed effects within the model explain the majority of variance in MEFS total score across participants. Overall, gender contributes to variance in EF but only during particular points in development. Maternal education level had a consistent influence on EF performance during early childhood.

3.4 Discussion

Within the present study, the goal was to explore the longitudinal stability of EF from the toddler period through early childhood using the same measure over time in one of the first longitudinal studies to do so.

Previous research suggests socio-economic factors including maternal education level are important for EF (Hackman et al., 2015; Lawson, Hook, & Farah, 2018). The results of the present study support this. Performance on the MEFS task at 30 months of age was influenced by maternal education level, an indicator of socio-economic status. Here, children with a more highly educated mother demonstrated higher MEFS total scores at 30 months of age. This effect of maternal education continued to be important at 78 months of age, however this relationship involved gender. For girls, the positive association between maternal education level and MEFS total score remained at 78 months of age. For boys, the inverse was found. Here, boys with a less educated mother received higher MEFS total scores at 78 months of age. It is important to note that by 78 months of age, children who are struggling with EF skills the most are more likely to have been identified in school, and offered intervention. These boys with a less educated mother were the ones demonstrating lower EF scores at 30 months of age; thus, it is possible these boys received additional support once entering school. To investigate this further, I plan to explore which children received interventions in our cohorts in future work.

Overall, maternal education level was consistently related to the MEFS total score, with a main effect of maternal education also being found within the longitudinal model of EF stability. Here, children with a more highly educated mother attained higher MEFS total scores regardless of age or gender. This supports previous research that indicators of parental education are strongly associated with EF development (Hackman et al., 2015; Waters et al., 2021). As discussed by Waters et al. (2021), the present study agrees with suggestions for increased access to education opportunities for parents, particularly those with a low-level of education, as a means of supporting children's cognitive development.

From examining this longitudinal model of EF, results show an expected increase in executive function from 30 to 78 months. Importantly, the fine-grained analysis of cohort two demonstrates a year-by-year increase in EF from 30 to 54 months of age, with a further age-related increase to 78 months of age. These results provide evidence that EF, measured on the MEFS task, is longitudinally stable over a 5-year period from as early as 30 months of age. The present study supports previous research indicating performance on DCCS-based measures is strongly associated with age (Doebel & Zelzo, 2015). Moreover, whilst EF predicted itself at each year, the level at which MEFS scores increased depended on gender. Both boys and girls see a similar age-related increase in EF, however girls demonstrate a steeper increase in EF at 42 months of age. Consistent with previous research (Schirmbeck et al., 2020; Palomino & Brudvig, 2022), girls demonstrated higher MEFS total scores at each age, but by 54 and 78 months of age, boys and girls EF scores were more comparable. To examine the trajectory of EF further, growth curve analyses could be modelled to produce a trajectory of average EF growth over this time period. This could allow researchers to identify children deviating from this developmental trajectory in order to provide intervention.

It is important to note that due to the Covid-19 pandemic, some MEFS assessments were conducted remotely. All guidelines from Reflection Sciences (n.d.) on how to conduct remote assessment were followed. However due to the nature of conducting this assessment online, reaction times may have been slightly reduced. Consequently, MEFS total scores may have been artificially reduced. Whilst this is not believed to be a highly impactful concern, I plan to conduct additional analyses including whether the assessment was conducted remotely in future work.

In conclusion, EF is longitudinally stable when accounting for maternal education level and gender. Children with a higher EF at 30 months of age demonstrated a higher EF across childhood through to 78 months of age.

Chapter 4

Visual Working Memory as a predictor of Executive Function in Early Childhood.

4.1 Introduction

Cognition in early childhood is strongly related to future outcomes, for example, visual cognition measured in early childhood is predictive of EF skills up to 11 years later (Rose et al., 2012). One aspect of visual cognition, VWM, has also been found to be particularly important for verbal and written communication (Daneman & Merikle, 1996). VWM is a sub-system of the working memory system that involves the ability to represent, hold, and manipulate visual information in a limited capacity over short periods of time (Cowan, 2001). This VWM system is implicated in EF, with representations of objects within the working memory system being important for EF skills (Buss & Spencer, 2014). These EF skills are consistently linked to academic achievement (McClelland et al., 2014). Recent research has suggested that to increase our understanding of these EF skills, we must understand the development of EF early, when the components of EF are also developing (Spencer et al., 2025).

Previous research has attempted to examine EF longitudinally from the early infancy period. For example, Broomell and Bell (2022) investigated EF from 5 months to 9 years of age. These researchers found that a composite of EF began to show stability from 24 months of age. This composite was created from performance on the DCCS task, the tongue task, the A-not-B task, and a crayon delay task at 24 months of age, with alternative tasks such as the day/night task being introduced at 36 and 48 months of age. From the 24 month time point, EF was found to be longitudinally stable, with each composite predicting the next age-based composite of EF. However, infant EF composites at 5 and 10 months of age, using only the A-not-B task, demonstrated no longitudinal stability. This contradicted their previous examinations of these data, where EF at 10 months of age demonstrated a continuous pattern of development to 6 years of age. These differences may have resulted from the use of a different dependent variable from the A-not-B task. Other work by Carlson et al. (2004) also suggests that EF may be stable and self-predictive from the second year of life onwards. In particular, Carlson et al. (2004) found a similar 24-month composite of EF significantly correlated with a 39-month EF composite.

Although this work is promising, there remains a limited ability to predict later EF from early infancy. This may stem from different EF-related components developing at different ages. By creating composites of EF during early infancy, we may wrongly be assuming that each component provides an equal contribution to EF across different stages of development. It is important to consider that each component of EF has its own developmental trajectory. Thus, to better our understanding of EF, it may be useful to investigate the relationships between component-level developmental trajectories and the development of EF. For instance, there are no longitudinal examinations investigating the co-development of VWM and EF across early childhood.

The goal of the present chapter is to investigate the co-development of the VWM and EF systems by looking at relationships between VWM from early infancy and toddler EF, utilising related VWM measures and the same EF task over time. I expect to find VWM measured using the previously used VWM_{PL} task in infancy and VWM_{CD} task in later childhood will both be predictive of EF on the MEFS task from 30 to 78 months of age. I expect that children who show a higher VWM ability will demonstrate a higher EF ability later. Based on the prior work reviewed in Chapter 3 showing that socio-economic status measured via maternal education influences EF development, I expect to find children with more highly educated mothers will demonstrate an overall higher EF ability. Consequently, I expect the influence of VWM on EF will be impacted by maternal education level.

4.2 Methods

4.2.1 Participants

151 children (76 girls) completed the visual working memory and executive function tasks. There were two cohorts. Cohort one began the study at six months and were tested for VWM at 6-, 18 , and 30 months and EF skill at 30 and 78 months old only. Cohort two began the study at 30 months and were tested for VWM at 30 , 42 , 54 months and EF skills at 30 , 42 , 54 , and 78 months old. Demographics are shown in Table 4.1. Average maternal education level was a Bachelor's Degree, and mean family income was 40342.68 (SD = 12000.99). Participants had normal or corrected-to-normal vision. Colour vision was examined through family history colour blindness risk, at-risk children were excluded. All participants were full-term infants.

This project was reviewed and approved by the Ethics Committee at NHS England. Parents signed an informed consent form on behalf of the child. Children received a toy and a t-shirt for participating at each lab visit. Parents were given £20 for each visit to the lab and £5 for home visits. At the home visit the child also received 3 toys totalling £5. The data reported here are a subset of a larger study examining the neural basis of visual working memory and attention in early development.

Data counts revealed that the 64 children from cohort one who completed the MEFS task in year 3, at 30 months of age, all completed the VWM_{PL} in year 1, at 6 months of age. Of these 64 children, 2 did not complete the VWM_{PL} task in year 2, at 18 months of age, and 2 did not complete the VWM_{PL} in year 3, at 30 months of age (see Table 4.2).

Cohort two completed the VWM_{CD} task in years 2 and 3, at 42 and 54 months of age. Cohort two also completed the MEFS task at 42 and 54 months of age, with an additional MEFS task completed in year 4, at 78 months of age. Of the 75 children from cohort two who completed the MEFS task in at least one year of participation, 3 did not complete the VWM_{CD} task at any age and were excluded. Of the total 72 children who completed both the MEFS and the VWM_{CD} task in at least one year of participation, 15 children did not complete the VWM_{CD} task and 14 did not complete the MEFS task in year 2, at 42 months of age. Of the 72 children who completed both tasks in at least one year of participation, a separate 15 children did not complete the VWM_{CD} task at 54 months of age, in year 3, and 6 did not complete the MEFS task at this age. Of the 72 children who completed both tasks in at least one year of participation, all completed the MEFS task in year 4, at 78 months of age (see Table 4.3).

Both cohorts completed the VWM_{PL} task at 30 months of age and the MEFS task at 78 months of age. Of the 137 children who completed the MEFS task at 78 months of age, 8 did not complete the VWM_{PL} task at 30 months of age (see Table 4.4).

Table 4.1
Demographic Characteristics

Variable	**Cohort One** N = 67	**Cohort Two** N = 84
Gender		
Boys	35 (52%)	40 (48%)
Girls	32 (48%)	44 (52%)
Maternal Education Level		
Left School before 16	1 (1.5%)	1 (1.2%)
GCSE/O Levels or equivalent	4 (6.0%)	14 (17%)
A Levels or equivalent	6 (9.0%)	11 (13%)
Trade Apprenticeship	0 (0%)	5 (6.0%)
Some University	5 (7.5%)	8 (9.5%)
Bachelor's Degree	33 (49%)	30 (36%)
Master's Degree	12 (18%)	11 (13%)
Doctorate or Professional Degree	6 (9.0%)	4 (4.8%)
Ethnicity		
White British	57 (85%)	75 (89%)
Asian	1 (1.5%)	0 (0%)
Black African	0 (0%)	1 (1.2%)
South African	2 (3.0%)	0 (0%)
White British and South American	2 (3%)	0 (0%)
White British and Asian	1 (1.5%)	2 (2.4%)
White European and Asian	1 (1.5%)	0 (0%)
White British and Black African	0 (0%)	2 (2.4%)
White British and Black Caribbean	0 (0%)	2 (2.4%)
White British and Other European	3 (4.5%)	2 (2.4%)

Table 4.2
VWM_{PL} and MEFS data counts and descriptive statistics, cohort one.

	Year 1			Year 2			Year 3		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
CP_C	64	0.64	0.18	62	0.66	0.20	62	0.63	0.15
CP_{NC}	64	0.45	0.18	62	0.43	0.19	62	0.44	0.19
MEFS Total Score							64	15.74	5.65
Total Participants	**64**			**62**			**64**		
Total N	64								

Table 4.3
VWM_{CD} and MEFS data counts and descriptive statistics, cohort two.

	Year 2			Year 3			Year 4		
Variable	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
K_{MAX}	57	1.28	0.68	57	2.06	0.68	72		
MEFS Total Score	58	50.3840.78	15.93	66	60.33	12.29	72	75.46	9.03
Total Participants	**58**			**66**			**72**		
Total N									

Table 4.4
VWM_{PL} and MEFS data counts and descriptive statistics across cohorts.

Variable	**N**	**Mean**	**SD**
CP_C at 30 months	129	0.65	0.16
CP_{NC} at 30 months	129	0.43	0.18
MEFS Total score at 78 months	137	73.27	9.56
Total N	137		

4.2.2 Procedure

For cohort one, participants completed the VWM_{PL} task outlined in Chapter 2 and the MEFS task outlined in Chapter 3. For cohort two, participants completed the VWM_{PL} and VWM_{CD} tasks outlined in Chapter 2 and the MEFS task outlined in Chapter 3.

4.2.3 Materials

For the VWM_{PL} and VWM_{CD} eyetracking tasks, an Eye-Link 1000 plus (SR Research, Ontario, Canada) was used (See Chapter 2, Materials for details). For the MEFS tablet task a 1st generation iPad Pro (12.9 inch) was used.

4.2.4 Method of Analysis

Given the stability of MEFS total score over time within our longitudinal data (see Chapter 3 for details), total score will be used as the measure of executive function performance for cross-task examinations. Visual working memory will be included in models as a predictor of EF, first using the measures from VWM_{PL} outlined in Chapter 2, CP_C and CP_{NC}. To follow the previous strategy used when applying these variables as predictors, TLT will be added alongside CP_C and CP_{NC}. VWM will then be examined as a predictor of EF using the K_{MAX} measure from VWM_{CD} (see Chapter 2 for details).

Where maternal education level was included in models, it was entered as a scaled numerical variable. Here, a maternal education of "left school at or before 16" was entered as 1, "GCSE/O levels or equivalent" as 2, "A Levels or equivalent" as 3, "Trade Apprenticeship" as 4, "Some University" as 5, "Bachelor's Degree" as 6, "Master's Degree" as 7, and "Doctorate or Professional Degree" as 8. Gender was also scaled to create a numerical variable, with boys being entered as - 0.5 and girls as 0.5 in all models.

As MEFS total score will be the dependent variable, the analytical strategy from Chapter 3 will be followed. Due to the longitudinal nature of the data resulting in hierarchical data, and the need to handle non-normally distributed data, generalized linear mixed models (GLMM) were used. The glmm function from the glmmTMB R package (Brooks et al., 2017) was used to enable the use of Student's t-distribution by using the t-family function in R. The summary function from the R package (R. C. Team, 2021) was used to provide regression coefficients. For significant predictors, the estimated magnitude and direction of the effect are reported. For models with a random intercept, a type III Wald Chi-squared test from the car package in R (Fox & Weisberg, 2019) was used to assess the contribution of each parameter in reducing residual deviance of the model. Due to the non-normal distribution of scores, total score was scaled in all models. Any variables entered as a predictor of EF were scaled and centred accordingly. At each stage, participants were removed for missing data. Normality was assessed by examining residuals from the DHARMa R package (Hartig, 2024) producing Q-Q plots and DHARMa residuals.

4.3 Results

4.3.1 Does VWM_{PL} during infancy predict EF later?

To begin to capture the longitudinal relationship between VWM and EF, I will first examine how the measures from the VWM_{PL} task in infancy relate to performance on the MEFS task later. This analysis will be conducted using data from cohort one only. Here VWM_{PL} data was examined at 6- and 18 months of age, and MEFS was examined at the earliest available time-point of 30 months of age.

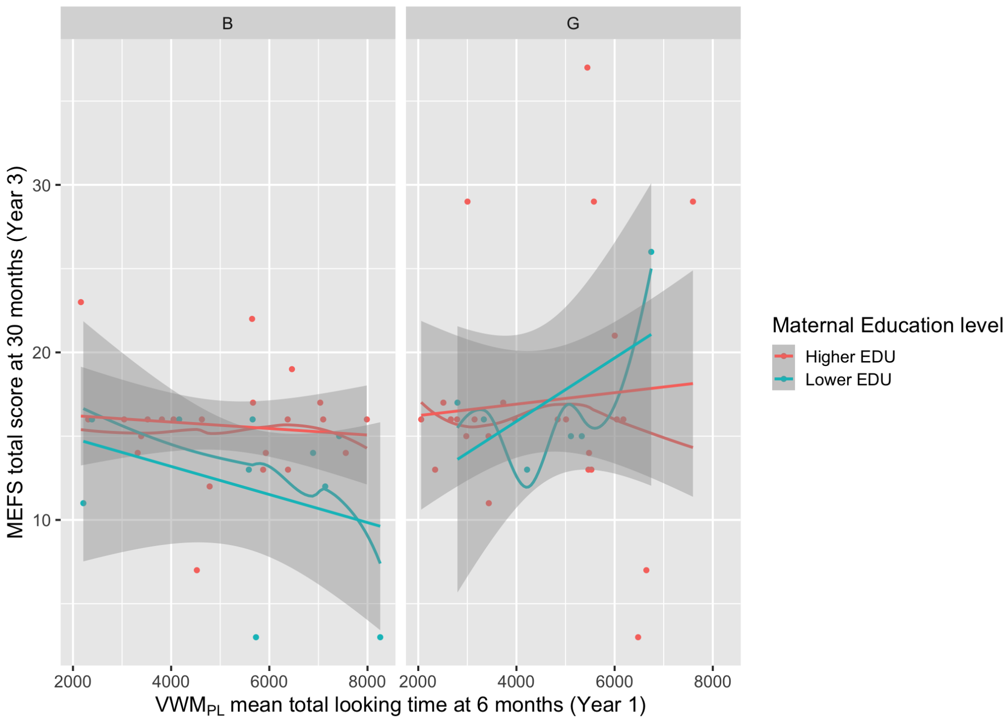
First, performance on the VWM_{PL} task at 6 months of age was examined as a predictor of EF at 30 months of age. A GLMM predicting MEFS total score at 30 months of age as a function of mean CP_C at 6 months of age, mean TLT at 6 months of age, maternal education level and gender was assessed. As the outcome variable contained only one score per participant, no random intercept was added. Scale parameters were fixed to control for overdispersion and stabilise model convergence. Maternal education level and gender were added as fixed effects.

One child was excluded for not having both a MEFS total score at 30 months of age and a CP_C at 6 months of age. 63 participants were included in the model, run as:

$$glmmTMB(Total\,Score\ \sim\ CP_C \times TLT \times Gender \times Maternal\,Education\,level)$$

There were no significant main effects found in the model (see Table 27). There was a significant interaction between gender and mean TLT at 6 months of age, β=0.02, z =2.19, p =.02874. This was superseded by an interaction between maternal education level, gender, and mean TLT at 6 months of age, β= - 0.03, z = - 2.58, p = .00991. As seen in Figure 4.1, girls showed a positive trend with TLT, where a higher TLT at 6 months of age predicted a higher MEFS total score at 30. months of age, particularly for girls with a less educated mother. This was inverse for boys, although the relationship were generally weaker. For boys, a higher TLT at 6 months of age predicted a lower MEFS total score at 30 months of age, particularly for boys with a less educated mother (see Table 28 for means). Overall, children with a more highly educated mother tended to demonstrate higher MEFS total scores, with less dependence on TLT at 6 months of age.

Figure 4.1
Graph showing MEFS total score at 30 month of age by mean TLT at 6-months of age, gender, and Maternal Education level category.



Note: The GLMM prediction is demonstrated alongside the linear model fit in order to aid visual depictions of effect directionality. Maternal education level is demonstrated here using a median split to categorise low and high levels, but maternal education was entered as continuous in all models. Here, a lower education represents below University level.

Next, performance on the VWM_{PL} task at 6 months of age was examined using the CP_{NC} measure as a predictor of EF at 30 months of age. A GLMM predicting MEFS total score at 30 months of age as a function of mean CP_{NC} at 6 months of age, mean TLT at 6 months of age, maternal education level and gender was assessed. The same 63 participants were included in the model, run as:

$$glmmTMB(Total\,Score \sim CP_{NC} \times TLT \times Gender \times Maternal\,Education\,level)$$

There were no significant main effects found in the model (see Table 29). There was a marginally significant interaction between maternal education and mean TLT at 6 months of age, β= - 0.01, z = - 1.67, p =.09598. This was superseded by an interaction between maternal education level, gender, and mean TLT at 6 months of age, β= - 0.03, z = - 2.59, p = .00953. This effect, seen in Figure 4.1, was therefore robust across both analyses.

