Structure and development of executive function in early adolescence

Structure and Development of Executive Function in Early Adolescence

Elizabeth Rebecca Booth

A thesis submitted for the degree of Doctor of Philosophy

Department of Psychological Sciences, Birkbeck, University of London

April 2024

Acknowledgements

My heartfelt thanks and gratitude go to all the wonderful people I've come across at Birkbeck and the wider group of universities across Bloomsbury. To have had the chance to explore all the big ideas about cognition, development and life in general has been a fabulous experience. To all in Michael's DNL group, the CEN PhD group, Iroise's iDCN lab: the discussions and ideas we shared have been so impactful on me and my work, my thanks to you all.

I would like to thank my supervisors for the good advice and endless patience you have shown me. Michael for your help with the big ideas, and Iroise for your help in getting to grips with the details!

Finally, and perhaps most importantly, I would like to thank my family, and especially my partner Harry for supporting me so well throughout all the years this thesis has taken me to complete. This thesis is dedicated to you. Love always.

The work presented in this thesis was supported by a U.K. Economic and Social Research Council studentship.

Abstract

Executive functions (EFs) are a set of over-arching cognitive functions that act to that to coordinate other mental processes, enabling goal-directed behaviour to take place. The exact structural relationship between components of EFs is as yet unclear, and the relationships of EFs to other factors such as socio-economic status are also still to be fully explored. The developmental pathways of EF components and the structural relationships between them, in particular during adolescence, also remain unclear.

This thesis uses data from the Study of Cognition, Adolescents and Mobile Phones - a longitudinal cohort study of over 6,600 high-school students across Greater London – to explore structure and development of EFs across adolescence. A computerised task battery was completed at two time points: baseline assessment took place during school Years 7-8 (ages approximately 10-13 years) and follow-up occurred during school Years 9-10 (ages approximately 13-16 years).

The first experimental chapter investigates the relationship between EFs and socio-economic status (SES) at baseline assessment, finding significant associations between overall SES and EF, and between some specific EF measures and aspects of SES. In some cases these relationships remain significant even when accounting for fluid intelligence. The second chapter uses multiple regressions and multi-level modelling to explore developmental trajectories of EF and fluid intelligence across early adolescence, and finds significant associations with age and task score for most of the cognitive tasks. The final chapter uses exploratory factor analysis to explore structural relationships between EF components at the two assessment points, and finds evidence supporting a three-factor model of EF within our cohort.

Table of Contents

Contents

Table of Tables

Table of Figures

Abbreviations used in this thesis

BDS Backward Digit Span

CFA Confirmatory factor analysis

CFT Cattell's Culture Fair Task

CPT Continuous Performance Task

EF Executive function

EFA Exploratory factor analysis

fMRI functional Magnetic Resonance Imaging

G General Intelligence

Gf Fluid General Intelligence

Gc Crystallised General Intelligence

IQ Intelligence Quotient

M Mean

MANCOVA Multiple analysis of covariance

MLM Multi-level modelling

N Number of participants / data points

PFC Prefrontal cortex

RF-EMF Radio-frequency electromagnetic fields

SCAMP Study of Cognition, Adolescents and Mobile Phones

SES Socio-economic status

SNR Signal-to-noise ratio

SWM Spatial Working Memory

TMT Trail Making Task

WM Working Memory

Chapter 1. Introduction

Executive functions (EFs) are over-arching cognitive processes that act to coordinate other mental processes. They allow goal-directed behaviour, including in novel or distracting situations (Diamond, 2013). EFs may also be described as higher-level cognitive functions, and are generally associated with frontal lobe activity (Stuss & Levine, 2002; Miyake & Friedman, 2012). Despite the large volume of research into executive function (EF) in the neuropsychological literature, a single precise definition remains difficult to pin down, and is yet to be formally operationalised (Jurado & Rosselli, 2007). Different researchers describe EF in terms of different cognitive functions, and many different terms are used in the literature for similar related concepts. A simpler definition may only include inhibition, working memory and switching, where more complex definitions may also encompass verbal reasoning, problem solving, planning, sequencing and / or multitasking (Chan et al., 2008a). EFs are approximately analogous to the activities carried out in Baddeley's central executive in his working memory model, or the supervisory attention system (SAS) in Norman and Shallice's work (Fournier-Vicente et al., 2008). By their nature as higher-level cognitive processes, EFs act to coordinate other cognitive process to achieve complex goal oriented actions. EFs are generally considered to be multifactorial, in that the overall EF umbrella consists of multiple separable components of cognitive functioning. These different components of EF must work together to complete complex tasks (Zelazo et al., 2008).

EFs have prolonged developmental trajectories, in keeping with the protracted development of frontal brain regions. Behavioural and functional development of the EF system extends across adolescence. EFs tend to develop from being more general processes to more complex and specific ones across childhood and adolescence (Diamond, 2013). There is some debate in the literature as to whether EF capability reflects the working of a single system (unitary models, e.g. Duncan, 2013); or coordinated action of multiple dissociable, but related, sub-systems of both (also known as unity-yet-diversity models, e.g. (Miyake et al., 2000). Factor analysis research has shown that structural models of EF change across development, and across adolescence in particular, with a general pattern of fewer, more general EF components earlier in childhood, with an increasing number of more specialised components in later adolescence. EFs are associated with a wide range of important life outcomes, such as wealth, academic achievement and health in adulthood.

This thesis will investigate three key areas related to EFs in adolescence. The first empirical chapter (**Chapter 3**) considers the relationships between EF and socio-economic status. **Chapter 4** explores the development of specific EF abilities in terms of task performance across early adolescence. **Chapter 5** uses factor analysis to explore the structure of EF at our two assessment time points in early adolescence, to investigate whether the factor structure matches those found in adulthood, and whether there are any developmental differences in EF structure across our assessment points.

Models of Executive FunctionAt least two broad explanations of the structure of EFs are plausible. EFs may be undifferentiated and unitary in nature, with performance in EF tasks reflecting the work of one system; or they may be fractionated in structure, representing the work of multiple components or sub-systems (Diamond, 2013). Most models of EF take at least some account of both unity and diversity of EF in their explanations.

1.1.1 Approaches in EF Research

Research into executive functioning takes two broad approaches: the first investigates cognitive deficits in patients with frontal lobe damage, and the second uses cognitive tasks that tap putative executive control functions in normally functioning individuals. The work of Norman and Shallice (1986) is a typical example of the first approach. Various tasks were used to analyse the impact of frontal lobe injuries on patients' cognitive functions. It was observed that patients showed specific difficulties in complex tasks involving planning and other EF functions, such as the Wisconsin Card Sort Task (WCST), while retaining ability to perform well in other cognitive tests including IQ tests (Norman & Shallice, 1986). The collection of the specific deficits in these patients was termed 'dysexecutive syndrome'. This work formed the basis for subsequent theories of the roles of the frontal lobes in EF activities. However, it has been noted that there is no simple 1:1 correspondence between specific lesion sites and pathology of individuals, meaning that the roles of frontal cortex in cognition are complex.

A second approach is to focus on non-pathological individuals, and use cognitive testing and brain scanning methods to investigate cognitive control functions across different conditions. An example of this is the work of Baddeley and Hitch (1974), which explored components of working memory with experimental methods, and described the 'central executive' as being a key component in working memory processes, working to maintain information in mind and make action decisions based on relevant information.

A wealth of theories of EF have been developed over the years since the earliest explorations of the central executive and dysexecutive syndrome. Some key theories of executive functioning arising out of these two broad methodological approaches, and of course combinations of these approaches, will now be discussed.

1.1.2 Supervisory Attentional System (SAS)

Arising out of research with patients with frontal lobe injuries, the supervisory attentional system (SAS) model proposes that executive functions act to coordinate attentional processes to generate new response schemas in novel situations, initiate responses, and monitor accuracy (Norman & Shallice, 1986). There are four levels of thought and behaviour in this model. Firstly, simple "cognitive or action units" correspond to basic actions and low-level cognitive abilities (e.g. perceive a stimulus; press a button). Secondly, "schemata" are nests of actions or cognitive abilities that form closely related groups, which are repeatedly triggered at the same time (e.g. press the brake and put car into neutral when stopping). Thirdly, a process of "contingency scheduling" takes place to trigger action schemas where a well-rehearsed schema is appropriate – this process does not involve EF control (e.g. at a red light trigger the car stopping schema). Finally, EF control is a (usually) conscious process enacted by the "supervisory attentional system" to decide on appropriate actions, and initiate the relevant contingency schedule process (e.g. plan how to get to a Doctor's appointment on time).

This model suggests EFs are used in a wide range of situations and are not tied to any particular stimulus type. The conceptual framework of EFs in the SAS include a wide variety of cognitive processes such as working memory and inhibition of incorrect responses; wider attentional processes including shifting, maintenance, division and selection of attention; and wider EFs such as problem solving, novel responses, response monitoring and error correction, general decision making, priming of anticipated responses, and planning.

The original SAS model could be critiqued as being largely descriptive in nature (Hommel et al., 2002). There is little explanation of how these processes carried out by the SAS might occur. The cognitive control or EF portion of the SAS is largely unitary in nature – in that all of these complex cognitive functions are carried out broadly within the frontal lobes, with little differentiation between sub-types of EF present in the original model. However, some patient studies showed participants with deficits within specific types of EF tasks, rather than across all EFs in general (Shallice & Burgess, 1991). This led to further work to explore more specific aspects of EF control and specific frontal lobe regions in more detail, which produced more differentiated models within the original SAS framework, with EF becoming increasingly conceptualised as a group of differentiated control functions located within sub-regions of the frontal lobes.

Gateway Hypothesis of Rostral Pre-frontal Cortex Function

Later work by Tim Shallice, Paul Burgess et al. has developed more detailed theories of frontal lobe function within the SAS framework, which has informed by studies of patients with specific cognitive deficits and by cognitive testing with non-pathological individuals. This work has attempted to identify the functions of more specific regions within frontal cortex, and expound theories of specific EF capabilities and relate these functions to regions of frontal cortex. One such theory is the "gateway hypothesis" of rostral PFC function (Burgess et al., 2007). Rostral PFC the very front section of the frontal cortex - anatomically defined as Brodmann area 10. It forms over 1% of the brain's overall volume. Burgess et al. (2007) suggests that this region performs a cognitive control function, balancing attention paid to incoming stimuli and non-stimulus related thoughts while both performing tasks and at rest. Burgess suggests that this region is heavily involved in prospective memory, including priming of expected responses and in performing delayed response tasks. Patients with specific lesions in this region show specific deficits in real-world tasks involving multi-tasking and in novel situations; and in the lab, tasks involving task-switching. Positron emission tomography (PET) studies indicate that regional cerebral blood flow (rCBF) is increased in rostral PFC when expecting a stimulus during prospective memory tasks (Burgess et al. 2007).

The gateway hypothesis suggests rostral PFC acts to guide behaviour in situations where the optimal course of action is unclear, and where multi-tasking or switching between responses is required. This hypothesis expands on the original SAS model, by suggesting mechanisms of action supporting purported cognitive actions of this particular region of the frontal lobes. Rostral PFC has close anatomical connections with ventrolateral and dorsolateral PFC – the gateway hypothesis suggests rostral PFC performs a role of selecting and switching between inputs from these systems. It suggests rostral PFC acts as a gateway, similar to a switch-point on a train-track, to select attentional focus at any given moment. It synthesises inputs from stimulus-related and stimulus-independent thought systems. It's attention selection processes can be influenced by stimulus-independent or conscious thoughts ("I need to focus on x…"), and/or by salience of stimulus inputs (Red light = Stop). The function of the rostral PFC is then to activate or bias towards particular "schemata" or "contingency schedules" to be triggered, then the processes which are enacted by these lower-level systems will be monitored to ensure ongoing accuracy. Burgess et al. (2007) suggests the PFC operates across multiple domains, i.e. it is independent of stimulus or input type.

This hypothesis goes some way to addressing the critique of the SAS being largely descriptive, in that it makes specific predictions as to how the system will operate in particular conditions. Evidence supporting this theory comes from fMRI studies which show increased rostral PFC activation in a task requiring switching between stimulus-oriented and stimulus-independent thought versus similar non-switching tasks (Gilbert et al., 2005).

1.1.3 Baddeley's Central Executive

An example of the second type of approach to EF research (i.e. using cognitive testing in non-pathological individuals) focussed on the "central executive", the control component of Baddeley and Hitch's Working Memory Model (Baddeley & Hitch, 1974). The central executive was postulated as a central control unit, allocating attention and processing the inputs from other units of working memory such as the visuo-spatial scratchpad and phonological loop. Original conceptions of the central executive were largely unitary in nature.

One criticism which could be levelled at the original description of the unitary central executive is that it appears to be little more than a 'homunculus', offering few specific explanations of mechanisms of action or testable predictions (Baddeley, 1996a). Baddeley revisited the working memory model and the central executive in 1996 (Baddeley, 1996a, 1996b). In this later work, he described fractionation in the working memory model overall, and suggested potential avenues for future research into the fractionation of central executive function. He suggested four potential components that central executive function might be differentiated into: dual task coordination; alternation between cognitive strategies (cognitive flexibility); selective attention; and activation and retrieval of information from the long-term memory.

1.1.4 Posner's Model of Attention and EF

Another model of executive function is that it forms part a broader set of attentional control processes (Posner & Petersen, 1990; later updated in Petersen & Posner, 2012). In their original models, Posner and Petersen emphasise three concepts about attention. Firstly, attentional systems are anatomically separate from cognitive processing systems. Secondly, attentional processes make use of specific networks of interlinked brain regions. Thirdly, specific aspects of attentional processes are linked to distinct networks of brain regions. These distinct neural networks carry out specific aspects of attentional processes.

Petersen and Posner (1990) describe three key attentional networks of brain regions: alerting, orientation and executive control. The alerting system is responsible for producing and maintaining vigilance and focus during tasks. The alerting neural network is largely right-lateralised; is based around projections of the locus coeruleus; and includes both cerebellar and right cerebral cortex regions. The orientation system allows particular sensory modality or locations to be prioritized, and for example, would be active in a location priming task where response times are reduced when stimulus locations are indicated prior to the event. The orienting neural network is based around the dorsal and ventral visuo-spatial attentional systems. The executive control system relates most closely to the concept of executive functions explored in this thesis. In their original conception of attention, the executive control network was described as being responsible for focal attention and target detection during tasks, and the brain regions involved were largely midline regions of the medial frontal cortex and anterior cingulate.

A later update of their models favours the conceptualisation of executive control as two separate sets of functions, with two separate neural networks. They suggest that top-down control and attention during tasks is controlled by the functional co-ordination and interaction of these two networks. Firstly, they describe a frontoparietal system which is related to noticing start-cues, task switching behaviour, initiation of task appropriate responses, and real-time monitoring and adjustment of performance during trials. A second network of midline structures and the anterior cingulate (ACC) allows maintenance of on-task focus across trials, and acts as a stable background for maintaining attentional focus throughout a task. This is in contrast to many cognitive control theories, which conceive of a single unified system of executive control in which lateral pre-frontal cortex is responsible for top-down control, while performance monitoring and maintenance of attention is guided by midline structures and the ACC, as part of a single attentional neural network.

In favour of their dual network model of executive control, Petersen and Posner (2012) cite evidence from lesion studies. More laterally located lesions can result in perseveration on outdated task requirements, and an inability to switch sets when required, while ability to carry out tasks correctly within a single sustained set is preserved. This, they argue, means that lateral brain regions are responsible for task switching, rather than controlling all task performance per-se. They also argue that their dual system model is more consistent with research into timing of neural activity across neural networks during executive control tasks. They cite research where fronto-parietal regions of the control network activate early in trials, where ACC-related and midline regions activate later during trials and may even begin activity post-trial to enable post-trial understanding of performance. However, this evidence does not appear to be entirely exclusive of a single-network with dual-function conception, though they argue we might expect closer timing of activity across a single network than is being observed in the cited timing studies. Posner’s model has also been applied to childhood developmental processes, and used to explain development of attentional and executive function processes (Rueda et al., 2004).

1.1.5 Duncan’s Multiple Demand Network

The multiple demand (MD) network supports a wide variety of cognitive functions (Camilleri et al., 2018). Duncan equates the function of the multiple demand network to general fluid intelligence (Gf). Gf refers to a general ability to perform a wide variety of cognitive functions, and is generally tested by for example matrix completion tasks, and other logic tasks similar to those used in Cattell's culture fair task (CFT) (Cattell & Cattell, 1960b). It can therefore be interpreted as a unitary explanation of EF – in that the same MD network acts to carry out various EF and cognitive tasks. Duncan et al. (2020) suggests that the multiple demand network is very similar to brain networks identified in other research, including task-positive network (Fox et al., 2005), cognitive control network (Niendam et al., 2012), or extrinsic mode networks (Hugdahl et al., 2015). Key brain regions shown to be active during executive control tasks in the MD network, and across all the above mentioned brain networks, include posterior-medial frontal cortex (pre-supplementary motor area and middle cingulate cortex), bilateral anterior insula, intraparietal sulcus, and posterior inferior frontal sulcus (Camilleri et al., 2018). Activity is sometimes also seen in the rostro-lateral PFC, but this is not as frequently observed across as many types of task (J. Duncan et al., 2020).

A summary paper by Duncan (2010) sets out the MD network theory. The MD network is a system of brain regions which co-activate in many different situations. The MD network is central to general cognitive control. Beyond a narrow definition of EF consisting of isolated component operations e.g. inhibition, working memory, attentional bias or switching between tasks, the MD network also works to organise and structure the required actions and thoughts associates with complex multi-step behaviour. Real world examples of this type of behaviour include things like cooking a meal or solving a maths problem. Goals are achieved by organising a series of sub-tasks, or multiple steps, which are each separately defined then sequenced together by activity in the MD network to form a coherent series of actions that work toward achieving an overarching goal.

Duncan (2013) further suggests that cognition is structured in attentional episodes. During these episodes, the multiple demand network aligns attention to segment cognition into small steps required to complete a task at hand (i.e. breaks down the current goal into many smaller sub-goals that are needed to achieve it), then integrates the required steps to produce a complete outcome aligned with the task goal. These processes of segmenting and integrating attention to small component steps is influenced by things such as overarching goals, memories, and knowledge of required subgoals to meet the overarching goal. In this context, EFs would be necessary to maintain task goals and rules in working memory; for the ability to switch between different sets of rules and strategies to achieve goals as required; for monitoring progress and changing behaviour to fit current stage of the task; and inhibiting inaccurate responses based on this monitoring.

In series of experiments using specific versions of matrix tasks combined with fMRI, Duncan et al. (2020) demonstrated that tasks requiring integration and segmentation of attention recruit regions of the multiple demand network, with different task types recruiting somewhat differing groups of brain regions in the network. This recent work suggests that the MD network is not as unitary as first described, rather, it has differential patterns of activity in different types of EF and cognitive tasks. Although the original description of the multiple demand network was somewhat unitary in nature, in that it described the overall functions of this system across many areas of cognition, this has been adjusted to include aspects of differentiation of neural function.

1.1.6 Unity-yet-diversity models of EF

Unity-yet-diversity models suggest that EF is made up of multiple separable, though related, components of cognitive functionality. One influential model of this type was proposed by Miyake et al. (2000). They described EF as consisting of three dissociable yet correlated components: set-shifting or switching ability; inhibitory control or inhibition; and updating working memory contents. This thesis uses a working definition of EF based on the unity-yet-diversity models of Miyake et al. (2000) and Diamond (2013), with EF conceptualised as three dissociable yet related components.

A commonly used technique to explore unity-yet-diversity models is Confirmatory Factor Analysis (CFA). CFA is a statistical technique which enables researchers to identify latent variables that underlie complex cognitive performance, such as that required for tasks used to asses EF. To conduct CFA, sets of tasks are carefully selected to target specific proposed EF components. A variety of tasks will usually be used to address each hypothesised component. These tasks will have different surface features (i.e. different specific requirements, with different lower-level cognitive processes underlying their performance such as visual perception or motor control abilities), but will all tap into some common element of some EF component. Task scores are then analysed using CFA. Using this technique, common variance among groups of tasks is extracted statistically, in order to identify whether certain pre-planned latent structures adequately describe performance across multiple tasks which ostensibly tap the same component. Various models are then tested to see which combination of latent variables best explains the observed variance. These models will test whether the latent variables are related to each other, or whether each variable is statistically separable from the others. Miyake et al. (2000) used nine tasks, with three intended to tap each hypothesised EF factor. They used CFA to whether their proposed three-factor model was preferable to a one-factor (unitary) model, and with any two-factor models. They found that in adults, a three-factor model best explained the observed variance, but that the three latent variables were significantly correlated to each other. This research supports the idea that EFs are fractionated, and that the underlying components are also somewhat related to each other (unity-yet-diversity).

Other studies have observed that a three-factor model of EFs is also present in atypical populations. CFA of data obtained from a questionnaire-based EF assessment (the Behavioural Rating Inventory of Executive Function or BRIEF scale) in a clinical population revealed an underlying pattern of three dissociable latent variables (Gioia et al., 2002). A three-factor model was preferable to one- or two-factor models. The observed latent structure was similar to that found by Miyake et al. (2000). This suggests that a model of EFs consisting of dissociable yet related components may be present in clinical populations also, and can be found by assessing participants using questionnaire measures of EF.

Miyake et al.'s (2000) research does not rule out other alternative fractionated models, as tests were selected to specifically address the three factors of interest in their study. Other research has taken a broader view of EF, using CFA to look at other cognitive processes which may also be considered as part of EF. For example, Fournier-Vicente et al. (2008) found that a five-factor model explained EF data in adults. These five factors were labelled as: verbal storage-and-processing coordination; visuospatial storage-and-processing coordination; selective retrieval; selective attention; and shifting. The research had included many and more varied tasks than in the Miyake research, for example they included tasks involving dual-task co-ordination. They had expected to find dual-task coordination as an additional EF factor in their modelling, however a five-factor model best explained their data (Fournier-Vicente et al., 2008). The possible addition of dual-task functions as an additional, potentially separable EF component is also supported by a review by Miyake and Freidman (2012).

Anderson (2002) also suggests that a unity-yet-diversity model explains executive function well. They discuss that EF in childhood develops across four key areas of flexibility, attentional control, goal setting and information processing. Anderson's model is similar to Miyake et al.'s model, in that these components are considered to be functionally separable, but at the same time, correlated with each other. The fact Anderson includes four components to EF does suggest that other alternate models are possible beyond the widely accepted three component model, and that other types of activity than just inhibition, working memory and switching are also related to EF.

Other tasks and areas of cognition could therefore be considered as being part of EF in this type of model, similar to other models of EF described above. In particular, real world examples of EF type cognitive control clearly requires more than just the three components of inhibition, working memory and switching – the need to identify appropriate actions and to plan an overarching approach to any real life task are also obviously important, and are not included in these types of model. Other factors than the three key components of EF that could be considered as being essential for EF may well exist, and including a wider variety of tasks in EF research might be able to identify these and the relationships they have to other EF components.

Current models take into account the idea of both unity and diversity of EFs (Miyake & Friedman, 2012). In confirmatory factor analysis research in particular, relationships between latent variables can be examined by considering the correlations between any identified factors. These have been found to be moderate in size (Miyake et al., 2000). The existence of significant correlations between factors suggests the factors are related to distinct, but not entirely independent, cognitive components. The presence of correlations greater than zero between EF factors further suggests that EF component processes rely on some common EF ability (unity of EFs); but the fact that components are not perfectly correlated indicates the dissociable nature of the factors, meaning that they are to some extent independent processes (diversity of EFs).

What is the shared variance between EF components in unity-yet-diversity models?

One critique of the differentiation or diversity of EF components hypothesis in EF is similar to one that can be applied to the discussion of intelligence in general. General fluid intelligence, or G, is defined as the shared variance among different cognitive measures. However the underlying substantive meaning of this is concept vague and not well-defined (K. Lee et al., 2013). Similarly, in an EF context, components of EF have shared variance – i.e. the latent EF variables identified in factor analysis research are commonly found to have significant correlations with each other. The meaning of this shared variance is not clear. Various ideas have been proposed in the literature.

The idea that common EF might reflect inhibition was proposed by Friedman and Miyake (2017). In this review paper, they described inhibition as underpinning the other two EF components, and rather than acting as a separate component at the same level of WM and switching. The idea that inhibition represents common EF processes has also been supported by the finding that inhibition processes are often necessary in order to successfully perform working memory tasks, so inhibition is thought to be underpinning other EF functions (Malagoli & Usai, 2015).

An alternative suggestion is that shared variance between EFs represents processing speed. Rose, Feldman and Jankowski (2011) found that children who had been born preterm and with low birthweight had cognitive deficits in EF tests, compared to their non-preterm peers, at age 11. They found that this deficit in EF performance was fully mediated by inclusion of processing speed as a covariate using structural equation modelling. This led to the conclusion that processing speed is a fundamental process underpinning EF performance, and that processing speed can be described as the underlying shared variance between EF components.

This type of unity-yet-diversity model is not incompatible with more neurally-based models, such as the SAS or multiple demand network theories of EF. Rather, these other theories can be used to explain aspects of the neural underpinning of EF task performance, while working in a unity-yet-diversity framework.

1.1.7 Model of EF used in this thesis

This thesis follows a unity-yet-diversity model of EFs, including inhibition, working memory and switching, as described by Miyake et al. (2000) and subsequently by Diamond (2013). The consensus in the literature currently is that a three-factor unity-yet-diversity model is probably the most plausible description of EF structure. However, the lack of a single, specific, operationalised definition of executive functioning remains an issue. For example, other elements of cognition may also be considered to be part of EF such as dual-task coordination (Baddeley, 1996b), or wider attentional processes (Norman & Shallice, 1986), but are not included in the Miyake et al. (2000) three-factor model.

It is important to consider both the unity and diversity of EF components when investigating EF structure, as the factors identified in these three-factor models are not entirely independent of each other. That is to say, most factor analysis research has found significant correlations between the latent components of EF. The three-factor conception of EF structure, although widely accepted in the literature, have not have been sufficiently replicated to draw absolute conclusions about the structure of EF in general. Furthermore, the structure identified in adults may not apply to other age groups, in particular children, adolescents, or older adults; and also may not apply to specific populations such as people with specific learning difficulties, Down's syndrome or other clinical populations. Of particular interest for this thesis, factor analysis research investigating the latent structure of EF during childhood and adolescence is reviewed in **Chapter 5**.

1.2 Components of EF

The fractionation of EFs into dissociable, but related, components is the generally accepted model in current EF research. As discussed above, factor analysis research characterises EFs as consisting of three inter-related, yet separable, component cognitive functions. This thesis uses a working definition of EF based on the unity-yet-diversity models of Miyake et al. (2000) and Diamond (2013), with EF conceptualised as three dissociable yet related components: inhibition, working memory and switching. These three factors will be discussed in turn below.

1.2.1 Inhibition

Inhibition or inhibitory control is the ability to prevent a dominant or prepotent response, where that response is not relevant or applicable to the current situation or task requirement (Diamond, 2013; Miyake et al., 2000). In a real-life example, a child could use inhibition to prevent his prepotent response to shout out an answer to a teacher's question straight away, and put his hand up and wait instead.

Inhibition can be conceptually subdivided into different types of inhibitory control. Response inhibition allows us to override automatic responses. The ability to inhibit automatic reactions allows us to carry out responses that are not usually required, for example, to look away from a presented stimulus rather than towards it in a Go/No Go task. Another form of inhibition would allow us to direct and focus attention on particular, selected external stimuli (e.g. tuning in to the voice of a particular person at a noisy cocktail party), and another would allow broader, and more long term, self-controlled behaviours (e.g. inhibiting a desire to eat unhealthily if we are trying to lose weight) (Diamond, 2013). These different aspects of inhibition are considered to reflect different specific cognitive components. One model of inhibition structure comes from Tiego et al. (2018), who proposed a hierarchical model of inhibitory control. Some aspects of inhibition, such as cognitive inhibition (ability to supress unwanted thoughts / memories) have been shown to be dissociable from other forms, such as inhibition of prepotent responses or avoiding external distractions (Friedman & Miyake, 2004). For a summary of some tasks that have been used to assess inhibition, see **Table 1.1.**

Miyake and Friedman (2012) have also suggested that inhibition might represent the 'common EF' – once covariance between EF components (the unity between EF factors) has been accounted for, there is no unique variance left for an inhibition-only factor to explain. This suggests that inhibition may be a common cognitive process underlying the other aspects of EFs, rather than a component of EF at the same 'level' as working memory or shifting. Further evidence supporting this view is that inhibitory control is required for WM performance, suggesting that inhibition may underpin other EF performance (Tiego et al., 2018).

1.2.2 Working memory or Updating

This is the ability to retain and manipulate information in mind over a short period of time, and to update what information is maintained as needed to address a given task (Diamond, 2013). Different researchers refer to this concept using different terms, for example updating (Miyake et al., 2000) or simply working memory (Purpura et al., 2017). Working memory is closely related to inhibition (Diamond, 2013). In order to complete tasks which require updating of working memory contents, it is necessary to inhibit automatic or prepotent responses to stimuli, and to inhibit other representations that might conflict with the required task information.

Tasks addressing the updating or working memory component generally involve some element of recalling information that is presented in a complex form, or monitoring information in a complex or continually altering situation. For example, the listening recall subtest of the AWMA (Automated Working Memory Assessment; Alloway et al., 2008) requires that participants recall the last words of a set of sentences, whilst also completing comprehension tasks (Purpura et al., 2017). See **Table 1.1** for other examples of WM tasks.