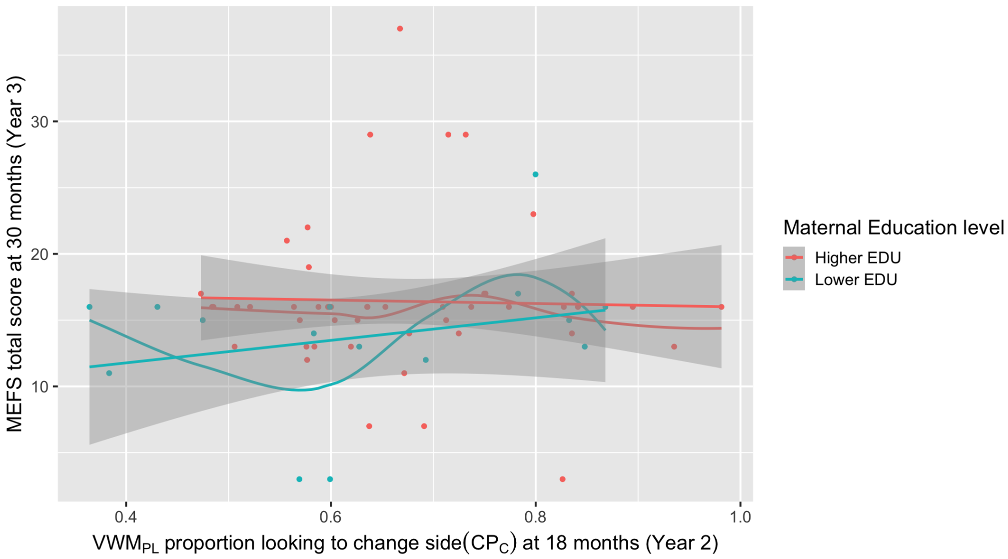
The extent to which early VWM is predictive of EF at 30 months of age was then further examined using the measures from the VWM_{PL} task at 18 months of age. A GLMM predicting MEFS total score at 30 months of age as a function of mean CP_C at 18 months of age, mean TLT at 18 months of age, maternal education level and gender was assessed. Maternal education level and gender were added as fixed effects. All 62 participants were included in the model, run as:

$glmmTMB(Total\,Score \sim CP_C \times TLT \times Gender \times Maternal\,Education\,level)$

There were no significant main effects of mean TLT at 18 months of age, gender, or maternal education level (see Table 30). There was a significant main effect of mean CP_C at 18 months of age on MEFS total score at 30 months of age, β = 0.09, z = 2.17, p =.02981. This was superseded by an interaction between maternal education level and mean CP_C at 18 months of age, β = - 0.12, z = - 2.61, p =.00917. As seen in Figure 4.2, for children with a less educated mother, a higher mean CP_C at 18 months of age predicted a higher MEFS total score at 30 months of age. Children with a less educated mother and a higher mean CP_C at 18 months of age demonstrated a mean MEFS total score at 30 months of age (M = 16.50, SD = 5.01) on par with that of children with a more highly educated mother. Children with a less educated mother who had a lower mean CP_C at 18 months of age showed the lowest mean MEFS total score at 30 months of age (M = 11.89, SD = 5.30).

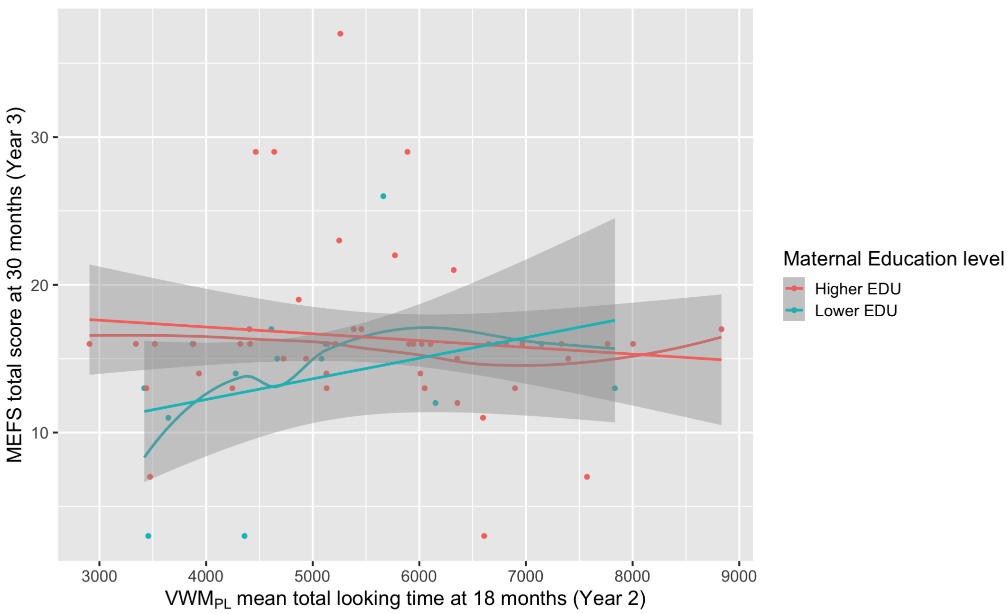
A significant interaction between maternal education level and TLT was also found, β = - 0.01, z = - 2.58, p =.00989. As seen in Figure 4.3, children with a less educated mother and a higher mean TLT at 18 months of age had a higher MEFS total score at 30 months of age. Overall, children with a more highly educated mother demonstrated higher MEFS total scores at 30 months of age, regardless of mean TLT at 18 months of age (see Table 31 for means).

Figure 4.2
Graph showing MEFS total score at 30 month of age by mean CP_C at 18 months of age and Maternal Education level category.



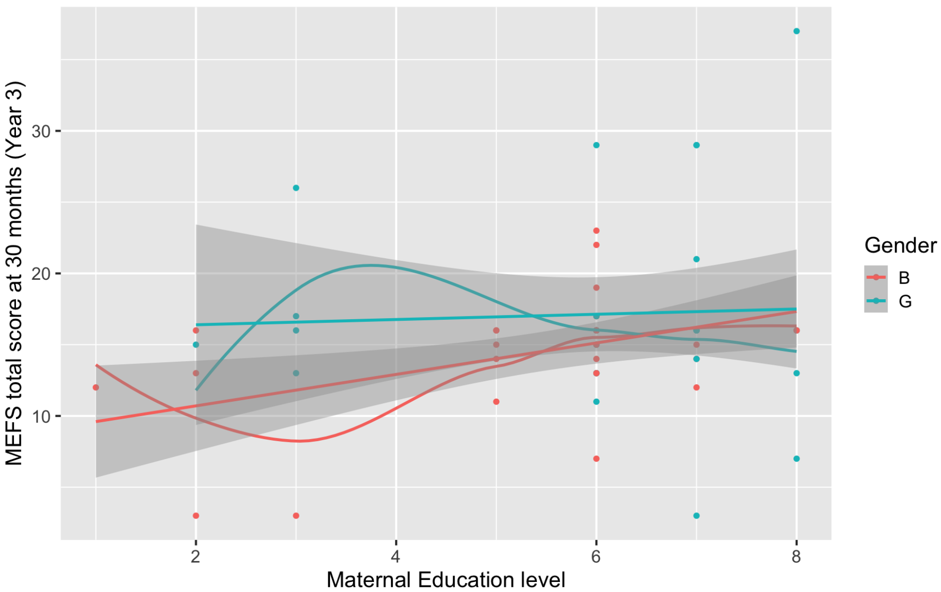
Note: The GLMM prediction is demonstrated alongside the linear model fit in order to aid visual depictions of effect directionality. Maternal education level is demonstrated here using a median split to categorise low and high levels, but maternal education was entered as continuous in all models. Here, a lower education represents below University level.

Figure 4.3
Graph showing MEFS total score at 30 month of age by mean TLT at 18 months of age and Maternal Education level category.



There was also a significant interaction between maternal education level and gender, β= - 0.03, z = - 2.55, p =.01065. As seen in Figure 4.4, a higher maternal education level predicted a higher MEFS total score at 30 months of age, particularly for boys.

Figure 4.4
Graph showing MEFS total score at 30 month of age by gender and Maternal Education level.



Note: The GLMM prediction is demonstrated alongside the linear model fit in order to aid visual depictions of effect directionality. Maternal education level is demonstrated here using a median split to categorise low and high levels, but maternal education was entered as continuous in all models. Here, a lower education represents below University level.

Finally, there was a significant interaction between maternal education level, gender, mean CP_C at 18 months of age and mean TLT at 18 months of age, β= - 0.31, z = - 3.03, p =.00244. This four-way interaction must be interpreted with caution due to a reduced sample size and the high number of predictors. For example, within this model there were only 10 boys with a lower educated mother, and 6 girls with a lower educated mother. When these groups are split further to account for TLT and CP_C level at 18 months, the data were highly unbalanced. Whilst GLMM is sensitive to unbalanced data, there is increased likelihood of over-fitting of the data. Given these concerns, I did not attemp to interpret the four-way interaction further. Note that the significant two-way interactions above were each contained within the four-way interaction.

Performance on the VWM_{PL} task at 18 months of age was then examined using the CP_{NC} measure as a predictor of EF at 30 months of age. All 62 participants were included in the model, run as:

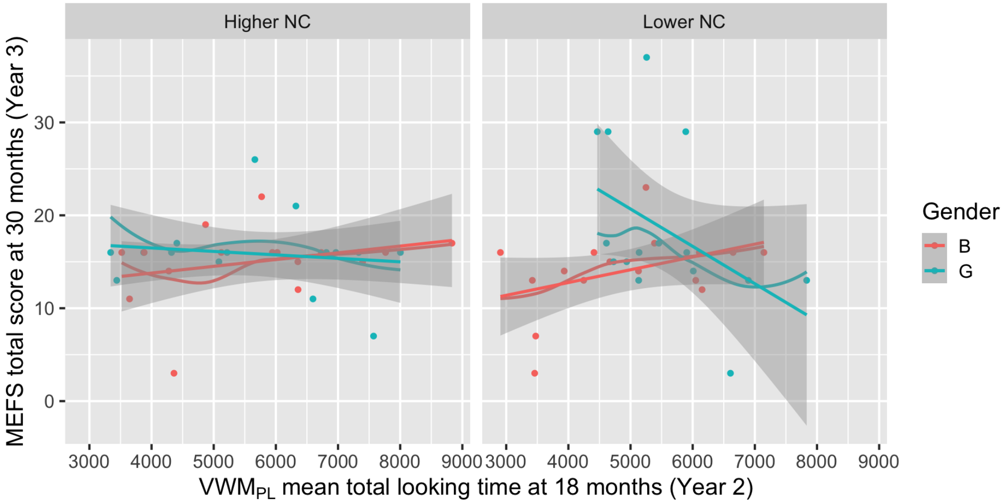
$glmmTMB(Total\,Score \sim CP_{NC} \times TLT \times Gender \times Maternal\,Education\,level)$

There were no significant main effects within the model (see Table 32). Consistent with the previous model, there was a significant interaction of mean TLT at 18 months of age and maternal education level, β= - 0.01, z = - 2.22, p =.0266. This is the same effect seen in Figure 4.3, showing consistency across models.

There was a significant interaction between gender and mean TLT at 18 months of age, β= - 0.02, z = - 2.48, p =.0131. There was also a significant interaction between mean CP_{NC} at 18 months of age and mean TLT at 18 months of age, β= 0.09, z = 2.03, p =.0428. These two-way interactions were superseded by a significant three-way interaction between gender, mean CP_{NC} at 18 months of age, and mean TLT at 18 months of age, β= 0.19, z = 2.22, p =.0267. As seen in Figure 4.5, for children with a higher mean CP_{NC} at 18 months of age, a higher TLT at 18 months of age predicted a higher MEFS total score at 30 months of age. For children with a lower CP_{NC} at 18 months of age, we see differences in the influence of TLT dependent on gender. For boys with a lower CP_{NC} at 18 months of age, a higher mean TLT at 18 months of age predicted a higher MEFS total score at 30 months of age (see Table 33 for means). For girls with a lower CP_{NC} at 18 months of age, this was inverse. It is important to note that girls with a lower mean CP_{NC} at 18 months of age demonstrated the highest MEFS total scores at 30 months of age.

Figure 4.5

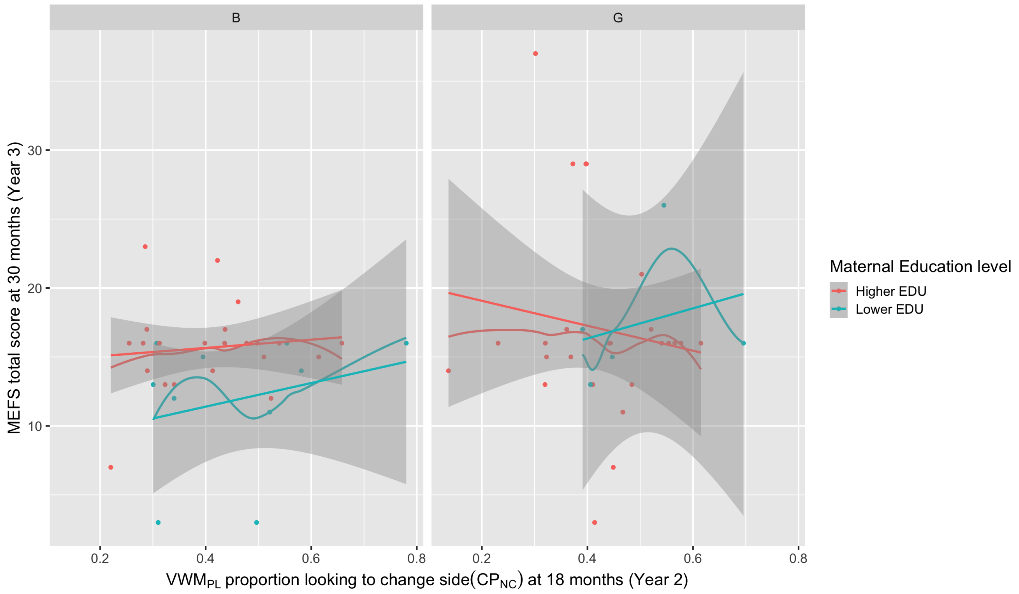
Graph showing MEFS total score at 30 month of age by gender, mean TLT at 18 months of age, and CP_{NC} category at 18 months of age.



Note: The GLMM prediction is demonstrated alongside the linear model fit in order to aid visual depictions of effect directionality. NC level refers to mean CP_{NC} at 18 months of age and is demonstrated here using a median split to categorise low and high levels, but CP_{NC} was entered as continuous in all models.

There was also a significant interaction of gender, maternal education level, and mean CP_{NC} at 18 months of age, β= - 0.28, z = - 2.06, p =.0397. As seen in Figure 4.6, a higher mean CP_{NC} at 18 months of age predicts a higher MEFS total score at 30 months of age for boys. Similarly, girls with a less educated mother who have a higher mean CP_{NC} at 18 months of age perform better on the MEFS task at 30 months of age (see Table 34 for means). For girls with a more highly educated mother, this relationships seems inverse. Here, a lower CP_{NC} at 18 months of age predicted a higher MEFS total score at 30 months of age.

Figure 4.6
Graph showing MEFS total score at 30 months of age by gender, Maternal Education level, and CP_{NC} category at 18 months of age.



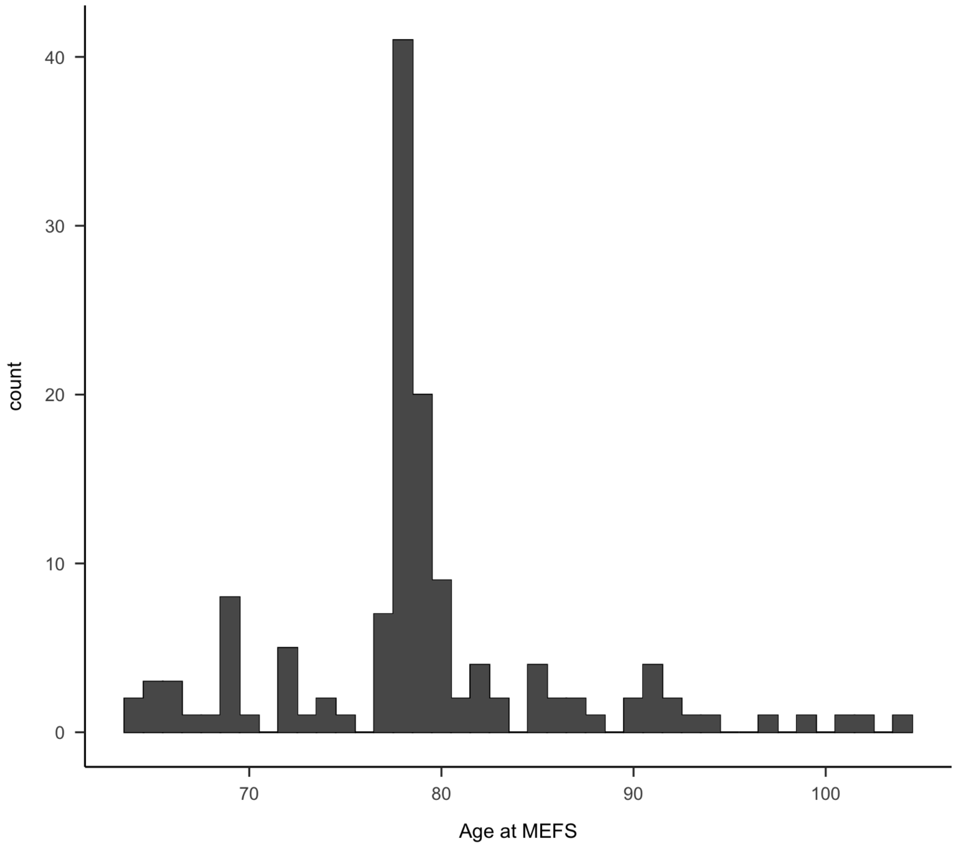
Note: The GLMM prediction is demonstrated alongside the linear model fit in order to aid visual depictions of effect directionality. Maternal Education level is demonstrated here using a median split to categorise low and high levels, but maternal education was entered as continuous in all models. Here, a lower education represents below University level.

Overall, there tend to be positive trends with the VWM_{PL} measures, where a higher TLT, CP_C, and CP_{NC} at 18 months of age predict a better MEFS total score at 30 months of age, particularly for children with a less educated mother.

4.3.2 Does toddler VWM_{PL} during predict EF later?

To continue to examine the longitudinal predictability of VWM, the data were examined across both cohorts. This allowed for a longer period of prediction. Both cohorts completed the VWM_{PL} task at 30 months of age and the MEFS task at 78 months of age. Consequently, I examined performance on the VWM_{PL} task from 30 months of age as a predictor of EF at 78 months of age. Given the use of MEFS total score at 78 months of age as an outcome variable, it is important to note that whilst the majority of children were 78 months of age during this year 4 examination, effects of covid-19 resulted in some children being tested at an older age (see Figure 4.7). Consequently, age at test will be included in models examining year 4 (78 month) MEFS as an outcome variable.

Figure 4.7
Histogram demonstrating age at test for MEFS total score in year 4.