1.2.3 Switching

This is the ability to flexibly alter response patterns, perhaps in response to changing circumstances, or in order to try a new solution or strategy in problem-solving. This factor may be known as set-shifting, switching or cognitive flexibility (Diamond, 2013; Miyake et al., 2000; Purpura et al., 2017)

This kind of ability is essential in everyday life, for example, if while a student is writing an essay her phone rings, she may switch from one set of task requirements and behaviours to a new set in order to pay attention to the phone call and respond appropriately. Tasks used to address this aspect of EF have included plus-minus task and trail making tasks (e.g. Tamnes et al., 2010); and card sort tasks such as the Wisconsin Card Sort Task (WCST) (Miyake et al., 2000) and the dimensional change card sort (DCCS) task (Purpura et al., 2017). Tasks generally involve switching between different sets of rules during the task (Monsell, 2003), with the rules either explicitly provided (DCCS) or not (WCST). See **Table 1.1** for other examples of switching related tasks.

1.3 Assessment of EF

A wide variety of tasks have been used to assess EF in empirical studies. Nyongesa et al. (2019) reviewed the literature to identify measures that have been used to assess EF during adolescence. Ten commonly used tasks make up close to half (44%) of reported EF measures, and a wide variety of over 300 different tasks make up the remaining 66%. This illustrates that a very wide variety of tasks have been used in the literature to assess EF. Digit span, trail making, and continuous performance tasks which are used in this thesis are amongst the ten commonly used paradigms.

Table 1.1 describes some examples of commonly used paradigms used to assess EF in childhood and adulthood. The table includes the main component of EF that is assessed by each task paradigm, and gives some examples of EF studies that have used each task.

In addition to task measures of EF, questionnaires have been used to assess EF. One example of this type of questionnaire is The Behavioural Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000). This questionnaire assesses EF in an everyday context, considering behaviours such as initiation, planning, inhibition, shifting, working memory and monitoring.

The Strengths and Difficulties Questionnaire (SDQ) is a well-validated, normed behavioural questionnaire, used to assess mental health of children up to 17 years (A. Goodman & Goodman, 2009). Scores on the difficulties portion of the questionnaire correlate with likelihood of psychological diagnosis (R. Goodman et al., 2000). The SDQ measures five subscales of behaviour: emotional symptoms, conduct problems, hyperactivity, peer relationship problems, and prosocial behaviours. Research using the SDQ alongside measures of EF has shown that SDQ scores correlate with results obtained from the BRIEF (Monnier et al., 2014), with self-regulation ability (Lakes, 2013), and EF capabilities in adolescence (Donati et al., 2021). This may indicate that the SDQ may act a proxy measure for real-life EF capabilities, or that childhood EF is related to mental health. This idea is further discussed in **Chapter 5**.

Table 1.1 Examples of Tasks Used to Assess Executive Functions

Key EF Component	Task Name	Brief Description	Examples of studies using this task
Complex EF	Wisconsin Card Sort Task (WCST)	Sort cards according to sets of rules, which change over time. Rule changes must be inferred from successes and failures. Used with adults.	(Huizinga et al., 2006a)
Complex EF	Dimensional Change Card Sort (DCCS)	Sort cards according to sets of rules, which change over time. Rule changes are explicitly stated. Used with children.	(Purpura et al., 2017) (Rosen et al., 2020) (Alfonso & Lonigan, 2021)
Complex EF	Towers of London / Towers of Hanoi	Participants must move varying sized hoops into a prescribed pattern across three stands, following rules such as no larger item may be put on top of a smaller one.	(Huizinga et al., 2006a) (De Luca et al., 2003) (Lehto et al., 2003)
Inhibition	Stroop	Participants state the colour of ink words are presented in. The written word is to be ignored. Some items are incongruent colour words, which acts as a distractor.	(Tamnes et al., 2010) (Alfonso & Lonigan, 2021)
Inhibition	Modified Stroop (Day/Night)	Participants must say 'day' when presented with a moon, and 'night' when presented with a sun. Used with younger children.	(Purpura et al., 2017); (Diamond et al., 2002)
Inhibition	Go/no-go	Participants must make a certain response to a 'Go' stimulus. In a few trials, a 'No-go' stimulus will appear instead, and the usual response should not be made.	(Malagoli & Usai, 2015)
Inhibition	Flanker	Participants must respond with the direction of a target item that appears on a screen. Incongruent distractors surround the target in distractor trials.	(Eriksen & Eriksen, 1974; K. Lee et al., 2013)
Inhibition	Simon task	Items appear in congruent or incongruent locations. In incongruent trials, the features, but not the location of item, must be attended to.	(Davidson et al., 2006; K. Lee et al., 2013)

Key EF Component	Task Name	Brief Description	Examples of studies using this task
Inhibition	Simon says game	Verbal instructions are given by the experimenter, but should only be carried out if they say “Simon says” first.	(Rosen et al., 2020)
Switching	Intradimensional / Extradimensional Set Shifting task (ID/ED)	Participants must choose the correct stimuli according to current rules while avoiding distractors. The rules change over time, and changes to rules must be inferred from trial successes and failures.	(De Luca et al., 2003)
Switching	Trail making task	Participants follow a trail of letter items, in alphabetical order, like a dot-to-dot puzzle. A switching version has number and letter items, and participants must alternate between numbers and letters in sequence.	(Salthouse, 2011)
Working memory	Corsi block tapping task	Blocks are tapped in an apparently random order by the experimenter. Participants must reproduce the order.	(Corsi, 1972; De Luca et al., 2003; Kessels et al., 2008))
Working memory	Backward digit span	A sequence of numbers is read aloud to participants. Participants repeat the sequence aloud, in reverse order.	(Alfonso & Lonigan, 2021; Kessels et al., 2008; Rosen et al., 2020)
Working memory	Spatial working memory task	Participants search for an occluded target by selecting from identical boxes. When they open the correct box, the target will move to a new location. Continues until the target has appeared and been found in each location.	(De Luca et al., 2003)

1.3.1 Issues with EF Assessment

The assessment of EFs is notoriously complex (Chan et al., 2008a). By their nature as higher-level cognitive functions, tasks intended to address EFs are necessarily embedded in some other lower-level cognitive process (Diamond, 2013). For example, in the Stroop task (**Table 1.1**), abilities in cognitive processes such as visual perception, colour recognition, articulation speed (or rehearsal speed), reading ability and English language ability all introduce variance in performance. Scores in the task are therefore not a pure measure of inhibitory processes alone. Specific task paradigms can therefore result in significant variance between tasks ostensibly testing the same EF construct, perhaps due to individuals' varying skill in the embedding task type (K. Lee et al., 2013). Variation in EF task results can also be caused by cognitive strategy use (Chan et al., 2008a). One example of this is in the spatial working memory task (**Table 1.1**), where the working memory load can be reduced by using an efficient strategy such as starting a search in the same location when searching for the next box within a set of trials. This kind of tactical strategy use by some participants raises a broader question around the construct validity of EF tests in general, i.e. whether a given task actually tests the same cognitive construct across participants if different people employ different strategies for completion (Meredith, 1993). Another potential interpretation of the same idea, that tactics and strategy use influence EF task scores, is that use of an efficient strategy is itself an example of using a higher-level EF skill such as planning, and the test is therefore a valid assessment of this construct, as people who have better planning skills and therefore employ an efficient strategy will score better in the task.

The task impurity problem is a major issue in assessing EFs. Tasks designed to tap into a single putative EF factor may in fact require use of multiple EFs. A good example of this is the WCST, a commonly used assessment of adult EF (**Table 1.1**). EFs needed to succeed on a given trial include: inhibition of prior learning to avoid perseverating on old rule sets; updating working memory contents to recall features of cards which have been correctly and incorrectly placed when working out a new rule set; and shifting between behavioural strategies to successfully sort cards within rule sections, then shifting focus to work out new rules when the rule set changes. Confirmatory factor analysis (CFA) is a statistical technique which is used to minimize the effects of the task impurity problem. This technique looks for latent variables that represent shared variance between tasks, and can produce models of underlying functionality structures for complex cognitive components such as EFs. This will be discussed further in **Chapter 5.**

When assessing task performance, different scoring techniques can also change the interpretation of results. For example, in the Tower of London task (**Table 1.1**), when considering total number of errors as the outcome measure, performance continues to improve into adulthood. However, if reaction time is used as the outcome measure, adult-like levels of performance are seen by around 13 years (Huizinga et al., 2006a).

Ecological validity is also an issue in EF assessment, in that performance in tests in lab conditions may not predict scores of the individual in other tests, or importantly, in real-life situations (Chan et al., 2008a). Assessment of EF in a lab environment may not reflect EF deficiencies that may be present in a person's real life. The lab testing environment is controlled by an experimenter, tasks are structured, assessment time is quite limited, and task requirements are clearly defined by an external source. Real-world EF is more complex, with multiple, flexible, on-going and short-term goals and no pre-defined way of achieving them, and is therefore not directly assessable through lab-based tasks. Furthermore, construct validity of certain EF tests is ill defined. It is not clear which aspect of EF is being assessed in some tasks, or whether they are more general assessments of some umbrella EF, or indeed of cognitive functioning more generally than EF. The Wisconsin Card Sort Task (WCST) is an example of an EF task that has this construct ambiguity, despite its common use as a measure of executive functioning (Nyhus & Barceló, 2009).

Reliability of measures is also an issue in EF research. Some studies find good test-retest reliability in tasks such as anti-saccade and flanker tasks (Paap & Sawi, 2016). However, other studies report only moderate to poor test-retest reliability for a variety of response inhibition tasks, with significant task practice effects evident for some of these (Müller et al., 2012). Novelty is important in EF assessment (Diamond, 2013). Since novelty of the task is a requirement for many EF tests, results may be invalid if participants repeat the same tasks multiple times (Chan et al., 2008). In fact one view of EF is that it is the cognitive component which is recruited in novel situations where routine responses or actions are not sufficient. EF is generally considered to be most relevant when the task is novel, making assessments of for example test-retest reliability more complex (Diamond, 2013; Miyake et al., 2000). Another aspect to reliability of measures over time is that participants may develop strategies for tasks they have previously been exposed to. Strategy selection is known to influence measures of executive functioning (Chan et al., 2008a). Many published EF studies do not report psychometric properties (reliability and validity) of their tasks at all (Nyongesa et al., 2019).

A review of the literature found that over 300 different tasks have been used to assess EF in adolescents, and that even within similar tasks, the exact paradigms and task measures also vary widely (Nyongesa et al., 2019). Of considerable concern for the field of research, they found scant evidence of reliability or validity of even commonly used EF measures. Studies often did not report psychometric properties (including reliability or validity estimates) of EF measures. For complex tasks in particular, tasks often illustrated poor reliability where estimates were provided (Nyongesa et al., 2019). Previous research has also found poor test-retest reliability of EF measures (Miyake et al., 2000). It remains unclear what exactly EF measures are measuring (validity), and furthermore, whether they are consistently able to measure this (reliability).

1.4 Development of EF Behavioural Capabilities

Improvements in EF performance are observed throughout childhood and adolescence (Prencipe et al., 2011a). Broadly, there is a developmental shift from more general EF processes towards greater specialization and modularity of specific EF components (Hwang & Luna, 2013). The developmental period of EF is prolonged compared with that of other cognitive domains; adult-like levels of EF performance are not generally reached until late adolescence or young adulthood (Diamond et al., 2002). The protracted period of behavioural EF development is functionally correlated with the similarly protracted period of neural development of the PFC and other key regions of the cognitive control network, as compared with other brain regions (Diamond, 2013). Firstly, I will discuss the behavioural development of EF abilities, focussing on the period of adolescence. The neural underpinning of EF in adulthood, and relationships between neural and behavioural development of EF, will be discussed later in this section.

1.4.1 General EF development

Earliest EF capabilities are present by around 6 months old (Diamond, 2013). EFs show rapid development through pre-school – and school readiness can be indicated by the ability to focus on tasks and inhibit distractions (Fitzpatrick et al., 2014). Continued development in EFs occurs through childhood. Younger children show similar deficits in performance on complex EF tasks (such as Towers of London tasks, see **Table 1.1**) as patients with PFC damage. Compared with typical adults, children require more moves to complete Tower of London tasks, and tend to perseverate on old rules in card sort tasks (Huizinga et al., 2006a). In a developmental study, peak performance in Towers of London task has been found to occur between age 15-19, with younger participants performing significantly worse than participants of this age group (De Luca et al., 2003).

By young adulthood or late adolescence, most EFs have reached a peak of performance, with some task types maturing earlier than others (Diamond, 2013). Further to a general improvement in EF across childhood and adolescence, specific components of EF have been shown to have distinct developmental trajectories during this period (Huizinga et al., 2006a). The development during childhood and adolescence of the three EF components considered in this thesis, namely inhibition, working memory and switching ability, are discussed below. For more detail on prior work looking at developmental trajectories of the specific EF tasks used in the SCAMP battery, see also the Introduction to **Chapter 4.**

1.4.2 Development of Inhibition

Inhibitory control is the earliest EF component to develop. Inhibitory control can be observed in pre-school children, using tasks such as the Day/Night task to assess young children (Best et al., 2009). See **Table 1.1** for a brief description of this paradigm. Inhibition performance improves through childhood. Inhibition has been found to reach a plateau in performance at adult-like levels in early to mid adolescence, depending on the specific paradigm and measures of inhibitory control considered.

In one study, adult level performance in an anti-saccade inhibition task was seen by age 14 (Luna et al., 2004). In a latent variable analysis, performance in an underlying inhibition factor was found to improve with age across childhood, with peak performance reached by around age 11-12 (Brocki & Bohlin, 2004). This inhibition factor encompassed measures of a continuous performance task (CPT): CPT disinhibition errors, CPT impulsivity, CPT inattentive impulsivity, and also commission errors on a Go/No-go task. Developmental shifts were observed between their second (7.6-9.5) and third age groups (9.6-11.5), with no further improvement up to the oldest group (11.6-13.6). These findings are similar to another study, which found inhibition performance reached adult levels by around 12 years (Huizinga et al., 2006a). Previous research has shown variation in the point at which performance maturity is reached – it is possible this is due to variation in development of specific aspects of inhibitory control that are being tapped by the different specific tasks or task measures being used in different studies. Taken together, these findings indicate that inhibitory control mechanisms reach adult-like performance levels at some point during early adolescence, with development in performance reaching a plateau some time between age 11 and 14.

In younger children, inhibition appears to play a key role in EF in general – inhibition ability is strongly indicative of children's performance in other EF tasks (Isquith et al., 2004). The development of inhibition skills are also strongly related to WM and processing speed development in children and adolescents (Luna et al., 2004; Urben et al., 2011). The relationship between inhibition and other EFs is not as strong in later childhood or adulthood as it is in early childhood however, suggesting that inhibition is of particular importance in young children's EF abilities.

1.4.3 Development of Working Memory

The fundamentals of working memory are present by the age of six, and perhaps even earlier (Diamond, 2013). Research indicates that behavioural development of working memory ability occurs throughout childhood and adolescence, and perhaps extending into young adulthood for some aspects of working memory. There is some debate over exactly how long WM development continues. Some studies have found that WM reaches adult-like levels in mid-adolescence, by around age 15. For example, performance in a backward digit span task (as measured by span capacity) reaches adult-like performance around age 15 (Huizinga et al., 2006a). However, in a study using factor analysis with a variety of EF tasks, it was found that a working memory latent variable did not reach adult levels until around 19 years of age (Luna et al., 2004).

Another developmental latent variable study by Brocki and Bohlin (2004) investigated a smaller age range, and investigated whether WM ability develops linearly from age 6-13 years. Using latent variable analysis on a battery of eight EF tasks, with multiple task measures for some of these tasks, they found an underlying component of WM/fluency improved with age across the period of ages 6-13. This latent component comprised task measures of verbal fluency, hand movements, Digit Span Forward, Digit Span Backward, a Stroop-like task, and time reproduction. Although they found no significant linear effect of age on overall WM performance, they found that the youngest group aged 6-7.5 performed worse than the older groups. Little development in WM was observed between age 7.6 and 11.5. Then, the oldest group aged 11.6-13 years also performed better than the others in the working memory component. This analysis suggests that there might not be a linear effect of age across childhood and adolescence, rather, there might be particular periods in early childhood and again in early adolescence where improvements happen more rapidly than other times. Other studies have however found linear development of WM ability in childhood and adolescence: Performance in a Backward Digit Span (BDS) task improved linearly from age 8 through to 15 years (Prencipe et al., 2011a); and performance in a variety of working memory tasks develops linearly from six to 15-years (Gathercole et al., 2004). It is notable that these studies did not extend into late adolescence or early adulthood; research that has considered a more protracted developmental period have found that some measures of WM continue to improve into young adulthood.

Other studies have considered whether different aspects of WM performance might show different developmental trajectories, and might therefore have different ages at which peak or adult-like performance is reached. De Luca et al. (2003) looked at EF development across the lifespan, from age 8 to 64 years. They used various measures of Corsi and Spatial Working Memory (SWM) task performance in their research. They found overall working memory capacity peaked in their 15-19 year old age group, whereas a strategic planning measure of working memory peaked in their 20-29 year old age group. This indicates that specific aspects of WM may have separate developmental trajectories, with different periods of maturation, and that some aspects of WM development may extend into adulthood. Spatial Working Memory capacity appears to have reached a plateau by around age 13 (De Luca et al., 2003). Working memory accuracy improves from age 8-12, reaching adult levels by age 13-15, whereas reaction time continues to show improvement until young adulthood (age 18-25) (Crone et al., 2006). Functional gains have been observed in an efficiency measure of working memory, between the ages of 15 and 19 years, with further increases until 20–29 years of age (De Luca et al., 2003). Taken together, these findings suggest that working memory accuracy and capacity might reach adult-like levels by mid-adolescence, around age 13-15, where other aspects of WM such as strategic planning may continue to develop into young adulthood, and may not peak until after age 19-20.

1.4.4 Development of Switching

Three to four year olds are able to switch between two simple tasks, illustrating that switching ability is present in pre-schoolers (Diamond et al., 2002). The ability to switch between increasingly complex tasks develops with age, and reaches adult accuracy levels in early adolescence (Best et al., 2009). However, other work has found by age 8, switching or cognitive flexibility is already at adult-like levels (De Luca et al., 2003). Lee et al. (2013) found that switch-cost, as measured by the relative speed in congruent vs incongruent conditions on Simon and Flanker tasks (see **Table 1.1** for a brief description of these paradigms) improved with age across the whole range of time between 8 and 15 years of age. They also saw a levelling off in improvements in switch cost within the Simon task from age 10 onwards, with statistically significant though small decreases in switch-cost after this age. This suggests that developmental improvements in switch-cost may be levelling off by around the time children reach age 10, but are still not complete by age 15.

It has been noted that there appears to still be a speed-accuracy trade off present in early adolescence (Huizinga et al., 2006a). 13 and 15 year olds show similar levels of accuracy to older participants in switching tasks, however they slow down their performance relative to both younger and older groups, perhaps in order to focus and achieve a good level of accuracy (Huizinga et al., 2006a). This suggests that perhaps capabilities to perform specific aspects of switching tasks, or the ability to perform well in specific paradigms, might develop at different rates, as different tasks rely on different components of cognition in the embedding task. This could perhaps explain why different research has shown different results in terms of development of switching ability.

1.4.5 Development of Structural Models of EF Across Adolescence

Structural research indicates that during early childhood, executive functions tend to operate as a singular construct, with increasing modularity and separability of EF function as age increases across childhood and early adolescence. For more detail on development of EF latent structures across adolescence, see the Introduction to **Chapter 5.**

1.5 Neural Basis of Executive Function

EF relies on the frontal lobes, in particular the prefrontal cortex (PFC) (Stuss, 2011). Early evidence for the link between PFC and EF came from research into patients with frontal lobe lesions and work investigating the specific impacts of brain lesions on patients. When damage was located exclusively within in the frontal lobes, a specific collection of impairments was observed dubbed the "dysexecutive syndrome" (Norman & Shallice, 1986). These patients were selectively impaired in their performance of tasks involving higher level cognitive abilities; specifically, tasks involving planning, inhibition, working memory and switching between rule sets, while other cognitive abilities remained largely intact. Double dissociation was observed, in that patients with loci of damage outside the frontal lobes were often not impaired on higher level cognitive tasks, but were impaired on other types of tasks; where patients with damage located in frontal regions performed poorly in EF tasks but not necessarily in other types of cognitive tasks. Neuroimaging evidence has since also indicated a strong link between PFC and EF, but has also revealed that functional networks across many brain regions also underlie EF performance, and furthermore that PFC also has roles in other cognitive functions than EF (Stuss, 2011). PFC function and EF are therefore considered separate but related concepts.

One influential theory around the neural underpinnings of EF is that the frontal lobes follow a broadly hierarchical structure along the rostro-caudal axis; with increasingly complex or abstract cognitive control functions supported by more anterior regions. Various versions of this hypothesis have been expounded (as covered in a review by Badre, 2008). A common thread amongst these theories is that more abstract control functions are supported by more anterior brain regions. The PFC is the very font region of the frontal lobes. It is defined as the area of cortex which lies anterior to the supplementary motor area, and covers approximately the front one-third of the total frontal lobe cortical area (Best et al., 2009). The PFC itself comprises several different regions, including anterior cingulate cortex (ACC), Ventrolateral PFC (VL-PFC) and dorsolateral PFC (DL-PFC), and rostro-lateral PFC (RL-PFC), and orbito-frontal cortex (OFC) (Zelazo et al., 2008). A slightly more complex version of the rostro-caudal complexity hypothesis outlines that the 'apex of abstraction' lies not at very front of the brain, but slightly lateral to this, in the RL-PFC (Badre & D'Esposito, 2009).

The brain overall can be conceptualised as a collection of neural networks, each of which consist of structurally and functionally interconnected brain regions (Richmond et al., 2016). The PFC is heavily interconnected with other brain regions (Anderson & Reidy, 2012). Neuroimaging research implicates a complex EF functional network across the whole brain. That is to say, although areas of the PFC are active during EF related tasks, EF functionality does not reside entirely within the PFC

(Crone & Dahl, 2012). One conceptualisation of the role of the PFC in the EF network is that this region plays a coordinating role, activating and inhibiting activity in other posterior cortical and subcortical regions via its functional connections in order to achieve particular goals at any given moment (Best et al., 2009). This suggests that the PFC plays a specific and important role in EF, and PFC functionality is essential to enable good EF task performance.

Increasingly refined models of the neural underpinnings of EF suggest heterogeneity of PFC activation during different types of EF activity. Inhibitory control is associated with activity in the orbito-frontal cortex (the most anterior segment of PFC) and dorsolateral PFC (Davidson et al., 2006; (Huizinga et al., 2006a). Inhibition may also be particularly associated with activity in the right inferior frontal gyrus (Aron et al., 2014). Lateral PFC (both ventro-lateral and dorso-lateral PFC) is especially associated with working memory performance (Crone & Steinbeis, 2017). Switching tasks tend to recruit more medial PFC regions, alongside some activity in lateral PFC (Davidson et al., 2006; Huizinga et al., 2006).

Different components of EF are furthermore associated with differential patterns of activation across the EF related neural network of cortical and sub-cortical brain regions in adults (Crone & Steinbeis, 2017). This suggests some level of diversity between the neural networks that underpin different EF components. The EF network involves various cortical regions such as the inferior frontal junction, premotor cortex, pre-supplementary motor area and the anterior cingulate, and subcortical structures such as the insula and cerebellum (Davidson et al., 2006). EF task performance is associated with functional activity in overlapping regions, in particular with activity in fronto-striatal regions and PFC. These findings implicate some level of unity of EF in terms of neural structures that underpin cognitive components of EF (Best et al., 2009).

1.6 Neural Development Associated with EF in Adolescence

The neural underpinning of EF continues to develop through adolescence, supporting continued behavioural EF improvements. As discussed, there is significant evidence showing that EF skills continue to develop throughout adolescence. Neural structures and activity patterns associated with EF task performance in adulthood also continue to change and develop throughout adolescence; these structures have protracted developmental periods compared with regions supporting other cognitive activities such as language use (Diamond, 2013). A general hypothesis of structural and functional PFC development supporting EF development has been supported by a range of studies, discussed below, and more specific hypotheses regarding specific aspects of functional and structural development of PFC and other brain regions supporting specific aspects of EF development have also been proposed.

The beginning of adolescence is marked biologically, by the onset of puberty, where the end of the period is marked culturally, by the attainment of a functional or independent role in society (Dumontheil, 2016). Broadly adolescence begins at a point as early as age 10 or 11, though the starting points of puberty vary between individuals, and ends around age 18-20. Adolescence is a period of significant neural and behavioural and neural malleability (Mills, 2014). The brain undergoes significant changes both structurally and functionally during adolescence, underpinning the significant development of EF during this period.

Both structural and functional changes to brain regions supporting EF can be observed across the period of adolescence (Crone & Dahl, 2012;). In general, two complementary process occur in terms of broad functional brain changes during adolescence. Firstly, functionality of local cortical regions become more specific and modular; i.e. functional specialisation increases with age across adolescence. Secondly, neural networks undergo a strengthening of structural and functional connectivity, and gain increased functional integration (Hwang & Luna, 2013). That is to say, connections between brain regions across functional neural networks associated with particular cognitive functions become strengthened across adolescence.

Regions associated with EF have a protracted developmental time period compared with other neural regions. Historically, observations that PFC damage results in similar patterns of deficits in cognitive control as in young children led to a theory that PFC development might underpin EF development. A general model of PFC development supporting cognitive control development can be proposed on this basis – but we can also be more specific about specific aspects of PFC development that might underpin specific aspects of EF (Crone & Steinbeis, 2017).

In terms of structural development of the PFC, there are significant changes that occur through late childhood and early adolescence. White matter volume increases overall across the brain, including in PFC, and grey matter volumes in the PFC decrease. White matter increases reflect increasing brain volumes as the child grows physically. The decreases in grey matter during late childhood and adolescence may reflect the pruning of synaptic connections in the brain during this period, which may be a process which reflects the plasticity of the brain and allows the brain to be restructured in response to experiences (Zelazo et al., 2008). Engelhardt et al. (2019) suggest that the neural architecture of EF is established by middle childhood.

The pattern of functional brain activation when undertaking EF tasks also changes across adolescence. Research has suggested a broad brush change from more diffuse activity across the whole of PFC in earlier childhood, to more focussed activity in subregions underpinning specific kinds of EF activity as the child develops through adolescence (e.g. Bunge et al., 2002). Another broad finding is that increasingly anterior regions of PFC are recruited as children get older (Zelazo et al., 2008).

Developmental neuroimaging studies have also more specifically linked the developmental trajectories of individual EF processes with the maturation of certain sub-regions of PFC (Crone & Steinbeis, 2017). For example, increasingly anterior activity in the PFC is associated with a commensurate improvement in an inhibitory task performance (Lamm et al., 2006). The behavioural development of working memory performance co-occurs with the functional maturation of lateral PFC, with the better working memory performance in older children being associated with increased activity in the superior frontal and inter-parietal cortex (Klingberg et al., 2002).

Klingberg et al. (2002) used fMRI to assess regions of neural activation during completion of a Spatial WM task. They assessed 14 participants aged 9-18. WM capacity was greater in their older participants. In terms of neural function, older participants had greater activation of superior frontal and intraparietal cortex regions than the younger children. Combined, these findings suggest that the changes in patterns of neural activity observed across adolescence might underpin the commensurate developmental increases in WM capacity observed over this period.

Further complexity is added to the hypothesis that developmental changes in EF task performance is underpinned by underlying neural network changes by the finding that people of different ages might show similar behavioural performance levels of accuracy or capacity, but can recruit different brain regions while undertaking the tasks. Ciesielski et al. (2006) used fMRI to investigate regions of neural activation during a visual working memory n-back task involving categorisation of visual stimuli. They compared neural activity between children aged 6-10 and young adults aged 20-28 years. Although the 10-year-old children showed similar behavioural performance accuracy to the adults, the neural networks recruited by the children and adults were significantly different. Children showed activation largely in the dorsal visual stream (usually associated with planning actions related to visual inputs) and sensory-motor pathways while undertaking the task, where adults show more activity in areas associated with the ventral visual stream (usually associated with object recognition), including ventral prefrontal cortex and inferior temporal networks. This suggests that patterns of functional recruitment of brain regions during EF tasks might change during adolescence, or at least some time between age 10 and age 20. This change in functional patterns of activity is not necessarily associated with changing behavioural performance of EF tasks. Furthermore, there has been indication that extensive practice in WM tasks in childhood increases the functional recruitment of more adult-like regions in a small scale pilot functional neuroimaging study (Jolles et al., 2012). This suggests functional patterns of brain activity associated with EF might be altered with experience. This indicates the relationship between structure and function of neural regions underpinning EF is not a simple one, and may be experience related as well as being associated with biological maturation processes.

1.6.1 Adolescence as a sensitive period for EF development

It has been proposed that adolescence might mark a sensitive period in EF development (A. Thompson & Steinbeis, 2020). A sensitive period in this context is a particular temporal window during which neural systems underpinning particular cognitive functions undergo structural and functional alterations in response to experience. During the sensitive period, experience in the related cognitive area will significantly alter the development of the underpinning neural systems, where experience outside of the sensitive period will result in reduced levels of responsivity, though some change is still possible outside of the temporal window. As opposed to non-sensitive period learning, a sensitive period is experience-expectant rather than experience-dependent; is time limited by maturational stage rather than occurring across the lifespan; is associated with the formation of a developing system rather than with reorganisation of extant systems; and is associated with specific biological processes including synaptic pruning and remodelling, neural myelination, and changes in specific neurotransmitter levels such as GABA and dopamine. There is indicative evidence that early childhood marks one sensitive period in EF development, and the significant changes in both neural architecture and behavioural improvements in EF during adolescence suggest that adolescence may also be a critical sensitive period for EF development.