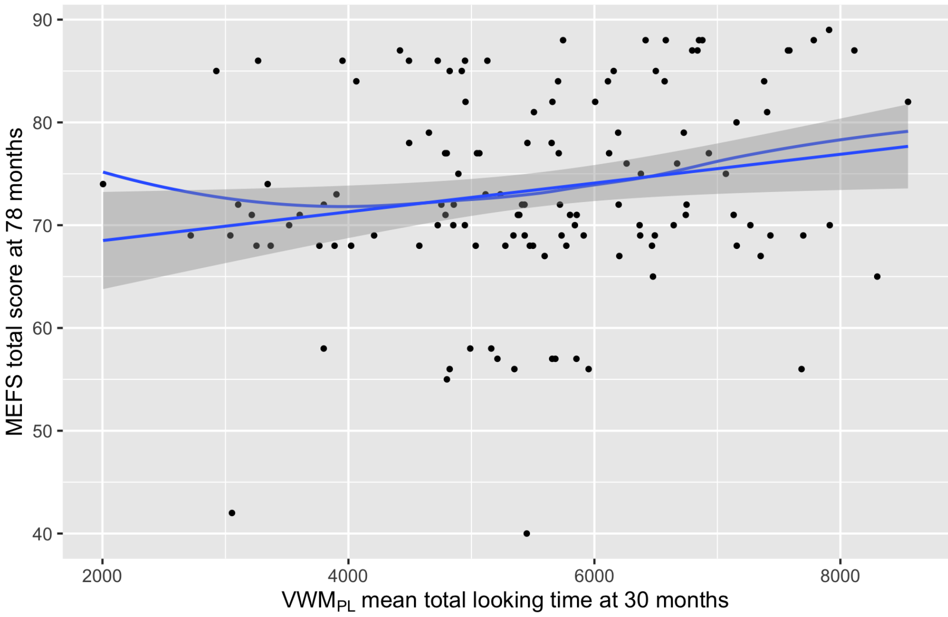


Performance on the VWM_{PL} task at 30 months of age was then examined using the CP_C measure as a predictor of EF at 78 months of age. A GLMM predicting MEFS total score at 78 months of age as a function of mean CP_{NC} at 30 months of age, mean TLT at 30 months of age, maternal education level and gender was assessed. To account for possible variation due to age at test, age at test of MEFS was added as a fixed effect with no interaction terms. All 129 participants with both a CP_C at 30 months of age and a MEFS total score at 78 months of age were included in the model, run as:

$glmmTMB(Total\,Score\ \sim\ CP_C \times TLT \times Gender$
$\times\ Maternal\,Education\,level\ +\ MEFS\,Age\,at\,test)$

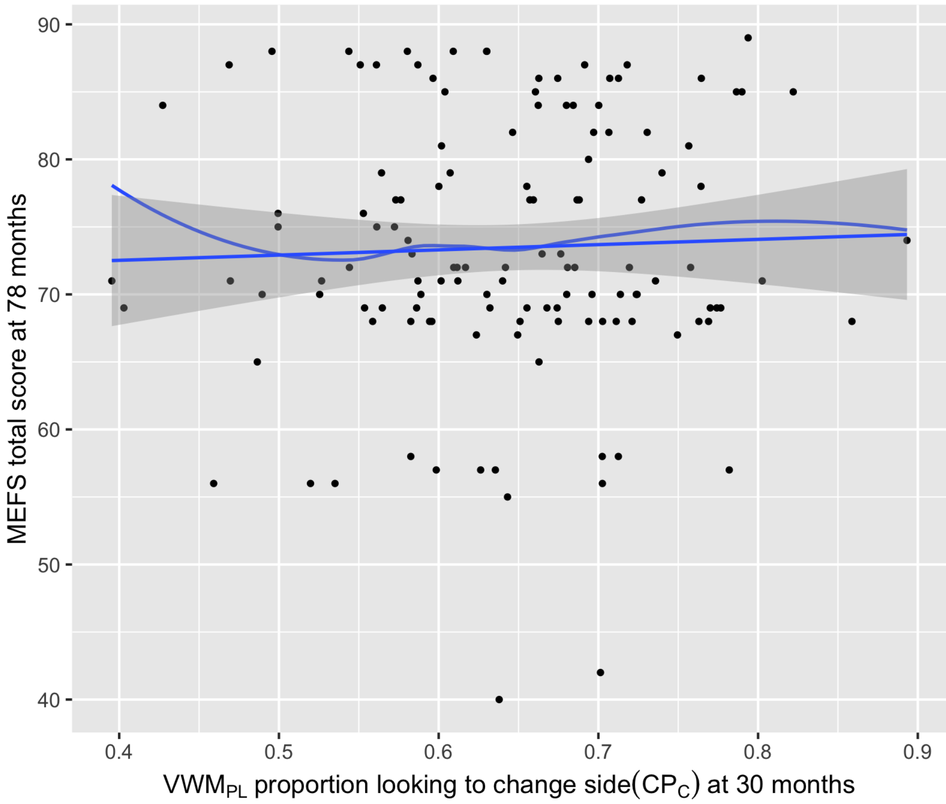
There was a main effect of mean TLT at 30 months of age on MEFS total score at 78 months of age, β= 0.02, z = 2.30, p =.0217. As seen in Figure 4.8, a higher mean TLT at 30 months of age predicted a higher MEFS total score at 78 months of age.

Figure 4.8
Graph showing MEFS total score at 78 month of age by mean TLT at 30 months of age.



There was also a significant main effect of mean CP_C at 30 months of age on MEFS total score at 78 months of age, β= 0.20, z = 2.108, p =.0350. As seen in Figure 4.9, a higher mean CP_C at 30 months of age predicted a higher MEFS total score at 78 months of age.

Figure 4.9
Graph showing MEFS total score at 78 months of age by mean CP_C at 30 months of age.



Note: The GLMM prediction is demonstrated alongside the linear model fit in order to aid visual depictions of effect directionality. Maternal Education level is demonstrated here using a median split to categorise low and high levels, but maternal education was entered as continuous in all models. Here, a lower education represents below University level.

There were no main effects of gender or maternal education level (see 35 for regression coefficients). There was a marginal interaction between gender and maternal education level, similar to that of the model of MEFS total score at 78 months presented in Chapter 3.

Performance on the VWM_{PL} task at 30 months of age was then examined using the CP_{NC} measure as a predictor of EF at 78 months of age. A GLMM predicting MEFS total score at 78 months of age as a function of mean CP_{NC} at 30 months of age, mean TLT at 30 months of age, maternal education level and gender was assessed. All 129 participants were included in the model, run as:

$$glmmTMB(Total\,Score\ \sim\ CP_{NC} \times TLT \times Gender$$
$$\times\ Maternal\,Education\,level\ +\ MEFS\,Age\,at\,test)$$

There was a main effect of mean TLT at 30 months of age on MEFS total score at 78 months of age, β= 0.02, z = 2.14, p =.0325. This is the same effect as can be seen in Figure 4.8, showing consistency across models. There were no main or interaction effects of mean CP_{NC} at 30 months of age, gender, or maternal education level (see Table 36). There was no main effect of age at test of MEFS.

4.3.3 Does VWM_{CD} predict later EF?

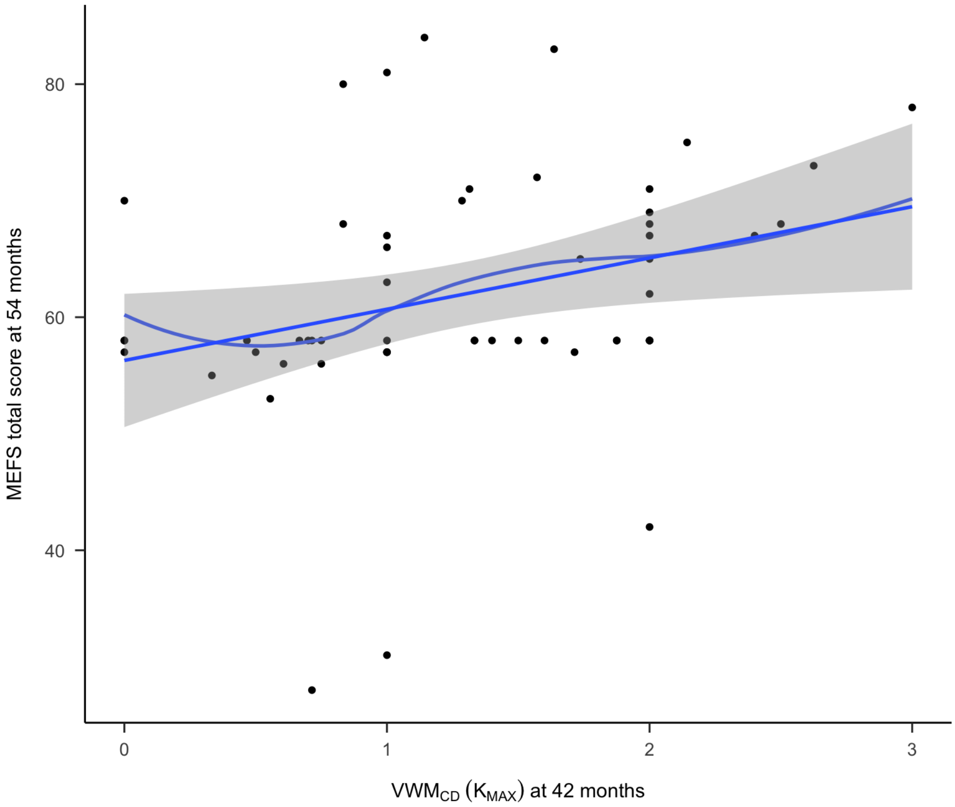
To further understand how VWM relates to later EF, the question of whether VWM capacity predicts later EF was assessed. Performance on the VWM_{CD} task at 42 and 54 months of age was examined, using data from cohort two only. Data collection in year 3, at 54 months of age, was impacted by the covid-19 pandemic similarly to the year 4 data. Consequently, age at test will be included within this model.

Firstly, VWM_{CD} at 42 months of age was examined as a predictor of EF at 54 months of age. A GLMM predicting MEFS total score at 54 months of age as a function of maximum capacity, K_{MAX}, at 42 months of age, maternal education level and gender was assessed. To account for possible variation due to age at test, age at test of MEFS was added as a fixed effect with no interaction terms. 17 participants were excluded for not having both a K_{MAX} at 42 months of age and a MEFS total score at 54 months of age. 55 participants were included in the model, run as:

$$glmmTMB(Total\,Score \sim K_{MAX} \times Gender \times Maternal\,Education\,level + Age\,at\,test)$$

There was a significant main effect of K_{MAX} at 42 months of age on MEFS total score at 54 months of age, β= 0.04, z = 2.17, p =.0302. As seen in Figure 4.10, a higher K_{MAX} at 42 months of age was related to a higher MEFS total score at 54 months of age. There was also a marginally signifiant main effect of gender, β= - 0.05, z = - 1.95, p =.0514. There were no main or interaction effects with maternal education level or AgeMEFS (see Table 37 for regression coefficients).

Figure 4.10
Graph showing MEFS total score at 54 months of age by K_{MAX} at 42 months of age.

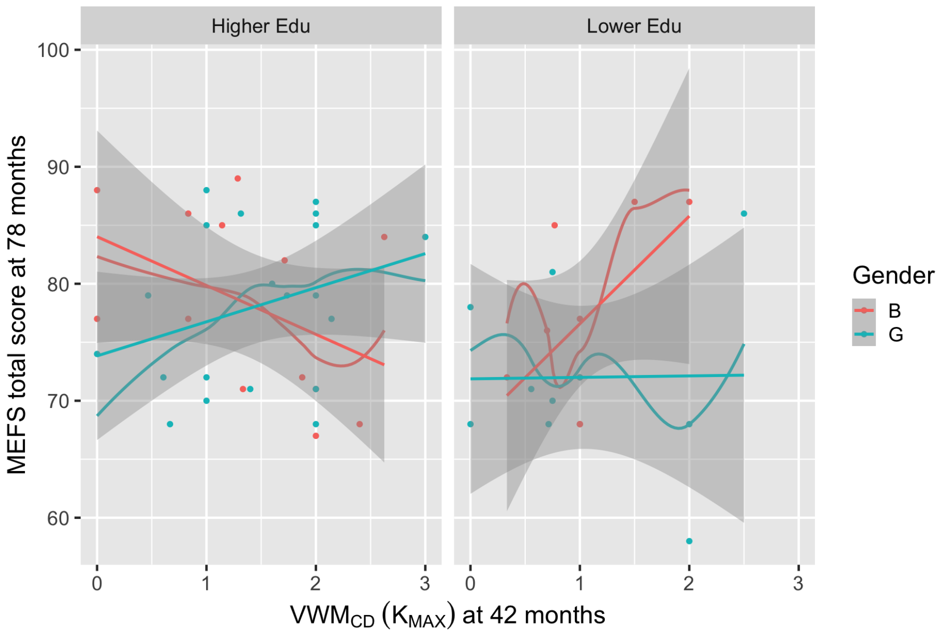


Next, VWM_{CD} at 42 months of age was examined as a predictor of EF at 78 months of age. A GLMM predicting MEFS total score at 78 months of age as a function of K_{MAX} at 42 months of age, maternal education level and gender was assessed. 20 participants were excluded for not having both a K_{MAX} at 42 months of age and a MEFS total score at 78 months of age. 52 participants were included in the model, run as:

$$glmmTMB(Total\,Score \sim K_{MAX} \times Gender \times Maternal\,Education\,level + MEFS\,Age\,at\,test)$$

There was no main effect of age at test (see Table 38 for regression coefficients). There was a significant main effect of gender on MEFS total score at 78 months of age, β= - 0.05, z = - 2.49, p =.01266, and a significant main effect of maternal education level on MEFS total score at 78 months of age, β= 0.02, z = 2.16, p =.03116. There was also a significant interaction of maternal education level and gender, β= 0.05, z = 2.80, p =.00518, and a significant interaction of maternal education level and K_{MAX} at 42 months of age, β= - 0.04, z = - 2.03, p =.04248. These effects were superseded by a significant interaction between maternal education level, gender, and K_{MAX} at 42 months of age, β= 0.16, z = 4.40, p <.001. As seen in Figure 4.11, a higher K_{MAX} at 42 months of age predicted a higher MEFS total score at 78 months of age for girls with a more highly educated mother and for boys with a less educated mother. This was inverse for boys with a more highly educated mother. For boys with a more highly educated mother a lower K_{MAX} at 42 months of age predicted a higher MEFS total score at 78 months of age. Girls with a less educated mother showed the least dependency on K_{MAX} at 42 months of age, with MEFS total scores at 78 months of age being comparable despite a higher or lower K_{MAX} at 42 months of age (see Table 39 for means).

Figure 4.11
Graph showing MEFS total score at 78 months of age by Maternal Education level, gender, and K_{MAX} at 42 months of age.



Note: The GLMM prediction is demonstrated alongside the linear model fit in order to aid visual depictions of effect directionality. Maternal Education level is demonstrated here using a median split to categorise low and high levels, but maternal education was entered as continuous in all models. Here, a lower education represents below University level.

Finally, VWM_{CD} at 54 months of age was examined as a predictor of EF at 78 months of age. A GLMM predicting MEFS total score at 78 months of age as a function of K_{MAX} at 54 months of age, maternal education level and gender was assessed. 20 participants were excluded for not having both a K_{MAX} at 54 months of age and a MEFS total score at 78 months of age. 52 participants were included in the model, run as:

$$glmmTMB(Total\,Score \sim K_{MAX} \times Gender \times Maternal\,Education\,level + MEFS\,Age\,at\,test)$$

There were no main effects or interactions of maternal education level, gender, or K_{MAX} at 54 months of age on MEFS total score at 78 months of age (see Table 40). There was also no main effect of age at test.

Overall, we see VWM_{CD} is predictive of later executive function, with K_{MAX} at 42 months predicting MEFS total score at 78 months when accounting for maternal education level and gender.

4.4 Discussion

Within the present chapter, the goal was to investigate the co-development of the VWM and EF in the first study to track these systems longitudinally from early infancy. The first research question was whether measures from a VWM task during early infancy could predict individual differences in toddler EF. Results showed that at both 6 and 18 months of age, a measure of visual exploration on the VWM_{PL} task, TLT, consistently predicted performance on the MEFS task at 30 months of age. This measure of visual exploration demonstrates children's ability to sustain attention to the task, and was previously found to be longitudinally stable from 6 to 54 months of age.

At 6 months of age, TLT was particularly important for girls with a less educated mother. In particular, a higher TLT predicted a higher MEFS total score for these girls. By contrast, girls with a more highly educated mother showed less dependency on their looking behaviour at 6 months of age. For boys with a less educated mother, an inverse relationship was found with a lower TLT at 6 months predicting a higher MEFS total score at 30 months of age. This is an unexpected finding, however by 18 months of age children with a less educated mother demonstrated a positive effect of TLT for both genders. This may suggest that boys with a less educated mother who struggle to maintain fixation to the task, resolve this behaviour by 18 months of age, attend to the task, and go on to obtain a higher MEFS total score at 30 months. In summary, results generally showed that better sustained attention at 6 months of age predicted higher EF at 30 months of age.

At 18 months of age, TLT predicted a higher MEFS total score at 30 months for all children with a less educated mother, although children with a less educated mother had lower MEFS total scores overall. Given the longitudinal stability of TLT shown previously, it is possible that better sustained attention at 18 months carries over to impact performance on the MEFS task with better sustained attention in that task as well. This may facilitate flexible dimensional attention to the appropriate features on the cards in the MEFS task, an ability that is proposed to underlie successful sorting on this task (Buss & Kerr-German, 2019). In addition, the CP_C measure at 18 months predicted a higher MEFS total score at 30 months of age, particularly for children with a less educated mother. A higher CP_C indicates the ability to detect and sustain attention to novelty. It may be that directing attention to novelty early facilitates the flexible allocation of attention later in the MEFS task.

More complex interactions were found with the CP_{NC} measure at 18 months of age. In general, higher CP_{NC} scores at 18 months predicted higher MEFS scores at 30 months, except for girls with a more highly educated mother. In addition, higher TLT at 18 months predicted higher MEFS scores at 30 months, except for girls with lower CP_{NC} scores. It is not clear why girls showed these inverse patterns. Girls with a more highly educated mother consistently obtained higher MEFS total scores. Moreover, there was a great deal of variation in MEFS total scores for the girls with a more highly educated mother who showed a lower CP_{NC} at 18 months of age. It may be that a higher maternal education level is acting as a protective factor, and some of these girls are able to arrive at higher EF skills despite poorer VWM earlier in development.

The next research question assessed whether measures of VWM during the toddler period were related to later EF, over a longer four year period. Results show that a higher mean TLT at 30 months of age was found to predict higher EF at 78 months of age. Once again, the ability to sustain attention to the task was important for later EF. A higher mean CP_C at 30 months of age was also important for EF at 78 months of age. Note that although CP_C showed less longitudinal stability in Chapter 2, this measure was longitudinally stable in year

1 which included the 30-month-olds from cohort 2. As discussed by Buss and Spencer (2014), DCCS-type tasks do not place high demands on the working memory system; however, it is critical to use VWM to detect and contrast the features on each target card to sort the features correctly. It has been suggested that the ability to orient attention within working memory contributes to the ability to form internal representations of objects (Scerif & Shimi, 2021). The present study may be demonstrating this: in the less demanding context of starting on the changing side, 30 month-old children who are better able to orient their attention to attend to the novel colours in the VWM_{PL} task, may be better able to utilise these skills at 78 months of age to successfully sort cards on the MEFS task.