In general, the plasticity of the brain decreases with age, with decreased changes in neural structure and functional patterns occurring in response to experience over time. One possibility is that pubertal hormones act to induce a period of increased neural plasticity, within certain neural circuits and under certain conditions, during adolescence (Laube et al., 2020). This could be a mechanism by which a sensitive period for EF development may be induced during adolescence.

Evidence for the theory that adolescence marks a specific sensitive period in EF development is so far somewhat limited, however, the fact that maturation of neural systems associated with EF occurs concurrently with behavioural gains in EF during adolescence indicates the possibility that adolescence could be a sensitive period for EF development. Training studies indicate that functional gains in some domains of EF are higher in adolescence than at other times, however, the findings in this area have been mixed, and with few studies collecting measures indicating biological pubertal stage of participants, it is hard to draw concrete conclusions from the current evidence in this area (Laube et al., 2020). It is possible also that specific aspects of EF could have individual sensitive periods; this possibility is yet to be investigated widely (A. Thompson & Steinbeis, 2020). Research considering the interaction of pubertal hormones and neural plasticity indicates that this could be a mechanism by which neural plasticity in certain domains of functionality may increase during adolescence, relative to childhood and adulthood (Laube et al., 2020).

1.7 EF and General Intelligence

There has been some debate in the literature as to whether EFs are a separate component of cognitive functioning, or whether they are simply aspects of cognition associated with general intelligence. Early research with patients with frontal lobe lesions suggested that it is possible to have dissociation between EF and more general cognitive ability. That is to say, specific patients were observed to have deficits in performing complex, planning related tasks, but retained good performance on other more general intelligence measures (such as the Weschler Adult intelligence Scale (WAIS), an IQ proxy measure) (Friedman et al., 2006). This can be interpreted as indicating that EFs are a differentiated cognitive construct to general intelligence, as performance in EF tasks can be selectively impacted by brain lesions, while retaining other more general cognitive abilities.

Other research has suggested that the common elements of tests of executive function are closely related to general intelligence or Spearman's g (Duncan et al., 1997). Spearman's g was proposed as a unitary explanation for the positive manifold effect: the observation that scores obtained from a great variety of psychological tests all share positive correlations with each other. The 'positive manifold' suggests that all psychological tests, including putative EF tests, are positively correlated with each other. A general intelligence process was proposed to explain this, which suggests that all kinds of cognitive processing are associated with a single broad capability, known as general intelligence (Spearman, 1904).

Spearman's concept of g has been further dissected, with two key differentiated cognitive elements identified with factor research by for example Cattell. These two components are general fluid intelligence (Gf) and general crystallised intelligence (Gc) (Horn & Cattell, 1966). Gf reflects general higher level mental abilities, such as reasoning and planning, and is employed to solve problems and complete a wide range of cognitive tasks. The concept of Gf is as opposed to General crystallised intelligence (Gc), which reflects knowledge that has been acquired through experiences (J. Duncan, 2013). Gf ability is often assessed by tasks such as Cattell's Culture Fair Task (CFT) (Cattell & Cattell, 1960a). This task is putatively free of the influence of particular cultural or life experiences, in that it should not: a) rely on specific knowledge to succeed; or b) be 'trainable' through practice of similar tasks. There has been some debate in the literature around how 'culturally fair' the culture fair task in fact is (Nenty & Dinero, 1981). It has also been found to be subject to practice effects in a similar way to most other psychological tests. In terms of development, research has indicated that differentiation between Gf and Gc increases across childhood (Horn & Cattell, 1966).

Moving to the relationship between EFs and these concepts of general fluid and crystallised intelligence, one suggestion is that EF is simply an example of an intelligence function that is not really separate to Gf. Some studies have found that variance in performance in EF tests may be accountable by changes in fluid intelligence alone – indeed, for some EF tests such as WCST, once fluid intelligence is accounted for, observed differences between patients with frontal damage and controls is removed. Thus EF may not be dissociable from general fluid intelligence in the performance of complex EF tasks (Roca et al., 2010). However, the WCST is not a simple test of EF alone. It suffers strongly from the task impurity problem (outlined earlier in this chapter **Section 1.3**), as it relies heavily on multiple EF components of working memory, task switching and inhibition, but also more broadly on other cognitive functions such as reasoning, planning, logic, visual processing and processing speed.

In other cases, EFs have been shown to be separable from general fluid intelligence (Friedman et al., 2006). For example, other cognitive tests (other than the WCST) such as Hotel and Proverbs tests reveal deficits which are dissociable from fluid intelligence loss alone in patients with frontal lobe damage (Roca et al., 2010). In a young adult sample, Freidman et al. found that some, though not all EF components are related to Gf. Updating working memory capacity was significantly related to scores in WAIS, but inhibition and shifting were not, and therefore appear to be separate from Gf (Friedman et al., 2006). Rice (2017) suggested that perhaps EF is acting a catch-all term, that needs more refinement in the literature.

1.8 Background to the SCAMP Study

1.8.1 Mobile Phone Usage

Mobile phone use has become ubiquitous in society over the last 10 years. In 2016, 93% of adults in the UK said they owned or used a mobile phone. Smartphones are by far the most common type of mobile phone (Ofcom, 2016). Smartphones are used in a huge variety of ways (Deloitte, 2017), and since 2016 are the most common method used to access the internet (Office for National Statistics, 2016). Children's mobile phone ownership and access has also increased significantly since 2014 when the SCAMP study began data collection (Ofcom, 2019). The potential impacts of mobile phone use on cognitive, psychological and health outcomes remain unclear. The ubiquity of mobile phone and other tech use makes this issue highly relevant to society.

Table 1.2 Percentage of Children in the UK with Mobile Phone Access

	2014				2019			
Age in years	5-15	5-7	8-11	12-15	5-15	5-7	8-11	8-15
Child has own mobile phone (%)	41	4	32	78	48	5	43	85
Household has a mobile phone and child uses (%)	14	19	18	6	19	28	24	7
Total with mobile access (%)	55	23	50	84	67	33	67	92

Note: data from Ofcom reports (Ofcom, 2014) and (Ofcom, 2019)

1.8.2 What are the concerns with mobile phone usage?

Exposure to RF-EMF from mobile phones might have health or cognitive impacts

Mobile phones emit low power radio-frequency electromagnetic radiation (RF-EMF) while switched on and in use. The only known biological effect of exposure to RF-EMF emitted by mobile phones is that of slight cell and tissue heating (Ahlbom et al., 1998). There is some concern that this could have effects on the functioning of cells which are exposed to RF-EMF – however the mechanism by which this might occur is currently unspecified.

The WHO considers mobile phones to be “possibly carcinogenic” - meaning that it is theoretically possible that mobile phones could increase the risk of cancer. However, the WHO acknowledge that there is no evidence suggesting a causal relationship so far. Early studies concluded that mobile phones are not carcinogenic (Blettner & Berg, 2000). More long-term research is required to rule out this possibility (World Health Organization, 2010). Regarding other health outcomes, a recent study in Switzerland looking at the long-term effects of mobile phones showed that higher mobile phone use was associated with a small increase in the incidence of non-specific health symptoms, such as headache and general feelings of illness – they conclude that this is not likely to be caused by RF directly, but may relate to some other aspect of mobile phone use (Schoeni et al., 2017).

Studies have shown that RF exposure from mobile phones might have some short-term effects on cognition. A meta-analysis concluded that short-term exposure to RF radiation improved reaction times in attention tasks, but decreased accuracy in memory tasks (Barth et al., 2008). However, the size of the effect was small, and did not persist over time. Long-term studies indicate that neither mobile phone use nor RF exposure cause any lasting cognitive impairments (Roser et al., 2016).

Mental health effects

Much literature has reported a link between levels of screen time and metal health issues. The pattern of this association is not simple: both very high and very low amounts of mobile phone use are associated with poorer mental health in adolescents, with moderate users reporting the best mental health (Przybylski & Weinstein, 2017). It is important to note that this literature largely consists of cross-sectional studies, which cannot determine causality. Long-term studies are required to better identify causality. The Health Effects Related to Mobile phonE use in adolescentS (HERMES) study has found some association between mental health and screen time, particularly between poorer mental health and high levels of social media use (Schoeni et al., 2017). Furthermore, the type, rather than just the amount, of mobile phone use is important. Problematic mobile phone use, with features such as dependency ("I feel lost without the phone") and loss of control ("I have tried to spend less time with the phone but have been unable to do so") is more strongly associated with mental health issues than the amount of mobile phone use alone (Roser et al., 2016).

Cognitive training and learning effects

Neuroplasticity allows the brain to adapt in response to our experiences. Any repeated activity will induce some neuroplasticity, effectively training the brain to become quicker or more accurate at that activity. Mobile phone use could induce neuroplasticity, particularly if they are used for a frequently repeated activity over time. Research into possible neural changes specifically due to phones is scarce. Research into video gaming shows that gaming can improve visual accuracy and performance in decision making tasks (Boot et al., 2011). Mobile phones are often used alongside other activities, and therefore their effects are not so simple to identify as there are other environmental factors alongside their use.

Effects of increased screen time

It has been proposed that increased screen time might mean reduced time spent doing other, potentially more beneficial, activities. The evidence is not clear however. Mobile phones are often used alongside other activities (Deloitte, 2017), suggesting people do not necessarily 'displace' other activities with mobile phone use. Studies have shown that increased screen time of other types, in this case watching TV, is associated with poorer language development (Kostyrka-Allchorne et al., 2017). However, this is correlational evidence only – the causality may be reversed (the poor language skills mean children watch TV more) or the relationship may be due to some other factor (such as parental engagement) that influences both TV watching and language development.

Mobile phones also carry the effects of the type of usage

Mobile phone use is not an end in itself – mobile phone use encompasses a huge variety of activities (Deloitte, 2017). The use of mobile phones carries with it any attendant risks or benefits of the activity for which we use them. For example, mobile phones are used to access social media, which carries with it the risk of harassment or cyber-bullying (Strickland & Dent, 2017). Use of mobile phones to access video games could potentially have beneficial effects on certain aspects of cognition (Boot et al., 2011), but may also have other attendant risks.

1.8.3 Potential Cognitive Effects of Mobile Phone Use

Meta-analysis suggests that short-term, direct exposure to RF-EMF can affect cognition, with improved reaction times in an attention task and an increased errors in working memory tasks (Barth et al., 2008). In a study among college students, EF performance decreased when smartphones were removed from participants by experimenters (Hartanto & Yang, 2016). The authors suggest removing the phone caused an increase in anxiety resulting in observed decrease in task performance. This suggesting mobile phones may have more general cognitive effects beyond those associated with direct exposure to RF-EMF.

The Mobile Radiofrequency Phone Exposed Users' Study (MoRPhEUS) found faster, but less accurate, response patterns in EF tasks were associated with greater mobile phone use (Abramson et al., 2009). The HERMES (Health Effects Related to Mobile phonE use in adolescentS) prospective longitudinal study found decreased performance in verbal and figural memory tasks associated with greater mobile phone use (Schoeni, Roser, & Röösli, 2015). The HERMES study focussed on RF-EMF exposure alone, using around 400 Swiss teenagers as participants. Compared with previous similar studies, the SCAMP considers a wider range of mobile phone use measures, and also a wider range of cognitive, behavioural and health outcomes. The SCAMP study also has a greater number of participants, which should provide greater power to detect smaller effect sizes, and likely will allow more covariates to be accounted for in analyses.

1.8.4 Why study this age group? Why focus on EF?

Children and teenagers may be particularly susceptible to effects of mobile phone use. The brain develops and matures across childhood and adolescence, reaching full adult maturity around age 20 (Tiego et al., 2018). Children's skulls are thinner, meaning they might absorb more RF-EMF radiation if exposed to it. If there are effects of RF-EMF on the brain, it is more likely that children would be more affected than adults receiving the same dosage. The pattern of brain development is affected by the environment, via neuroplastic responses. By repeated exposure to particular stimuli will 'train' the brain to become quicker or more accurate at that repeated activity. As mobile phone use is very high in teenagers (**Table 1.2**), the type of activity they do may be at such a level as to train their brains to become better at those activities.

Both structural and functional changes to brain regions supporting EF occur across the period of adolescence (Dumontheil, 2016; Crone & Dahl, 2012). EFs are at a critical point of development during adolescence, when the brain regions supporting higher level thinking skills are maturing. The continued behavioural development of EF throughout adolescence, when mobile phone and technology use is high, makes it a likely target for a cognitive domain which may be affected by mobile phone use during this period.

1.8.5 Motivation for SCAMP and this thesis

The long-term effects of mobile phones are not fully understood. In addition to potential impacts of direct RF-EMF exposure, mobile phones also carry the effects of the activities they are used for, such as video gaming, online gambling, online harassment via social media, etc. Research has indicated that very high and very low levels of mobile phone use may be associated with poorer mental health during adolescence. Specific types of use, such as frequent use of social media, may have greater associations with mental health in adolescents. Mobile phone and other technology does not have clear-cut negative consequences for cognition: in fact, some types of use may be associated with specific cognitive benefits. SCAMP is investigating mobile phone use in teenagers, in a large-scale longitudinal sample, in order to try to identify potential impacts on health, wellbeing, academic, cognitive and / or behavioural outcomes. This thesis will provide a summary of the cognitive development of EF in the SCAMP cohort, to enable future research to explore associations between these EF measure developmental outcomes and other factors assessed by SCAMP, including mobile phone and / or video game use. By nature of their protracted developmental periods, EFs are likely to be more susceptible to environmental influence than other aspects of cognition (Lawson & Farah, 2017).

Work completed as part of this thesis has importance for the understanding of EF development during early adolescence in a general sense. Furthermore, individual task and composite measures of cognitive performance which have been created here will be useful in future studies using the SCAMP data. For example, studies considering the associations between bilingualism, SES and cognitive outcomes, and of access to green and blue space in the local environment have already been published using measures created during work on this thesis (Filippi et al., 2022; Maes et al., 2021). Future work will explore the relationship between mobile phone and other technology use and cognitive development in the SCAMP cohort. It is hoped that the work undertaken here will therefore be able to contribute to exploration of the impacts of technology use on cognitive development. Initially, it should help to identify in detail associations between socio-economic status and EF, to identify the developmental trajectories of EF abilities in adolescence, help to validate whether EF's underlying structure in adolescence changes over time, and whether the structure of EF in adolescence is similar to that in adulthood.

1.9 Thesis Structure

Chapter 1 explored the theoretical background to the empirical chapters, including the concept of EF, structure and development of EF, neural structures underpinning EF, and explored the motivation for SCAMP overall and the empirical work in this thesis.

Chapter 2 describes the methods used in the SCAMP data collection processes, task administration procedures, task measures for the cognitive and questionnaire data used in this thesis, and the data cleaning and processing techniques applied to the data.

Chapter 3 investigates associations between socio-economic status and executive functions in early adolescence, using MANCOVA and multiple regression analysis.

Chapter 4 uses multiple regressions and multilevel modelling to describe the developmental trajectories of executive function and general fluid intelligence, between our baseline and follow-up assessment points in early adolescence.

Chapter 5 uses factor analysis methods to investigate the structure of EF at both assessment points, exploring changes in EF structure during early adolescence.

Chapter 6 discusses the overall conclusions and limitations of the work included in this thesis, and relates findings to previous literature.

Chapter 2.
Methodology: The Study of Adolescents, Cognition and Mobile Phones (SCAMP)

This chapter describes the methodology for the Study of Adolescents, Cognition and Mobile Phones (SCAMP), from which the data for the following chapters are drawn. This chapter firstly gives a broad overview of the study and its aims. Participant characteristics and recruitment processes are described. There is a summary of the measures collected in the SCAMP study, both within the main assessment battery and the additional sub-studies. General data collection procedures are described, and specific details of administration of each cognitive task are provided. Task measure selection and data cleaning processes are summarised for each task. Finally, a table shows the number of data points for each task which are to be used in the subsequent chapters of this thesis.

2.1 Overview of SCAMP

SCAMP is a prospective longitudinal cohort study. Its key aim is to investigate whether use of mobile phones, other electronic devices, and exposure to radio-frequency electromagnetic waves (RF-EMF) are associated with differences in cognitive, behavioural, educational, and health outcomes during early adolescence. Pupils in 39 schools across the greater London area were assessed at baseline during school Years 7-8 (N = 6,680; age range = 9.62 - 15.41 years, age M = 12.07 years; SD = 0.47), and at follow-up during school Years 9-10 (N = 5,138; age range = 10.93 - 19.15 years; M age = 14.26 years; SD = 0.51). These age ranges include all reported figures at the time of testing, however some of these data will be excluded for analysis as the reported ages are not in the expected ranges for their school years. A total of N = 3,787 participants completed the assessments at both time points – this is only participants that were able to be matched based on their reported characteristics during testing, see **Chapter 2.7** for more detail on how the matching across time points was carried out.

Assessments consisted of a main assessment battery, and multiple optional enhancements. The main battery was a school-based, computerised assessment battery consisting of nine cognitive tasks and five questionnaire sections. The same tasks and questionnaires were included at both time points. Cognitive tasks assessed general fluid intelligence, executive functions, visuospatial processing, and speech processing. Questionnaires in the main battery covered topics including socio-economic status, languages spoken, mobile phone ownership and usage habits, gaming habits, leisure activities, home environment, and mental and physical health. In addition to the main assessment battery, participants and their parents were invited to complete various optional enhancements, which included additional online questionnaires (for parents and children), consent for data linkage to educational and health records. Participants from a sub-sample of schools were invited to provide biological samples (BioZone) or to allow monitoring of personal RF-EMF exposure. **Table 2.1** summarises the measures collected in the main battery and the various optional enhancements.

Table 2.1 Measures in the SCAMP main battery and optional enhancements

Area assessed	Specific Measures	Data Source: Main battery section	Data Source: Additional data
Cognition	Fluid intelligence (Cattell Culture Fair) Speech processing (Speech in Noise) Task-switching (Trail Making) Sustained attention / distractibility (Continuous Performance) Inhibition (Trail Making; Continuous Performance; Spatial Working Memory) Working memory (Corsi; Backward Digit Span; Spatial Working Memory) Subitisation range (Enumeration) Mental Rotation	Cognitive Tasks	
Technology use			
Mobile phone	Ownership / use of others' mobile phone Problematic mobile phone use behaviours Night-time and pre-sleep mobile phone use Type of mobile phone Age first used a mobile phone Usual location of mobile phone when carried and when in use	Questionnaire	Parent questionnaire; Child questionnaire
	Call frequency and duration Internet use frequency and duration Method of internet access (wifi / network) Messaging type and frequency	Questionnaire	Parent questionnaire; Child questionnaire; Network data
	Mobile phone policy School internet access (is Wifi present; pupil internet access policy) Laptop and computer use at school	School provided	
Cordless phone	Call frequency and duration Location of base station	Questionnaire	Parent questionnaire

Area assessed	Specific Measures	Data Source: Main battery section	Data Source: Additional data
Other device use	Frequency and duration of device use at home / school Laptop Desktop computer Tablet / E-book reader Smart TV Media player Games console / portable gaming device	Questionnaire	Parent questionnaire
Other technology use	Frequency and duration of use at home and school Email TV Internet Social media Music (headphones / speaker) Wifi at home	Questionnaire	Parent questionnaire
Video gaming	Frequency and duration of play Types of game played Who played with / alone Type of equipment used for gaming	Questionnaire	Parent questionnaire
Demographics	Ethnicity Religion Was English first language learned Which languages are spoken at home	Questionnaire	
Socio-economic status	School type (independent or state)	School-provided	
	Parental occupation Parental education Home postcode (for Carstairs estimates (O. Morgan & Baker, 2006))	Questionnaire	
Environmental factors	Smoking in the home Travel methods to school Living nearby busy road Green and blue space access and use	Questionnaire	Parent questionnaire

Area assessed	Specific Measures	Data Source: Main battery section	Data Source: Additional data
	Cooking, windows and ventilation at home Mould and damp at home		Parent questionnaire
	Noise exposure at home (indoor and outdoor)		Child questionnaire
Health and wellbeing Health-related quality of life	KIDSCREEN-10	Questionnaire	
Anxiety	Generalized Anxiety Disorder assessment (GAD-7; Spitzer et al., 2006)	Questionnaire (follow-up only)	
Depression	The Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001)	Questionnaire (follow-up only)	
Sleep	Length, latency, quality, disturbances	Questionnaire	
Health	Hearing and tinnitus Headaches	Questionnaire	
	Cyber-bullying	Questionnaire (follow-up only)	
	Disabilities, long-term illnesses or medical conditions Prescriptions, medications, therapy Previous physical trauma, major surgery		Parent questionnaire
	Puberty status Body image		Child questionnaire
School-related health and wellbeing	Learning disabilities Special educational needs (SEN) Attention deficit hyperactivity disorder Giftedness		Parent questionnaire
Behaviours	Self-efficacy Impulsivity: Domain-Specific Impulsivity Scale for Children (Tsukayama et al., 2013) Musical instruments		Child questionnaire
	Sport and physical activity	Questionnaire (follow-up only)	Child questionnaire
	Smoking, alcohol, cannabis consumption	Questionnaire	
	Diet	Questionnaire	Child questionnaire;

Area assessed	Specific Measures	Data Source: Main battery section	Data Source: Additional data
			Parent questionnaire
	Leisure activities		Parent questionnaire
Strengths and difficulties	Pro-social behaviour, emotional symptoms, conduct problems, hyperactivity or inattention, peer relationship problems: Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997)	Questionnaire	
RF-EMF exposure (estimates to be modelled using these measures)	Mobile phone use Other electronic device use Other technology use	Questionnaire	Parent questionnaire; Child questionnaire; Network data
	Other lifestyle and environmental factors	Questionnaire	
	Personal, home and school environment RF-EMF exposure levels (Exposimeter readings) Home and school pollution markers		Personal and environment monitoring
	Android phone data (X-MobiSense (Goedhart et al., 2015) Duration, frequency and type of calls, messaging, data use (Wi-Fi and network)		Android App ‘X-MobiSense’
Biological information	Saliva and urine samples (to assess pubertal status, stress hormones, DNA, pollution exposure chemical markers) Forced vital capacity, waist circumference, body mass index (BMI)		Bio-Zone
	Height, weight	Questionnaire	Bio-Zone
External data-linkage permissions	Educational records (including school exam results, KS2 and KS3 results, Cognitive Ability Test results, info from national pupil database, Special Educational Needs) Health records (including primary care data, Hospital Episode Statistics, patient care, birth records) Mobile phone usage (mobile network provider data)		Parental consent form

2.2 Main Assessment Battery

The main assessment battery included the participant questionnaires and the cognitive tasks. The assessment lasted around an hour and was carried out in schools in place of a normal lesson during the school day. Testing was conducted using Psytools software (Delosis, 2017), which schools pre-installed on their computers. At least one experimenter was present during each session (alongside the class teacher) in order to answer questions, encourage engagement, and resolve any technical issues that arose. Participants were allowed to ask the experimenter for clarification during the assessments - otherwise they were encouraged to work alone and in silence. The level of noise and other situational information (such as number of pupils present and type of lighting) was recorded for each session by the experimenter.

Assessments were conducted on computers in class groups, which enabled speedy testing of a large number of students simultaneously. Recent research shows that digital assessments of school age children (age 7-13) carried out in groups in a classroom environment provide reliable and valid estimates of children's cognitive performance, which correlate strongly with teacher assessments (Bignardi et al., 2020). Research has suggested that measurements of working memory obtained from classroom-administered tests may actually have greater ecological validity than those obtained from controlled individual testing, in that they better predict academic achievement, however, scores obtained from classroom settings may be lower than those obtained in individual settings (Bos & Weijer-Bergsma, 2020).

There are intrinsic constraints associated with collecting data in a classroom environment. The level of control is reduced compared with one-to-one testing, which is generally considered to be the 'gold-standard' for psychological testing. For example, some level of talking and interaction between students was practically inevitable, despite instructions to work in exam conditions. The order of testing was therefore staggered to discourage discussion of the tasks and questionnaires. The presence of an experimenter and a teacher in the testing rooms did allow some level of control over the battery administration, especially compared with remote or internet-based testing.

The level of noise during the assessments varied across schools and individual classrooms, in part due to the behaviour of the participants, but also due to the differing environmental noise levels around the schools. Acute effects of classroom noise on task performance are unclear, with negative impacts on cognitive task performance being found in some studies (Joseph et al., 2018), and no impacts found in others (Kanerva et al., 2019) – it is therefore possible that noise levels during testing may have had some influence on performance.

Technical issues arose for some participants, such as issues with sound production or headphone failure, which caused some delays and loss of data which might have been avoided if testing had taken place in a lab environment. These issues are to be weighed against the fact that large amounts of data were collected from a wide sample of participants, who were in a familiar environment. The use of computerised tasks meant there was no variability in task presentation, or task instructions. Issues relating to the presentation of specific cognitive tasks are discussed further in **Section 2.6**.

The tasks included in the battery were the same at both time points. It is expected that there will be improvements in the task scores between baseline and follow-up testing. There is evidence that some improvement in working memory test scores may occur due to practice effects. For example, Cacciamani et al. (2018) found practice effects with a period of 6 months between administrations in adult patients with mild cognitive impairments. Practice effects were observed with three-week period between testing in pre-school children (Müller et al., 2012). In adults, practice effects have been observed when re-testing after a delay of 12 months (Basso et al., 1999). EF abilities tend to have greater practice effects than tasks that rely on crystallised intelligence functions (Suchy et al., 2017). It is therefore possible that practice effects may account for at least part of any observed developmental changes in EF task score for those participants who complete both assessment points, and this should be considered in any analysis.

There will also likely be improvements in test scores due to neural and functional development during early adolescence. Most of the cognitive tasks assess some element of executive function – these tests were selected on the basis that we are likely to see changes in performance in early adolescence because of the prolonged development of executive functions across adolescence.

A fixed testing order was used. Key cognitive tasks and questionnaires were completed earlier in the testing order. This was done to ensure that key measures were completed by as many individuals as possible, and to maximise the chances of having a complete data set for at least a subset of the tasks (as opposed to having considerable missing data for all the measures). Since this is a longitudinal study, using the same fixed task order across both time points means we are more likely to have data for the same tasks at both time points, improving comparability of individuals' performance across time. The fixed order further ensures that differences between individuals' task performance are not due to task order effects. The task order was intended to improve efficiency of presentation (e.g. the CPT is the longest task, so this was placed last in the sequence). Although a fixed order risks conflating order effects (fatigue, boredom, etc.) with task performance, and could also result in potentially problematic structure to the missingness of data, it was thought that the apparent advantages were greater than these disadvantages. The inclusion of additional questionnaires and cognitive tasks at the end of the assessment battery also aimed to keep participants who completed the assessment quickly engaged on tasks rather than risking them distracting their peers once they had completed the key measures.

Cognitive tasks were interspersed with questionnaire sections. This was done to minimise boredom from doing a long set of questionnaires in one go, and to reduce cognitive fatigue from completing multiple cognitive tasks in succession. Two staggered presentation sequences were used to slightly vary the stage that participants were at within a session, in order to discourage 'copying' or other group behaviour effects. Participants were randomly allocated to order A or B. At follow-up, *n* = 2,840 previous participants who were recognised by the automated system were assigned to the same order as they completed at baseline. The original presentation orders are shown in the first two columns of **Table 2.2**. During baseline testing, some participants had difficulty in accessing the Speech-in-Noise (SPIN) task due to audio problems, which caused some delays while these issues were addressed. This task was therefore moved to later in the testing sequence to minimise the impact of this issue, while still retaining this important assessment of temporal cortex function at a relatively early point in the battery. The last two columns of **Table 2.2** show these adjusted presentation orders. At baseline, *n* = 3,852 participants used the original task orders and *n* = 2,820 used the adjusted orders. All follow-up testing used the adjusted orders.

Table 2.2 Original and adjusted presentation orders of the SCAMP assessment battery

Original Order A	Original Order B	Adjusted Order A	Adjusted Order B
1. Trail making task	1. Trail making task	1. Trail making task	1. Trail making task
2. Backward digit span	Questionnaire A	2. Backward digit span	Questionnaire A
Questionnaire A	2. Backward digit span	Questionnaire A	2. Backward digit span
3. Spatial working memory	Questionnaire B	3. Spatial working memory	Questionnaire B
Questionnaire B	3. Spatial working memory	Questionnaire B	3. Spatial working memory
4. Speech in noise	Questionnaire C	4. Enumeration	Questionnaire C
Questionnaire C	4. Speech in noise	Questionnaire C	4. Enumeration
5. Cattell's culture fair task	Questionnaire D	5. Cattell's culture fair task	Questionnaire D
Questionnaire D	5. Cattell's culture fair task	Questionnaire D	5. Cattell's culture fair task
6. Enumeration	Questionnaire E	6. Speech in noise	Questionnaire E
Questionnaire E	6. Enumeration	Questionnaire E	6. Speech in noise
7. Corsi	7. Corsi	7. Corsi	7. Corsi
8. Mental rotation	8. Mental rotation	8. Mental rotation	8. Mental rotation
9. Continuous performance task	9. Continuous performance task	9. Continuous performance task	9. Continuous performance task

2.2.1 Additional data and add-on sub-studies

Additional measures regarding school characteristics and policies were collected from schools and teachers. Students and parents were invited to complete additional online questionnaires. Permission to allow data-linkage to health records, educational records and data from mobile phone providers was sought via consent forms. Participants were also invited to download a mobile phone app to track their mobile phone usage, including duration, frequency and type of calls, messaging and internet use. The app, XMobiSense (Goedhart et al., 2015), was available for Android phones only. Twelve schools participated in the Bio-Zone sub-study, where biological samples of saliva and urine and measurements of height, weight and other health indicators were collected from consenting participants. These participants were also invited to complete a further environmental monitoring sub-study, where measures of pollution and RF-EMF levels in and around their home environment were collected, and they carried a personal RF-EMF exposimeter for a 24 hour period. **Table 2.1** has a summary of all the measures collected in SCAMP so far; for more details, see Toledano et al. (2018).