The final question evaluated whether VWM capacity was related to performance on the MEFS task across childhood. Firstly, this was assessed over a one year period. Results show that a higher maximum capacity on the VWM_{CD} task at 42 months of age predicted a higher MEFS total score at 54 months of age. This was then assessed over a three year period. Generally, a higher maximum capacity at 42 months of age also predicted a higher MEFS total score at 78 months of age, with one exception—boys with a more highly educated mother. For these boys, an inverse relationship between maximum capacity and MEFS total score was found. This may demonstrate that whilst a higher capacity may enable children to perform better on the MEFS task, this is not a requirement for success. As discussed previously, the working memory requirements of the MEFS task are not overly challenging (Buss & Spencer, 2014). As the working memory requirements in the MEFS task are lower, a higher capacity may not be necessary to succeed in the task. These boys with a more highly educated mother continuously demonstrate relationships where measures of sustained attention are important for their EF. Consequently, boys with a higher educated mother may rely less on VWM capacity and more on the ability to attend to the task, and the correct features in the task. So, for these boys being able to sustain attention and having a more highly educated mother may help them overcome a lower VWM capacity.

It is also important to note that girls with a less educated mother showed little dependency on capacity at 42 months of age for performance on the MEFS task at 78 months of age. These girls demonstrated lower MEFS total scores regardless of capacity. This may also support the proposal that VWM capacity does not fully determine success on the MEFS task.

By 54 months of age, maximum capacity was not found to be related to EF at 78 months of age. This may support the idea that the capacity itself is not as important for MEFS, and instead previous relationships are a reflection of the multiple systems engaged during the VWM_{CD} task that are also important within the MEFS task. However, by 54 months of age there is also much less variance in capacity scores, with all children being at ceiling level in SS1. Given the reduced sample size and lower variance for capacity estimates, the statistical power in the present study may not have been enough to assess the differences between children with a lower or higher capacity estimate at this age.

It is important to note that the reported effects were consistently stronger for children with a less educated mother, particularly boys. Whilst children with a less educated mother were found to show lower MEFS scores across almost all ages, these relationships may demonstrate that having a good VWM and ability to sustain attention from an early age is an important protective factor for these children. In conclusion, the present study finds that the development of VWM and EF are related across multiple ages in early childhood. Generally, measures of VWM were positively related to later EF from early infancy. In the final chapter, I place these findings in the context of the broader literature.

Chapter 5

General Discussion

5.1 Summary and Integration of Findings

Throughout this thesis, I have aimed to investigate how the VWM system is integrated with emerging EF skills in early development. To assess this, it was necessary to first understand the development of the VWM system in, to my knowledge, the first longer-term longitudinal study to do so from early infancy. It was also necessary to assess the longitudinal stability of EF across early childhood using a consistent measure over time. After these assessments were made, it was possible to examine the co-development of VWM and EF to begin to understand how these systems interact over time. In the present chapter, I will review the key findings from each chapter and integrate these findings with the broader literature to address how this thesis contributes to our understanding of EF. I will consider challenges and limitations within this project, including a discussion of the statistical methods used. I will finish the discussion by considering the real world implications of the findings presented within this thesis.

In chapter 1, a review of the literature unveiled discontent with the current component-based approach to examining EF over development. Importantly, this component approach is informed by the adult literature. When applied to EF in childhood, there is little consistency across findings. This may, in part, be due to the statistical methods used. Confirmatory factor analysis methods of examining EF have repeatedly been criticised for their overzealous approach to latent variables (Miller et al., 2012). As discussed, multiple researchers have argued for a re-conceptualisation of EF, moving beyond this latent variable approach, particularly early in development where multiple systems are likely co-developing together (Spencer et al., 2025). In order to provide an informed re-conceptualisation, we must first understand more about the role of multiple systems in EF from early infancy. However, the literature review made it clear that whilst one of the systems heavily involved in EF, working memory, can be tracked from early infancy using the sub-component of VWM, there have been no studies of the co-development of VWM and EF in the literature. Having identified this clear gap, the goal of the thesis was to understand the co-development of VWM and EF from early infancy.

In chapter 2, I used two VWM tasks: a preferential looking task to be used from infancy and a canonical change detection task to anchor our understanding of the preferential looking task in childhood. I then conducted the first longitudinal examination of VWM from early infancy. This allowed me to assess the replicability of findings from prior cross-sectional work.

I began by assessing the longitudinal stability of measures from the VWM_{PL} task. Findings revealed that two measures from this preferential looking task are stable from 6 to 54 months of age. The first measure was TLT, a measure of sustained attention to the task. The second measure was CP_{NC}. This CP_{NC} measure captures a child's ability to detect and sustain attention to novelty in a resource-demanding context. In particular, this measure assesses a child's detection of the changing side when they start the task on the non-changing side. Consequently, the child must consolidate each item in working memory, detect that no change is occurring, release fixation from the non-changing side and switch to the changing side. Once on this changing side, the child must then consolidate the new display in working memory, detect any changes, and update working memory as new colours are presented. This measure showed longitudinal stability where the previous widely used measure of change preference, CP_{10} did not.

The next measure from this task, CP_C, assesses the child's ability to detect change when starting on the changing side. This measure showed longitudinal stability from year 1 to year 2, but not from year 2 to year 3. It is possible the lack of longitudinal stability from year

2 to year 3 reflects the simpler demands placed on the child when they start looking at the changing side in the VWM_{PL} task. That is, this measure might show less subject-specific variance in year 3 because it was relatively easy for older children to remain on the changing side when starting there. I acknowledge, however, that this is likely not the complete story as cohort 1 failed to show longitudinal stability in this measure from 18 to 30 months, while cohort 2 showed longitudinal stability from 30 to 42 months. Thus, there is variability in the developmental trajectory of this measure. More generally, however, results from analyses of the VWM_{PL} task generally showed that VWM is stable longitudinally from infancy to later childhood, particularly when examined in resource-demanding contexts.

An assessment of cross-task relationships revealed that the data did not replicate the findings of Simmering (2016) within a longitudinal sample. This is likely due to the limitations of the measures used. When re-examining these models using the longitudinally stable measures of CP_{NC} and TLT, I found that both measures were positively related to performance on the VWM_{CD} task. These findings add further support for the use of the new 'first-look' measures for assessing performance on the VWM_{PL} task. Going forward, future research should make use of these measures, replacing the previous measures which do not account for the differing levels of demand based on the context of the child's looking behaviour.

Importantly, the CP_{NC} and TLT measures predicted performance on the VWM_{CD} task up to two years later, although some of these interactions were complex. Interestingly, these findings revealed the impact of maternal education level on VWM. The effect of maternal education level was consistent across all models predicting maximum VWM capacity, including in interactions with CP_{NC} and TLT. This aligns with previous research indicating that the VWM system is influenced by maternal education level, with children with a less educated mother demonstrating difficulty in suppressing distraction on the VWM_{PL} task (Wijeakumar et al., 2019). The findings from this thesis add to these prior findings, highlighting that the effect of maternal education level extends to measures of VWM capacity.

In chapter 3, I introduced the literature assessing the longitudinal stability of EF from early childhood. Whilst studies have attempted to examine EF from early childhood, longitudinal studies have typically used different tasks across different ages. Thus, the goal of chapter 3 was to examine the longitudinal stability of EF from early childhood using a consistent measure of EF at all ages. Thus, I used the MEFS task and studied the developmental trajectory of EF from 30 to 78 months in the same longitudinal cohorts studied in chapter 2.

Using the MEFS task, EF was found to be longitudinally stable from 30 to 78 months of age. At each age, age-related developments in EF were reflected by an increase in MEFS total score, and individual differences in the total score were predictive over development across both cohorts of children. Thus, EF is longitudinally stable from 2.5 to 6.5 years of age and MEFS provides a robust measure of individual differences in EF that can be used consistently across this age range. These findings further support previous research suggesting that EF is stable from the second year of life (Carlson et al., 2004).

Importantly, maternal education level was found to be an important predictor of MEFS total score. Generally, children with a more highly educated mother had higher MEFS total scores across all ages. This aligns with previous research showing parental education is strongly associated with EF development (Hackman et al., 2015; Waters et al., 2021). The implications from this finding are discussed below.

The present study also clarified that effects of gender on MEFS performance were isolated to the 42-month time point. Girls and boys scores on the MEFS task were comparable, except at 42 months of age where girls showed a much sharper increase in MEFS performance than boys. By 54 months of age, the boys showed a similar sharp increase resulting in MEFS scores being more comparable from this age. Whilst the present study did find an influence of gender on EF at 78 months of age, this was discussed in relation to maternal education level effects and the possibility that some children were receiving interventions. This remains an important issue to investigate in future work with this cohort.

It is useful to note here that there is disagreement regarding the nature of the MEFS task. Some researchers have viewed this as a cognitive flexibility task, while other researchers emphasise the involvement of multiple systems including working memory to remember the rules and inhibitory control to inhibit the prepotent response (see Spencer et al. 2025). The interpretation of this task is important when placing these findings in the context of chapter 4 where I looked at the co-development of VWM and EF.

Chapter 4 explored how the trajectories of VWM and EF co-develop across early development. Integrating the measures used previously in this thesis, chapter 4 assessed this question in three stages. First, I explored the relation between VWM in early infancy and toddler EF, at the first possible point of assessment using the MEFS task. Next, I assessed the relation between VWM at 30 months of age and MEFS at 78 months of age. Finally, I examined the relationship between VWM measured on the VWM_{CD} task and EF from 42 to 78 months of age. Here, I will integrate and discuss these findings, considering theories of EF and prior research.

Results showed consistent relationships between measures from the VWM_{PL} task in infancy and EF at 30 months of age. At 6 and 18 months of age, TLT positively predicted MEFS total score in most instances. At 18 months of age, this relationship was particularly important for children with a less educated mother. This TLT measure is an indication of the ability to sustain attention to a task. Results also demonstrated that this TLT measure at 30 months of age was positively predictive of MEFS total score at 78 months of age. TLT may be important in two ways. First, a low TLT may demonstrate an inability to suppress distraction. Poor suppression of distraction has been linked to poorer performance on the VWM_{PL} task (Wijeakumar et al., 2019). Whilst every effort was made to ensure the research environment did not contain distractions, children may still have been distracted by other items in the room, such as their own clothing or a piece of reflective equipment. Children with a low ability to suppress distraction would be expected to perform worse on the MEFS task, as they may fail to inhibit distraction from the prepotent response on post-switch trials. Thus, a higher TLT and a higher MEFS total score may be explained through a better ability to suppress distraction in both cases. An alternative explanation is that a low TLT performance may be an indicator of a VWM system that is unable to cope with the task demands. Children may become overwhelmed when attempting to process the information on each screen, leading to a reduced TLT as the child releases fixation to reduce stress on the VWM system. Further investigations are necessary to tease these explanations apart. One method that may assist in doing this is creating quantitative simulations to model children's behaviour.

Previous simulations of a dynamic field model have provided insights into the requirements in DCCS-type tasks. Buss and Kerr-German (2019) showed that dynamic field models that were able to succeed on the DCCS task were those able to attend to the dimensions on the card, build up a stronger memory trace of the relevant dimension, and engage this dimensional representation when prompted on test trials. The representation of feature dimensions within VWM in this model was important for success on DCCS-type tasks, such as the MEFS task. The VWM system facilitates the detection and contrast of features on each target card to enable the correct sorting of the features (Buss & Spencer, 2014).

Findings from this thesis provide support for this role of 'feature contrast' within the VWM system in the MEFS task. At 18 and 30 months of age, the CP_C measure was found to be important for later EF. A higher CP_C at 18 months of age predicted a higher MEFS total score at 30 months of age, particularly for children with a less educated mother. A higher CP_C at 30 months of age also predicted a higher MEFS total score at 78 months of age. As discussed in Chapter 2, this CP_C measure does not place the VWM system in a highly demanding context, similar to the less demanding VWM context of the MEFS task. Within this less demanding context, children who can correctly contrast features and orient attention to novel colours from 18 and 30 months of age go on to be successful in the MEFS task at 78 months of age. In both the VWM_{PL} and MEFS task, children must orient their attention to the correct features. The CP_C measure captures the ability to orient attention to novel colours, whereas the MEFS task captures the ability to orient attention to the correct dimension. Thus, the relationship between these measures may demonstrate a commonality in the multiple systems involved in both tasks.

This provides support for the role of VWM in EF in the manner described by dynamic systems models (Buss & Spencer, 2014; Buss & Kerr-German, 2019). Indeed, a recent dynamic field model of visual exploration and word learning has shown that scene representations and word-feature representations, such as dimension labels like 'colour', are linked (Spencer et al., 2025). Children's learning and representations of individual feature words, such as for individual colours, map onto words such as 'colour' to drive dimensional attention when cued, for example, when asked to 'play the colour game' in the MEFS task. The features attended to in the task are driven by the words used in the task. Within the dynamic field model (Spencer et al., 2025), this feature information is consolidated within working memory, feeding down to a contrast layer in the model. This is the same layer that detects changes, such as those presented in the VWM_{PL} and VWM_{CD} tasks (see Simmering, 2016). Here, we see the same systems involved in the detection of changes in VWM and the top-down attention necessary for the MEFS task. This may explain the strong links between the CP_C measure and later performance in MEFS reported here.

Further support for the role of VWM in EF skill was demonstrated from the CP_{NC} measure; however this measure was only found to be robustly related to EF at 18 months of age. Results show that, in general, a higher CP_{NC} at 18 months predicted higher MEFS total score at 30 months. This measure captures a child's ability to orient attention to novel colours when starting on the non-changing side. As discussed previously, this requires increased resources from the VWM system. Whilst this measure was important for future EF at 18 months of age, no relationship was found from 30 months age. This may reflect the reduced working memory demands of the MEFS task. More generally, these findings show how investigating multiple measures of performance in the VWM task can reveal more complex relationships between VWM and EF. This contrasts with the CFA approach which typically assesses a single measure from each task.

The final research question evaluated within this thesis was whether measures of VWM capacity were related to EF skill across childhood. Results showed that VWM capacity at 42 months of age was positively predictive of EF at both 54 and 78 months of age. Whilst there was no relationship between VWM maximum capacity at 54 months of age and EF at 78 months of age, this was discussed in relation to limitations of statistical power. However, it is also possible this, again, reflects the lower working memory demands of the MEFS task. There was a complex interaction found for boys with a more highly educated mother who obtained higher MEFS total scores with a lower maximum capacity at 42 months of age. These findings may reflect that capacity itself is not critical for later EF, although this conclusion must be anchored to how EF was assessed in the present study using MEFS. It is possible that other EF tasks that place more demands on working memory capacity would show stronger links to working memory capacity measures earlier in development.

In many of the analyses reported here, maternal education level was seen to consistently influence VWM capacity and EF. From 30 to 54 months of age, children with a more highly educated mother showed higher MEFS total scores and higher VWM capacity estimates. It is possible that maternal education level acts as a protective factor with some children, that is, these children are able to obtain higher level of EF skills in spite of poorer VWM. One important factor here may be that mothers with a higher education may themselves have better EF skills and model such skills during interactions with the child. For example, Kao et al. (2018) demonstrated that parents who performed better on EF tasks had children who also performed better on EF tasks. However, no relationship with parental education level was found in this study. Other studies have found that highly educated parents promote children's cognitive development in a number of ways, including engaging in more stimulating activities, spending more time with their children, and engaging in more complex speech patterns (Landry et al., 2006). Other researchers have suggested that correlations between a parent's education level and child outcomes may be the result of parental characteristics that lead parents to be both a good student and good parent (see Duncan & Magnunson, 2012 for a review). Thus, there are many factors that could explain why a higher maternal education level supports a child's EF, however this interaction does have important implications discussed below.

Overall, the present study found that measures of VWM were positively related to later EF from early infancy. These findings suggest that VWM is a key system that enables us to keep track of what is where in the world, an ability that feeds into later EF skills. For instance, Buss and Spencer (2014) emphasise that this ability to track features in the world is important in EF tasks as demonstrated through modelling of behaviour from the DCCS task.

5.2 Challenges and Future Work

Although many aspects of the findings from this thesis are compelling, it is important to acknowledge several limitations. First, I note that data from this thesis was part of a larger scale project examining the emergence of VWM. Consequently, children took part in a number of other tasks. This likely led to some fatigue and certainly impacted children's ability to complete all tasks at all ages. Furthermore, fNIRS data was collected during the VWM task. This involved the children wearing a cap which may have been distracting and may have led to increased discomfort and data loss. In order to limit the impact of missing data, analyses were conducted using aggregate measures and analyses focused on year on year predictions. Whilst this successfully allowed me to examine the longitudinal data whilst reducing the impact of missing data, some of the analyses still suffered from limited power. To push beyond these limitations, future work should conduct examinations of growth curve analyses with this data set (Mirman, 2014) that model group effects in each measure over multiple years. Differences between individuals in the context of this model of overall growth can then be identified. Children who deviate from the 'typical' growth pattern can be identified and the specific reasons for this deviation explored in more detail.

This thesis was further limited by using statistical methods that focused on single outcome measures rather than looking at multi-variate patterns across measures. Whilst the majority of interactions described in this thesis were simple, there were a number of complex interactions that are yet to be fully understood. Due to previously reported correlations between the CP_C and CP_{NC} measures (Forbes et al., in prep), predictive relationships with these measures were unable to be explored due to the likelihood of multicollinearity. As discussed in Chapter 2, a method that may allow further understanding of these complex interactions involving both measures is cluster analysis (Mooi & Sarstedt, 2011). Conducting such an analysis would confirm suspicions that certain complex interactions were related to each other. For example, I proposed that the finding that girls with a less educated mother who demonstrated a lower TLT at 30 months of age and yet obtained a higher maximum capacity at 42 months of age may be related to a previous interaction between TLT and CP_{NC}. These explanations cannot be confirmed without grouping children based on their behaviour across measures and evaluating performance as a result of that behaviour. For example, a cluster analysis may reveal that the children with a lower TLT and a lower CP_{NC} demonstrated a higher level of CP_C, and this was more important for a higher capacity. This would be a beneficial analysis for future work to consider.

Many of the conclusions in this thesis would be strengthened with the inclusion of other measures, for example measures that examine infant attention or inhibitory control. Examining these measures alongside the current measures presented in this thesis would allow future research to explore the involvement of these other systems in more detail, and this may provide additional insights into the emergence of EF in early development. A key challenge here, however, is that it is not clear we have measures of other candidate components, such as inhibitory control that can be used across multiple ages and are longitudinally predictive. Future work should build on current understanding of inhibitory control to construct these measures. It is important to note that the robustness of the VWM_{PL} task is relatively unique, and there are few tasks which are able to demonstrate longitudinal stability from infancy through childhood. Future work should attempt to replicate these findings, to provide further evidence for these longitudinally stable relationships.

An important question that arose from Chapter 1, and is yet to be answered within this thesis, is whether the candidate components of EF are separable. This question arises from literature taking a strong latent variable approach, which identifies common sources of variance and labels the non-shared variance as specific components based on commonly studied processes. As discussed heavily within this thesis, this approach is problematic. The alternative is to understand how multiple processes co-develop over development and support each other. This was the approach taken in this thesis, aligning with the approach used in recent modelling research (Spencer et al., 2025). This may pave the way for a move from the component style thinking towards one of an integrated systems perspective. Instead of trying to compare tasks using one measure of a latent variable of EF or VWM, this thesis embraced the approach of reviewing multiple measures. Whilst at the surface level, the tasks used appear to be examining different processes, when we look deeper we see multiple related systems are involved in success on both VWM tasks and the MEFS task. For example, representations of colour features are important to be able to detect change on the two VWM tasks. Within the MEFS task, these feature representations are important for successful sorting. Suppressing distraction has also been demonstrated as important for the VWM tasks (Wijeakumar et al. 2019), in a similar manner to the necessary suppression of distraction from the non-probed dimension in the MEFS task. To further understand how these multiple integrated systems can lead to higher-order EF skills, and understand how the same system functions in these two VWM tasks, we should utilise models of integrated neural systems such as the dynamic field model (Spencer et al., 2025). For instance, can we create models with individual differences in parameters to demonstrate high vs low performers within each measure? From these models, would we then be able to predict differences in EF performance in a similar manner to reproduce / simulate the data presented here? Future work should probe how we can use these integrative neural architectures to model these longitudinal data directly.

Given the push for the use of neural models, it is also important to consider the neural systems that underlie changes in VWM and EF. As aforementioned, the data within this thesis was a part of a larger scale longitudinal study that included fNIRS data at each year for the VWM task. Future analyses should bring together these behavioural findings with this fNIRS data to understand more about the changing neural systems underlying VWM over time, and how these changes may relate to the interactions found between VWM and EF in this thesis.

5.3 Real World Implications

This research provides support for the use of preferential looking tasks in examining VWM from early infancy. Given the importance of working memory and EF for children's future academic success, these early measures may allow identification of at-risk children in infancy. More data would be needed to replicate these findings and create categories of performance across measures, but it may be possible in future work to use these measures to identify 'at risk' infants. For instance, it may be possible to use the VWM_{PL} task to generate a 'cognitive growth score' indicating whether the child's performance was on track or of concern in the context of age-specific norms. The VWM_{PL} task would also have to be adapted into a more transportable set up, such as on a laptop, so that the task could be taken to children's centres or general practitioner surgeries for assessment purposes. Doing this may allow the use of this task to identify children with difficulties in engaging VWM from as young as four to six months of age. This would allow intervention measures focussed on supporting VWM to be put in place from early infancy, for example parenting interventions focussed on nurturing parent-child interaction that have been shown to improve child cognition (Landry et al., 2008).

Moreover, the findings of this thesis consistently show that a stronger VWM supports children with a less educated mother. However, the nature of this relationship is not yet fully understood. It is clear that maternal education level influences child executive functioning. Consequently, this thesis provides support for intervention strategies that call for greater access to education for parents, in order to improve child cognition (Waters et al., 2021). One example of a possible intervention strategy is family learning programmes. Many councils in the United Kingdom offer Family Learning, Literacy, and Numeracy (FLN) courses (see Cara & Brooks, 2012). These courses aim to improve literacy and numeracy skills of less educated parents, in hopes of improving their confidence to engage in educational activities with their children. The findings of this thesis offer support for such programmes, in hopes that increasing parental education levels provides a good environment for the development of stronger EF skills.

5.4 Conclusions

To conclude, this thesis provides important insights to the VWM and EF literatures. Firstly, I demonstrated that VWM is longitudinally stable from early infancy through childhood in the first longitudinal study to examine this. These findings emphasise the importance of utilising new measures of visual cognition that account for differing task demands dependent on the child's own looking behaviour. Secondly, this thesis concluded that EF shows stability from 2.5 years of age to 6.5 years of age in, to my knowledge, the first longitudinal study to examine this using a consistent measure over early childhood. Third, evidence from this thesis shows that the development of VWM and EF are related from infancy through early childhood with early measures of VWM predicting later EF outcomes. Consequently, we may be able to identify children who will struggle with EF using performance on the VWM preferential looking task during early infancy. This could help us identify at-risk infants early in development, facilitating the delivery of effective interventions during infancy.

References

Ahmed, S. F., Tang, S., Waters, N. E., & Davis-Kean, P. (2019). Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. *Journal of Educational Psychology*, *111*(3), 446–458. https://doi.org/10.1037/edu0000296

Alloway, T. P., & 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. https://doi.org/10.1016/j.jecp.2009.11.003

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417-423. DOI: 10.1016/S1364-6613(00)01538-2

Baddeley, A. (2020). Working memory. In *Memory* (pp. 71-111). Routledge. ISBN: 9780429449642

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation Vol. 8* (pp. 47–89). New York: Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, **67**(1), 1–48. doi:10.18637/jss.v067.i01.

Bauer, P. J., & Zelazo, P. D. (2014). The National Institutes of Health Toolbox for the assessment of neurological and behavioral function: A tool for developmental science. *Child Development Perspectives*, *8*(3), 119-124. https://doi.org/10.1111/cdep.12080

Bell, M. A. (2012). A psychobiological perspective on working memory performance at 8 months of age. *Child development*, *83*(1), 251-265. https://doi.org/10.1111/j.1467-8624.2011.01684.x

Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. In *Child Development* (Vol. 81, Issue 6, pp. 1641–1660). https://doi.org/10.1111/j.1467-8624.2010.01499

Blakey, E., & Carroll, D. J. (2018). Not all distractions are the same: Investigating why preschoolers make distraction errors when switching. *Child development*, *89*(2), 609-619. https://doi.org/10.1111/cdev.12721

Blankenship, T. L., Slough, M. A., Calkins, S. D., Deater-Deckard, K., Kim-Spoon, J., & Bell, M. A. (2019). Attention and executive functioning in infancy: Links to childhood executive function and reading achievement. *Developmental Science, 22*(6). https://doi.org/10.1111/desc.12824

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot'and 'cool'executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*(3), 337-349. https://doi.org/10.1016/j.ecresq.2009.06.001

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM (2017). "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling." *The R Journal*, **9**(2), 378–400. doi:10.32614/RJ-2017-066.

Brooks, P. J., Hanauer, J. B., Padowska, B., & Rosman, H. (2003). The role of selective attention in preschoolers' rule use in a novel dimensional card sort. *Cognitive Development, 18*(2), 195–215. https://doi.org/10.1016/S0885-2014(03)00020-0

Broomell, A. P., & Bell, M. A. (2017). Inclusion of a mixed condition makes the day/night task more analogous to the adult Stroop. *Developmental neuropsychology*, *42*(3), 241-252. https://doi.org/10.1080/87565641.2017.1309655

Broomell, A. P., & Bell, M. A. (2022). Longitudinal development of executive function from infancy to late childhood. *Cognitive Development*, *63*, 101229. https://doi.org/10.1016/j.cogdev.2022.101229

Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). The differentiation of executive functions in middle and late childhood: A longitudinal latent-variable analysis. *Intelligence, 47,* 34–43. https://doi.org/10.1016/j.intell.2014.08.010

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228. https://doi.org/10.1080/87565640801982312

Buss, A. T., & Kerr-German, A. (2019). Dimensional attention as a mechanism of executive function: Integrating flexibility, selectivity, and stability. *Cognition*, *192*, 104003. https://doi.org/10.1016/j.cognition.2019.06.015

Buss, A. T., & Spencer, J. P. (2014). The emergent executive: A dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, *79*(2), vii. https://doi.org/10.1002/mono.12096

Buss, A. T., Fox, N., Boas, D. A., & Spencer, J. P. (2014). Probing the early development of visual working memory capacity with functional near-infrared spectroscopy. *Neuroimage*, *85*, 314-325. https://doi.org/10.1016/j.neuroimage.2013.05.034

Cara, O & Brooks, G; (2012) Evidence of the Wider Benefits of Family Learning: A Scoping Review. (pp. pp. 1-31). Department for Business, Innovation and Skills: London, UK.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

Carlson, S. M. (2009). Social origins of executive function development. *New directions for child and adolescent development*, *2009*(123), 87-98. https://doi.org/10.1002/cd.237

Carlson, S. (2021). *Minnesota Executive Function Scale: Technical Report.* Excelsior, MN: Reflection Sciences, Inc. https://reflectionsciences.com/wp-content/uploads/Technical-Report-April_20_2021.pdf

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology, 40*(6), 1105–1122. https://doi.org/10.1037/0012-1649.40.6.1105

Carlson, S. M., & Zelazo, P. D. (2014). Minnesota Executive Function Scale: Test manual. St. Paul, MN: Reflection Sciences.

Carroll, D. J., Blakey, E., & FitzGibbon, L. (2016). Cognitive flexibility in young children: Beyond perseveration. *Child Development Perspectives*, *10*(4), 211-215. https://doi.org/10.1111/cdep.12192

Cheie, L., MacLeod, C., Miclea, M., & Visu-Petra, L. (2017). When children forget to remember: Effects of reduced working memory availability on prospective memory performance. *Memory & cognition*, *45*, 651-663. https://doi.org/10.3758/s13421-016-0682-z

Cheng, C., Kaldy, Z., & Blaser, E. (2020). Coding of featural information in visual working memory in 2.5-year-old toddlers. *Cognitive development*, *55*, 100892. https://doi.org/10.1016/j.cogdev.2020.100892

Chevalier, N., Sheffield, T. D., Nelson, J. M., Clark, C. A., Wiebe, S. A., & Espy, K. A. (2012). Underpinnings of the costs of flexibility in preschool children: The roles of inhibition and working memory. *Developmental neuropsychology*, *37*(2), 99-118. https://doi.org/10.1080/87565641.2011.632458

Ciesielski, K. T., Harris, R. J., & Cofer, L. F. (2004). Posterior brain ERP patterns related to the go/no-go task in children. *Psychophysiology*, *41*(6), 882-892. https://doi.org/10.1111/j.1469-8986.2004.00250.x

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS biology*, *10*(3), e1001293. https://doi.org/10.1371/journal.pbio.1001293

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *In Trends in Cognitive Sciences* (Vol. 7, Issue 12, pp. 547–552). Elsevier Ltd. https://doi.org/10.1016/j.tics.2003.10.005

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(1), 87-114. doi: 10.1017/S0140525X01003922

Cuevas, K., Swingler, M. M., Bell, M. A., Marcovitch, S., & Calkins, S. D. (2012). Measures of frontal functioning and the emergence of inhibitory control processes at 10 months of age. *Developmental cognitive neuroscience*, *2*(2), 235-243. https://doi.org/10.1016/j.dcn.2012.01.002

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic bulletin & review*, *3*(4), 422-433. https://doi.org/10.3758/BF03214546

Darki, F., & Klingberg, T. (2015). The role of fronto-parietal and fronto-striatal networks in the development of working memory: a longitudinal study. *Cerebral cortex*, *25*(6), 1587-1595. https://doi.org/10.1093/cercor/bht352

Devine, R. T., Ribner, A., & Hughes, C. (2019). Measuring and Predicting Individual Differences in Executive Functions at 14 Months: A Longitudinal Study. *Child Development, 90*(5), e618–e636. https://doi.org/10.1111/cdev.13217

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching a. *Annals of the New York Academy of Sciences*, *608*(1), 637-676. https://doi.org/10.1111/j.1749-6632.1990.tb48913.x

Diamond A. (2002). Normal development of prefrontal cortex from birth to young adulthood: cognitive functions, anatomy, and biochemistry. Principles of Frontal Lobe Function DT Stuss, RT Knight 466–503 London: Oxford Univ. Press

Diamond, A. (2013). Executive functions. *Annual review of psychology, 64*(1), 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating the dimensions aids the ability to switch. Developmental

Neuropsychology, 28, 689-729. http://dx.doi.org/10.1207/s15326942dn2802_7

Dick, A. S., Overton, W. F., & Kovacs, S. L. (2005). The development of symbolic coordination: Representation of imagined objects, executive function, and theory of mind. *Journal of cognition and development*, *6*(1), 133-161. https://doi.org/10.1207/s15327647jcd0601_8

Doebel, S. (2020). Rethinking executive function and its development. *Perspectives on Psychological Science*, *15*(4), 942-956. https://doi.org/10.1177/174569162090477

Doebel, S., & Zelazo, P. D. (2015). A meta-analysis of the Dimensional Change Card Sort: Implications for developmental theories and the measurement of executive function in children. *Developmental Review*, *38*, 241-268. https://doi.org/10.1016/j.dr.2015.09.001

Duncan, G. J., & Magnuson, K. (2012). Socioeconomic status and cognitive functioning: moving from correlation to causation. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 377-386. https://doi.org/10.1002/wcs.1176

Duncan, R. J., McClelland, M. M., & Acock, A. C. (2017). Relations between executive function, behavioral regulation, and achievement: Moderation by family income. *Journal of Applied Developmental Psychology*, *49*, 21-30. https://doi.org/10.1016/j.appdev.2017.01.004

Fiske, A., & Holmboe, K. (2019). Neural substrates of early executive function development. In *Developmental Review* (Vol. 52, pp. 42–62). Mosby Inc. https://doi.org/10.1016/j.dr.2019.100866

Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. https://www.john-fox.ca/Companion/.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of experimental psychology: General*, *133*(1), 101. https://doi.org/10.1037/0096-3445.133.1.101

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive development*, *10*(4), 483-527. https://doi.org/10.1016/0885-2014(95)90024-1

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Developmental Psychology, 47*(2), 404–416. https://doi.org/10.1037/a0021065

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental psychology*, *50*(6), 1698. https://doi.org/10.1037/a0036633

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychology, 20*(6), 713–736. https://doi.org/10.1080/09297049.2013.857650

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology, 40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 312–7 years old on a stroop-like day-night test. *Cognition*, *53*(2), 129-153. https://doi.org/10.1016/0010-0277(94)90068-X

Goldstein, S., & Naglieri, J. A. (Eds.). (2014). *Handbook of executive functioning.* Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-8106-5

Gooch, D., Thompson, P., Nash, H. M., Snowling, M. J., & Hulme, C. (2016). The development of executive function and language skills in the early school years. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 57*(2), 180–187. https://doi.org/10.1111/jcpp.12458

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental science*, *18*(5), 686-702. https://doi.org/10.1111/desc.12246

Hartig ,F. (2024). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.4.7, https://github.com/florianhartig/dharma.

Hendry, A., Jones, E. J. H., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. In *Developmental Review* (Vol. 42, pp. 1–33). Mosby Inc. https://doi.org/10.1016/j.dr.2016.06.005

Hendry, A., & Scerif, G. (2023). Moulding environmental contexts to optimise neurodiverse executive function performance and development: A goodness-of-fit account. *Infant and Child Development*, *32*(5), e2448. https://doi.org/10.1002/icd.2448

Heyes, S. B., Zokaei, N., & Husain, M. (2016). Longitudinal development of visual working memory precision in childhood and early adolescence. *Cognitive Development*, *39*, 36-44. https://doi.org/10.1016/j.cogdev.2016.03.004

Holmboe, K., Bonneville-Roussy, A., Csibra, G., & Johnson, M. H. (2018). Longitudinal development of attention and inhibitory control during the first year of life. *Developmental Science, 21*(6). https://doi.org/10.1111/desc.12690

Howard, S. J., Okely, A. D., & Ellis, Y. G. (2015). Evaluation of a differentiation model of preschoolers' executive functions. *Frontiers in Psychology, 6*(MAR). https://doi.org/10.3389/fpsyg.2015.00285

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions?. *Child and adolescent mental health*, *7*(3), 131-142. https://doi.org/10.1111/1475-3588.00024

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2009). Tracking executive function across the transition to school: A latent variable approach. *Developmental neuropsychology*, *35*(1), 20-36. https://doi.org/10.1080/87565640903325691

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental neuropsychology*, *20*(3), 573-591. https://doi.org/10.1207/S15326942DN2003_2

Jaroslawska, A. J., Gathercole, S. E., Logie, M. R., & Holmes, J. (2016). Following instructions in a virtual school: Does working memory play a role?. *Memory & Cognition*, *44*, 580-589. https://doi.org/10.3758/s13421-015-0579-2

Johansson, M., Marciszko, C., Brocki, K., & Bohlin, G. (2016). Individual differences in early executive functions: A longitudinal study from 12 to 36 months. *Infant and Child Development*, *25*(6), 533-549. https://doi.org/10.1002/icd.1952

John, A. M. S., Kibbe, M., & Tarullo, A. R. (2019). A systematic assessment of socioeconomic status and executive functioning in early childhood. *Journal of experimental child psychology*, *178*, 352-368. https://doi.org/10.1016/j.jecp.2018.09.003

Johnson, J. S., Simmering, V. R., & Buss, A. T. (2014). Beyond slots and resources: Grounding cognitive concepts in neural dynamics. *Attention, Perception, & Psychophysics*, *76*, 1630-1654. https://doi.org/10.3758/s13414-013-0596-9

Kao, K., Nayak, S., Doan, S. N., & Tarullo, A. R. (2018). Relations between parent EF and child EF: The role of socioeconomic status and parenting on executive functioning in early childhood. *Translational Issues in Psychological Science*, *4*(2), 122. http://dx.doi.org/10.1037/tps0000154

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin, 144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental science*, *6*(5), 449-467.

Kochanska, G., Coy, K. C., & Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child development*, *72*(4), 1091-1111. https://doi.org/10.1111/1467-8624.00336

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child development*, *67*(2), 490-507. https://doi.org/10.1111/j.1467-8624.1996.tb01747.x

Kuznetsova ,A., Brockhoff, P.B., Christensen, R.H.B., (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, **82**(13), 1–26. doi:10.18637/jss.v082.i13.

Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: Establishing early foundations for social, communication, and independent problem-solving skills. Developmental Psychology, 42(4), 627–642. https://doi.org/10.1037/0012-1649.42.4.627

Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental science*, *21*(2), e12529. https://doi.org/10.1111/desc.12529

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*(1), 59–80. https://doi.org/10.1348/026151003321164627

Lowery, K., Nikam, B., & Buss, A. T. (2022). Dimensional label learning contributes to the development of executive functions. *Scientific Reports*, *12*(1), 11008. https://doi.org/10.1038/s41598-022-14761-2

Lunt, L., Bramham, J., Morris, R. G., Bullock, P. R., Selway, R. P., Xenitidis, K., & David, A. S. (2012). Prefrontal cortex dysfunction and 'Jumping to Conclusions': Bias or deficit?. *Journal of neuropsychology*, *6*(1), 65-78. https://doi.org/10.1111/j.1748-6653.2011.02005.x

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in psychology*, *5*, 599. https://doi.org/10.3389/fpsyg.2014.00599

McKay, C. A., Shing, Y. L., Rafetseder, E., & Wijeakumar, S. (2021). Home assessment of visual working memory in pre-schoolers reveals associations between behaviour, brain activation and parent reports of life stress. *Developmental science*, *24*(4), e13094. https://doi.org/10.1111/desc.13094

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A Latent Variable Approach to Determining the Structure of Executive Function in Preschool Children. *Journal of Cognition and Development, 13*(3), 395–423. https://doi.org/10.1080/15248372.2011.585478

Miller, S. E., & Marcovitch, S. (2015). Examining executive function in the second year of life: Coherence, stability, and relations to joint attention and language. *Developmental Psychology, 51*(1), 101–114. https://doi.org/10.1037/a0038359

Mirman, D. (2014). Growth Curve Analysis and Visualization Using R (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315373218

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current directions in psychological science*, *21*(1), 8-14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & 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. https://doi.org/10.1006/cogp.1999.0734

Mooi, E., & Sarstedt, M. (2011). A concise guide to market research: The process, data, and methods using IBM SPSS statistics. Springer. DOI: 10.1007/978-3-642-12541-6

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British journal of psychology*, *81*(2), 111-121. https://doi.org/10.1111/j.2044-8295.1990.tb02349.x

Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *40*(3), 255-265. https://doi.org/10.1002/dev.10033

Müller, U., Steven Dick, A., Gela, K., Overton, W. F., & Zelazo, P. D. (2006). The role of negative priming in preschoolers' flexible rule use on the dimensional change card sort task. *Child development*, *77*(2), 395-412. https://doi.org/10.1111/j.1467-8624.2006.00878.x

Munakata, Y., Morton, J. B., & Yerys, B. E. (2003). Children's perseveration: attentional inertia and alternative accounts. *Developmental Science*, *6*(5).