2.3 Participants

The SCAMP cohort consists of pupils attending 39 schools across Greater London and the surrounding area. Schools were contacted if they met the following eligibility criteria: located in the Greater London metropolitan area; a Year 7 headcount of over 200 in state schools or over 50 in independent schools; not primary, infant, junior or middle schools; not pupil referral units or secure units. Of the 206 eligible schools, 35 agreed to participate. A further 8 schools (mostly from just outside the Greater London area) contacted the research team independently to participate. Four schools dropped out prior to the initial assessments taking place. This resulted in a final N = 39 participating schools at baseline (Toledano et al., 2018). Parents of pupils in these schools were then contacted via the school with information packs about the study and consent forms. Participation was on an opt-out basis, meaning that pupils or their parents could choose not to participate at any point. If they chose not to participate in the school-based assessment, pupils were given something to do by their teacher. Of the N = 7,375 Year 7 pupils registered at the schools, N = 6,680 pupils completed baseline assessment between November 2014 and July 2016, and N = 5,138 pupils completed follow-up between September 2016 and August 2018. Parental opt-outs, absentees on the day of assessment, technical issues, withdrawals and non-assents by the participants account for the drop-outs. Sample demographics were largely representative of the target area population (Toledano et al., 2018). Numbers and socio-demographic characteristics of participants whose data are presented in this thesis are described in **Table 2.3.** Exclusions are described in **Section 2.7.**

Table 2.3 Socio-demographic information for SCAMP participants at baseline and follow-up

	Baseline		Follow-up	
Final *N* (after exclusions on age)	*N* = 6,591		*N* = 5,116	
	M	***SD***	***M***	***SD***
Age [a]	12.05	0.48	14.62	0.52
Sex	***N***	***%***	***N***	***%***
Male	3132	47.52	2339	45.72
Female	3459	52.48	2777	54.28
Ethnicity	***N***	***%***	***N***	***%***
White (British, Irish, Other)	2658	44.46	2119	46.00
Black (Caribbean, African, Other)	969	16.21	695	15.09
Asian (Indian, Pakistani, Bangladeshi, Chinese, Other)	1667	27.89	1299	28.20
Any Mixed Race	2	0.03	26	0.56
Other	682	11.41	468	10.16
Socio-Economic Status Measures				
School Type	***N***	***%***	***N***	***%***
Independent	1472	22.32	1283	26.09
State	5122	77.68	3635	73.91
Father Education	***N***	***%***	***N***	***%***
Attended University	2750	68.92	2805	41.89
Did not attend University	1240	31.08	2113	58.11
Father Occupation [b]	***N***	***%***	***N***	***%***
1 Routine occupations	307	6.21	171	4.49
2 Semi-routine occupations	466	9.43	273	7.17
3 Lower supervisory / technical	388	7.85	272	7.14
4 Small employer / own account worker	1347	27.25	1042	27.35
5 Intermediate occupations	315	6.37	160	4.20
6 Lower managerial / professional	613	12.40	551	14.46
7 Higher professional	1064	21.52	926	24.30
8 Large employer / higher managerial	444	8.98	382	10.03
0 Never worked / Long term unemployed	-	-	33	0.87
Mother Education	***N***	***%***	***N***	***%***
Attended University	2642	63.51	2899	58.95
Did not attend University	1518	36.49	2019	41.05
Mother Occupation [b]	***N***	***%***	***N***	***%***
1 Routine occupations	245	5.74	117	3.27
2 Semi-routine occupations	827	19.36	394	11.01
3 Lower supervisory / technical	110	2.58	78	2.18
4 Small employer / own account worker	194	4.54	272	7.60
5 Intermediate occupations	535	12.53	352	9.84
6 Lower managerial / professional	1152	26.97	807	22.56
7 Higher professional	985	23.06	943	26.36
8 Large employer / higher managerial	223	5.22	208	5.81
0 Never worked / Long term unemployed	-	-	406	11.35

[a] Age (in years) is at time of completion of the computerised assessment battery at each time point.

[b] Occupations classified according to ONS NSSEC-8 categories; these values are re-coded such that higher values indicate higher socio-economic status. At baseline category '0' were classified as missing data as numbers were considered very low.

2.4 Ethics and Consent

The North West - Haydock Research Ethics Committee approved the original SCAMP study protocol and subsequent amendments (ref 14/NW/0347). The study is conducted in accordance with the Declaration of Helsinki (1964 and later revisions). Ethical approval for secondary data analysis carried out in this thesis was provided by the Birkbeck Department of Psychological Sciences Ethics Committee. Consent was obtained on an opt-out basis: schools provided initial consent for testing to take place during school time, then participants and parents were informed about the study via letters and information packs sent home from schools. They were told that they could withdraw from the study at any point. Consent forms were included in the information pack, along with requests to allow data-linkage to external data sources (health and educational records; mobile phone network data); to complete additional online parent and child questionnaires; and to participate in optional study extensions (see **Table 2.1**).

2.5 Questionnaires

Questionnaires in the main assessment battery were presented in five blocks, interspersed with the cognitive tasks. A fixed presentation order was used to ensure the most important topics were more likely to be completed by placing these at the beginning of the assessment. Around 250 questions were included in total. Block A covered personal details (name, address, date of birth), mobile phone ownership and use, and languages spoken. Block B covered other device use and video gaming. Block C covered internet access, social media, sleep (duration, quality, interruptions), habits around listening to music, hearing issues, and mother's and father's education and occupation. Block D covered medical conditions, mental and physical health and wellbeing (Kidscreen-10, SDQ, PHQ-9, GAD-7) and experiences of bullying. Block E covered experience of green and blue space, outdoor activities, religion, height and weight, headaches, eating and caffeine consumption, and alcohol, cigarette and cannabis use. The questionnaires measured key predictors (e.g. mobile phone use patterns), outcomes (e.g. mental and physical health), and important co-variates (e.g. socio-economic status). Measures collected in the questionnaires are summarised in **Table 2.1**.

Chapter 3 investigates the association of SES and EF, using SES measures collected from the questionnaire. A local deprivation measure (Carstairs index; Morgan & Baker, 2006) is calculated for the participant's home postcode reported in Block A. Parental occupations and education levels are assessed in Block C. More details about these measures can be found in **Chapter 3**.

Chapter 5 makes use of responses to the 11-17 self-report version of the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). This was presented in questionnaire Block D. The SDQ assesses four domains of difficulties: emotional symptoms, conduct problems, hyperactivity problems and peer problems; and one strength domain: prosocial behaviours. There are five questions for each domain with twenty-five questions in total. The complete SDQ items can be found in **Appendix A.** Items are presented as statements such as "I am easily distracted" or "I find it difficult to concentrate". Participants rate how much each statement is true for them, with 0 = Not True, 1 = Somewhat True and 2 = Certainly True. The key measures used are sum scores in each of the five subscale domains. Reverse-scored items within the subscales were recoded. Each subscale therefore has a minimum score of zero, and a maximum of 10.

2.6 Cognitive Tasks

The cognitive tasks in the SCAMP battery assessed aspects of executive function, fluid intelligence, sustained attention, speech processing, and visuospatial perception. These areas were chosen as they are (i) likely to display development during early adolescence; (ii) likely to show individual differences; and (iii) likely areas of cognition whose development could be affected by RF-EMF exposure, or mobile device use. **Table 2.4** summarises the tasks used and their cognitive domains.

Table 2.4 Cognitive tasks in the SCAMP battery

Task	Cognitive Domain	Grouping for Analyses
Trail making	Cognitive flexibility; inhibition	Executive function
Backward digit span	Working memory	Executive function
Spatial working memory	Working memory; inhibition	Executive function
Corsi	Working memory	Executive function
Cattell's culture fair	Non-verbal fluid intelligence	Stand-alone task
Continuous performance	Sustained attention; distractibility	Stand-alone task
Speech in noise [a]	Speech and language processing	Stand-alone task
Enumeration [a]	Visual attention	Visuospatial perception
Mental rotation [a]	Visuospatial rotation ability	Visuospatial perception

[a] Task not discussed in this thesis

2.6.1 Trail Making Task (TMT)

A Measure of Executive Function: cognitive flexibility and inhibition components

This task is a computerised version of the TMT used in neuropsychological studies (e.g. Tombaugh, 2004), which was originally published in 1944 (Army Individual Test Battery, 1944). The task is traditionally presented in a pen-and-paper format with two parts: in part A, participants follow a trail of labelled circles alphabetically; and in part B, participants follow a trail of alternating numbers and letters. Participants are instructed to complete the trail as quickly as possible without lifting their pen from the paper. In our computerised version, participants instead clicked the stimuli in sequence. Further details of the presentation method are included in the Procedure section below.

Cognitive Components Required for Task Performance

Successful performance in the TMT requires many cognitive processes, including visual attention, scanning, speed of processing, working memory, switching, and inhibition of incorrect responses (Sánchez-Cubillo et al., 2009; Tombaugh, 2004). Various measures of performance may be obtained from the task, each of which reflect different components of cognition. Direct measures of performance in parts A and B include total time to complete each sub-task; or number of incorrect responses. If a digital presentation method is used, a wider variety of performance metrics may be obtained, such as number of time the pen is lifted or number and duration of pauses (Dahmen et al., 2017; Salthouse, 2011). Research comparing results from the TMT with other cognitive tests suggests that performance in part A is related to global processing speed, motor speed and visual perception ability, while part B is more closely related to inhibitory control, working memory, visuospatial sequencing ability, and executive function (Allen et al., 2012; Fellows et al., 2017). Measures of difference in completion time between the two parts is thought to reflect a higher level EF component than either part alone, as the difference in speed between the two conditions at least in part accounts for differences in processing and motor speed coordination. Thus measures that combine performance in the two parts of the task, such as difference or ratio measures, are thought to more closely reflect set-shifting or switching ability (N. R. Lee et al., 2014).

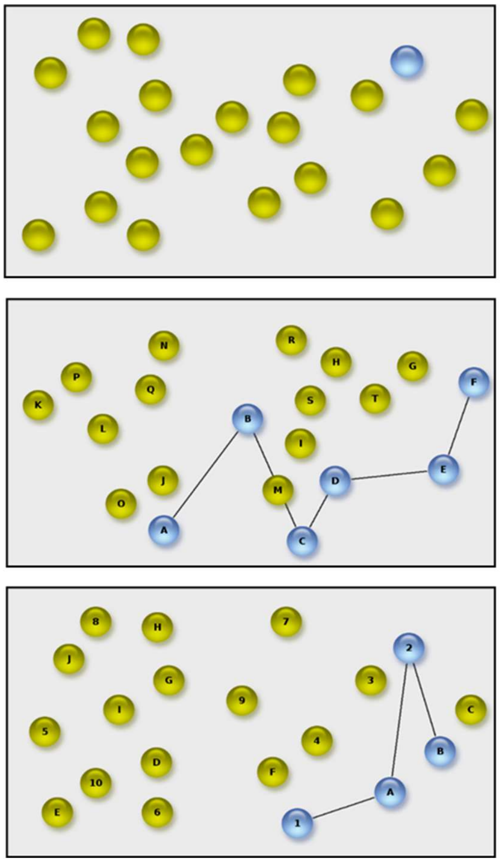
Development and Associations with Demographic Characteristics

Sarsour et al. (2011) used TMT with age 8-12 year old children, showing that the task is appropriate for use in this age group in addition to in adulthood. However they found no significant association with age within this age group. Other research has shown that TMT performance is associated with age, gender, education level and general intelligence (Kowalczyk et al., 2001). It might therefore be useful to take these factors into when interpreting TMT performance. Development of switching ability and TMT task performance is further discussed in **Chapter 4.**

Procedure

This task was placed first in the battery as it provides a range of useful measures of processing speed, visuospatial processing, motor sequencing and a well-validated measure of executive function. In addition, the task instructions are clear and the task is quite entertaining and varied, making this task an engaging start of the assessment. Three sub-tasks were included: the Dots condition, the Letters condition and the Number and Letters or Switching condition. These are illustrated in **Figure 2.1.**

Figure 2.1 Illustration of the three TMT sub-tasks. Each condition has twenty items to complete.



1. Dots condition

Participants click the blue dots as they light up in sequence

2. Letters condition

Participants click labelled dots in alphabetical order

3. Numbers and Letters condition

Participants click labelled dots in alternating numerical and alphabetical order (1-A-2-B...)

Each sub-task began with a short video demonstration alongside written instructions. Then the participant completed a short practice with six dots. Task instructions were repeated on screen if an incorrect location was clicked during practice. In the main trials, text feedback was given to highlight the correct starting point if they clicked on the wrong location initially; then an aversive sound was made if they made a later incorrect response. Throughout the entire task, if the participant made no response for 20 seconds a warning was displayed, advising them that the task will be abandoned in a further 10 seconds if they do not make a response. Then if there was still no response the sub-task ended and the next condition began.

Participants completed three dot-to-dot sub-tasks. Locations the participant clicked correctly were connected by lines (forming the “trail”). The first sub-task participants completed is not part of the standard TMT. In this Dots condition, one dot appears highlighted in blue. Participants click on the highlighted dot; then a new dot is highlighted to be clicked; this repeats until all dots are in the completed trail (**Figure 2.1**). In the Letters condition, equivalent to the standard TMT Part A, participants clicked on labelled dots in alphabetical order (**Figure 2.1**). In the Number and Letters condition, equivalent to the TMT Part B, participants clicked on labelled dots in alternating numerical and alphabetical order: 1-A-2-B… (**Figure 2.1**). There were twenty dots to complete in each sub-task. Dot locations were pseudo-randomly distributed across the screen, with each sub-task having different dot locations. All participants saw the same location distribution within a sub-task.

We used a computerised version of the TMT task. Scores from other digital versions of the TMT have been shown to be strongly correlated with results from paper-and-pen versions. Fellows et al., (2017) found correlations of R=.53 for TMT-A and .80 for TMT-B total times; p's < .001. Digital and paper-and-pen versions of the TMT have been shown to share similar patterns of association with other cognitive test results (Dahmen et al., 2017). However, this previous research used a tablet and stylus presentation, which might be considered more similar to paper-and-pen than the mouse-click version used in the SCAMP battery.

Task Measures

Various measures of task performance were recorded automatically: time to complete parts A and B of the task, response time for each individual mouse click (for both correct and incorrect responses), the number of clicks made to incorrect dots, and the number of clicks elsewhere in the screen (non-dot clicks). TMT performance is usually estimated with some measure of time to complete parts A and / or B, or occasionally the number of errors made in each part.

The key difference in terms of cognition between parts A and B is that part B requires switching between two modalities. In order to estimate EF ability, specifically of the switching component of EF, it is therefore useful to take a relative measure comparing performance in parts A and B. Relative measures are 'purer' measures of EF performance, specifically of switching or cognitive flexibility, than direct measures of performance in part B alone (Sánchez-Cubillo et al., 2009). Various relative measures may be calculated: a simple difference score (B-A); a ratio (B:A); a proportion (B-A)/A)); or the residual of B after B has been regressed on A (BrA) (Fellows et al., 2017). Analysis comparing results from relative measures of TMT with other cognitive tests indicate that simple difference (B-A) associates with overall processing speed, which is similar to a direct measure of time to complete part A (Salthouse, 2011). B:A ratio measures reduce the influence of other demands, such as psycho-motor skill or general processing speed, and better captures cognitive flexibility and executive function than direct part B measures (Kowalczyk et al., 2001; Salthouse, 2011; Sánchez-Cubillo et al., 2009). B:A ratio measures and BrA residual measures have similar patterns of association with other cognitive performance; they are most closely associated with general fluid intelligence and not significantly associated with memory, speed or vocabulary measures (Salthouse, 2011). The BrA residual measure has lower variance than B:A ratio measure, and is therefore preferable if sample sizes are sufficient (Salthouse, 2011). A version of a BrA residual measure was also used by Lee et al.

(2013) as it is thought to reflect a purer measure of task switching ability than other measures of TMT performance.

Here, we calculated the unstandardized residual BrA as the first key EF measure of the TMT, this is used in **Chapter 3** in the analyses of the associations of cognition with SES. This measure was calculated by running a regression with time to complete part A as predictor, and time to complete part B as outcome, then saving out the unstandardized residuals. Higher scores indicate greater cost of switching (i.e. higher score = relatively poorer switching performance).

However, as the residuals are centred on zero and are calculated across the whole sample of participants, it makes it impossible to estimate any absolute improvement in performance by a given individual across multiple time points using this measure. Rather, any increase in score between two time points would reflect a relative improvement in switching ability compared to the other participants in the sample. Given that in our study the samples at the two time points are not identical (as many participants are missing data from one or other time point), improvements in a BrA residual score would not be easily interpretable. Therefore we used a proportion score (B-A)/A in the chapters (**4 & 5**) where we considered scores across both time points.

A proportion measure is the second key measure of TMT we calculated, this is used in **Chapters 4 and 5** in the developmental trajectory and EF structural analyses. This was calculated by working out the difference between time to complete parts A and B, and dividing the this by time to complete part A, i.e. (B-A)/A. Higher scores indicate poorer switching performance.

Data Cleaning

A sub-set of N = 358 participants completed the letters and switching conditions twice, with a different array to the first iteration. Because of time constraints, the task was changed early in baseline testing to include only one array of each sub-task. Analyses in this thesis include only the first array attempted for each sub-task for all participants.

Participants were excluded if they made an excessive number of clicks away from the dot stimuli elsewhere on the screen, i.e. non-dot clicks. A few of these could be considered as normal as the test is administered using mouse clicks, for example, if a participant is hurrying they might just miss the intended stimulus with the mouse. However, very high numbers of non-dot clicks would indicate that the participant was just clicking randomly and not attempting to do the task properly. Participants were excluded if they made over 20 non-dot clicks in any single sub-task. This is equivalent to the number of stimuli on each screen, so they made a total of more than one error for every stimulus item. Participants were also excluded if they did not complete all three sub-tasks, such as if they failed to make a response for over 30 seconds and the task timed out, or any other reason such as quitting the whole battery part way through. Numbers of participants excluded on these criteria and final good N's are found in **Table 2.5** at the end of this chapter**.**

2.6.2 Backward Digit Span (BDS)

A Measure of Verbal Working Memory

The BDS is a widely used and well-validated measure of verbal working memory (Richardson, 2007). Backward and forward digit span tests are among the oldest and most widely used assessments of working memory, and are commonly used in neuropsychological assessment batteries, such as the Automated Working Memory Assessment (AWMA) (Alloway et al., 2008). In the traditional presentation of this task, a sequence of numbers is read aloud to a participant, who then repeats the sequence back in reverse order. The BDS task relies on working memory, as it requires participants to manipulate information held in short-term memory, and also other cognitive processes including attention, auditory encoding and auditory processing. Overall, BDS is considered a measure of working memory (Sarsour et al., 2011).

Development

BDS capacity shows small but statistically significant development during adolescence (Prencipe et al., 2011a). Although performance peaks in adulthood, during the 30s, relatively small differences in overall backward digit span capacity are observed between adolescents and adults (Alloway & Alloway, 2013). Adult participants have a maximal backward digit span of between 3 and 8 items, with an average capacity between 5 and 6 (Gregoire & Van Der Linden, 1997). Sarsour et al. (2011) used Digit span (forward and backward) with age 8-12 year olds, and found no significant association with age in this group with the total number of items correct, suggesting that development of this particular task measure within this early portion of adolescence is quite limited.

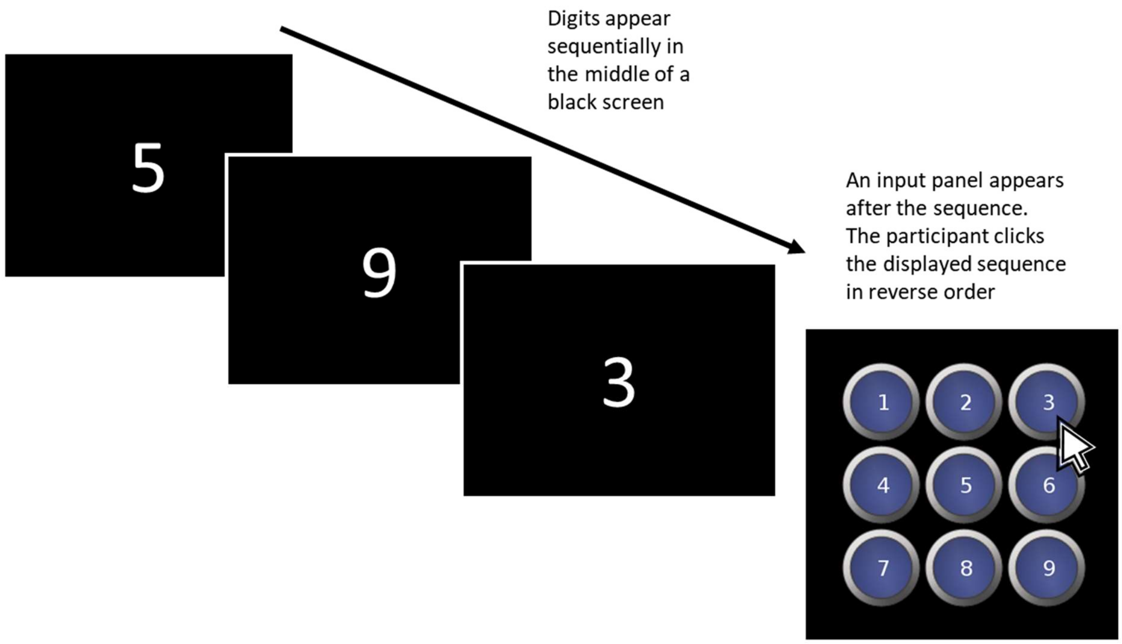
Procedure

In the standard BDS presentation, trials are presented in blocks of increasing difficulty, with two or more trials at each level of difficulty. A certain number of correct responses is required in order to pass each level. Working memory span may then be estimated by taking the number of items in the longest sequence with a correct response, or the total number of trials with correct responses (Kessels et al., 2008). However, a 'total correct' score may not accurately reflect participants' working memory capacity; rather it conflates maximal span length with consistency in performance across the task. For example, a participant who fails one trial each of 4, 5, 6 and 7 items and both trials of 8 length would have the same 'total correct' score as someone who passes all trials up to level 5 then fails both trials of 6 items. An alternative measure is to calculate a product score which takes into account both maximum span length and number of trials correct (i.e. total correct multiplied by maximum correct trial length) (Kessels et al., 2008). This measure may better reflect overall performance on the task than either contributing measure alone. A further issue with estimating BDS performance is that differing test schedules (e.g. more repeats at each level, or different numbers of correct responses required to pass a level) are used in research, making task performance difficult to compare directly across studies.

The standard sequential procedures with multiple trials at each length are quite time consuming and inefficient, especially where participants have a relatively high span threshold. Adaptive staircase procedures can improve both presentation time, and provide more trials of lengths close to the participants' span capacity threshold to better estimate their true maximal span (Woods et al., 2011). In the SCAMP battery the BDS was presented in an adaptive staircase format to speed up presentation (see Staircase Procedure section below). A further disadvantage of the traditional verbal presentation is that variance in tone, speed, and intonation of the reader can affect results (Woods et al., 2011). Interpretation of participant's verbal response can also be difficult (for example if the participant has a speech impediment, or speaks very quickly or quietly), and may result in inaccurate response recording (Raiford et al., 2010).

Here we used a computerised task presentation method. On each trial, participants saw a sequence of individual digits displayed in large white text on a black screen. Each digit was shown for 1000ms. After each sequence was shown, a numerical response grid appeared, and participants reproduced the sequence in reverse order by clicking with their mouse on the numbered buttons (**Figure 2.2**). This presentation format allowed standardisation of stimuli presentation and response recording. Additional information about responses could also recorded, such as response times. However, responses are potentially vulnerable to mouse-slips or response delays, especially if the participant is unfamiliar with using computers. This method is also somewhat different to the standard procedure in that it does not purely assess verbal working memory, since the stimuli are not presented verbally. While it is expected that participants transformed the visual presentation of numbers into a verbal code, the task does involve an additional visual component compared to the standard task version, both in terms of how the stimuli were presented and how the responses were made.

Figure 2.2 Illustration of a three item trial in the backwards digit span task. Participants see a sequence of digits (5, 9, 3) displayed on screen, then click the sequence in reverse order (3, 9, 5) on the input panel



Task instructions were displayed first. During testing, the instructions were modified a few times because the participants found it difficult to grasp the task. From battery Version 2 onwards, the instructions were made clearer to emphasise that participants should reverse the number order, as many participants were making errors where they were repeating the number sequence in forward order. This may mean there is poorer performance observed in Battery Version 1 participants, compared to participants who did other task versions. The final instructions presented to participants are in **Appendix B.**

Following the instructions, participants then began a set of practice trials: two sequences of two digits and one sequence of three digits. If the participant made an error during practice, a red cross appeared; an instruction prompt was shown again; then the trial was repeated up to three times. If they failed any practice trial three times the task was abandoned. Otherwise, the main task began. The main task had a minimum sequence length of three digits, and a maximum length of nine digits. No feedback was given during the main task.

Staircase procedure

A staircase procedure (Levitt's procedure) was used to reduce the time taken on this task. In the main task, participants began with a sequence of three digits. If participants succeeded on the initial trial, sequence length increased initially by three digits, then after the next success by two digits, then one digit, and continued to increase by one until either three successes at the maximum level of nine digits had occurred, or until an error was made. The staircase inverted after an error: the next trial sequence decreased initially by two digits, then by one after each subsequent failure, until another success occurred, or until the participant failed three trials of the initial starting length. After the next success, the staircase inverted again and increased the sequence by one digit. After the next failure, another inversion occurred, then the task finished. The stimuli used at each level were drawn from a list of 10 options selected in pseudo-random order such that two participants taking the same route through the procedure will have seen exactly the same stimuli on each trial.

This test was placed second in the battery as it is a commonly used test of working memory, which shows improvement during early adolescence, exhibits significant individual differences, and is a predictor of academic performance.

Task measures

The key task measure used in this thesis is the average of the mean level passed and the mean level failed. To calculate this, we first calculated the mean load of all the trials the participant had passed, and then the mean load of all the trials they failed. We then took the mean of these two averages.

This measure was used because our task presentation used a staircase procedure. We therefore cannot use 'standard' task measures (e.g. percentage accuracy) as levels were not presented progressively due to time constraints. A threshold value can be calculated by averaging the sequence length of the participants' last 4 inversion points, however some participants did not get a threshold estimate. These are the poorer performers who failed out of the task after failing at floor level three times in a row. Using this threshold estimate would therefore skew the distribution of who we have data for. The alternative measure of performance which we used is to take an average of the mean level passed and the mean level failed. As correlation between this and the threshold measure is very high (R=.978; p< .0005) this second measure was used to be able to retain data from the poorer performers.

Data Cleaning

Participants who failed the practice were excluded, as were participants who passed the practice but did not pass any trials in the main task. No estimate of performance in the main task was possible for either of these groups. No other exclusions were made. Numbers of participants excluded and final good N's are in **Table 2.5.**

2.6.3 Spatial Working Memory (SWM)

A Measure of Visuospatial Working Memory

The SCAMP battery included two tasks of visuospatial working memory. The first was adapted from the Spatial Working Memory task of the CANTAB battery (Luciana & Nelson, 1998). The task requires participants to search for a series of target locations, while remembering where previous targets were located and which locations have already been searched within a trial. Task measures include the number and types of errors made during the search process, and the use of a specific search strategy. The SWM therefore assesses both visuospatial working memory and a higher order executive function components related to strategy. Self-ordered tasks (where participants decide the order in which they search an array) typically show more prolonged development than simpler types of working memory tasks.

One strategy for success is to use a consistent ordered search pattern across all searches within a block, e.g. searching the items in order from furthest left to right or in a clockwise manner, and eliminating locations from this predetermined sequence as targets are found. This kind of method reduces the working memory demand of the task (Luciana & Nelson, 1998). Strategy use reflects a higher level executive function of planning ability (Owen et al., 1990). A strategy use measure may be calculated from the number of searches that begin with the same starting location within each block. A higher score indicates lower rate of use of this strategy (Luciana & Nelson, 1998).

Performance in the CANTAB version of this task develops across adolescence, with significant improvement between the ages of 10 and 15, and optimal performance being reached during early adulthood (20-29 years). Strategy use shows a slightly delayed developmental trajectory compared with error count measures, suggesting that higher level EF skills continue to develop later on into adolescence (De Luca et al., 2003). There is also an indication from previous research that males perform slightly better than females in this task, perhaps due to a general relative strength in spatial processing ability (De Luca et al., 2003).

Procedure

This test was the third cognitive task performed by the participants. In the original computerised version of this task, participants searched for tokens hidden within boxes on a screen (Luciana & Nelson, 1998). Participants clicked on boxes to ‘open’ them and find which one contained the token. When a token was found, another would be hidden in a new box. Each box contained a token exactly once during the task, and the task ended once all the tokens had been found (Robbins et al., 1998). In the SCAMP version of the task, participants searched for a phone which was ‘ringing’ amongst an array of identical images of phones; each phone would ring exactly once within a block; and the task ended once all the phones had been answered in all blocks. Visuospatial working memory is necessary to recall which phones have been clicked within an individual search trial, and also which have already rung within a block, to avoid clicking items multiple times.