Nguyen, T., & Duncan, G. J. (2019). Kindergarten components of executive function and third grade achievement: A national study. *Early Childhood Research Quarterly*, *46*, 49-61. https://doi.org/10.1016/j.ecresq.2018.05.006

Nigg, J. T. (2000). On Inhibition/Disinhibition in Developmental Psychopathology: Views from Cognitive and Personality Psychology and a Working Inhibition Taxonomy. *Psychological Bulletin, 126*(2), 220–246. https://doi.org/10.1037/0033-2909.126.2.220

Palomino, C. I., & Brudvig, A. (2022). Examining the role of demographic characteristics, attachment, and language in preschool children's executive function skills. *Early Child Development and Care*, *192*(12), 1967-1981. https://doi.org/10.1080/03004430.2021.1958803

Papageorgiou, K. A., Smith, T. J., Wu, R., Johnson, M. H., Kirkham, N. Z., & Ronald, A. (2014). Individual differences in infant fixation duration relate to attention and behavioral control in childhood. *Psychological science*, *25*(7), 1371-1379. https://doi.org/10.1177/0956797614531295

Perner, J., & Lang, B. (2002). What Causes 3-year-olds' Difficulty on the Dimensional Change Card Sorting Task? *Infant and Child Development, 11*(2), 93–105. https://doi.org/10.1002/icd.299

Perone, S., Molitor, S. J., Buss, A. T., Spencer, J. P., & Samuelson, L. K. (2015). Enhancing the executive functions of 3-year-olds in the dimensional change card sort task. *Child Development*, *86*(3), 812-827. https://doi.org/10.1111/cdev.12330

Perone, S., Plebanek, D. J., Lorenz, M. G., Spencer, J. P., & Samuelson, L. K. (2019). Empirical tests of a brain-based model of executive function development. *Child development*, *90*(1), 210-226.https://doi.org/10.1111/cdev.12885

Perone, S., Simmering, V. R., & Buss, A. T. (2021). A Dynamical Reconceptualization of Executive-Function Development. *Perspectives on Psychological Science, 16*(6), 1198–1208. https://doi.org/10.1177/1745691620966792

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Reflection Sciences. (n.d.). *Deliver MEFS remotely.* Reflection Sciences. https://reflectionsciences.com/remote-administration-of-the-mefs/

Reyes, L. D., Wijeakumar, S., Magnotta, V. A., Forbes, S. H., & Spencer, J. P. (2020). The functional brain networks that underlie visual working memory in the first two years of life. *NeuroImage*, *219*, 116971. https://doi.org/10.1016/j.neuroimage.2020.116971

Roberts Jr, R. J., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental neuropsychology*, *12*(1), 105-126. https://doi.org/10.1080/87565649609540642

Roderer, T., Krebs, S., Schmid, C., & Roebers, C. M. (2012). The Role of Executive Control of Attention and Selective Encoding for Preschoolers' Learning. *Infant and Child Development, 21*(2), 146–159. https://doi.org/10.1002/icd.738

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of Infant Cognition for Executive Functions at Age 11. *Psychological Science, 23*(11), 1345–1355. https://doi.org/10.1177/0956797612444902

Ross-sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child development*, *74*(6), 1807-1822. https://doi.org/10.1046/j.1467-8624.2003.00639.x

Schirmbeck, K., Rao, N., & Maehler, C. (2020). Similarities and differences across countries in the development of executive functions in children: A systematic review. *Infant and Child Development*, *29*(1), e2164. https://doi.org/10.1002/icd.2164

Schneegans, S., Spencer, J.P., Schöner, G., Hwang, S. & Hollingworth, A. (2014). Dynamic interactions between visual working memory and saccade target selection. Journal of Vision, 14, 1-23. https://doi.org/10.1167/14.11.9

Schutte, A.R. & Spencer, J.P. (2009). Tests of the dynamic field theory and the spatial precision hypothesis: Capturing a qualitative developmental transition in spatial working memory. Journal of Experimental Psychology: Human Perception and Performance, 35, 1698-1725. https://doi.org/10.1037/a0015794

Schutte, A.R., Spencer, J.P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development. Child Development, 74, 1393-1417. https://doi.org/10.1111/1467-8624.00614

Shimi, A., & Scerif, G. (2022). The influence of attentional biases on multiple working memory precision parameters for children and adults. *Developmental Science*, *25*(3), e13213. https://doi.org/10.1111/desc.13213

Silva, C., Sousa-Gomes, V., Fávero, M., Oliveira-Lopes, S., Merendeiro, C. S., Oliveira, J., & Moreira, D. (2022). Assessment of preschool-age executive functions: A systematic review. *Clinical Psychology & Psychotherapy*, *29*(4), 1374-1391. https://doi.org/10.1002/cpp.2718

Simmering, V. R. (2016). I. Working memory capacity in context: Modeling dynamic processes of behavior, memory, and development. *Monographs of the Society for Research in Child Development*, *81*(3), 7-24. https://doi.org/10.1111/mono.12249

Simmering, V. R., & Perone, S. (2013). Working memory capacity as a dynamic process. *Frontiers in psychology*, *3*, 567. https://doi.org/10.3389/fpsyg.2012.00567

Simpson, A., & Riggs, K. J. (2006). Conditions under which children experience inhibitory difficulty with a “button-press” go/no-go task. *Journal of Experimental Child Psychology*, *94*(1), 18-26. https://doi.org/10.1016/j.jecp.2005.10.003

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, *6*(4), 392-412. https://doi.org/10.1111/1467-7687.00295

Spencer, J. P., Forbes, S. H., Naylor, S., Singh, V. P., Jackson, K., Deoni, S., ... & Kumar, A. (2023). Poor air quality is associated with impaired visual cognition in the first two years of life: a longitudinal investigation. *ELife*, *12*, e83876. https://doi.org/10.7554/eLife.83876

Spencer, J. P., Buss, A. T., McCraw, A. R., Johns, E., & Samuelson, L. K. (2025). Integrating attention, working memory, and word learning in a dynamic field theory of executive function development: Moving beyond the ‘component’ view of executive function. *Developmental Review*, *75*, 101182. https://doi.org/10.1016/j.dr.2024.101182

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly journal of experimental psychology*, *59*(4), 745-759. https://doi.org/10.1080/17470210500162854

Steelandt, S., Thierry, B., Broihanne, M. H., & Dufour, V. (2012). The ability of children to delay gratification in an exchange task. *Cognition, 122*(3), 416–425. https://doi.org/10.1016/j.cognition.2011.11.009

Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., ... & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, *38*(4), 388-402. https://doi.org/10.1016/S0028-3932(99)00093-7

Sutherland, C., Hare, D., Johnson, P. J., Linden, D. W., Montgomery, R. A., & Droge, E. (2023). Practical advice on variable selection and reporting using Akaike information criterion. *Proceedings of the Royal Society B*, *290*(2007), 20231261. https://doi.org/10.1098/rspb.2023.1261

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *cortex*, *49*(10), 2674-2688. https://doi.org/10.1016/j.cortex.2013.06.007

Thelen, E. (2001). Dynamic mechanisms of change in early perceptual-motor development. In J. L. McClelland & R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives* (pp. 161–184). Lawrence Erlbaum Associates Publishers.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104

Vandervert, L. R., Schimpf, P. H., & Liu, H. (2007). How working memory and the cerebellum collaborate to produce creativity and innovation. *Creativity Research Journal*, *19*(1), 1-18. https://doi.org/10.1080/10400410709336873

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and aging*, *25*(2), 343. https://doi.org/10.1037/a0017729

Wagenmakers, E.J., Farrell, S. (2004). AIC model selection using Akaike weights. Psychonomic Bulletin & Review 11, 192–196. https://doi.org/10.3758/BF03206482

Waters, N. E., Ahmed, S. F., Tang, S., Morrison, F. J., & Davis-Kean, P. E. (2021). Pathways from socioeconomic status to early academic achievement: The role of specific executive functions. *Early Childhood Research Quarterly*, *54*, 321-331. https://doi.org/10.1016/j.ecresq.2020.09.008

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, *43*(7), 352-360. https://doi.org/10.3102/0013189X14553660

Wei, Z., Wang, X. J., & Wang, D. H. (2012). From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. *Journal of Neuroscience*, *32*(33), 11228-11240. https://doi.org/10.1523/JNEUROSCI.0735-12.2012

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental neuropsychology*, *4*(3), 199-230. https://doi.org/10.1080/87565648809540405

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental psychology*, *44*(2), 575. https://doi.org/10.1037/0012-1649.44.2.575

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of*

Experimental Child Psychology, 108(3), 436–452. https://doi.org/10.1016/j.jecp.2010.08.008

Wijeakumar, S., Kumar, A., Delgado Reyes, L. M., Tiwari, M., & Spencer, J. P. (2019). Early adversity in rural India impacts the brain networks underlying visual working memory. *Developmental Science*, *22*(5), e12822. https://doi.org/10.1111/desc.12822

Wijeakumar, S., Forbes, S. H., Magnotta, V. A., Deoni, S., Jackson, K., Singh, V. P., ... & Spencer, J. P. (2023). Stunting in infancy is associated with atypical activation of working memory and attention networks. *Nature human behaviour*, *7*(12), 2199-2211. https://doi.org/10.1038/s41562-023-01725-3

Wijeakumar, S., & Spencer, J. (2020). A dynamic field theory of visual working memory. In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working memory: State of the science* (pp. 358–388). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0013

Willoughby, M. T., Wirth, R. J., Blair, C. B., & Family Life Project Investigators. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment, 24*(2), 418–431. https://doi.org/10.1037/a0025779

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in psychology*, *5*, 213. https://doi.org/10.3389/fpsyg.2014.00213

Yamamoto, N., & Imai-Matsumura, K. (2019). Gender differences in executive function and behavioural self-regulation in 5 years old kindergarteners from East Japan. *Early Child Development and Care*, *189*(1), 56-67. https://doi.org/10.1080/03004430.2017.1299148

Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends in cognitive sciences*, *8*(1), 12-17. DOI: 10.1016/j.tics.2003.11.001

Zelazo, P. D. (2015). Executive function: Reflection, iterative reprocessing, complexity, and the developing brain. *Developmental Review*, *38*, 55-68. https://doi.org/10.1016/j.dr.2015.07.001

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). NIH toolbox cognition battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78*(4), 16–33. https://doi.org/10.1111/mono.12032

Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). Executive Function: Implications for Education. NCER 2017-2000. *National Center for Education Research*. IES: http://ies.ed.gov/ncer/pubs/20172000/

Zelazo, P. D., & Carlson, S. M. (2023). Reconciling the context-dependency and domain-generality of executive function skills from a developmental systems perspective. *Journal of Cognition and Development*, *24*(2), 205-222. https://doi.org/10.1080/15248372.2022.2156515

Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood: III. Study 2: Rule complexity and stimulus characteristics in executive function. *Monographs of the Society for Research in Child Development, 68*(3), 48–64. https://doi.org/10.1111/j.0037-976X.2003.00263.x

.1 Appendices

.1.1 Appendix A

Table 1

Regression results for predicting CP_{10} from year 1 to year 2.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	0.52***	0.00	115.65	[0.52, 0.53]		<.001
$CP_{10}1$	- 0.21*	0.09	- 2.28	[- 0.39, - 0.03]	1.004	.0246
Age	- 0.00	0.01	- 0.46	[- 0.02, 0.01]	1.031	.64
Gender	0.00	0.01	0.20	[- 0.02, 0.02]	1.004	.84
Maternal Ed	0.01	0.00	1.51	[- 0.00, 0.02]	1.004	.13
$CP_{10}1$:Gender	0.18	0.18	0.97	[- 0.19, 0.54]		.33
$CP_{10}1$:Maternal Ed	0.08	0.13	0.60	[- 0.18, 0.34]		.55
Gender:Maternal Ed	0.01	0.01	0.95	[- 0.01, 0.03]		.35
$CP_{10}1$:Gender:Maternal Ed	0.08	0.26	0.31	[- 0.43, 0.60]		.75
N	138					
R^2	0.05					
Adj R^2	- 0.01					

Note: CP_C1 refers to mean CP_C in year 1. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.2 Appendix B

Table 2

Regression results for predicting CP_{10} from year 2 to year 3.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	0.52***	0.00	109.10	[0.51, 0.53]		<.001
$CP_{10}2$	0.01	0.10	0.13	[- 0.18, 0.20]	1.003	.90
Age	- 0.01	0.01	- 1.30	[- 0.03, 0.01]	1.022	.20
Gender	- 0.00	0.01	0.16	[- 0.02, 0.02]	1.003	.88
Maternal Ed	- 0.00	0.01	- 0.59	[- 0.01, 0.01]	1.003	.56
$CP_{10}2$:Gender	0.17	0.19	0.88	[- 0.21, 0.55]		.38
$CP_{10}2$:Maternal Ed	- 0.03	0.10	- 0.24	[- 0.23, 0.18]		.81
Gender:Maternal Ed	- 0.01	0.01	- 0.61	[- 0.03, 0.01]		.54
$CP_{10}2$:Gender:Maternal Ed	- 0.16	0.21	- 0.75	[- 0.57, 0.26]		.45
N	124					
R^2	0.02					
Adj R^2	- 0.04					

Note: $CP_{10}2$ refers to mean CP_{10} in year 2. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.3 Appendix C

Table 3

Regression results for predicting CP_{NC} from year 1 to year 2.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	0.42***	0.01	39.47	[0.40, 0.44]		<.001
$CP_{NC}1$	0.24**	0.09	2.76	[0.07, 0.42]	1.003	.007
Age Cohort	- 0.00	0.02	- 0.12	[- 0.05, 0.04]	1.023	.91
Gender	0.01	0.02	0.43	[- 0.03, 0.05]	1.003	.67
Maternal Ed	0.00	0.01	0.07	[- 0.02, 0.02]	1.003	.94
$CP_{NC}1$:Gender	- 0.05	0.18	- 0.28	[- 0.40, 0.30]		.78
$CP_{NC}1$:Maternal Ed	0.09	0.10	0.97	[- 0.09, 0.27]		.34
Gender:Maternal Ed	- 0.01	0.02	- 0.24	[- 0.05, 0.04]		.81
$CP_{NC}1$:Gender:Maternal Ed	- 0.13	0.19	- 0.71	[- 0.50, 0.23]		.48
N	138					
R^2	0.07					
Adj R^2	0.02					

Note: $CP_{NC}1$ refers to mean CP_{NC} in year 1. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.4 Appendix D

Table 4

Regression results for predicting CP_{NC} from year 2 to year 3.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	0.42***	0.01	36.91	[0.40, 0.44]		<.001
$CP_{NC}2$	0.23*	0.09	2.54	[0.05, 0.40]	1.004	.0123
Age	- 0.02	0.02	- 0.72	[- 0.06, 0.03]	1.030	.47
Gender	- 0.00	0.02	- 0.18	[- 0.05, 0.04]	1.004	.86
Maternal Ed	- 0.01	0.01	- 1.09	[- 0.04, 0.01]	1.004	.28
$CP_{NC}2$:Gender	0.16	0.18	0.90	[- 0.19, 0.51]		.37
$CP_{NC}2$:Maternal Ed	0.09	0.09	0.99	[- 0.09, 0.26]		.32
Gender:Maternal Ed	- 0.01	0.02	- 0.28	[- 0.06, 0.04]		.78
$CP_{NC}2$:Gender:Maternal Ed	- 0.14	0.17	- 0.81	[- 0.49, 0.20]		.42
N	124					
R^2	0.08					
Adj R^2	0.02					

Note: $CP_{NC}2$ refers to mean CP_{NC} in year 2. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.5 Appendix E

Table 5

Regression results for predicting CP_C from year 1 to year 2.

Predictor	*B*	SE	*t*	95% CI [LL, UL]	VIF	*p*
(Intercept)	0.65***	0.01	59.40	[0.62, 0.67]		<.001
CP_C1	0.22*	0.10	2.21	[0.02, 0.42]	1.006	.0292
Age	- 0.02	0.02	- 0.88	[- 0.06, 0.03]	1.041	.38
Gender	- 0.00	0.02	- 0.04	[- 0.04, 0.05]	1.006	.97
Maternal Ed	0.01	0.01	0.81	[- 0.01, 0.03]	1.006	.42
CP_C1:Gender	0.16	0.20	0.80	[- 0.24, 0.56]		.43
CP_C1:Maternal Ed	0.02	0.11	0.18	[- 0.19, 0.23]		.85
Gender:Maternal Ed	0.02	0.02	1.05	[- 0.03, 0.07]		.30
CP_C1:Gender:Maternal Ed	0.04	0.22	0.17	[- 0.39, 0.47]		.86
N	138					
R^2	0.05					
Adj R^2	- 0.01					

Note: CP_C1 refers to mean CP_C in year 1. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.6 Appendix F

Table 6

Regression results for predicting CP_C from year 2 to year 3.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	0.63***	0.01	70.49	[0.61, 0.65]		<.001
CP_C2	0.08	0.07	1.07	[- 0.06, 0.21]	1.003	.29
Age	- 0.01	0.02	- 0.60	[- 0.05, 0.02]	1.024	.55
Gender	- 0.02	0.02	- 1.18	[- 0.06, 0.01]	1.003	.24
Maternal Ed	- 0.01	0.01	- 0.81	[- 0.03, 0.01]	1.003	.42
CP_C2:Gender	0.06	0.14	0.43	[- 0.22, 0.34]		.67
CP_C2:Maternal Ed	0.12	0.08	1.58	[- 0.03, 0.28]		.12
Gender:Maternal Ed	0.03	0.02	1.71	[- 0.01, 0.07]		.09
CP_C2:Gender:Maternal Ed	0.29	0.16	1.83	[-0.59,		.07
N	124					
R^2	0.10					
Adj R^2	0.04					

Note: CP_C2 refers to mean CP_C in year 2. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.7 Appendix G

Table 7

Regression results for predicting TLT from year 1 to year 2.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	5.40***	0.12	46.55	[5.18, 5.63]		<.001
TLT 1	0.25**	0.08	3.20	[0.10, 0.41]	1.012	.00173
Age	- 0.17	0.24	- 0.71	[- 0.65, 0.31]	1.084	.47
Gender	0.07	0.23	0.29	[- 0.39, 0.53]	1.012	.77
Maternal Ed	0.13	0.13	1.07	[- 0.12, 0.39]	1.012	.29
TLT 1:Gender	0.17	0.15	1.07	[- 0.14, 0.47]		.29
TLT 1:Maternal Ed	- 0.01	0.09	- 0.10	[- 0.19, 0.17]		.92
Gender:Maternal Ed	- 0.25	0.25	- 1.03	[- 0.74, 0.23]		.31
TLT 1:Gender:Maternal Ed	- 0.06	0.18	- 0.33	[- 0.42, 0.30]		.74
N	139					
R^2	0.10					
Adj R^2	0.05					

Note: TLT 1 refers to mean TLT in year 1. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.8 Appendix H

Table 8

Regression results for predicting TLT from year 2 to year 3.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	5.6***	0.11	51.06	[5.38, 5.82]		<.001
TLT 2	0.28**	0.08	3.38	[0.12, 0.45]	1.003	.00101
Age	0.41	0.22	1.90	[- 0.02, 0.85]	1.023	.06
Gender	- 0.14	0.22	- 0.62	[- 0.57, 0.30]	1.003	.54
Maternal Ed	0.12	0.12	1.01	[- 0.12, 0.36]	1.003	.32
TLT 2:Gender	0.01	0.17	0.07	[- 0.32, 0.34]		.95
TLT 2:Maternal Ed	0.08	0.09	0.88	[- 0.10, 0.27]		.38
Gender:Maternal Ed	- 0.35	0.24	- 1.49	[- 0.82, 0.12]		.14
TLT 2:Gender:Maternal Ed	0.09	0.19	0.48	[- 0.28, 0.47]		.63
N	124					
R^2	0.17					
Adj R^2	0.12					

Note: TLT 2 refers to mean TLT in year 2. LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.9 Appendix I

Table 9

Regression results for predicting A' prime by year and set size, using A' prime in set size 3 and year 3 as the criterion.

Predictor	***B***	**SE**	***t***	**df**	***p***
(Intercept)	0.56***	0.03	20.70	318.01	<.001
Year	0.12***	0.01	11.50	286.28	<.001
Set size 1	0.17***	0.03	5.03	254.61	<.001
Set size 2	- 0.05	0.03			.19
Year:Set size 1	- 0.04	0.01	- 2.95	253.55	<.001
Year:Set size 2	0.02	0.01	1.52	251.93	.13
N	67				
Observations	327				
Pseudo R^2	0.59				
Pseudo R^2 (Fixed effects)	0.40				
AIC	- 543.83				
BIC	- 513.51				

Note: * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.10 Appendix J

Table 10

Regression results for predicting A' prime by year, set size, maternal education level, and gender.

Predictor	*B*	SE	*t*	df	*p*
(Intercept)	0.85***	0.01	99.07	54.85	<.001
Year	0.12**	0.01	11.52	282.66	<.001
Set Size	- 0.08***	0.01	- 12.96	246.56	<.001
Gender	- 0.01	0.02	- 0.48	54.85	.63
Maternal Ed	0.02**	0.01	2.77	56.38	.00769
Year:Set Size	0.03*	0.01	- 1.06	246.61	.01003
Year:Gender	- 0.01	0.02	- 0.31	282.66	.74
Year:Maternal Ed	- 0.01	0.01	- 1.06	285.29	.29
Set Size:Gender	- 0.03*	0.01	- 2.29	246.56	.02305
Set Size:Maternal Ed	0.01*	0.01	2.149	247.42	.03258
Gender:Maternal Ed	0.01	0.02	0.65	56.38	.52
Year:Set Size:Gender	- 0.00	0.02	- 0.04	246.61	.97
Year:Set Size:Maternal Ed	- 0.01	0.01	- 1.04	247.44	.30
Year:Gender:Maternal Ed	0.01	0.02	0.39	285.29	.70
Set Size:Gender:Maternal Ed	0.01	0.01	0.45	247.42	.65
Year:Set Size:Gender:Maternal Ed	0.00	0.03	0.16	247.44	.88
N	67				
Observations	327				
R^2 (Total)	0.60				
Pseudo R^2 (Fixed effects)	0.43				
AIC	- 473.74				
BIC	- 405.52				

Note: * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.11 Appendix K

Table 11
Regression results for predicting K_{MAX} by year, maternal education level, and gender, using K_{MAX} in year 3 as the criterion.

Predictor	***B***	**SE**	***t***	**df**	***p***
(Intercept)	1.67***	0.07	23.68	54.67	<.001
Year 2	- 0.41***	0.05	- 7.64	48.83	<.001
Maternal Ed	0.26***	0.07	3.58	52.95	<.001
Gender	- 0.10	0.14	- 0.74	54.67	.46
Year 2:Maternal Ed	- 0.00	0.05	- 0.09	46.88	.93
Year 2:Gender	0.12	0.11	1.13	48.83	.26
Maternal Ed:Gender	- 0.02	0.14	- 0.14	52.95	.89
Year 2:Maternal Ed:Gender	0.10	0.11	0.95	46.88	.35
N	67				
Obs	115				
R^2(Total)	0.56				
R^2(Fixed effects)	0.34				
AIC	263.21				
BIC	290.66				

Note: * indicates p < .05, ** indicates p < .01, *** indicates p < .001.

.1.12 Appendix L

Table 12
Hierarchical regression results for predicting K_{MAX}.

Predictor	***B***	**SE**	***t***	**df**	***p***	R^2 (Total, Fixed)
Model 1						0.56, 0.24
(Intercept)	-0.31	0.27	-1.15	62.84	.25	
Year	0.79***	0.10	7.62	53.76	p<.001	
Model 2						0.60, 0.28
(Intercept)	0.79	0.71	1.12	89.71	.27	
Year	0.78***	0.10	7.68	51.04	p<.001	
CP_{10} SS6	- 1.95*	0.76	- 2.57	95.55	.0116	
CP_{10} SS4	0.29	0.78	0.38	78.07	.71	
CP_{10} SS2	-0.41	0.80	-0.50	96.91	.61	
Model 3						0.61, 0.31
(Intercept)	0.39	0.76	0.51	88.43	.61	
Year	0.79***	0.11	7.39	59.55	p<.001	
CP_{10} SS6	- 2.15**	0.76	- 2.85	95.51	.0054	
CP_{10} SS4	0.54	0.82	0.67	84.32	.51	
CP_{10} SS2	- 0.14	0.81	- 0.17	96.58	.87	
Switch rate SS6	- 0.03	0.43	- 0.07	80.44	.94	
Switch rate SS4	- 0.52	0.41	-1.26	106.92	.21	
Switch rate SS2	0.76*	0.34	2.24	101.28	.0274	
	Random Effects	*SD*				
Model 1	Participant Code (Intercept)	0.46				
	Residual	0.53				
Model 2	Participant Code (Intercept)	0.46				
	Residual	0.52				
Model 3	Participant Code (Intercept)	0.45				
	Residual	0.51				
N	67					
Obs	115					

Note: * indicates p < .05, ** indicates p< .01, *** indicates p <.001.

.1.13 Appendix M

Table 13

Regression results for Model 4, predicting K_{MAX} from year, CP_{10}, and TLT.

Predictor	***B***	**SE**	***t***	**df**	***p***
(Intercept)	0.40	0.72	0.55	94.09	0.58
Year	0.73***	0.10	7.01	52.92	<.001
CP_{10} SS6	- 2.07**	0.75	- 2.77	97.43	0.007
CP_{10} SS4	0.04	0.78	0.06	81.21	0.96
CP_{10} SS2	- 0.40	0.80	- 0.49	96.83	0.62
TLT SS6	0.03	0.07	0.48	100.81	0.63
TLT SS4	0.14	0.07	1.96	101.45	0.05
TLT SS2	- 0.05	0.07	- 0.72	92.90	0.47
Random Effects	*SD*				
Participant code (Intercept)	0.41				
Residual	0.53				
N	67				
Obs	115				
R^2 (Total)	0.58				
R^2 (Fixed effects)	0.33				

Note: * indicates p < .05, ** indicates p < .01, *** indicates p < .001.

.1.14 Appendix N

Table 14

Regression results for Model 5, predicting K_{MAX} from year, CP_{NC}, and TLT.

Predictor	***B***	**SE**	***t***	**df**	***p***
(Intercept)	-1.36	0.44	-3.06	95.06	0.003
Year	0.82***	0.11	7.33	53.66	<.001
CP_{NC} SS6	-0.00	0.29	-0.01	84.11	0.99
CP_{NC} SS4	-0.32	0.34	-0.95	99.78	0.35
CP_{NC} SS2	0.91*	0.39	2.35	90.81	0.02
TLT SS6	0.03	0.07	0.37	103.27	0.71
TLT SS4	0.16*	0.08	2.18	102.55	0.03
TLT SS2	-0.07	0.07	-0.93	97.70	0.35
Random Effects	*SD*				
NIHCode (Intercept)	0.36				
Residual	0.56				
N	65				
Obs	113				
R^2 (Total)	0.53				
R^2 (Fixed effects)	0.34				

Note: * indicates p < .05, ** indicates p < .01, *** indicates p < .001.

.1.15 Appendix O

Table 15

Regression results for predicting K_{MAX} at 42 months from 30-month mean CP_C, 30-month mean TLT, maternal education level, and gender.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	1.22***	0.11	11.01	[0.99, 1.44]		<.001
Maternal Ed	0.28**	0.11	2.61	[0.06, 0.50]	1.34	.01237
Gender	0.09	0.22	0.39	[- 0.36, 0.53]	1.38	.70
CP_C	0.82	1.29	0.63	[- 1.78, 3.42]	1.85	.53
TLT	- 0.02	0.11	- 0.21	[- 0.25, 0.20]	2.70	.84
Maternal Ed:Gender	0.23	0.22	1.05	[- 0.21, 0.66]		.30
Maternal Ed:CP_C	0.01	1.23	0.00	[- 2.48, 2.50]		1.00
Gender:CP_C	- 3.31	2.58	- 1.29	[- 8.51, 1.89]		.21
Maternal Ed:TLT	0.08	0.13	0.58	[- 0.19, 0.34]		.57
Gender:TLT	- 0.26	0.22	- 1.18	[- 0.71, 0.19]		.24
CP_C:TLT	0.48	1.11	0.43	[- 1.76, 2.73]		.67
Maternal Ed:Gender:CP_C	- 1.87	2.47	- 0.76	[- 6.85, 3.11]		.45
Maternal Ed:Gender:TLT	0.25	0.26	0.95	[- 0.28, 0.78]		.35
Maternal Ed:CP_C:TLT	0.14	1.48	0.10	[- 2.84, 3.13]		.92
Gender:CP_C:TLT	1.50	2.22	0.67	[- 2.99, 5.99]		.50
Maternal Ed:Gender:CP_C:TLT	1.41	2.96	0.48	[- 4.56, 7.38]		.64
N	58					
R^2	0.29					
Adj R^2	0.04					

Note: LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p < .01, *** indicates p < .001.

.1.16 Appendix P

Table 16

Regression results for predicting K_{MAX} at 42-months from 30-month mean CP_{NC}, 30-month mean TLT, maternal education level, and gender.

Predictor	B	SE	t	95% CI [LL, UL]	VIF	p
(Intercept)	1.19***	0.10	11.73	[0.98, 1.39]		<.001
Maternal Ed	0.25*	0.11	2.25	[0.03, 0.48]	1.83	.0299
Gender	- 0.02	0.20	- 0.11	[- 0.43, 0.39]	1.46	.91
CP_{NC}	- 0.69	0.85	- 0.81	[- 2.40, 1.02]	1.44	.42
TLT	0.03	0.08	0.42	[- 0.13, 0.20]	1.87	.67
Maternal Ed:Gender	0.24	0.22	1.07	[- 0.21, 0.69]		.29
Maternal Ed:CP_{NC}	0.34	0.83	0.42	[- 1.33, 2.01]		.68
Gender:CP_{NC}	3.54*	1.69	2.09	[0.12, 6.96]		.0426
Maternal Ed:TLT	0.13	0.11	1.22	[- 0.09, 0.35]		.23
Gender:TLT	- 0.24	0.16	- 1.47	[- 0.57, 0.09]		.15
CP_{NC}:TLT	1.55*	0.73	2.11	[0.07, 3.03]		.0409
Maternal Ed:Gender:CP_{NC}	0.50	1.65	0.30	[-2.84, 3.83]		.77
Maternal Ed:Gender:TLT	0.28	0.21	1.33	[- 0.15, 0.72]		.19
Maternal Ed:CP_C:TLT	0.68	0.98	0.69	[- 1.30, 2.66]		.49
Gender:CP_{NC}:TLT	1.24	1.47	0.84	[- 1.73, 4.20]		.40
Maternal Ed:Gender:CP_{NC}:TLT	- 1.18	1.96	- 0.60	[- 5.14, 2.78]		.55
N	58					
R^2	0.44					
Adj R^2	0.24					

Note: LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p < .01, *** indicates p < .001.

.1.17 Appendix Q

Table 17

Means and Standard Deviations of K_{MAX} Scores at 42-months by gender and CP_{NC} category at 30-months.

Gender	CP_{NC} **Cat**	**Mean**	**SD**
B	Higher NC	0.96	0.70
B	Lower NC	1.47	0.61
G	Higher NC	1.39	0.82
G	Lower NC	1.23	0.70

Note. B = Boys, G = Girls. SD = Standard Deviation. CP_{NC} category was calculated using a median split.

.1.18 Appendix R

Table 18

Means and Standard Deviations of K_{MAX} Scores at 42-months by TLT category and CP_{NC} category at 30-months.

TLT Cat	CP_{NC} Cat	Mean	SD
Higher TLT	Higher NC	1.29	0.85
Higher TLT	Lower NC	1.23	0.64
Lower TLT	Higher NC	1.07	0.72
Lower TLT	Lower NC	1.41	0.69

Note. SD = Standard Deviation. CP_{NC} and TLT category were calculated using a median split.

.1.19 Appendix S

Table 19

Regression results for predicting 54-month K_{MAX} from 30-month mean CP_C, 30-month mean TLT, maternal education level, and gender.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	2.02	0.10	20.98	[1.83, 2.21]		<.001
Maternal Ed	0.24*	0.11	2.17	[0.06, 0.50]	1.74	.03543
Gender	- 0.24	0.19	- 1.25	[- 0.36, 0.53]	1.35	.22
CP_C	1.16	1.08	1.08	[- 1.78, 3.42]	1.68	.29
TLT	0.02	0.10	0.23	[- 0.25, 0.20]	2.64	.82
Maternal Ed:Gender	- 0.21	0.22	- 0.98	[- 0.21, 0.66]		.33
Maternal Ed:CP_C	1.31	1.27	1.04	[- 2.48, 2.50]		.31
Gender:CP_C	0.13	2.15	0.06	[- 8.51, 1.89]		.95
Maternal Ed:TLT	0.07	0.12	0.59	[- 0.19, 0.34]		.56
Gender:TLT	0.08	0.21	0.37	[- 0.71, 0.19]		.71
CP_C:TLT	- 1.98	1.28	- 1.55	[- 1.76, 2.73]		.13
Maternal Ed:Gender:CP_C	0.66	2.53	0.26	[- 6.85, 3.11]		.79
Maternal Ed:Gender:TLT	0.30	0.25	1.21	[- 0.28, 0.78]		.23
Maternal Ed:CP_C:TLT	0.85	1.32	0.64	[- 2.84, 3.13]		.52
Gender:CP_C:TLT	- 3.47	2.55	- 1.36	[- 2.99, 5.99]		.18
Maternal Ed:Gender:CP_C:TLT	2.53	2.64	0.96	[- 4.56, 7.38]		.34
N	58					
R^2	0.38					
Adj R^2	0.15					

Note: LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p < .01, *** indicates p < .001.

.1.20 Appendix T

Table 20

Regression results for predicting K_{MAX} at 54-months from 30-month mean CP_{NC}, 30-month mean TLT, maternal education level, and gender.

Predictor	***B***	**SE**	***t***	**95% CI** [LL, UL]	**VIF**	***p***
(Intercept)	2.05***	0.10	20.67	[1.85, 2.25]		<.001
Maternal Ed	0.28*	0.12	2.29	[0.03, 0.52]	2.11	.0273
Gender	- 0.42*	0.20	- 2.11	[- 0.82, - 0.02]	1.41	.0412
CP_{NC}	- 0.15	0.95	- 0.16	[- 2.06, 1.77]	1.81	.88
TLT	- 0.12	0.09	- 1.34	[- 0.30, 0.06]	1.89	.19
Maternal Ed:Gender	- 0.10	0.24	- 0.39	[- 0.59, 0.39]		.70
Maternal Ed:CP_{NC}	0.22	0.97	0.22	[- 1.75, 2.18]		.83
Gender:CP_{NC}	1.11	1.90	0.59	[- 2.72, 4.94]		.56
Maternal Ed:TLT	0.19	0.11	1.73	[- 0.03, 0.41]		.09
Gender:TLT	- 0.01	0.18	- 0.06	[- 0.37, 0.34]		.95
CP_{NC}:TLT	0.02	0.74	0.03	[- 1.48, 1.52]		.98
Maternal Ed:Gender:CP_{NC}	- 0.95	1.94	- 0.49	[- 4.87, 2.98]		.63
Maternal Ed:Gender:TLT	0.59*	0.22	2.67	[0.14, 1.04]		.01083
Maternal Ed:CP_{NC}:TLT	0.64	0.97	0.66	[-1.32, 2.61]		.51
Gender:CP_{NC}:TLT	1.64	1.49	1.10	[-1.36, 4.65]		.28
Maternal Ed:Gender:CP_{NC}:TLT	2.02	1.95	1.04	[-1.91, 5.94]		.31
N	58					
R^2	0.37					
Adj R^2	0.14					

Note: LL and UL indicate the lower and upper limits of a confidence interval, respectively. * indicates p < .05, ** indicates p < .01, *** indicates p < .001.

.1.21 Appendix U

Table 21

Means and Standard Deviations of K_{MAX} at 54-months by Maternal Education category, gender, and TLT category at 30-months.

Maternal Ed cat	Gender	TLT Cat	Mean	SD
Higher Edu	B	Higher TLT	2.43	0.53
Higher Edu	B	Lower TLT	2.45	0.67
Higher Edu	G	Higher TLT	2.12	0.70
Higher Edu	G	Lower TLT	1.97	0.55
Lower Edu	B	Higher TLT	1.74	0.56
Lower Edu	B	Lower TLT	1.78	0.95
Lower Edu	G	Higher TLT	1.40	0.53
Lower Edu	G	Lower TLT	2.27	0.73

Note. B = Boys, G = Girls. SD = Standard Deviation. Maternal Ed and TLT category were calculated using a median split.

.1.22 Appendix V

Table 22

Generalized Linear Mixed Model results for predicting MEFS total score at 30-months from Maternal Education level and gender.

Predictor	*B*	SE	*z*	*p*
(Intercept)	0.15***	0.00	40.87	< .001
Maternal Education level	0.01.	0.00	1.90	.058
Gender	0.01	0.01	1.17	.243
Maternal Education level:Gender	0.00	0.01	0.03	.974
logLik	217.4			
σ^2	0.0017			

Note. AIC = - 424.9, BIC = - 410.2. logLik = log-likelihood. σ^2 = Dispersion parameters for Student's t. *p < .05, **p < .01, ***p < .001.