An array of identical images of phones was presented across a screen (**Figure 2.3**). The locations were pseudo-random; all participants saw the same arrays. In each trial, one phone is 'ringing' – this is the target item. Participants search for the ringing phone by clicking on phones to 'pick them up'; feedback is displayed as to whether they have been successful in answering the ringing phone after each click. Each search ending with a successful identification of the target constitutes one trial. A block includes all phones that can ring within the array.

The instructions for the task were presented during a demonstration video, which illustrated the concept that once a particular phone has been found to be ringing then it will not ring again. A practice with four items followed. If between search errors – i.e. if the participant returned to a box which has previously been found to contain a token – were made during the practice, additional instructions were presented and the practice was repeated. If three unsuccessful attempts at practice occurred, the task was abandoned. If the participant made no between search errors during any round of the practice, the main task began.

The main task consisted of four blocks with four, six, eight, then ten phones to find. Feedback was shown after each target was located. A counter at the top indicated how many ringing phones had been found so far within the block, and how many remained to be found (**Figure 2.3**). On each trial, if the participant made twice as many errors as there were items (e.g. eight errors within a four item display) the active phone was indicated with an arrow and the trial was failed; then the block continued with a new target phone. If the participant made every possible between search error the trial was also failed. If all searches in a block were failed, the task was then abandoned. Otherwise the task progressed until the final ten item block was completed.

Figure 2.3 Illustration of an eight-item block in the Spatial Working Memory task. Participants click on phones until they find the one that is ringing in each trial. Each phone rings once per block. The feedback bubble indicates the participant has successfully identified the ringing phone in this trial. The counter at the top indicates they have now identified two out of eight phones within this block.

Task measures

To estimate performance, it is useful to consider the task as a series of 'searches'. Each search consists of the series of clicks leading up to the discovery of a target item. Two key types of error may occur during the task. A within-search error (WSE) occurs if the participant clicks the same item twice within a single search. A between-search error (BSE) occurs if the participant returns to a box which has previously been found to contain a token. These two types of search errors reflect different time-scales of spatial working memory; within search errors reflect a shorter time-scale, and between search errors reflect a slightly longer time-scale. Errors may also be both WSE and BSE: if a participant clicks an item which has already been the target location two times within a single search trial, this second erroneous click is an error both within the current search, and also between searches. In the analyses in this thesis, these 'both errors' are counted as a single error in our total errors measure, and add would one to the number of WSE and BSE if these were considered separately.

A strategy use measure was also calculated, based on that used in Luciana & Nelson (1998). The strategy we considered was using a set search pattern within a block, i.e. starting with same location each time, which would reduce the memory load required between searches. Higher scores indicate lower rate of use of this strategy. The score was calculated as follows: If the participant begins with the same start location for every search trial within a block, they would score the minimum value of 0 for that block. Where the participant begins a subsequent search after the first search within a block with the a new start location, we add one to their strategy score total. We do not add one to the score if the first item they click within a search happens to be the target location, as this is still a valid use of the strategy. Total strategy score is summed across the blocks of 6, 8 and 10 items. There are likely other possible useful strategies that could be employed, but it has been suggested that the most efficient method of completing the searches is to tart from the same location and follow the same pattern through all the items for each search (Lehto et al., 2003).

The measures used in this thesis are:

1. Total errors score, which includes both within and between search errors. A higher number indicates poorer performance.

2. Strategy score, equivalent to the total number of excess changes in start location made across the blocks of 6, 8 and 10 items. A higher number indicates less consistent use of the assessed strategy, and therefore poorer strategic thinking.

Data Cleaning

In battery version 1 participants repeated the blocks with 4, 6, and 8 items. These repeats were removed from version 2 onwards due to testing time constraints. Any repeats were excluded during data cleaning.

The same exclusions are applied in baseline and follow-up data. For participants who completed multiple repeats at same level (battery version 1 only) we only use their first attempt at each level for the analysis. We also excluded participants if did not reach the final Level 10 as it would be impossible to prorate for performance in only the easier levels. It was decided that it would be too complex to pro-rate the number of errors based on a smaller number of items seen by the participant at the loads 4, 6 and 10, as proportionally more errors may be made on level 10 as the memory load is greater. These participants did not reach L10 as they failed on an earlier level – this was often due to time out as they did not click the mouse for an extended period, and were no longer participating in the task. We also excluded participants who are outside 3.29SD of mean in total errors measure, as some participants had made very large numbers of errors (for example one participant at baseline made over 150 non-dot clicks). Numbers of participants excluded on the different criteria and final good N's are found in **Table 2.5.**

2.6.4 Corsi Block Span Task

A Test of Visuospatial Working Memory

The Corsi block tapping task (Corsi, 1972) is a classic test of visuospatial working memory span. Developmental research has shown that performance in Corsi-type spatial span tasks continues to develop across early adolescence, with significant improvements in performance occurring between ages 10 and 15, and peak performance being reached in early adulthood (Luciana & Nelson, 2002), (De Luca et al., 2003) (Dumontheil & Klingberg, 2012). In addition, performance in this task has been shown to predict maths performance (Dumontheil & Klingberg, 2012). This task was therefore included in the SCAMP battery to provide a more direct measure of visuospatial working memory than in the more complex, self-ordered search SWM task.

Procedure

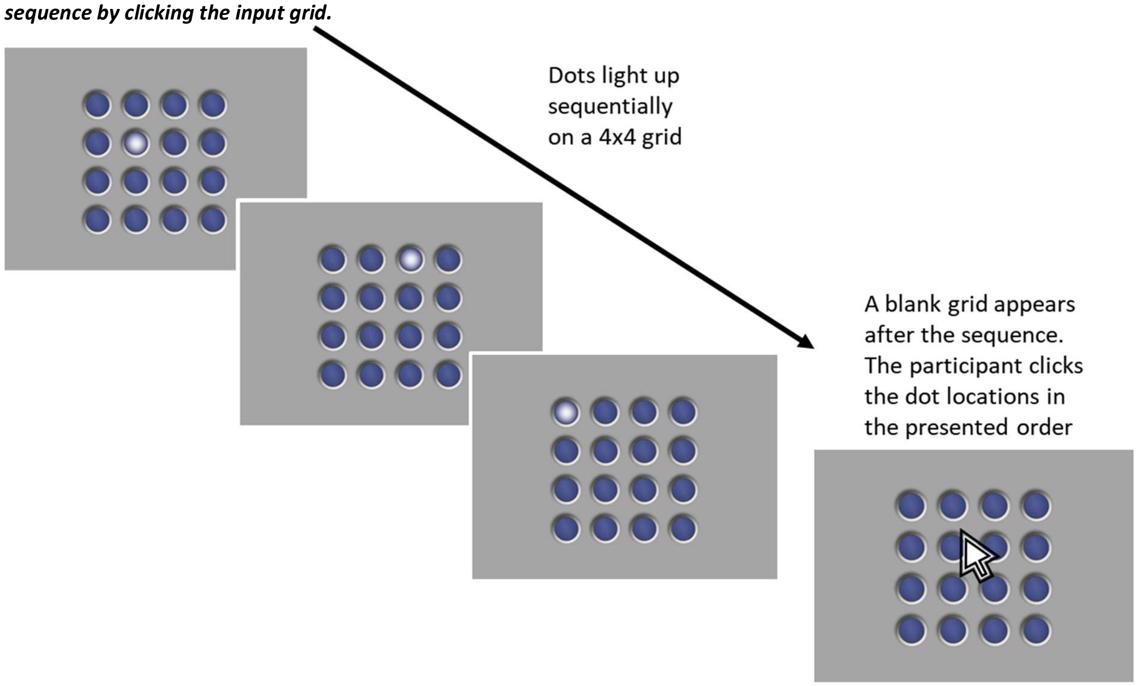
As the SWM phones task also assessed visuospatial working memory cognitive component, and has an additional strategic component, the Corsi task was placed towards the end of the battery as an additional measure. The Corsi span is a widely used and well-validated assessment of visuospatial working memory; it has been used to predict maths performance in school; and is short to complete, so it was considered a useful extra task to include towards the end of the battery.

In the original version, the experimenter tapped blocks in a particular sequence, then the participant tapped that same sequence (Corsi, 1972). A computerised version of the Corsi task is used in the CANTAB battery (Luciana & Nelson, 1998). The version in SCAMP has a 4 x 4 grid of dots which light up in a sequence in place of the boxes. Computerised versions of this task have been used in previous research (e.g. De Luca et al., 2003; Dumontheil & Klingberg, 2012). The SCAMP version of the task is similar to the block span task in the Automated Working Memory Assessment (AWMA) (Alloway et al., 2008).

Instructions were shown first, followed by a short animation demonstrating the procedure. Participants then completed a single practice trial with three locations to remember. To proceed to the main task, correct responses for all three locations in the correct order was required. If an error was made, feedback was displayed, and the sequence was repeated. Five attempts at practice were allowed; if all were failed the task was abandoned.

In each trial, sixteen blue circles were displayed in a 4x4 grid on a grey background (**Figure 2.4**). A series of circles lit up in white one-by-one. Participants were asked to reproduce the sequence of locations they saw in the same order by clicking on the relevant locations in a response grid. If they did not respond within 3 s, a “don’t know” button appeared – if this was clicked the trial ended.

Figure 2.4 Illustration of a three-item trial in the Corsi task. Participants see a sequence of dot locations, then repeat the sequence by clicking the input grid.



The minimum trial sequence length was three items, and the maximum nine. During the main task, if an error was made or the “don’t know” button was pressed, the trial ended and a red cross appeared, and the next trial in the staircase procedure was initiated. To speed up task administration, the main task followed a similar staircase procedure to the BDS task. The Levitt staircase procedure was used, starting at load 3, with an initial step size of 3. In the first block up to 4 trial failures are ignored. After the first block just one trial is administered at a load level. Success at a trial increases the load by the current step size. Failure reduces the load by the current step size. At reversal points (when the participant answers wrongly and then correctly or vice versa) the step size is reduced by 1 until the step size reaches 1. The task continues until there have been 4 reversals at a step size of 1. The stimuli used at each level are selected in order such that two participants taking the same route through the procedure will have seen exactly the same stimuli on each trial. The load range is constrained to 3-9, i.e. after a wrong answer on load 3 participants stay on load 3, and after a correct answer on load 9 participants stay on load 9, until participants have done three trials on that load, at which point the task is ends.

Task Measures

The approach taken here is the same as for the BDS task. This measure was used because the task used a staircase procedure. We cannot use a 'standard' task measures (e.g. percentage accuracy) as load did not increase progressively. A threshold estimate can be produced by taking an average of the load levels of their final 4 reversals – however the poorest performers who fail due to repeated errors on the lowest load level do not have sufficient data to estimate this threshold score (N=634 at baseline). Using this threshold estimate would skew the distribution of who we have data for. Therefore we use an alternative measure of performance, which is to take an average of the mean level they passed and the mean level they failed. As correlation between this average and the threshold measure is very high ($R = .969$; $p < .0005$); and the average measure was able to retain data from the poorest performers.

The key task measure used in this thesis is the average of the mean level passed and the mean level failed. We first calculated the mean level of all the trials they passed, and then the mean level of all the trials they failed. The key variable is the mean of these two averages.

Data cleaning

The same exclusions are applied in baseline and follow-up data. Participants who failed the practice are excluded – as they did not attempt the main task. Participants who passed the practice but did not pass any trials of main task also excluded, as they failed the main task entirely. No true estimate of performance is therefore possible. Numbers of participants excluded on the different criteria and final good N's are found in **Table 2.5.**

2.6.5 Cattell's Culture Fair Task (CFT)

A test of general fluid intelligence

This is a computerised version of a standardised visuospatial reasoning test, which assesses fluid intelligence (Cattell & Cattell, 1960b). Cattell distinguished between crystallised intelligence, which relies on specific prior knowledge, and fluid intelligence, which is a more general ability to reason and form solutions to novel problems. The culture fair test was devised to test fluid intelligence. It is intended to not rely on any particular prior knowledge to complete it. The test was intended to reduce the ethnic and cultural biases in IQ tests of the time, which largely assessed culturally-specific crystallised knowledge (Brown, 2016). However, cross-cultural research suggests that the CFT is not itself completely free of cultural bias, with participants from different counties (Nigeria and USA) exhibiting different patterns of performance on the sub-tasks (Nenty & Dinero, 1981). The key score is the total number of items correct in each sub-task. The total score reflects the participant's fluid intelligence, and scores from the standard version of the CFT may be used as an IQ proxy according to published norms (J. Duncan et al., 2017).

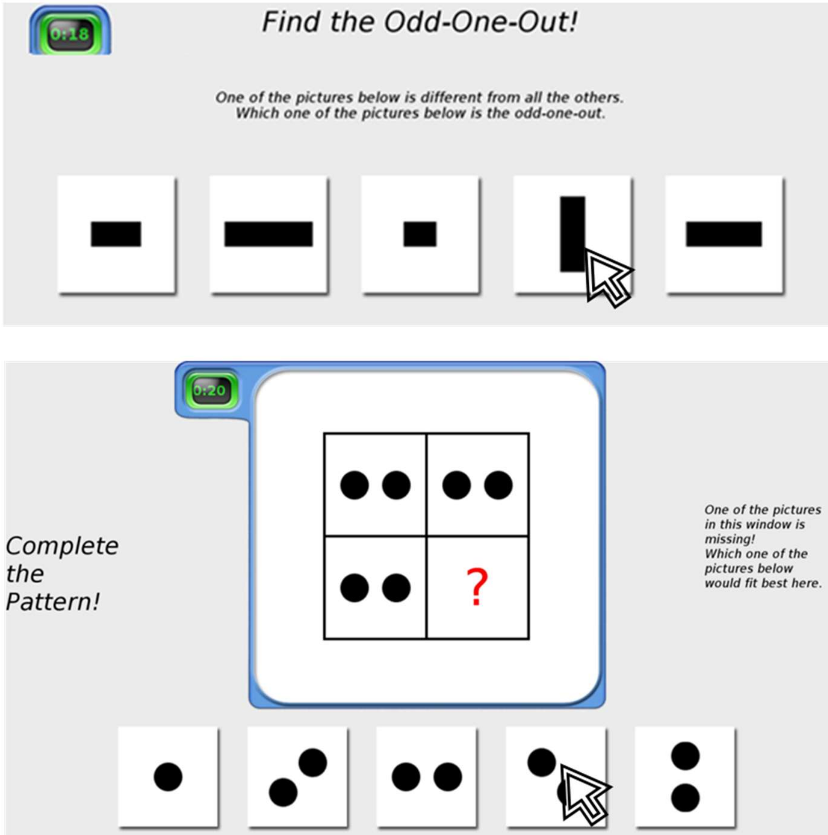
Procedure

This is a test of general cognitive ability, particularly targeting fluid intelligence, using a standardised visuospatial reasoning task. Scores are often used as a proxy for IQ. The CFT was included in the test battery as a measure of interest to teachers, parents and policy makers. CFT scores are also useful as they predict academic performance. The task was placed relatively early in the battery, but after the first three tasks assessing EF.

The standard CFT is administered as a pen-and-paper task with four sub-tasks: (1) linear completion, which asks participants to choose the item which best completes a linear sequence; (2) odd-one-out, which asks participants to choose the item which is not like the others; (3) matrix pattern completion, which asks participants to choose the item which best completes the matrix; (4) match the shape, which asks participants to choose which complex shape contains the simpler target shape. Each sub-tasks needs to be completed within a particular time limit.

The SCAMP battery originally contained sub-tasks 1, 2 and 3 (as sub-task 4 was deemed too complex to display clearly on a screen). Sub-task 1 (what comes next?) was removed from the battery early on in testing due to time constraints, and as scores form this sub-task were reasonably well correlated with the other two sub-tasks. Scores of sub-task 1 were not included in any of the analyses in this thesis. The procedures for the two sub-tasks 2 and 3, as used in the SCAMP battery, were as follows. Stimuli were taken from Cattell's Culture Fair Test Form A Scale 2 (Cattell & Cattell, 1960b). In sub-task 2 (odd-one-out) five images were presented (**Figure 2.5**). Participants clicked on the image that did not fit with the others. Sub-task 3 (complete-the-pattern) was similar to the Raven's Matrices test. Participants selected an item (out of five options) that best completed a 2 x 2 grid with one empty square (**Figure 2.5**).

Figure 2.5 Illustration of Culture Fair Task. The 'Odd-One-Out' section comes first, then the 'Complete the Pattern' section. In each section, participants complete as many trials as possible within three minutes, up to fifteen trials.



1. Odd-one-out

Participants click one image out of five options that is different from the others

2. Complete the pattern

Participants click on the image that best completes the block shown above from a selection of five images

Each sub-task was preceded by an animation which demonstrated a sample trial and indicated which was the correct answer. A single practice trial was then completed for that sub-task, with corrective feedback given until the correct answer was chosen. The main task began with the odd-one-out sub-task, then the complete-the-pattern sub-task. The participant had three minutes to complete as many items as they could out of 15 possible trials in each sub-task. Participants were allowed to skip a trial, then if they still had time at the end of presentation of the remaining trials, the skipped trials were re-presented until the time had elapsed or all trials had been attempted. Trials within each section were presented in order of increasing difficulty. An early termination rule was used: if the participant failed four out of five items, which indicated that they were responding at chance level, the assumption was made that they would continue to perform at chance and the task was terminated. Participants completed both sub-tasks; even if they had early termination in the first sub-task, they still progressed to the next one.

Task Measures

The standard measure of CFT performance is the total number of items correct in all 4 subtasks of the standard task. Subtask 4 from the original Cattell Culture Fair Task was not suitable for computer presentation so was not presented in the SCAMP battery. Sub-task 1 was only presented to some participants (those in battery version 1). All participants were presented a battery including subtasks 2 + 3. The task measure used in this thesis is the total number of correct trials on sub-tasks 2 + 3.

Data Cleaning

Participants who completed only one subtask are excluded – there were not many participants who did this and it would be likely to give a somewhat skewed estimate if we were to prorate for only the subtask they completed. he same exclusion criteria are applied at baseline and follow-up. Numbers of participants excluded on the different criteria and final good N's are found in **Table 2.5.**

2.6.6 Continuous Performance Task (CPT)

The CPT is a commonly used assessment of sustained attention, vigilant attention, or cognitive control (Brocki et al., 2010; Morandini et al., 2020; Iselin & DeCoster, 2009). It is essentially an extension of a basic go/no-go task, where participants are instructed to respond to one type of stimulus and not respond to another. The SCAMP assessment uses the A-X version of the CPT. In the AX-CPT, a series of letters is shown individually and sequentially to participants, who are instructed to make a response only when they see an A followed by an X (Halperin et al., 1991). Task parameters vary significantly in the literature. Often a very high number of trials is used, for example Iselin & DeCoster (2009) had 400 trials with a total task duration of around 30 minutes. Varying ratios of targets are also used – in Iselin and DeCoster's study 70% of trials were correct targets, which was intended to induce high levels of expectancy bias and therefore high numbers of commission errors, whereas Halperin et al. had only 10% of stimuli as correct targets. In SCAMP, we used 54 trials, of which 9 were target sequences.

By definition, this task requires maintaining attention in a sustained manner while stimuli are presented at a slow rate, while inhibiting impulsive responses. The task is therefore intentionally relatively long, and fairly boring. Success in the task requires various cognitive processes. Principally, it requires sustained vigilance, i.e. the ability to maintain attention to perform a simple, un-challenging task for a prolonged period of time (Morandini et al., 2020). Success also relies on aspects of executive function, including working memory to maintain the target items and task requirements in mind, and to update mental representations of stimuli; and inhibitory control to prevent commission errors (Iselin & DeCoster, 2009).

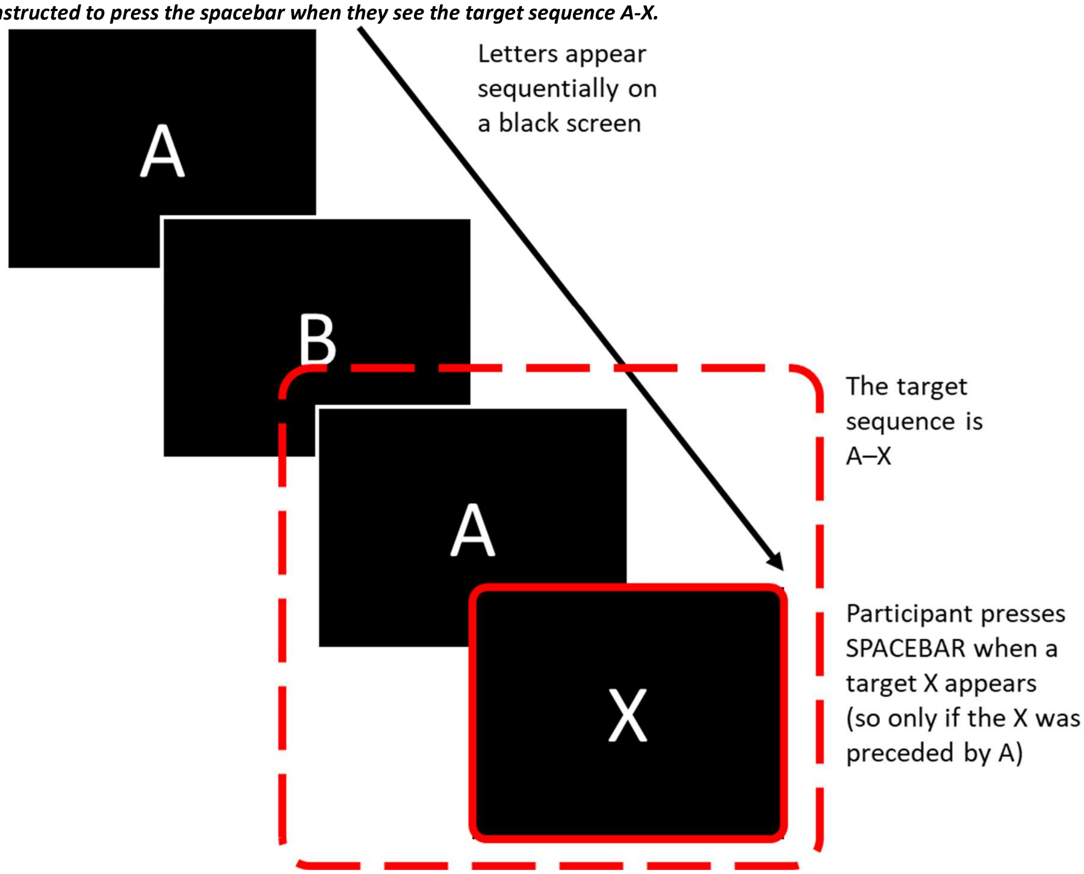
The number and type of errors is a common method of estimating CPT performance. Errors may be broadly classified as omission errors, where a target is missed, and commission errors, where a response is made to a non-target stimulus (Brocki et al., 2010). Different error types reflect different aspects of cognition. Omission errors may reflect inattention, as participants fail to maintain sufficient attention to the task to respond correctly, or a failure of working memory, as participants have forgotten the target item or task instructions (Halperin et al., 1991). Commission errors are generally considered to reflect some aspect of impulsivity or disinhibition. In a factor analysis study looking at executive function structure in adolescents with ADHD, omission errors in a CPT task loaded onto a different factor (inattention) than commission errors (inhibition) (Barkley et al., 2001).

Sub-types of commission errors may each reflect different aspects of cognition: response to an A alone reflects impulsivity; response to an X which was not preceded by A reflects disinhibition; response to a letter other than X which was preceded by A reflects inattention; and random errors occur when the participant responds at any other time (Halperin et al., 1991). Mean reaction times for correct responses may also be used to estimate performance, with faster reaction times indicating better performance and better sustained attention ability (Brocki et al., 2010).

Procedure

Instructions were displayed first, then participants were shown a short animation which demonstrated to press the spacebar when an X appeared which had been directly preceded by an A. A practice block followed, consisting of 18 trials with three target A-X sequences. Practice was failed if either an omission error occurred, or if two commission errors occurred. If this happened, the instructions were re-displayed and the practice repeated. Participants were allowed three attempts at practice. If they failed the practice all three times, the task was abandoned. Otherwise the main task then began. Individual letters appeared in a white font at the centre of a black screen. Each letter was displayed for 460ms, and a slightly jittered fixation period of around 2500ms occurred between successive letters. The total presentation time for the main task was around three minutes. A fixed series of 54 letters were displayed in two blocks with a short pause between blocks, participants were told to take a short break if they wanted, then press the spacebar to begin the second block. There were nine target A-X sequences in total. Because of its length, and because some participants in a pilot study found this task quite boring and disengaged with the whole assessment battery afterwards, this task was put in the last position of the prioritised list.

Figure 2.6 Illustration of AX-continuous performance task. A series of letters appears on a black screen. The participant is instructed to press the spacebar when they see the target sequence A-X.



Task Measures

Omission and commission errors are moderately strongly correlated, *R*=.319. The overall commission errors total is strongly correlated with the sub-groups of commission errors, i.e. the subcategories of impulsive, inattentive, disinhibited, random errors (*R*'s between .593 to .840). The subcategories of errors are also moderately to strongly correlated with each other.

We decided to analyse the omission and commission separately, as a. they aren't overly highly correlated to each other; b. they have a different meaning (inattention vs. disinhibition) (Barkley et al., 2001). In this thesis the numbers of omission errors and commission errors are used as the key task measures of the CPT task.

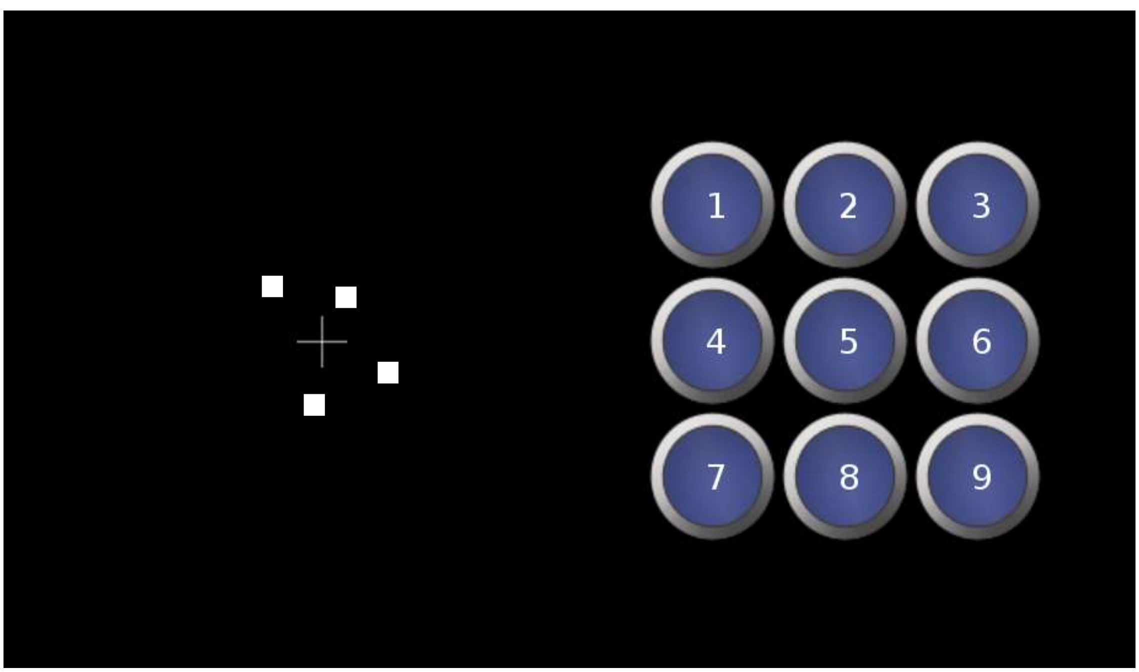
Data Cleaning

We considered excluding participants if it was clear that they not paying attention to the task at all. There are 9 AX targets in the whole task, and 54 non-target items. First, we considered whether participants were making excessive numbers of random-type errors, which would indicate they were pressing the response key entirely at random. At baseline, two participants made 10 or more random errors; these were excluded. Next we considered whether participants were not responding at all. No participants missed all the target items, suggesting they were at least responding correctly some of the time. Finally, checked for any unusual values in the commission error measure. At time 1, one participant made over 100 erroneous button presses (almost two per item displayed on average) – this participant was also excluded. No such unusual values were observed for the follow-up data. Numbers of participants excluded on the different criteria and final good N's are found in **Table 2.5.**

2.6.7 Enumeration task

This task is not analysed in this thesis so only a brief description is provided. This task assesses visual attention and specifically provides an estimate of subitisation range. Subitisation is the ability to very quickly 'recognise' the number of items present in a group; as opposed to counting the items consciously. Subitisation is only possible with small numbers of items, with most adults having a subitisation span of around five items (Green & Bavelier, 2006). Our task is a version the task used in Green and Bavelier (2006). A group of white squares was presented very briefly (50ms) on a black background (**Figure 2.7**). A response keypad then appeared, and participants were instructed to click the number of items they saw displayed. There were 36 trials with between 1 and 9 items each presented in pseudorandom order.