.1.23 Appendix W

Table 23

Generalized Linear Mixed Model results for predicting MEFS total score at 78-months from Maternal Education level and gender.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.73***	0.01	92.89	< .001
Maternal Education level	0.01	0.01	0.56	.570
Gender	0.00	0.02	0.14	.890
Maternal Education level:Gender	0.0357*	0.02	2.21	.0272
logLik	131.1			
σ^2	0.0069			

Note. AIC = - 252.1, BIC = - 237.5. logLik = log-likelihood. σ^2 = Dispersion parameters for Student's t. *p < .05, **p < .01, ***p < .001.

.1.24 Appendix X

Table 24

Generalized Linear Mixed Model results for predicting MEFS total score at 78-months from MEFS at 30-months, Maternal Education level, and gender.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.74***	0.01	87.25	< .001
Total Score (30)	0.88***	0.22	3.98	< .001
Maternal Education level	- 0.00	0.01	- 0.39	.698
Gender	- 0.00	0.02	- 0.10	.924
Total Score (30):Maternal Education level	- 0.16	0.20	- 0.81	.420
Total Score (30):Gender	- 0.45	0.44	- 1.01	.314
Maternal Education level:Gender	0.03	0.02	1.87	.062
Total Score (30):Maternal Education level:Gender	- 0.59	0.40	- 1.48	.139
logLik	121.0			
σ^2	0.0065			

Note. AIC = - 224.1, BIC = - 198.9. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.25 Appendix Y

Table 25

Generalized Linear Mixed Model results for predicting MEFS total score from year, Maternal Education level, and gender, using total score (TS) in year 4 as criterion.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.49***	0.01	74.45	< .001
TS Year1	- 0.33***	0.01	- 45.78	< .001
TS Year2	- 0.07***	0.01	- 5.60	< .001
TS Year3	0.13***	0.01	16.41	< .001
Maternal Ed	0.03***	0.01	3.91	< .001
Gender	0.05***	0.01	3.35	.0008
TS Year1:Maternal Ed	- 0.02*	0.01	- 2.33	.020
TS Year2:Maternal Ed	0.01	0.01	1.26	.208
TS Year3:Maternal Ed	0.01	0.01	0.84	.403
TS Year1:Gender	- 0.04*	0.02	- 2.34	.019
TS Year2:Gender	0.17***	0.03	5.82	< .001
TS Year3:Gender	- 0.06***	0.02	- 4.07	< .001
Maternal Ed:Gender	0.02	0.01	1.30	.195
TS Year1:Maternal Ed:Gender	- 0.00	0.01	- 0.26	.797
TS Year2:Maternal Ed:Gender	- 0.00	0.02	- 0.12	.908
TS Year3:Maternal Ed:Gender	- 0.01	0.02	- 0.70	.487
logLik	258.7			
σ^2	0.0026			
Random Effects	*SD*	Variance		
Participant code (Intercept)	0.033	0.001		

Note. AIC = - 479.3, BIC = - 410.7. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. TS refers to Total Score. *p < .05, **p < .01, ***p < .001.

.1.26 Appendix Z

Table 26
Means and Standard Deviations of MEFS total score by Year and Gender

Year	Gender	Mean	SD
1	B	14.77	4.23
1	G	16.78	7.07
2	B	33.88	16.86
2	G	46.21	14.90
3	B	60.00	14.88
3	G	60.56	9.42
4	B	75.42	10.23
4	G	75.49	8.02

Note. B = Boys, G = Girls. SD = Standard Deviation.

.1.27 Appendix AA

Table 27

Generalized Linear Mixed Model results for predicting MEFS Total Score at 30-months of age from mean CP_C at 6-months of age, mean TLT at 6-months of age, Maternal Education level, and gender.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.16***	0.01	24.97	< .001
Maternal Education level	0.01	0.01	1.28	.199
Gender	0.01	0.01	0.87	.383
CP_C	- 0.00	0.06	- 0.05	.961
TLT	0.01	0.01	1.55	.121
Maternal Education level:Gender	0.00	0.01	0.10	.919
Maternal Edcuation level:CP_C	0.08	0.08	1.02	.307
Gender:CP_C	- 0.10	0.12	- 0.83	.405
Maternal Education level:TLT	- 0.01	0.01	- 1.37	.172
Gender:TLT	0.02*	0.01	2.19	.02874
CP_C:TLT	0.03	0.03	1.00	.317
Maternal Education level:Gender:CP_C	0.09	0.16	0.57	.569
Materanl Education level:Gender:TLT	- 0.03**	0.01	- 2.58	.010
Maternal Education level:CP_C:TLT	- 0.03	0.04	- 0.58	.561
Gender:CP_C:TLT	0.07	0.07	1.04	.301
Maternal Education level:Gender:CP_C:TLT	- 0.07	0.09	- 0.73	.468
logLik	104.7			
σ^2	0.0014			

Note. AIC = - 175.4, BIC = - 139.0 logLik = log-likelihood σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.28 Appendix BB

Table 28

Mean and Standard Deviations of MEFS Total Score at 30-months of age by Maternal Educational Level, Gender, and TLT Category at 6-months of age.

Maternal Education level category	Gender	TLT category	Mean	SD
Higher EDU	B	Higher TLT	16.09	2.70
Higher EDU	B	Lower TLT	15.18	3.79
Higher EDU	G	Higher TLT	18.00	10.16
Higher EDU	G	Lower TLT	16.36	3.97
Lower EDU	B	Higher TLT	10.86	5.52
Lower EDU	B	Lower TLT	14.33	2.89
Lower EDU	G	Higher TLT	20.50	7.78
Lower EDU	G	Lower TLT	15.25	1.71

Note. Maternal Education and TLT categories were determined using a median split. Maternal Education level and TLT were both entered in the model as continuous variables.

.1.29 Appendix CC

Table 29

Generalized Linear Mixed Model results for predicting MEFS Total Score at 30-months of age from Maternal Education level, Gender, mean CP_{NC} at 6-months of age, and mean TLT at 6-months of age.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.15***	0.01	21.74	< .001
Maternal Education level	0.01	0.01	1.00	.318
Gender	0.02	0.01	1.06	.289
CP_{NC}	0.02	0.06	0.42	.677
TLT	0.00	0.00	0.91	.361
Maternal Education level:Gender	- 0.01	0.02	- 0.42	.675
Maternal Education level:CP_{NC}	0.05	0.07	0.68	.495
Gender:CP_{NC}	0.14	0.11	1.24	.217
Maternal Education level:TLT	- 0.01.	0.01	- 1.67	.096
Gender:TLT	0.01	0.01	1.45	.147
CP_{NC}:TLT	0.02	0.03	0.56	.573
Maternal Education level:Gender:CP_{NC}	0.07	0.15	0.49	.621
Maternal Education level:Gender:TLT	- 0.03**	0.01	- 2.59	.00953
Maternal Education level:CP_{NC}:TLT	0.06	0.04	1.30	.194
Gender:CP_{NC}:TLT	0.01	0.06	0.08	.936
Maternal Education level:Gender:CP_{NC}:TLT	0.17.	0.09	1.94	.0519
logLik	105.5			
σ^2	0.0014			

Note. AIC = - 177.1, BIC = - 140.7. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.30 Appendix DD

Table 30

Generalized Linear Mixed Model results for predicting MEFS Total Score at 30-months of age from Maternal Education level, Gender, mean CP_C at 18-months of age, and mean TLT at 18-months of age.

Predictor	***Estimate***	**SE**	***z***	***p***
(Intercept)	0.16***	0.01	31.21	< .001
Maternal Education level	0.01	0.01	1.08	.280
Gender	0.02	0.01	1.56	.119
CP_C	0.09*	0.04	2.17	.02981
TLT	0.00	0.00	0.19	.853
Maternal Education level:Gender	- 0.03*	0.01	- 2.55	.01065
Maternal Education level:CP_C	- 0.12**	0.05	- 2.61	.00917
Gender:CP_C	0.12	0.08	1.46	.144
Maternal Education level:TLT	- 0.01**	0.01	- 2.58	.00989
Gender:TLT	- 0.02.	0.01	- 1.95	.05105
CP_C:TLT	- 0.02	0.03	- 0.69	.490
Maternal Education level:Gender:CP_C	- 0.08	0.09	- 0.87	.387
Maternal Education level:Gender:TLT	0.00	0.01	0.25	.806
Maternal Education level:CP_C:TLT	- 0.08	0.05	- 1.60	.109
Gender:CP_C:TLT	- 0.01	0.06	- 0.19	.851
Maternal Education level:Gender:CP_C:TLT	- 0.31**	0.10	- 3.03	.00244
logLik	109.4			
σ^2	0.0011			

Note. AIC = - 184.8, BIC = - 148.7. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.31 Appendix EE

Table 31

Mean and Standard Deviations of MEFS Total Score at 30-months of age by Maternal Educational Level and TLT Category at 18-months of age.

Maternal Education level category	TLT category	Mean	SD
Higher EDU	Higher TLT	16.27	6.28
Higher EDU	Lower TLT	16.62	5.04
Lower EDU	Higher TLT	16.60	5.55
Lower EDU	Lower TLT	12.30	5.186521

Note. Maternal Education and TLT categories were determined using a median split. Maternal Education level and TLT were both entered in the model as continuous variables.

.1.32 Appendix FF

Table 32

Generalized Linear Mixed Model results for predicting MEFS Total Score at 30-months of age from Maternal Education level, gender, mean CP_{NC} at 18-months of age, and mean TLT at 18-months of age.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.16***	0.01	26.58	< .001
Maternal Education level	0.01	0.01	1.52	.127
Gender	0.02	0.01	1.59	.113
CP_{NC}	0.06	0.06	1.04	.298
TLT	- 0.00	0.01	- 0.32	.746
Maternal Education level:Gender	- 0.02	0.01	- 1.36	.174
Maternal Education level:CP_{NC}	- 0.11	0.07	- 1.64	.101
Gender:CP_{NC}	0.17	0.12	1.43	.152
Maternal Education level:TLT	- 0.01*	0.01	- 2.22	.0266
Gender:TLT	- 0.02*	0.01	- 2.48	.0131
CP_{NC}:TLT	0.09*	0.04	2.03	.0428
Maternal Education level:Gender:CP_{NC}	- 0.28*	0.14	- 2.06	.0397
Maternal Education level:Gender:TLT	- 0.00	0.01	- 0.05	.963
Maternal Education level:CP_{NC}:TLT	- 0.07	0.06	- 1.23	.218
Gender:CP_{NC}:TLT	0.19*	0.09	2.22	.0267
Maternal Education level:Gender:CP_{NC}:TLT	- 0.04	0.12	- 0.35	.728
logLik	105.7			
σ^2	0.0014			

Note. AIC = - 177.4, BIC = - 141.3. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.33 Appendix GG

Table 33

Mean and Standard Deviations of MEFS Total Score at 30-months of age by gender, CP_{NC} category at 18-months of age, and TLT category at 18-months of age.

Gender	CP_{NC} category	TLT category	Mean	SD
B	Higher NC	Higher TLT	16.11	2.62
B	Higher NC	Lower TLT	13.57	5.26
B	Lower NC	Higher TLT	15.00	2.00
B	Lower NC	Lower TLT	13.40	5.36
G	Higher NC	Higher TLT	16.13	5.74
G	Higher NC	Lower TLT	15.57	1.27
G	Lower NC	Higher TLT	17.75	10.54
G	Lower NC	Lower TLT	19.14	6.84

Note. CP_{NC} and TLT categories were determined using a median split. Both CP_{NC} and TLT were entered into the model as continuous variables.

.1.34 Appendix HH

Table 34

Mean and Standard Deviations of MEFS Total Score at 30-months of age by gender, Maternal Education level category, and CP_{NC} category at 18-months of age.

Gender	Maternal Education Category	CP_{NC} category	Mean	SD
B	Higher EDU	Higher NC	16.36	2.50
B	Higher EDU	Lower NC	15.00	3.82
B	Lower EDU	Higher NC	12.00	5.43
B	Lower EDU	Lower NC	11.80	5.17
G	Higher EDU	Higher NC	15.08	3.45
G	Higher EDU	Lower NC	18.92	9.27
G	Lower EDU	Higher NC	19.00	6.08
G	Lower EDU	Lower NC	15.00	2.83

Note. Maternal Education level and CP_{NC} categories were determined using a median split. Both CP_{NC} and Maternal Education level were entered into the model as continuous variables.

.1.35 Appendix II

Table 35

Generalized Linear Mixed Model results for predicting MEFS Total Score at 78-months of age from mean CP_C at 30-months of age, mean TLT at 30-months of age, Maternal Education level, and gender.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.55***	0.12	4.49	$< .001$
CP_C	0.20*	0.09	2.11	.0350
TLT	0.02*	0.01	2.30	.0217
Gender	-0.00	0.02	-0.15	.881
Maternal Education level	0.01	0.01	1.52	.129
AgeMEFS	0.00	0.00	1.38	.167
CP_C:TLT	-0.06	0.08	-0.75	.455
CP_C:Gender	-0.10	0.18	-0.53	.596
TLT:Gender	0.02	0.01	1.22	.221
CP_C:Maternal Education level	-0.10	0.10	-0.92	.357
TLT:Maternal Education level	-0.00	0.01	-0.21	.833
Gender:Maternal Education level	0.03.	0.02	1.89	.0595
CP_C:TLT:Gender	-0.23	0.17	-1.37	.170
CP_C:TLT:Maternal Education level	0.13	0.10	1.23	.220
CP_C:Gender:Maternal Education level	-0.36	0.22	-1.65	.100
TLT:Gender:Maternal Education level	-0.02	0.02	-1.19	.233
CP_C:TLT:Gender:Maternal Education level	0.07	0.19	0.37	.708
logLik	133.30			
σ^2	0.00581			

Note. AIC = -230.5, BIC = -179.0. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.36 Appendix JJ

Table 36

Generalized Linear Mixed Model results for predicting MEFS Total Score at 78-months of age from Maternal Education level, Gender, mean CP_{NC} at 30-months, and mean TLT at 30-months of age.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.56***	0.12	4.74	< .001
CP_{NC}	0.01	0.07	0.15	.885
TLT	0.02*	0.01	2.14	.0325
Gender	0.01	0.02	0.42	.671
Maternal Education level	0.01	0.01	1.26	.209
MEFS Age at test	0.00	0.00	1.40	.162
CP_{NC}:TLT	0.07	0.06	1.07	.284
CP_{NC}:Gender	0.06	0.14	0.44	.658
TLT:Gender	0.01	0.01	0.98	.329
CP_{NC}:Maternal Education level	0.05	0.08	0.65	.516
TLT:Maternal Education level	-0.00	0.01	-0.43	.666
Gender:Maternal Education level	0.03	0.02	1.86	.063
CP_{NC}:TLT:Gender	0.09	0.13	0.74	.459
CP_{NC}:TLT:Maternal Education level	- 0.08	0.07	- 1.11	.267
CP_{NC}:Gender:Maternal Education level	0.13	0.15	0.87	.385
TLT:Gender:Maternal Education level	- 0.03	0.02	- 1.58	.114
CP_{NC}:TLT:Gender:Maternal Education level	- 0.11	0.15	- 0.74	.457
logLik	128.70			
σ^2	0.00635			

Note. AIC = -221.5, BIC = -170.0. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.37 Appendix KK

Table 37

Generalized Linear Mixed Model results for predicting MEFS Total Score at 54-months of age from Maternal Education, Gender, and K_{MAX} at 42-months of age.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	- 0.12	0.52	- 0.22	.826
Maternal Education level	0.02	0.01	1.65	.098
Gender	- 0.05.	0.03	- 1.95	.051
K_{MAX}	0.04*	0.02	2.17	.0302
MEFS Age at test	0.01	0.01	1.39	.166
Maternal Education level:Gender	- 0.00	0.03	- 0.02	.981
Maternal Education level:K_{MAX}	0.00	0.02	0.21	.833
Gender:K_{MAX}	0.02	0.04	0.63	.529
Maternal Education level:Gender:K_{MAX}	0.08	0.05	1.77	.076
logLik	57.50			
σ^2	0.00548			

Note. AIC = -95.0, BIC = -75.0. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.38 Appendix LL

Table 38

Generalized Linear Mixed Model results for predicting MEFS Total Score at 78-months of age from Maternal Education level, Gender, and K_{MAX} at 42-months of age.

Predictor	*B*	SE	*z*	*p*
(Intercept)	0.77***	0.01	71.61	< .001
Maternal Education level	0.02*	0.01	2.16	.03116
Gender	- 0.05*	0.02	- 2.49	.01266
K_{MAX}	0.02	0.01	1.20	.232
Age at test	0.00	0.00	1.53	.126
Maternal Education level:Gender	0.05**	0.02	2.80	.00518
Maternal Education level:K_{MAX}	- 0.04*	0.02	- 2.03	.04248
Gender:K_{MAX}	- 0.01	0.03	- 0.50	.619
Maternal Education level:Gender:K_{MAX}	0.17***	0.04	4.40	< .001
logLik	70.10			
σ^2	0.00332			

Note. AIC = -120.2, BIC = -100.7. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.

.1.39 Appendix MM

Table 39

Mean and Standard Deviations of MEFS Total Score at 78-months of age by Gender, Maternal Education level category, and K_{MAX} category at 42-months of age.

Gender	Maternal Education Category	$K_{\mathbf{MAX}}$ Category	Mean	SD
B	Higher Edu	Higher K	76.56	8.35
B	Higher Edu	Lower K	82.00	5.83
B	Lower Edu	Higher K	87.00	0.00
B	Lower Edu	Lower K	74.00	5.97
G	Higher Edu	Higher K	79.42	6.57
G	Higher Edu	Lower K	76.00	7.27
G	Lower Edu	Higher K	70.67	14.19
G	Lower Edu	Lower K	72.57	5.03

Note. Maternal Education level and K_{MAX} categories were determined using a median split. Both K_{MAX} and Maternal Education level were entered into the model as continuous variables.

.1.40 Appendix NN

Table 40

Generalized Linear Mixed Model results for predicting MEFS Total Score at 78-months of age from Maternal Education level, gender, and K_{MAX} at 54-months of age.

Predictor	***B***	**SE**	***z***	***p***
(Intercept)	0.78***	0.01	57.72	< .001
Maternal Education level	- 0.01	0.01	- 0.39	.695
Gender	- 0.01	0.02	- 0.29	.771
K_{MAX}	0.02	0.02	1.06	.290
MEFS Age at test	- 0.00	0.00	- 0.42	.673
Maternal Education level:Gender	0.01	0.03	0.28	.776
Maternal Education level:K_{MAX}	- 0.03	0.02	- 1.63	.104
Gender:K_{MAX}	0.03	0.03	0.77	.442
Maternal Education level:Gender:K_{MAX}	0.00	0.03	0.15	.882
logLik	63.70			
σ^2	0.00455			

Note. AIC = -107.3, BIC = -87.6. logLik = log-likelihood. σ^2 = Dispersion parameter for Student's t. *p < .05, **p < .01, ***p < .001.