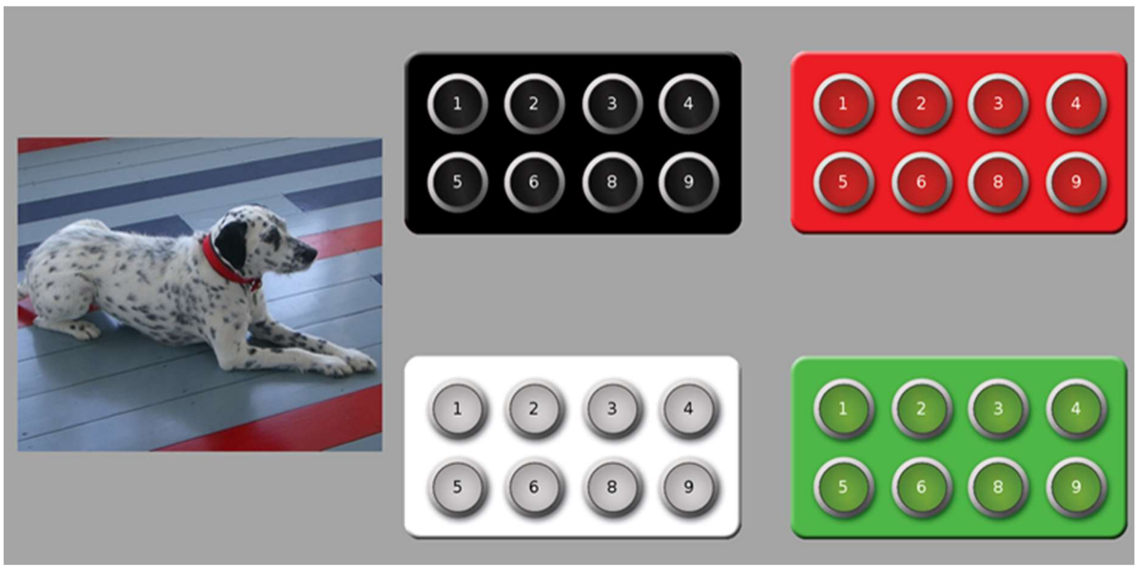
Figure 2.7 Example of a four-item stimulus in the Enumeration task.



2.6.8 Speech in Noise (SPIN)

This task is not analysed in this thesis, so only a brief description is provided. This task targets speech processing. Participants heard a recorded voice saying for example: "show the dog where the *red three* is"; with a different *colour number* combination for each trial. The participant then clicked the appropriate button on screen (**Figure 2.8**). The audio stimuli were masked with white noise at specified signal-to-noise ratios. Stimuli of varying difficulty were played following a staircase procedure similar to that used in the BDS and Corsi tasks, until the participant reached a plateau level of difficulty.

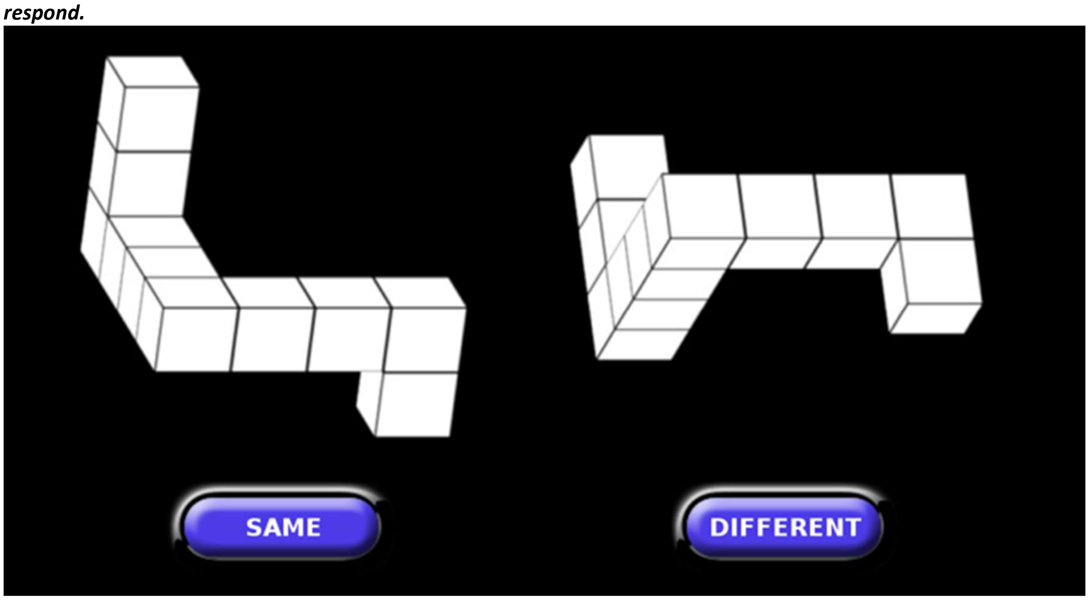
Figure 2.8 Illustration of the response screen in the SPIN task. An audio stimulus was played, e.g. "Show the dog where the red three is"; participants clicked the appropriate colour number button.



2.6.9 Mental rotation task (MR)

This task is not considered in this thesis, so only a brief description is provided. The task assesses visuospatial manipulation ability, specifically mental spatial rotation ability. The task consisted of two white 3D block images presented simultaneously on a black screen, adapted from the Peters & Battista (2008) mental rotation stimuli library. The shapes were either the same 3D model rotated in space (Same trials), or mirror images of each other also rotated in space (Different trials) (**Figure 2.9**). Participants clicked a response button below the image pair to indicate whether they thought the items were the same or different blocks.

Figure 2.9 Illustration of a 'Same' trial in the Mental Rotation task. Participants click the 'SAME' or 'DIFFERENT' button to respond.



2.7 Summary of Data Cleaning Procedures

2.7.1 Merging Data for Participants Who Moved Computers

Some participants moved computers during task administration, mainly because of technical issues such as faulty equipment. These participants were assigned new ID numbers on the new computer. Data for these multiple IDs had to be merged into the same record during initial data processing. Matching participant IDs were identified by searching for records with the same name and birthdate within the same school, then the data were merged into a single record, with data from the later computers recorded as additional attempts at the task battery.

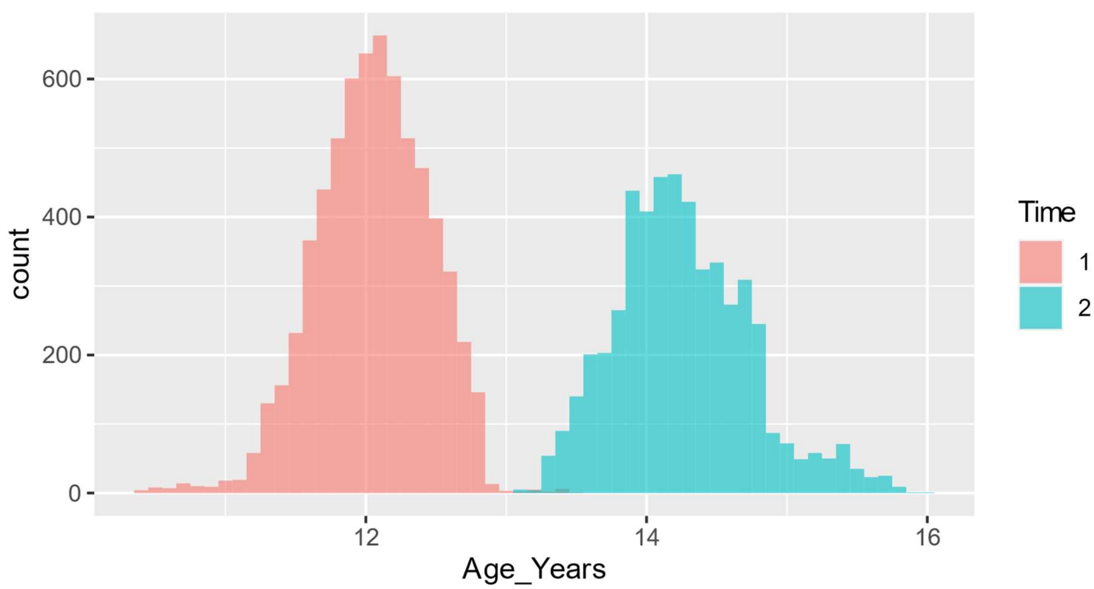
2.7.2 Attempt Selection

Where participants had completed the same task or questionnaire multiple times within an assessment, data from their first complete attempt were taken. Any subsequent attempts were filtered out of the final cleaned dataset.

2.7.3 Exclusions based on Age at Time of Testing

Data were excluded for participants whose reported age at the time of testing was far outside the intended target age groups. Participants reported their birthdate as part of the computerised assessment, and their age in days at assessment was calculated. The assessments were intended to capture pupils in school Years 7-8 at baseline (who are usually aged between 11 and 13 years) and Years 9-10 at follow-up (usually aged between 13 and 15 years). Some extremely high and low ages were reported, with a range at baseline between 9.62 - 15.41 years, and at follow-up a range between 10.93 - 19.15 years. This indicated that either some participants had incorrectly input their birthdate, or that some pupils from incorrect year groups had been sampled. Data distributions were inspected to ascertain sensible cut-off values. We did not wish to simply exclude data from all participants who were outside the original intended range, as they may legitimately be in the intended school year, e.g. by being held back or promoted a year, however we did not want the extreme outliers to skew analyses, especially where age might be used as a covariate. At baseline, cut-offs were set to exclude participants whose reported ages were below 10.4 or above 13.6, this excluded 81 participants; a further 8 participants were also excluded as they did not provide age data. This left a final N = 6,591. At follow-up, cut-offs were set to exclude those outside 13.0 and 16.0 years, this excluded 22 participants and gave a final N = 5,116. The distribution of ages of participants in each assessment point are shown in **Figure 2.10.** Final N's for each task following these exclusions are provided in **Table 2.5** at the end of this chapter.

Figure 2.10 Distribution of participant ages at baseline (Time 1) and follow-up (Time 2)



2.7.4 Matching Participants Across Assessment Points

An auto-matching algorithm was used during the follow-up assessment. In the main battery, participants first entered their first and last names and date of birth. Where these were an exact match to an existing baseline record, that participant was assigned the same ID number and the same task order as at baseline. Following data collection, data were manually inspected to identify any additional matches. Records where last name and date of birth matched ,and first names were very with similar sounding (e.g. Jon and John), small typos (e.g Jane and Hane), or common abbreviations (e.g. Will and William) were matched manually. These processes overall resulted in a total of N=3,715 participants matched across time points. Task-Based Data Cleaning and Final N's

Finally, data for each task were cleaned, as described in the relevant task description sections earlier in this chapter. **Table 2.5** summarises the numbers of participants who were excluded on age, and each of the specific task exclusion criteria, and gives the resulting final *N'*s for each task at each assessment point.

2.7.5 Creating Z-scored and Transformed Measures

For easy comparison of effects across the different tasks, we created standardised scores for each task variable within each time point. For the baseline data, we Z-scored the data by centring the mean to 0 and setting the standard deviation to 1. For the follow-up data, we created transformed variables by using the raw mean and standard deviation from the baseline data to create a transformed variable similar to z-score, i.e. we took away the baseline mean and divided by the baseline standard deviation for each task score respectively. This was done in order to be able to show any progression from baseline to follow-up – if we had simply taken the Z-scores within baseline and follow-up separately, we would not have been able to demonstrate any progression in scores in Chapter 4 in particular. These Z-scored and transformed measures were used in most analyses – see the relevant Methods sections in each experimental chapter for which version of task measures were used.

Table 2.5 Final Ns per task at each assessment point

	Baseline				**Follow-up**				**Both Times**
Task	Original *N*	Exclusion Criterion	Excluded *N*	Final *N*	Original *N*	Exclusion Criterion	Excluded *N*	Final *N*	Final *N*
	6680	Age	89	**6591**	5138	Age	22	**5119**	**3715**
TMT	6678	Age	89	**6424**	5026	Age	19	**4918**	
		Excessive non-dot clicks	30			Excessive non-dot clicks	33		
		Timeout	122			Timeout	47		
		Other Incomplete	13			Other Incomplete	9		**3526**
BDS	6657	Age	80	**6165**	5065	Age	20	**4864**	
		Failed Practice	482			Failed Practice	177		
		Other Incomplete	11			Other Incomplete	4		**3379**
SWM	6592	Age	88	**6300**	4846	Age	19	**4624**	
		Failed practice	127			Failed practice	110		
		Did not reach final level	27			Did not reach final level	42		
		Excessive Duration	45			Excessive Duration	41		
		Other Incomplete	5			Other Incomplete	10		**3281**
Corsi	3905	Age	17	**3791**	2156	Age	8	**2093**	
		Failed Practice	46			Failed Practice	15		
		Other Incomplete	51			Other Incomplete	40		**1258**
CFT	5887	Age	79	**5808**	4720	Age	19	**4701**	**3133**
CPT	1831	Age	2	**1573**	1096	Age	1	**912**	
		Failed Practice	0			Failed Practice	0		
		Incomplete	255			Incomplete	183		**400**
		Excessive commission errors	1						

Chapter 3.
Associations Between Socio-Economic Status and Executive Function in the SCAMP Baseline Sample

3.1 Abstract

Previous research has established an association between socio-economic status (SES) and executive function (EF). Higher SES is generally associated with better EF scores. Although some studies have found specific links between individual measures of SES and particular components of EF, there have been insufficient studies in this area to describe these links completely. It also remains unclear from the literature whether associations between SES and EF exist over and above associations with fluid intelligence. SES and EF are independently significant predictors of academic achievement and other important life outcomes, and there is evidence that EF ability may act as a mediator between SES and academic achievement. It is therefore of interest to obtain a better understanding of associations between SES and EF during development.

This chapter first investigates general associations between SES and EF, then aims to identify specific associations between aspects of these two constructs. It then investigates whether unique associations between SES and EF exist over and above associations with fluid intelligence. Data are taken from n = 6,591 participants in the SCAMP baseline sample. Multivariate analysis reveals significant association between three SES measures (school type, father's occupation and postcode deprivation) and overall EF. These associations remain significant after accounting for fluid intelligence, with effect sizes slightly reduced. Univariate analyses reveal significant relationships between some specific SES measures and individual EF measures. Almost all of these specific associations remain significant after accounting for fluid intelligence. In both sets of analyses, school type is the SES measure most closely associated with EF. These results suggest firstly, that there are unique associations between SES and EF over and above associations with fluid intelligence, and secondly, that there are specific patterns of associations between individual aspects of SES and individual EF measures.

3.2 Introduction

Socio-economic status (SES) in early life is associated with a variety of important outcomes, including academic achievement, literacy and numeracy skills, and general cognitive ability in later life (Devine et al., 2016; Blair & Raver, 2015; Kaplan et al., 2001). Higher SES is generally associated with better outcomes across all these categories. EFs independent of SES also have associations with a variety of life outcomes (Diamond, 2013). EFs predict academic achievement more widely (Huizinga et al., 2018), and individual differences in childhood EF predict wider life outcomes in adulthood, including emotional traits, employment status, drug use, obesity and alcohol consumption (Robson et al., 2020; Smith-Spark, Moss & Dyer, 2016). EF also predicts socio-emotional outcomes later in life (Sasser et al., 2015).

Within cognitive outcomes, associations between SES and EF have been noted in a variety of tasks and studies (e.g. meta-analysis from Lawson et al., 2018), and SES-related disparities in EF may also be greater than for other cognitive areas, suggesting there could be specific associations between SES and EF beyond associations of SES with more general cognitive abilities (Noble et al., 2005). Furthermore, some studies have noted that specific aspects of SES are differentially associated with specific EF components (Sheridan et al., 2017), suggesting different levels of association between certain measures of SES and EF. However, not all studies in this area have found SES-related EF disparities (e.g. Wiebe et al., 2008; Gregoire & Van Der Linden, 1997). In particular, the detail of specific relationships between aspects of SES and EF requires further exploration. It also remains unclear from the literature whether SES is associated with EF over and above SES associations with other cognitive areas. Literature exploring the associations between SES and cognition, and in particular the literature investigating SES and EF will be summarised in this introduction.

Research has found that early SES predicts a range of later life outcomes (e.g. Sirin, 2005). However, the mechanisms by which SES might have impacts on such a variety of later outcomes remain unclear. Specific aspects of SES may causally influence cognitive and academic outcomes through a range of more or less direct pathways. SES may exert influence over critical features of the child's developmental environment, causing them to travel down certain developmental pathways, in turn influencing their unique pattern of cognitive, psychological and social outcomes. Research into the mechanisms by which SES exerts this influence remains in its infancy, and many potential areas are yet to be fully explored – some potential mechanisms by which SES may influence specific cognitive outcomes in EF is also included in this introduction.

A better understanding of how SES may affect EF is of particular interest because there is evidence that EF abilities can mediate the effects of SES on academic achievement. For example, a study found that children's executive function ability mediates the impacts of parental factors associated with SES on children's educational achievement outcomes, while general cognitive ability had no such mediating effect (Devine et al., 2016). Working memory ability predicts maths achievement in school, over and above the effects of intelligence, affective variables and family factors (Monette et al., 2011). Lawson and Farah (2017) found that children's EF, but not verbal memory, mediates the effect of parental education and income on academic achievement across a two-year time period.

The aim of this study was to investigate associations between specific aspects of SES and executive functions during adolescence. While pre-natal and preschool effects of SES on cognition have been relatively widely investigated, less work has been done in later development. As discussed in **Chapter 1**, executive functions continue to mature during adolescence, and effects of SES are likely to continue to be observed at this point.

3.2.1 Defining and Measuring SES

Conceptually, definitions of SES vary across the literature. One useful definition is that SES reflects a person's access to financial, social or cultural resources (Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress, 2012). One area of general agreement is that SES is can be assessed across three key areas: household income, education level, and occupation status (Lawson et al., 2018). Wider definitions of SES may also include social prestige or relative perceived societal status, neighbourhood factors such as deprivation, or access to resources in the home or at school (Alves et al., 2017; Eamon, 2005; Jeynes, 2002; O. Morgan & Baker, 2006).

SES is assessed in a multitude of different ways in research. Researchers might use a single measure as a proxy for overall SES, or might use multiple measures of SES, which are often combined in some way to give a single composite measure of SES. There is little consensus, however, around exactly which measures of SES should be used, or how multiple measures should be combined to give a single score (Bradley & Corwyn, 2002). For example, within the UK, two different SES composite measures are used for official purposes. In England and Wales, the Carstairs local area deprivation score is used. This is a weighted average across four criteria (percentage of male unemployment, lack of car ownership, overcrowding and low status occupations in the local area) (O. Morgan & Baker, 2006). In Scotland, the Scottish Index of Multiple Deprivation (SMID) is also a weighted average local area deprivation measure, which combines over 30 specific measures across seven domains (income, employment, education, health, access to services, crime and housing) (Payne, 2006). **Table 3.1** lists some of the wide variety of measures that have been used in the literature to assess SES.

Table 3.1 Examples of measures which may be used to assess socio-economic status

SES component	Measures	Examples of studies using this measure
Family income / wealth	Income-to-needs ratio	Sarsour et al. (2011)
	How much money do you have in savings?	Sarsour et al. (2011)
	Annual income	Decker et al.(2020), Sirin (2005)
Occupational prestige	Hollingshead Index	Sarsour et al. (2011)
	ONS NSSEC categories	Office for National Statistics
Parental or own education level	Number of years in education	Sarsour et al., (2011)
	Highest educational level reached	Sarsour et al., (2011)
Deprivation of local area	Carstairs	Morgan & Baker (2006)
	Scottish Index of Multiple Deprivation	Payne (2006)
Home environment	Quality of physical environment	Rosen et al. (2019)
	Cognitive stimulation	Rosen et al. (2019)
	Violence exposure	Rosen et al. (2019)

SES may be best considered as multiple independent, though related, components, rather than as a unitary construct. The three key SES components (income, occupation status and education) have different degrees of stability over time, and exhibit different patterns of associations with health and academic outcomes (Sirin, 2005). This may indicate that these three key SES indicators reflect separate underlying constructs. Another issue with measuring SES is that specific indicators of the same component can be associated with different patterns of outcomes. For example, paternal occupation in a child's early life is more closely associated with their later life health outcomes than is maternal occupation (Pinilla et al., 2017). Mother's education has been found to be a stronger predictor of cognitive performance than father's education (Gutman et al., 2003). It is important therefore to investigate whether different measures of SES contribute differentially to any observed associations between SES and EF.

Finally, it is important to note that SES and poverty are distinct, although related, concepts. Those who experience poverty are those who do not have sufficient financial resources to live adequate day-to-day lives. In the UK, the current government definition of poverty is households earning under 60% of the median wage (Devine et al., 2016). Some researchers use the concepts of low SES and poverty somewhat interchangeably (e.g. Lipina & Posner, 2012), and much research looking at associations between SES and cognitive outcomes simply compares extreme SES groups (categorising participants as low or high SES, or those in poverty or not) (e.g. Noble, Norman, & Farah, 2005). This is problematic because it fails to capture potential variation in the middle of the scale, and cannot identify where along SES gradients in outcomes occur (D'angiulli, Lipina, et al., 2012). Research using a range of SES levels has indicated that there are gradients in terms of academic and cognitive outcomes across the whole SES spectrum, rather than just between high and low SES groups, or between those in poverty and not Lawson et al., 2018). It is therefore important to investigate where along the SES spectrum differences lie, rather than considering only low vs high groupings. The analyses performed for this chapter will use measures of parental occupation, school type, parental education, and postcode deprivation with values across the spectrum of SES.

3.2.2 Associations between SES and academic and cognitive outcomes

There is significant evidence that students from lower SES households are more likely to do worse in school than those from higher SES households (Blair & Raver, 2015; Devine et al., 2016; Sirin, 2005). Cross-cultural research has revealed similar findings in various countries including Cambodia, Vanuatu, Mongolia – where relative SES within the countries predicted later academic achievement (Sun et al., 2018). Exposure to lower SES neighbourhoods earlier in childhood has a greater impact on academic outcomes than exposure during adolescence (Chetty et al., 2016). Parental and school based SES factors both contribute to the disparities in educational outcomes – a study using latent variable analyses found parental SES (education and occupation) and school-related SES (community and school type) were both associated with academic achievement, but only family SES influenced cognitive performance (Alves et al., 2017). Disparities are present even at the start of formal education, with children from lower SES households being less likely to score well on measures of school readiness (such as self-regulation) at school entry (Blair & Raver, 2015). This suggests that school quality is not the only factor affecting students' outcomes, as differences are already present at school entry. One theory that has been proposed is that preschool EF abilities may moderate some aspects of associations between SES and early academic outcomes.

SES is associated with a wide range of cognitive outcomes. Children from lower SES backgrounds are more likely to experience cognitive delays and emotional problems (Brito & Noble 2014). Performance on multiple types of memory tasks (both long and short term) is related to SES (see Herrmann & Guadagno, 1997 for a review). SES gradients predict around 30% of variance in language ability tests (Noble et al., 2007). A study of over 700 children found associations between family income and their memory and language abilities, and that anterior hippocampal volumes also correlated with family income (Decker et al., 2020). Some studies have found that specific aspects of SES in childhood predicts specific cognitive abilities or specific aspects of complex life outcomes.

Quality of environment is associated with accuracy in a memory guided attention task, and violence exposure in childhood is associated with poorer associative memory (Rosen et al., 2019). This suggests that different aspects of SES may be associated with specific cognitive outcomes, and this is an area of interest to investigate.

3.2.3 Associations between SES and EF

Since executive function has a prolonged developmental period, lasting well into adolescence (see **Chapter 1** for review), it may be particularly sensitive to effects associated with differing SES levels . Associations between various EF measures and SES have been found. Children from disadvantaged backgrounds performed worse in a digit span working memory task (Globerson, 1983). Higher SES students showed better performance in attention control tasks in primary school children (Mezzacappa, 2004). Poorer performance of children in poverty on an A-not-B switching task has also been observed (Lipina et al., 2005). Multiple measures of EF were found to be associated with parental education level in a sample of over 600 South American children (Ardila et al., 2005). Rosen et al., (2020) used three measures of EF assessing cognitive flexibility, working memory and inhibition in a 18-month longitudinal study of children age 5-8. They found SES as measured by household income and highest parental education was associated with all performance in all three tasks at baseline, however, SES was not associated with improvements in performance over the 18 month follow-up period. This study suggests that SES is be associated with EF performance, and that the effect of SES does not alter over time.

In a meta-analysis of association between SES and EF, SES accounted for between 2.6% and 7.8% of variance in EF, depending on the type of SES and EF measures used (Lawson et al., 2018). Some studies however have found no relationship between SES and EF measures (e.g. Wiebe et al., 2008; Gregoire & Van Der Linden, 1997). From the results of the Lawson et al. meta-analysis, it seems possible that SES and EF are only weakly related, with small effect sizes, and perhaps the studies finding null results may have lacked the power to detect these small effects.

Evidence is mixed regarding whether associations between SES and EF may be greater than with other aspects of cognition. Some studies report a greater association between SES and EF than with other cognitive domains. For example, in a study with low and high SES kindergarteners Noble et al. (2005) found that EF and language abilities were more closely associated with SES than other cognitive measures were. Other studies have found that cognitive areas other than EF are more closely related to SES - for example, in a study with first-grade pupils in the USA, language, visuo-spatial and EF abilities were found to be significantly associated with SES. While around 6% of variance in working memory was explained by SES, a much larger effect size was found for language ability, with SES accounting for around 30% of the variance (Noble et al., 2007). In terms of associations with other life outcomes, one aspect of EF (working memory ability) has been shown to predict maths achievement in school, over and above the effects of intelligence, affective variables and family factors (Monette et al., 2011).

Furthermore, it is possible that specific aspects of the lower SES household have different patterns of association with EF and neural structure and functional recruitment supporting EF activity. For example, in a study considering the EF-related behavioural and neurobiological impacts of SES among adolescents, low parental education was found to be associated with poorer WM performance and altered neural activity patterns when undertaking WM tasks (Sheridan et al., 2017). Child neglect was also found to be associated with parentally reported EF issues. However, another aspect of SES, namely level of threat (measured by community violence and abuse) was found to not be associated with EF in the adolescent participants (Sheridan et al., 2017).

Evidence has suggested that associations between SES and EF occur across a range of SES levels. Strongest effects have often been found at the lower end of the SES scale (Hackman et al., 2010), or between low and middle SES groups (Noble et al., 2007). However, significant EF and other cognitive differences have been found across all levels of SES (Sarsour et al., 2011). Often, studies investigating associations between SES and EF have included participants from a limited range of SES, or have reduced their measures down to a small number of SES classifications (two or three) in analysis (Lawson et al., 2018). This reduces the ability to identify potential effects that exist between particular levels of SES. This study will investigate associations between SES and cognition across a wide range of SES levels, which may help to better identify the pattern of SES - EF relationships.

Previous studies investigating associations between SES and EF have often only used a single measure of EF, indeed in their meta-analysis paper, Lawson Hook and Farah (2018) found only six studies out of twenty-five where EF was estimated by multiple measures. Studies using only a single measure of EF cannot capture the complexity of any relationships between individual measures of EF and SES. Furthermore, many previous studies are lacking in SES variability and/ or classified their participants into low and high SES categories only. 15 out of 25 studies identified by Lawson Hook and Farah (2018) were considered to have ‘meaningful’ or ‘substantial’ SES variability. This makes it difficult to identify where along the SES spectrum EF differences might lie, and may result in null findings if samples of participants of the SES levels were differences exist are not included. Lawson et al. also suggest that most previous research studies have also lacked power to control for fluid intelligence or IQ , so it is impossible to identify from the literature whether SES and EF have any unique association over and above that between SES and IQ. As discussed in **Chapter 1,** EF and fluid intelligence are closely related topics, and the ability to control for IQ or another fluid intelligence measure would be useful to see whether SES and EF have any unique variance once IQ is accounted for. Another issue in SES - EF research is that very poorly performing participants are often excluded, for example if they cannot complete the tasks, or if there are too few participants of a particular SES category to compare with others. Exclusion criteria may therefore unintentionally exclude the lowest SES participants, so it is possible that the range of SES in the sampled participants is reduced, leading to inaccurate conclusions about the effects of SES.

3.2.4 EF as a Mediator of links between SES and later life outcomes

EFs are also independently predictive of later life outcomes, such as mental and physical well-being. It has been proposed that executive functions may act as a mediator between parental factors such as SES and academic achievement. In a longitudinal study, Devine et al. (2016) demonstrated that EFs mediate the effect that parental factors associated with SES (such as parenting style and quality of home learning environment) have on children's early academic achievement. General cognitive ability had no such mediating effect. A study conducted in the USA by Lawson and Farah (2017) showed that EF mediated links between SES and longitudinal maths progress in children aged between 6 and 15 years. Another longitudinal study, this time of 5-6 year olds, found that EF at baseline assessment explained SES differences in academic achievement at follow-up – and suggested that early EF may act as a mediating factor of associations between SES and later academic achievement (Rosen et al., 2020).

In a study that attempted to replicate these findings in a cross-cultural context, found differing results by gender and location: EF skills mediated associations between SES and numeracy skills in UK males, but not males in Hong Kong or females in the UK (Ellefson et al., 2020). Another cross-cultural study showed that EF was a mediator of associations between SES and academic achievement in children in some locations but not others: in Mongolia, EF mediated associations between SES and language, literacy and maths skills; in Vanuatu, EF mediated SES-maths and language disparities; whereas in Cambodia, EF was not a significant mediator of any links between SES and academic achievement (Sun et al., 2018).

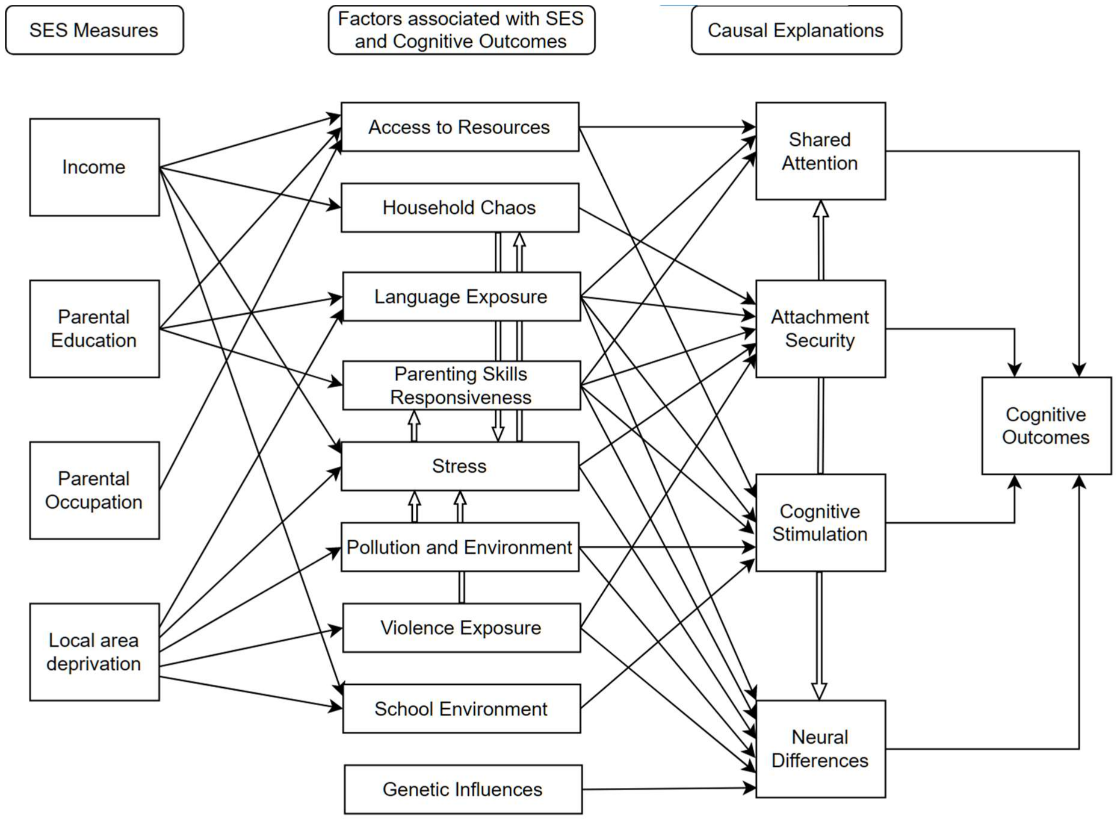
3.2.5 Possible mechanisms driving SES-EF relationships

Early SES is associated with a variety of cognitive and other important outcomes in later life. In a review paper, Duncan and Magnusson (2012) highlight that intervention and quasi-experimental studies have shown that directly increasing household income (i.e. by giving lower income families money) and improving maternal education can improve cognitive outcome measures in children.

These findings indicate that direct intervention to increase specific measures of SES can have causal effects on children's cognitive development. However, the mechanisms of such effects remain unclear. What is it about having more money in the household or a better education for the mother that causes a child to have better cognitive outcomes? Put another way, what are the possible mechanisms by which SES might impact on cognitive development?

Overall links between SES and cognitive outcomes have been well-established in the literature, as already discussed in this section. However, the mechanisms by which these associations might occur are as yet underexplored, with a few general theories in the literature and few specific studies exploring the extent of the potential mechanistic links. To attempt a summary of the literature, we have created a network diagram (**Figure 3.1**) to summarise where possible links exist between specific SES measures, intermediary factors that might be directly altered by these individual SES indicators, and then how these factors might be grouped into more general explanatory groups and have causal links to cognitive outcomes. Firstly, we will cover research at a lower level of explanation, exploring aspects of SES and the specific factors that differ between individuals or households of different SES levels, and highlighting associations between SES measures and these factors. Secondly, we will discuss a higher level of explanation using more general theories to link these lower-level causal factors to broader developmental processes, which might form the basis for mechanistic explanations of SES-cognition relationships. The interrelations amongst the different SES measures and amongst the possible causative factors make drawing out individual pathways very complex. Evidence for some of these possible routes is discussed in this section where this was available – others of these potential routes are speculative, and are yet to be explored in the literature.

Figure 3.1 Possible routes of causation linking SES with Cognition



[a.] ***Arrows indicate directions of possible causal relationships between SES, cognitive outcomes, and associated factors.***

Income

Income here is used as a shorthand for any household or individual parental monetary measures, including total money available to household, household and/or parental income, or poverty measures. Greater income is associated with increased access to resources (Sarsour et al., 2011), reduced household stress (Eamon, 2005) and decreased household chaos (Rosen et al., 2020). These links could be causal, i.e. the income itself can directly influence these factors, and these factors might be mechanistic pathways by which SES exerts influence on cognitive outcomes.

Access to resources is made easier with higher income, by being able to buy more child-related items for the household and afford to take more trips out of the house, both of which are associated with improved cognitive outcomes (Sarsour et al., 2011). Stress and household chaos can be reduced by having a greater income by reducing likelihood of some potential stressors such as money worries, and reducing the chances of certain measures of household chaos such as frequent house moves. It takes cognitive and emotional resources to deal with difficult life circumstances which may result from low income or poverty, perhaps reducing parental capacity to interact and intervene with the child and their developmental processes (Eamon, 2005), or impacting on the child's neural developmental processes via biological mechanisms such as increased cortisol levels and dysregulation of the HPA axis (Andrews et al., 2020). Some measures of household chaos (such as frequent house moves or household structure changes) are negatively associated with cognitive outcomes (Rosen et al., 2020).

There are also potential causal links between income and other measures of SES, for example, higher income might be used to live in more desirable areas, thus reducing local area deprivation. Parental occupation and education might cause increases to household income.

Parental education

Mother's education is associated with parenting quality (P. L. Morgan et al., 2009). Parenting quality here was measured by an observation of the quality of interactions between child and caregiver in play situations, and by a self-report assessment of aspects of home life such including aspects of cognitive stimulation: parental activities (frequency of responding to and interacting with child, taking child on errands or to public places), access to resources (e.g. books and toys in the home), and safety and supportiveness of home environment. This suggests that parental education, in particular mother's education, might be causative of improvements in the child's environment that might subsequently lead to improvements in their cognition (see later section on Cognitive Stimulation theory for more evidence for these factors' relationships to cognitive outcomes). Intervention studies have shown that a parental education programme about parenting skills can improve EF outcomes in children, suggesting that parenting skill might be a factor in causing the link between parental education levels and cognitive outcomes (Neville et al., 2013). Coupled together, these findings suggest one possible route by which parental education might influence cognitive outcomes – better educated parents have higher quality interactions with the child, which in turn is associated with improved cognition.

Children of better educated parents are exposed to more complex language in the home. Parental education and occupation levels account for significant variance in language development, by as early as 18 months old (Fernald et al., 2013). Parental education may be also causally associated with factors such as access to resources. For example, better educated parents are more likely to have more books and toys in the household (P. L. Morgan et al., 2009). This is perhaps as they better understand the importance of these factors in improving children's language outcomes, or perhaps because higher parental education is itself associated with greater household income thus providing improved access to resources. Furthermore, parents with longer educational backgrounds are more likely to be more involved in learning at home, and this involvement is positively associated with children's academic performance in school. This effect is over and above that of child's intelligence

(Topor et al., 2010). These causal links are however somewhat speculative, and deserve further investigation.

Parental occupation

As already mentioned, this measure could be considered as having influence in determining household income, so therefore it might be indirectly causally associated with the factors listed in the income section above. Parental occupation, along with education, is shown to be associated with more complex language exposure in the home, and with language development of young children (Fernals et al., 2013). However other mechanistic links between occupation class and cognitive factors remain unclear. One issue at hand is the inter-related nature of various SES measures. Occupation class is strongly correlated with household income, and with parental education. Occupation itself may well not be causative of the observed associations between parental occupation and children's cognition rather, perhaps improved parental education results in increased occupation class, therefore the causal relationship is in fact between parental education and children's cognition; or that higher parental occupation class results in increased household income, which could then impact on the above factors acting to improve children's cognition. Again, the potential mechanistic links here are somewhat speculative and deserve further investigation.

Local area deprivation

Higher local area deprivation is associated with increased pollution exposure (Evans, 2004), reduced access to greenspace in the local environment (Maes et al., 2021), increased violence exposure (Lambert et al., 2017) and might also be associated with increased experience of stressful events, changes in language exposure (outside of the household in particular), and the type of school environment a child might experience.

Lambert et al. (2017) investigated the association between violence exposure and EF and emotional control abilities in 16-17 year olds. They found violence exposure was negatively associated with emotional inhibition (hot EF), but not with neutral cognitive control abilities (cool EF). Violence exposure and cognitive stimulation deprivation were associated with poorer performance on an emotional inhibition task, and poorer automatic emotion regulation. As a mechanism, they suggest that increased exposure to threat and the lack of a safe environment makes it more difficult for children to discriminate between threat and safety cues, so they do not as frequently practice extinction of fear responses when these are not appropriate, which in turn affects their development of broader emotion control processes.

Environmental issues including pollution, exposure to toxins, and lack of access to green space are increased in more deprived areas. Air quality in the local environment tends to vary in a positive relationship with socio-economic status. This trend has been observed across England, with more deprived areas having greater levels of particulate matter and nitrogen dioxide pollution in the air (Fecht et al., 2015). A recent review of the literature on associations between air quality and cognition suggests that both particulate matter and nitrogen oxides are associated with executive function (in particular working memory) in children (R. Thompson et al., 2023). A meta-analysis considering effects in adults showed that small increases in exposure to nitrogen dioxide (NO_2) and small particulate matter ($PM_{2.5}$) were associated with small decreases in cognitive battery performance (R. Thompson et al., 2023). Increased rates of asthma and other breathing issues, could be a causal mechanism by which air quality could impact on cognitive developmental outcomes, but this mechanism is speculative.

Lower SES children are also more likely to have increased levels of lead in their blood. Lead is a neurotoxin that is not excreted by the body, therefore builds up over time. Increased levels of lead in the blood is known to affect IQ, academic achievement, and reading ability (Evans, 2004). The mechanism by which lead exposure might lead to decreases in cognitive abilities could be via some kind of interference with neurodevelopmental processes, but remains unclear.

Access to greenspace, and in particular woodland, in the local area has also been shown to be associated with cognitive and mental health outcomes, in a paper that used our same SCAMP cohort as its sample (Maes et al., 2021). The mechanistic pathways by which the association between greenspace and cognition might occur is as yet unknown.

Shared Attention

One potential mechanism of action of SES upon later cognitive development is via development of joint attention abilities early in life. Joint attention refers to the ability of a child to respond to cues for shared attention from others, such as redirecting attention based on another's eye gaze or pointing. Responses to joint attention cues are a foundational developmental milestone, and are implicated in subsequent development of language and a variety of cognitive abilities. For example, responding earlier in infancy to more subtle joint attention cues positively predicts later language and cognitive developments (Tomasello et al., 2005). Other research has shown that executive functioning in later adolescence is negatively associated with childhood attentional problems (Friedman et al., 2007). Joint attention could therefore be considered a precursor to EF development.

SES is associated with development of joint attention abilities in infants aged 8-18 months (Reilly et al., 2021). Specifically, family income (as measured as a percentage of the federal poverty line in the USA) and parental education (self-reported furthest education level achieved by participating parent) were positively associated with ability of the infant to respond to more subtle joint attention cues, after accounting for age and depressive and anxiety symptoms in the parent. Although not specifically investigated in this study, the authors suggest three strands of mechanisms by which SES might influence development of shared attention ability: home environment, parenting behaviour, and language exposure. Both measures of SES used (income and parental education) might influence home environment in ways that are already identified as influencing development, for example, income could: increase cognitive stimulation in the environment (e.g. number of books and toys); reduce household chaos; and reduce stress levels. Parental education could influence parenting behaviour, and the complexity of language exposure the child receives. Another study suggests that earlier maternal SES and development of pointing ability predict later language ability at 18 months, suggesting a role for joint attention skills in the association between SES and later cognitive abilities (McGillion et al., 2017).

The development of shared attention mechanisms might therefore be a pathway between these lower-level factors that SES has some kind of direct impact upon and resultant cognitive outcomes. However, specific evidence linking language exposure and parenting behaviours to development of joint attention abilities is mixed, and therefore these mechanisms are somewhat speculative and need further investigation. Development of shared attention mechanisms could also be influenced by level of cognitive stimulation (or perhaps levels of some specific aspects of cognitive stimulation) in the household, such as by the quality and amount of caregiver interactions.

Attachment Security

There is significant evidence that household factors that are related to attachment are also related to children's EF development. Experience of adversity early in life is associated with poorer EF abilities in children who have been adopted internationally from orphanages (Hostinar et al., 2012). Household stress also affects the quality of relationships between parents and children, and also is associated with children's later EF ability (Hackman et al., 2010). In a meta-analysis paper, household chaos was found to correlate negatively with EFs in childhood. Instability in the household was a significant predictor of EFs ($r=.17$), and household disorganisation was also significant but showed a much smaller relationship ($r=.06$) (Andrews et al., 2020). As a mechanism, Andrews et al. suggest that attachment security might be disrupted by household chaos and poorer parenting behaviours which might in turn affect EF development – though this hypothesised link between attachment security and EF is somewhat speculative and requires more investigation.

Cognitive Stimulation

Various factors that may be grouped together under the umbrella of cognitive stimulation are known to vary with SES. The factors include things like access to resources, such as books, toys and experiences; quality and quantity of language exposure and parental interactions; and quality of schooling. The theory has good surface parsimony, in that if we reduce the cognitive stimulation received by a child it would appear likely that their cognitive development would be negatively impacted through neuroplasticity processes as they are not receiving the required inputs for their cognition to develop normally.

Access to resources is one aspect of cognitive stimulation that has been shown to vary with SES, and have associations with cognitive outcomes. Families with higher SES tend to have greater numbers of books and toys within the household (Christensen et al., 2014). Furthermore, children from parents with greater household income are also likely to access a greater number of activities outside the home, such as sports and music lessons, visits to the museum or park, or joining parents on errands. (Sarsour et al., 2011). Research has shown associations between cognition and measures of home environment quality. Often in research, access to a wide range of child-related resources are grouped together into an overarching construct or factor. Christensen et al. (2014) used factor analysis on five home environment measures and found these grouped together as a single factor, and found this overall factor moderated the relationship between SES and general cognition. As access to multiple resources are often correlated with each other, is therefore difficult to disentangle any potential individual effects of access to any particular resources.

To relate access to resources back to the theory of cognitive stimulation, children who have greater access to resources will experience increased cognitive stimulation by using these resources. For example, having more books in the household might lead to the child being read to more often; which increases the stimulation of neural regions associated with reading, speech and language, focus and attention; which in turn results in which improvements in language ability, better and/or earlier development of shared attention abilities, and also may increase the attachment to their caregiver by experiencing more frequent positive interactions. So the book on the shelf does nothing in and of itself – but having access to the book increases the frequency of it being used in daily life, which in turn might exert influence over a wide variety of developmental pathways.

Language exposure in the home is another aspect of cognitive stimulation that varies with SES and might impact on cognitive development. Quantity and complexity of parental language exposure in the pre-school years vary with SES, with a greater variety of words and more complex sentence structures used in higher SES households (Hoff, 2003). Variations in maternal language exposure associated with SES are in turn associated with differences in language and vocabulary development in early life (Hoff, 2003). Vocabulary diversity and complexity of maternal language in early childhood have also been shown to be associated with children's later EF. These aspects of maternal language were also associated with children's vocabulary diversity, which itself mediated the association between maternal language and children's EF (Daneri et al., 2019). This suggests that maternal language and the resulting language abilities of the child are associated with their EF development. Other studies have also indicated that EF and language skills are positively associated across childhood (Gooch et al., 2016). These findings suggest a possible mechanism of action linking family SES and children's EF skills: increased parental language complexity in early childhood occurs in higher SES households, and this impacts on children's early language abilities, which in turn could impact their EF development.

Quality of school environment also varies with SES, and might influence cognitive stimulation and impact upon cognitive developmental processes. Independent schools generally have improved classroom factors, with better pupil behaviour, discipline and more positive relationships between teachers and pupils (Crosnoe et al., 2004a) along with better access to classroom resources, computers, smaller class sizes and higher funding per pupil (Crosnoe et al., 2004b). Crosnoe et al. indicate these factors are associated with improved academic outcomes. This could provide a mechanistic explanation of how school type might affect student outcomes.

Research has investigated the effects of independent vs. state schooling on academic outcomes, with somewhat mixed findings. Private schools show significantly better results at GCSE (Ndaji et al., 2016). However, the cohorts attending private and state schools are quite different at entry, with those attending private schools more likely to be higher SES students and have higher prior academic ability. Once students' prior academic ability, local area deprivation level of home address and gender at school entry were accounted for, the impact of private education on academic outcomes was reduced, though GCSE outcomes were still significantly better than at state schools (Ndaji et al., 2016). Other research in the UK comparing the effects of faith-based primary schools found that the greater performance of pupils in these schools was entirely explained by pre-existing characteristics at school entry (Gibbons & Silva, 2011).

In terms of cognitive outcomes, large-scale longitudinal research in Australia (Nghiem et al., 2015) found no significant differences in cognitive abilities between pupils attending state or private primary schools, once pupil's abilities at school entry were accounted for. These findings suggest that cognitive processes are not directly improved by independent school attendance. Alves et al. (2017) used SEM latent variable analyses to compare effects of parental variables (parent's schooling and socio-economic level) vs school effects (community and school type). They found both latent variables had impact on academic achievement, but only family influenced cognitive performance. Family factors explained a greater proportion of variance in cognitive and academic outcomes in primary school than did the type of school attended – further suggesting that it is not the school type itself that influences cognitive development, rather the pre-existing and continuing family effects that differ between independent and state school cohorts.

Sarsour et al. (2011) studied associations between SES (assessed with a composite measure of income, wealth, maternal education and parental occupation) and three domains of EF – cognitive flexibility, inhibition and working memory – in children aged 8-12, and found that family SES was associated with all three domains of EF. They investigated whether home environment was a mediator of these SES-EF effects. Home environment was assessed across eight areas: physical environment, enrichment activities, parental responsivity, encouragement of maturity, emotional climate/acceptance, learning materials and opportunities, family companionship, and family integration. A composite home environment measure was associated with all three components of EF, and overall home environment quality also partially mediated the association between SES and inhibition. Specific aspects of the home environment, namely access to a variety of learning materials and enrichment activities, were also associated with EF, and of the specific home environment measures assessed in the study they found that enrichment activities, parental responsivity and family companionship partially mediated associations between SES and working memory and inhibition, but no aspect of the home environment mediated the association between SES and cognitive flexibility. The authors suggest that a mechanism by which these better home environments result in improved cognitive abilities due to increased cognitive stimulation.

One issue with a cognitive stimulation explanation of SES-cognition links is that the term cognitive stimulation is somewhat ill-defined, with different researchers using widely varying measures of this concept. It is not clear whether all of the aspects of cognitive stimulation are tapping the same concept; or indeed whether specific elements of cognitive stimulation in early life have dissociable impacts on later cognition.

Neural Differences

SES is associated with neurodevelopmental differences which may underpin the relationships observed between SES and cognition. There is significant evidence linking SES with neural structure and function (Hackman & Farah, 2009). Firstly in terms of structure, total cortical surface area is reduced amongst children from lower income families, and children whose parents have had fewer years of education (Noble et al., 2015). At the lower end of the income scale, greater differences in cortical area were observed for relatively small increases in income, where at the top end of the scale the same increase in income showed smaller increases in cortical area. This suggests those at the lowest end of the income scale exhibit the greatest sensitivity to SES in terms of cortical area. In a structural MRI study, lower SES, in particular lower parental education, was associated with lower total cortical surface area in 14-19 year old participants (Judd et al., 2020).

The picture around total brain or cortical volumes is somewhat unclear – some studies find increased volumes of grey matter in certain brain areas among higher SES groups (Noble et al., 2005), and lower volumes of grey matter in other regions (Jednoróg et al., 2012, Jensen et al., 2015). Greater cortical thickness (across the whole brain) has been found to be associated with higher family income (Noble et al., 2015). Grey matter volumes in the hippocampus and frontal regions are associated with maternal education and occupation (Jednoróg et al., 2012). Experience of adversity and stressful events early in life is associated with lower grey matter volumes in the anterior cingulate cortex, and higher grey matter volume in the Precuneus measured during early adulthood (Jensen et al., 2015).

Reduced white matter tract volumes have also been noted in children from lower SES backgrounds (Ursache & Noble, 2016). Lower SES participants with reduced white matter tracts performed worse in cognitive flexibility tasks in this study. However, interestingly, higher SES children who had lower white matter tract volumes still showed good performance in cognitive flexibility tasks. This suggests that high and low SES participants may not use the same neural pathways when completing demanding tasks. Overall, analyses looking at associations between SES and the structure of specific brain regions suggest the greatest SES gradients in structure are found in brain regions associated with cognitive control, language, spatial skills and numeracy (Lipina & Posner, 2012; Noble et al., 2015).

Functionally, functional Magnetic Resonance Imaging (fMRI) and Event Related Potential (ERP) studies indicate that lower SES participants recruit different brain regions to higher SES participants while undertaking tasks involving attention and novel rule learning, with reduced activity in prefrontal regions and greater activity in other areas associated with worse performance in novel rule tasks (Hackman et al., 2010). Changes to patterns of neural activity in regions associated with EF have been found to be associated with certain aspects of SES. Lower parental education has been shown to be associated with lower efficiency of neural recruitment during a WM task (Sheridan et al., 2017). In an fMRI study, Sheridan et al. (2012) indicate that neural activity in the PFC underpinning performance in a complex EF task (a stimulus-response mapping task, similar to the DCCS see **Table 1.1**) differs between high and low SES children. Lower SES children had greater activation in frontal regions and the amygdala than did higher SES children in novel task conditions, where EF would be most required. The altered PFC activation patterns reflected poorer EF task performance of the lower SES children (Sheridan et al., 2012).

Preadolescent children from lower SES backgrounds show different patterns of Electro-Encephalogram (EEG) and ERP activity in a selective attention task: despite similar behavioural performance in the task, lower SES children showed a preference to attend equally to relevant and irrelevant environmental cues, and exhibited evidence (via increased activity in frontal regions) of more effortful control being exerted during the task (D'angiulli, Van Roon, et al., 2012). SES-related differences in functional recruitment of brain regions is consistent with other research that shows that even when task performance is equal, lower and higher SES participants display differences in neural activity while performing EF tasks (Hackman & Farah, 2009).

Given the above findings, it seems likely that neural differences are likely involved in causal pathways by which SES is associated with cognition. However, the mechanisms by which these neural differences actually occur are not clear. Two factors that relate to both neural changes and SES levels have been identified in the literature: stress and pollution exposure. These will be discussed below. It is also the case that via general neuroplasticity, any of the other discussed causal pathways (such as cognitive stimulation, attachment security or shared attention) might also influence developing neural pathways that underpin cognition, but the mechanisms associating these remain unclear.

Across their lifetimes, people growing up in lower SES households report more frequent experiences of stressful life events, such as home moves or changes in household make-up, and it has been suggested that the biological response over time to these stressors might explain associations between SES and health and cognitive outcomes (Hackman & Farah, 2009). Importantly, chronic stress exposure is thought to lead to dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis, in turn leading to neural changes in PFC and other regions underpinning EF, such as the hippocampus which is heavily involved in memory function, and the amygdala which is strongly implicated in emotional processing (hot EF) (Andrews et al., 2020).

As discussed in the above section on local area deprivation, children from lower SES households tend to have higher levels of lead in the blood (Evans, 2004). Higher concentrations of lead in the blood during childhood causes damage to developing neural structures in the PFC, hippocampus and cerebellum (Sanders et al., 2009) – important areas of the brain underpinning a wide range of cognition including EF. Thus exposure to pollutants in the local area and the subsequent disruption of neural developmental pathways might be one mechanism by which SES impacts on cognition.

Genetic influences

Finally, it is possible that genetics may play a role in how SES influences cognitive and life outcomes. The suggestion that genes may play a role in development of links between SES and other life outcomes does not suggest that genes are entirely determinative of outcomes, or that genes are the same as destiny (Sarsour et al., 2011). One study that has linked genes to SES and life outcomes is by Judd et al. (2020). They found that SES was correlated (r= .27) with frequency of certain genetic patterns across the genome, specifically *EduYears*-GPS. *EduYears*-GPS is a genome-wide polygenic score that has been found to be associated with academic and cognitive outcomes (around 9% of variance in academic achievement at age 16 is explained by *EduYears*-GPS), elements of brain structure, and with family SES (with around 7% of variance in SES is explained by *EduYears*-GPS) (Selzam et al., 2017). Another study showed that family SES modifies the genetic heritability rate of children's IQ, with SES accounting for significant portions of variance in IQ among lower SES households but very little of the variation in IQ in high SES households (Turkheimer et al., 2003). The causal mechanism of exactly how small genetic differences could exert influence over cognition remains unclear, but it is likely that neural differences would be involved in this pathway.

Deficiency or adaptation to different environments?

Many papers exploring the cognitive impacts of SES often explore areas in which the lower SES participants exhibit poorer performance in tasks. This is often related (explicitly or implicitly) to the idea that children from lower SES households are in some way deficient – they experience a deficit in performance relative to their higher SES peers. Another interpretation of the same data might indicate that children from lower SES households display adaptive behaviours to a different environment to that the higher SES households. For example, lower SES children are more likely to experience unpredictable or chaotic environments at home. This could result in an adaptive strategy by which the children are more vigilant to a variety of cues which could indicate potential disruption or threat in the environment, and must exert greater cognitive control to ignore task-irrelevant stimuli. This could result in either worse accuracy or longer response times in inhibition tasks, either due to greater attention being paid to irrelevant stimuli, or to more effortful control mechanisms being required to focus on the single task in hand and ignore irrelevant stimuli. This is an important consideration when interpreting data that explores the association between SES and EF.

3.2.6 Why study the relationship between SES and EF?

Given the heterogeneity of previous research, it remains unclear exactly what the nature of the link between SES and EF is. This is an important area to investigate for a number of reasons. SES is a potential source of individual differences, and cognitive developmental science is interested in identifying these kinds of factors generally (Foulkes & Blakemore, 2018). More broadly in developmental science, we should probably consider impact of SES on our participants, so understanding the magnitude and specific areas of cognition that are affected by SES, and by what specific aspects of SES are they affected, will be useful so that these measures could be used as covariates in future research. There has been a move in cognitive science recently to try to extend our findings beyond 'WEIRD' (Western, educated, industrialized, rich and democratic) participants, and the consideration of the impacts of SES is of importance in this endeavour (Henrich et al., 2010).

EF predicts a wide variety of life outcomes, and SES independently does too. Research has indicated that SES and EF are also related, and that EF skills may act as a mediator of SES related outcomes, with better EF reducing the impact of low SES on things like academic achievement. If we could identify the nature of links between SES and EF more precisely, it might be possible to design interventions aimed at improving EF in low SES groups, in order to improve more general life outcomes. The reverse idea is also possible: identification of specific SES components which have significant links to cognition could suggest targets for intervention studies – if certain SES measures are better at predicting cognitive outcomes, we could target these SES areas for intervention to improve outcomes. One intervention study found that lower SES children showed poorer selective attention skills, and that parents in these families also showed poorer attention-related parenting skills - this led to the development of a targeted attention-based intervention for parents and children which improved the children's selective attention (Neville et al., 2013). Outcomes of research investigating links between SES and EF may highlight the impacts SES has on individual development, which may have wider policy implications.

Effects of poverty on neural and behavioural development have been noted widely, but it is still unclear whether these effects exist across the whole spectrum of SES, or if instead effects are driven only by differences between for example very low SES and others, or between those in poverty and those not. The prolonged development of EF may make it particularly susceptible to changes due to SES effects across childhood and adolescence (Best et al., 2011). EF ability in early life has far reaching associations, and knowing more about the factors that influence the development of EF would be of interest to developmental science generally.

3.2.7 Aims

SES and EF are both independently predictors of children's academic success and other important life outcomes, and they are also correlated with one another. However, previous research has not been able to rule out whether the association between SES and EF may be explained by associations with cognitive ability more generally. Mediation research has suggested a specific role of EF in mediation of the effects of SES on academic outcomes, whereas more general cognitive ability does not have this mediating effect, making it of interest to research associations between SES and EF in more detail and to assess to what extent associations may be general and not specific to EF.

Both SES and EF are complex concepts and are measurable in a variety of ways. This study uses multiple measures of both SES and EF and attempts to unpick any specific associations between aspects of SES and EF. SES was measured through parental occupation, parental education, local postcode deprivation and school type, while EF was measured through TMT, BDS, Corsi, SWM, and CPT. The SES measures used cover a wide range of SES levels, allowing investigation of the associations of SES and EF across the spectrum of SES. The data used are from the first data collection in the SCAMP cohort, when participants were aged around 11-12 years old in school years 7-8. Analyses of this cross-sectional data first assessed associations between specific EFs and specific SES measures and investigated where in the SES spectrum were differences in EF observed, and second assessed whether these associations remained significant when covarying for fluid intelligence. This study provides sufficient power to account for CFT in the analysis of EF-SES links, which has not been common in the literature. It aims to identify specific areas of EF that are most associated with SES in general, and to see if there are specific measures of SES that are associated with specific aspects of EF. The study will consider participants across a wide range of SES, rather than considering only those in poverty vs. not in poverty.

3.3 Methods

3.3.1 Participants

Data are taken from the baseline SCAMP assessments. Exclusions based on age and specific task-based exclusions were applied (described in **Chapter 2**). The largest sample for the analyses presented in this chapter is n = 6,591, with a complete case sample of n = 1,428. Cross-tabular n's for cognitive tasks and SES measures analysed here are shown in **Table 3.2**.

Table 3.2 Participant numbers with data for cognitive tasks and socio-economic status measures

Complete Case *n* = 1,428	Total *n* [a]	TMT *n* [a]	BDS *n* [a]	SWM *n* [a]	Corsi *n* [a]	CFT *n* [a]	CPT *n* [a]
Total *n* [a]	6,591	6,424	6,084	6,300	3,791	5,808	1,574
School Type	6,576	6,409	6,072	6,228	3,781	5,796	1,571
Carstairs	6,379	6,226	5,913	6,111	3,717	5,643	1,555
Father Occupation	4,944	4,828	4,635	4,802	3,100	4,607	1,334
Mother Occupation	4,271	4,187	4,017	4,145	2,690	3,981	1,201
Father Education	3,990	3,897	3,740	3,872	2,540	3,710	1,126
Mother Education	4,160	4,065	3,894	4,038	2,598	3,856	1,164

[a] Final n's for each task. BDS: backward digit span; CFT: Cattell's culture fair test; Corsi: Corsi block task; CPT: continuous performance task; SWM: spatial working memory; TMT: trail making task.

3.3.2 Measures

Cognitive measures

EF is estimated from five cognitive tasks: Trail making task (TMT), backward digit span (BDS), spatial working memory (SWM), Corsi block task, and continuous performance task (CPT). Scores in Cattell's culture fair test (CFT) are also used as an estimate of fluid intelligence. Task administration and data cleaning procedures described in detail in **Chapter 2**; a brief reminder of the key task measures is provided below.

The trail making task (TMT) assesses cognitive flexibility or switching. The key measure is calculated by regressing response time in the switching condition on response time in the letters condition. Lower (more negative) scores indicate better performance.

The backwards digit span and Corsi span tasks assess verbal and visuospatial working memory respectively. The key measures for these tasks are estimates of span capacity, with the score calculated as the average of the mean trial length of trials they passed and mean trial length of trials they failed. Higher scores indicate better performance.

The key measures for the spatial working memory (SWM) task are the number of errors committed across the task, and a strategy use measure equivalent to the number of unnecessary changes in search starting location. Lower scores indicate better performance on both measures.

The continuous performance task (CPT) measures sustained attention and inhibitory control. The key measures are: the number of omission errors made (how many targets the participant failed to press the space bar for), reflecting inattention; and the number of commission errors made (how many times the participant pressed the space bar outside of targets), reflecting lower inhibitory control. Lower scores indicate better performance.

The Cattell culture fair task (CFT) is a test of visuospatial reasoning and a proxy for fluid intelligence. The key measure is the total number of items correct across the whole task. Higher scores indicate better performance.

Socio-economic status measures

SES is assessed by six measures: school type; education level of each parent; occupation category of each parent; and the Carstairs postcode deprivation estimate. School type is classified as State or Independent school. This variable was reported by the school. For the parental education measures, participants reported whether their parents attended university in the main assessment questionnaire. Participants also reported their parents' occupations during the main assessment questionnaire. Adolescent reporting of parental occupation level is considered reasonably accurate (Lien et al., 2001), and previous studies have also used adolescent assessment of parental occupation to estimate SES (Richter et al., 2006). Participants were asked a series of questions on parental occupation taken from the National Statistics Socio-economic classification (NS-SEC) (Office for National Statistics, n.d.). Responses were coded as per the 8-class version of the NS-SEC (D. Rose & Pevalin, 2003). Values were then re-coded for this analysis such that a higher number indicates a higher SES category (final codes used in analysis are in **Table 3.3**). Data were coded as Missing (NA) both where the researcher could not tell which category to use from the participant's response or if the participant selected the 'never worked or long term unemployed' response, as there were very small numbers of this type of response in the baseline sample.

Table 3.3 Socio-economic status (SES) parental occupation coding categories

SES Level	Occupation Description
8	Large employers and higher managerial and administrative occupations
7	Higher professional occupations
6	Lower managerial, administrative and professional occupations
5	Intermediate occupations
4	Small employers and own account workers
3	Lower supervisory and technical occupations
2	Semi-routine occupations
1	Routine occupations
NA	Never worked or long-term unemployed

Carstairs postcode deprivation is an estimate of the deprivation level of the postcode area of the participant's home address, relative to the area sampled by the SCAMP study. Home address was reported by the participant during the main assessment. A deprivation index score was estimated for each participant using the Carstairs index (O. Morgan & Baker, 2006). Carstairs index values are calculated nationally for each postcode area, by taking weighted Z-score composite of four key economic indicators (**Table 3.4**) for each geographic area across England, using 2011 census data (Morris & Carstairs, 1991). Scores were then normalised across the SCAMP study sample area, and individual values relative to these normalised scores were calculated for each participant's reported home postcode. Data were then categorised into quintiles, with 1 = least deprived and 5 = most deprived.

Table 3.4 Economic indicators used to calculate postcode deprivation in the Carstairs index

Indicator	**Description**
Male unemployment	The proportion of economically active males seeking or waiting to start work
Lack of car ownership	The proportion of all persons in private households which do not own a car
Overcrowding	The proportion of all persons living in private households with a density of more than one person per room
Low social class	The proportion of all persons in private households with an economically active head of household in partly skilled or unskilled occupations, according to ONS-NSSEC classifications

3.3.3 Missing Data: Multiple Imputation by Chained Equations (MICE)

Data were imputed using Multiple Imputation by Chained Equations (MICE). This is a principled approach to dealing with missing data. It makes use of all the available data to estimate missing values by performing a series of regressions for each individual variable, using all other variables in the MICE model to estimate any missing values. This is preferable to other methods of handling missing data such as analysing complete cases only, since sample size and statistical power are increased, or mean imputation, since other variables in the model are used to make more accurate estimates of missing values (Azur et al., 2011).

Imputation was carried out in R version 4.0.0, using packages mice_3.9.0 and tidyr_1.1.0. First, missingness pattern of the data was inspected. This revealed over 50% missing data for the CPT task, which was therefore omitted from the MICE procedure. Other measures broadly fit the assumptions of independence and of Missing at Random (MAR). All variables to be included as predictors, outcomes or covariates in the analyses (other than the CPT measures) were included in the MICE procedure (Buuren & Groothuis-Oudshoorn, 2010). Interactions between variables were not included as these were not being investigated in the subsequent analyses (Azur et al., 2011). Multiple imputation was used here. Ten MICE iterations were used to create five imputed datasets (Buuren & Groothuis-Oudshoorn, 2010). The ten iterations were used to achieve adequate model convergence on the missing element estimates within each imputed dataset. Checks of the imputed datasets were made, ensuring adequate model convergence and plausibility of imputed values.

Statistical analyses are then carried out on each of the five imputed datasets individually, and results of these multiple analyses are then combined or pooled in order to reconcile inconsistent results – see Statistical Analyses **Section 3.3.4** for details of pooling methods used for each analysis.

3.3.4 Statistical Analyses

In a first stage of analysis, a between-subjects six-way multivariate analysis of covariance (MANCOVA) was run. MANCOVA is a multivariate omnibus test, which tests for differences between levels of independent variables on a linear combination of multiple dependent variables, whilst also accounting for covariates. An advantage of MANCOVA over conducting multiple univariate tests is that MANCOVA takes into account correlations between multiple dependent variables, has greater power than using multiple univariate tests, and does not inflate the chance of a type 1 error which using multiple univariate tests would (Field, 2013).

MANCOVA was run on five EF measures: TMT, BDS, Corsi, SWM errors and SWM strategy score. Age and was entered as a covariate. Independent variables were the six SES measures (school type, postcode deprivation, each parent's occupation and each parent's education). MANCOVA was carried out on complete cases only, n = 1,428, as SPSS did not have functionality to provide pooled MANCOVA statistics for imputed data. The CPT task was omitted as significantly fewer participants completed this task (**Table 3.5**). Pillai's Trace was selected as this is relatively robust to differences in group sizes (Field, 2013).

Next, univariate follow-up tests of association between specific measures of SES and individual EF task measures were carried out with a series of multiple regressions. Models were conducted for each outcome variable separately: the seven EF measures (TMT, BDS, Corsi, SWM errors, SWM strategy, CPT omission errors and CPT commission errors). A separate model was also run to see whether fluid intelligence (CFT score) was associated with SES measures also – CFT will be used as a covariate in a second set of models to see whether relationships between SES and EF are still present after accounting for CFT.

Six SES measures were entered as predictors (school type, postcode deprivation, each parent's occupation and each parent's education). Models were constructed in two steps: Step 1 with the covariate age, and Step 2 with this covariate plus the six SES variables, using the Enter method. Analyses for the two CPT measures were carried out on the original complete dataset for this task , n = 1,124, as data for this task were not able to be imputed. Analyses for the TMT, BDS, Corsi, SWM errors, SWM strategy and CFT were carried out with imputed data, n = 6,591. Coefficient estimates for the imputed data analyses were pooled following Rubin's rules (Rubin, 1987). Estimates of R^2 and significance were pooled by simple averaging across the imputed dataset analyses – this method is used as it is straightforward to calculate, and provides reasonable estimates of R^2 values, with a slight conservative bias (Ginkel, 2019).

A final set of analyses was conducted to assess whether unique associations exist between SES and EF task performance beyond associations with general intelligence. The above described multivariate and univariate tests were repeated with CFT score as an additional covariate.

3.4 Results

3.4.1 Descriptive Statistics

Descriptive statistics and cross-tabular *N*s for the cognitive and SES measures, are presented in **Table 3.5**. The general trend is for higher SES to be associated with higher scores in all the tasks, however, a notable exception to this occurs when looking at the parental occupation categories, where the poorest performance is sometimes observed closer to middle value of the scale. **Table 3.6** shows the correlations between the outcome variables used in this Chapter. Although almost all the outcome variables are significantly correlated to each other, there are no major issues with multi-collinearity, as the R^2 values are sufficiently low.

Table 3.5. a. Descriptive statistics of performance on the trail making (TMT), backward digit span (BDS), spatial working memory (SWM) and Corsi tasks, divided by levels of socio-economic status measures

	TMT[a]			BDS			SWM Errors[a]			SWM Strategy[a]			Corsi		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
School Type															
State	4,944	1,571	17,208	4644	3.97	0.87	4849	30.57	13.82	4849	10.75	2.86	2771	4.92	0.77
Independent	1,465	-5,267	11,398	1428	4.50	0.86	1439	24.35	12.25	1439	11.63	5.42	1010	5.22	0.77
Carstairs Quintile [b]															
1	910	-4,319	13,073	886	4.36	0.89	897	25.98	12.72	897	11.48	4.90	582	5.21	0.81
2	929	-2,329	14,176	888	4.20	0.89	911	26.61	12.75	911	10.97	4.27	587	5.03	0.76
3	1,099	-859	15,064	1059	4.15	0.90	1075	27.76	12.67	1075	10.79	3.39	668	5.00	0.77
4	1,337	1,010	16,432	1280	4.01	0.86	1305	29.84	13.75	1305	10.79	3.18	785	4.98	0.79
5	1,951	2,615	17,749	1800	3.97	0.88	1923	31.69	14.33	1923	10.87	2.92	1095	4.90	0.75
Father Occupation [c]															
1	300	1,712	15,032	285	3.91	0.81	291	29.63	13.45	291	10.84	3.64	171	4.95	0.75
2	445	875	16,932	430	4.04	0.91	452	29.81	14.11	452	10.88	3.61	259	4.96	0.70
3	379	1,569	17,076	350	4.01	0.79	374	29.77	13.31	374	10.8	3.29	236	4.90	0.73
4	1,310	925	15,947	1,244	3.99	0.86	1307	30.66	13.46	1307	10.97	3.24	823	4.88	0.77
5	312	-396	17,289	295	4.21	0.87	309	28.22	13.18	309	11.43	3.79	194	5.07	0.82
6	604	-860	15,407	585	4.24	0.91	599	27.39	12.97	599	11.18	3.88	401	5.04	0.75
7	1,041	-3,151	13,642	1,024	4.31	0.90	1035	25.97	13.38	1035	10.48	3.77	714	5.20	0.80
8	437	-3,752	12,015	422	4.27	0.86	435	26.45	12.75	435	10.78	3.94	302	5.15	0.82
Mother Occupation [c]															
1	237	-604	17,726	216	3.87	0.82	232	32.70	13.56	232	11.18	2.95	132	4.79	0.70
2	805	-302	14,476	764	4.05	0.87	801	29.18	13.43	801	10.89	3.26	489	4.89	0.75
3	107	548	13,364	105	3.89	0.84	103	30.04	12.07	103	11.26	2.55	54	4.89	0.67
4	190	-232	15,039	178	4.13	0.95	191	29.65	14.69	191	10.34	3.04	122	5.09	0.76
5	529	-1,178	15,679	509	4.18	0.90	526	28.95	13.24	526	10.96	3.82	350	5.07	0.78
6	1,131	-985	14,830	1,093	4.20	0.91	1117	27.33	13.04	1117	10.81	3.78	750	5.06	0.78
7	968	-1,628	15,588	936	4.26	0.90	959	26.99	13.33	959	10.61	3.40	642	5.10	0.80
8	220	-2,767	12,996	216	4.17	0.90	216	26.64	12.15	216	10.43	3.05	151	5.02	0.75
Father Education															

	TMT[a]			BDS			SWM Errors[a]			SWM Strategy[a]			Corsi		
Did Not Attend University	1,209	556	15,986	1,152	4.00	0.87	1191	29.80	13.02	1191	11.08	3.41	758	4.96	0.74
Attended University	2,688	-1,377	15,622	2,588	4.23	0.90	2681	27.71	13.6	2681	10.88	3.87	1782	5.08	0.79
Mother Education															
Did Not Attend University	1,479	491	16,307	1,411	4.02	0.88	1468	30.07	13.18	1468	11.03	3.49	919	4.96	0.76
Attended University	2,586	-1,486	15,359	2,483	4.23	0.90	2570	27.57	13.40	2570	10.91	3.91	1679	5.08	0.79

a Lower scores indicate better performance in these tasks.
b Higher Carstairs quintile indicates greater level of deprivation.
c Occupations classified according to ONS NSSEC-8 categories; recoded such that higher values indicate higher socio-economic status.

Table 3.5.b. Descriptive statistics of performance on Cattell's culture fair test (CFT) and continuous performance task (CPT), divided by levels of socio-economic status measures.

	CFT			CPT Omission[a]			CPT Commission[a]		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
School Type									
State	4,449	12.86	3.93	991	0.66	1.16	991	5.57	7.30
Independent	1,347	14.64	3.44	580	0.36	0.78	580	3.21	3.58
Carstairs Quintile [b]									
1	831	14.19	3.62	278	0.35	0.78	278	3.23	3.66
2	857	14.17	3.67	270	0.39	0.85	270	4.15	5.14
3	997	13.63	3.78	292	0.51	0.92	292	4.21	7.10
4	1,212	13.10	3.89	328	0.66	1.24	328	5.07	6.38
5	1,746	12.54	3.90	387	0.73	1.19	387	6.18	7.36
Father Occupation [c]									
1	280	12.75	3.79	68	0.88	1.14	68	6.16	6.91
2	418	12.89	3.55	96	0.60	1.05	96	5.56	6.00
3	359	12.77	3.89	88	0.58	1.03	88	5.32	6.57
4	1,251	12.95	3.88	293	0.65	1.09	293	5.66	6.44
5	298	13.40	3.68	80	0.51	1.01	80	4.92	6.03
6	574	13.91	3.90	204	0.49	0.93	204	3.85	4.85
7	1,000	14.13	3.69	366	0.37	0.83	366	3.37	4.98
8	427	14.19	3.77	139	0.56	1.25	139	3.83	4.62

	CFT			CPT Omission[a]			CPT Commission[a]		
Mother Occupation [c]									
1	230	12.33	3.57	47	0.79	1.32	47	5.81	5.53
2	761	13.06	3.81	182	0.56	1.06	182	5.04	5.85
3	101	12.66	3.39	16	0.44	0.73	16	6.12	5.64
4	182	13.31	4.34	42	0.40	1.11	42	4.40	5.24
5	505	13.75	3.64	156	0.49	1.04	156	4.78	5.56
6	1,071	13.69	3.76	348	0.44	0.84	348	4.02	6.97
7	921	13.89	3.65	339	0.54	1.08	339	3.80	5.18
8	210	14.20	3.93	71	0.51	0.91	71	6.11	8.16
Father Education									
Did Not Attend University	1,152	12.88	3.76	300	0.56	0.94	300	5.81	6.73
Attended University	2,558	13.80	3.85	826	0.46	0.95	826	3.86	5.03
Mother Education									
Did Not Attend University	1,393	12.84	3.70	354	0.53	1.00	354	5.11	6.36
Attended University	2,463	13.91	3.85	810	0.51	1.02	810	4.13	5.23

a Lower scores indicate better performance in these tasks.
b Higher Carstairs quintile indicates greater level of deprivation.
c Occupations classified according to ONS NSSEC-8 categories; recoded such that higher values indicate higher socio-economic status.

Table 3.6 Pearson's correlations between the outcome variables used in MANCOVA and Multiple Regression Analyses

Variable		1	2	3	4	5	6	7	8
1) TMT	R^2	-							
	N	6424							
2) BDS	R^2	**-.250****	-						
	N	5964	6084						
3) SWM Errors	R^2	**.222****	**-.283****	-					
	N	6173	5880	6305					
4) SWM Strategy	R^2	**.063****	**-.070****	**.451****	-				
	N	6173	5880	6305	6305				
5) Corsi	R^2	**-.252****	**.313****	**-.282****	**-.151****	-			
	N	3731	3560	3704	3704	3791			
6) CFT	R^2	**-.198****	**.352****	**-.290****	**-.140****	**.279****	-		
	N	5678	5406	5646	5646	3791	5808		
7) CPT Omission	R^2	**.116****	**-.126****	.139**	**0.050**	**-.142****	**-.131****	-	
	N	1562	1520	1548	1548	1569	1572	1572	
8) CPT Commission	R^2	**.162****	**-.183****	**.188****	**.099****	**-.223****	**-.178****	**.326****	-
	N	1562	1520	1548	1548	1569	1572	1572	1572

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

3.4.2 MANCOVA: Associations between SES and EF measures combined

Multivariate analyses showed significant main effects for three SES measures: school type (F(5, 1401) = 16.3, p < .001; V = .055, father's occupation (F(35, 7025) = 1.71; p = .006; V = .042) and postcode deprivation (F(20, 5616) = 1.76; p = .019; V = .025) indicating there were significant differences between levels of these SES variables in the combined EF outcome measures, when accounting for age. School type had the largest effect size (partial η^2 = .055), followed by father occupation (partial η^2 = .008) and postcode deprivation (partial η^2 = .006). Age, was significantly related to the linear combination of the EF measures (p <.001), with effect size between partial η^2 = .027. When CFT was entered as an additional covariate, the three significant SES effects remained significant at p < .005, and effect sizes were similar though slightly reduced. The CFT as a covariate was a significant predictor and had the largest effect size of all variables, (F(5, 1400) = 31.85; p <.001; V = . 102; partial η^2 = . 102). Results are summarised in **Table 3.7**, and **Appendix C** has complete results for these models including for the CFT and Age covariate statistics.

Table 3.7 Results of MANCOVA analysis of differences between levels of SES measures in combined EF outcomes, after accounting for age

***n* = 1,428**	**df**	**F**	**Pillai's Trace (*V*)**	***p***	**partial η^2**
School Type	5, 1401	16.33	0.055	**<.001**[a]	0.055
Father Occupation	35, 7025	1.71	0.042	**0.006** [a]	0.008
Mother Occupation	35, 7025	1.27	0.031	0.135	0.006
Postcode Deprivation	20, 5616	1.76	0.025	**0.019** [a]	0.006
Father Education	5, 1401	1.07	0.004	0.375	0.004
Mother Education	5, 1401	1.37	0.005	0.232	0.005

[a] Bold = Significant, and Remained significant at p<.05 after also covarying for CFT

3.4.3 Multiple regression analyses: Associations between specific measures of EF and SES

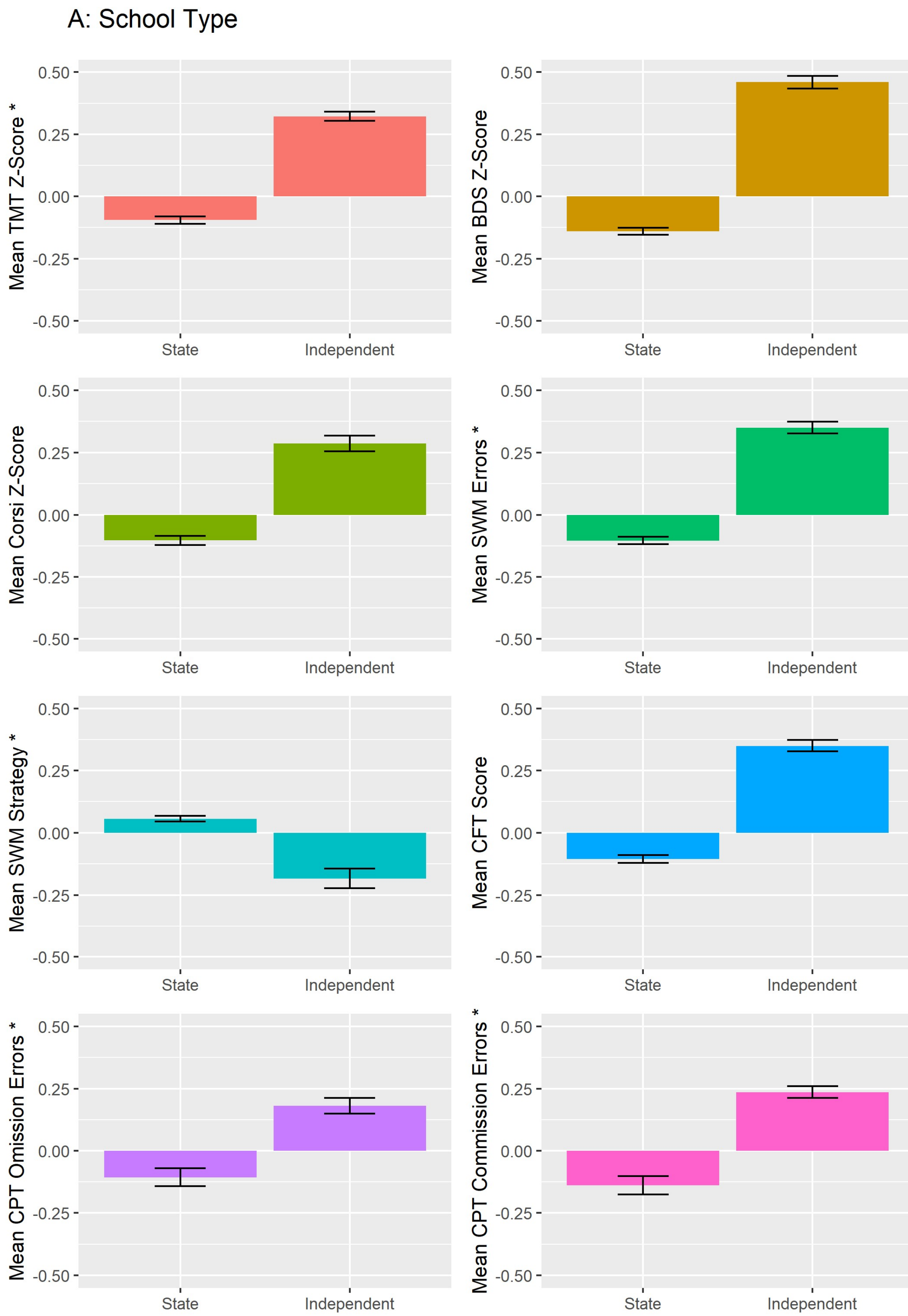
Multiple including SES measures as predictors explained significantly more in variance in all the of the individual EF and CFT measures, after accounting for the effects of age. Higher SES backgrounds predicted better overall EF. **Figure 3.1** shows the relationships between specific EF and SES measures. The contribution of the individual SES variables to each model is shown in **Table 3.8**.

In the first set of follow-up regressions, school type was significantly associated with all seven EF measures, and also with CFT score. Students at Independent schools performed significantly better than their peers in State schools. Postcode deprivation level was significantly associated with TMT, BDS, SWM errors, both CPT error measures, and with CFT score. Participants from areas with greater deprivation performed worse on these tasks. Father occupation was significantly associated with TMT, BDS, and SWM strategy. Mother occupation was associated with BDS and SWM Strategy. with Higher level occupations were associated with better task outcomes. Father education was only associated with BDS and CPT omission errors, and mother education was only associated with BDS and CFT. Here participants whose parents did attend university scored higher on these tasks.

The multiple regression exploring CFT score as an outcome variable, CFT score was significantly positively associated with father occupation, school type, and postcode deprivation, and mother education. When including CFT as an additional covariate, most of the associations observed between EFs and SES remained significant after also accounting for CFT – all the associations between TMT and SES measures (father occupation, school type and postcode deprivation) remained significant; for SWM Strategy all measures remained significant (i.e. those between SWM Strategy and father occupation, mother occupation, and school type); for SWM errors postcode deprivation and school type remained significant; for Corsi the only significant finding remained (with school type); and CPT commissions most associations remained significant (school type, postcode deprivation and father education), but the association with mother's occupation did not remain significant.

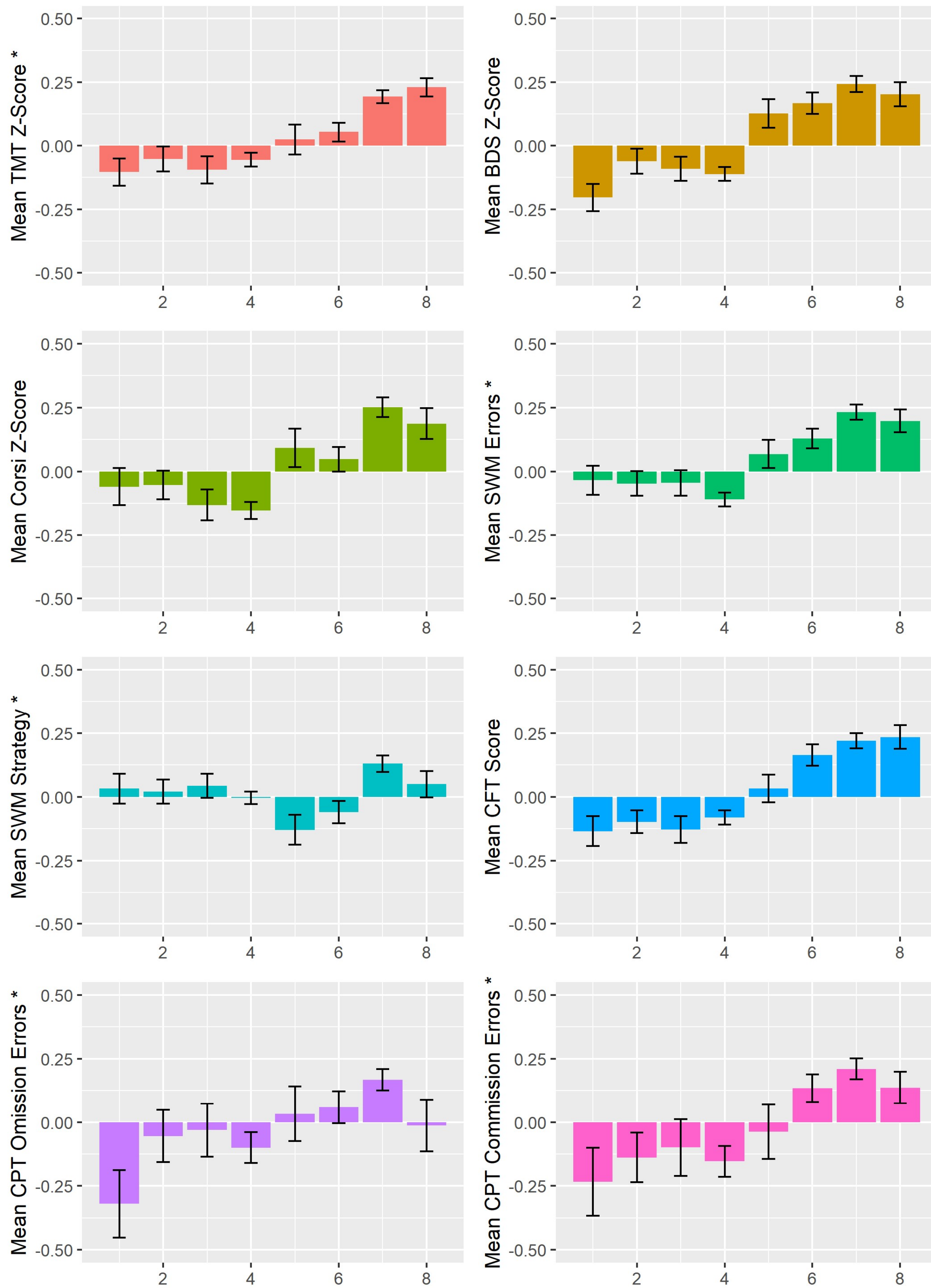
Two tasks were exceptions to this pattern: for with BDS task, here only the association between school type and BDS score remained significant (with previous associations with mother and father occupation, postcode deprivation and mother and father education becoming non-significant) after accounting for CFT score too; and for CPT omission errors both the previously significant associations became non-significant after accounting for CFT score (school type and postcode).

Figure 3.2 Relationships between executive function and socioeconomic status (SES) measures



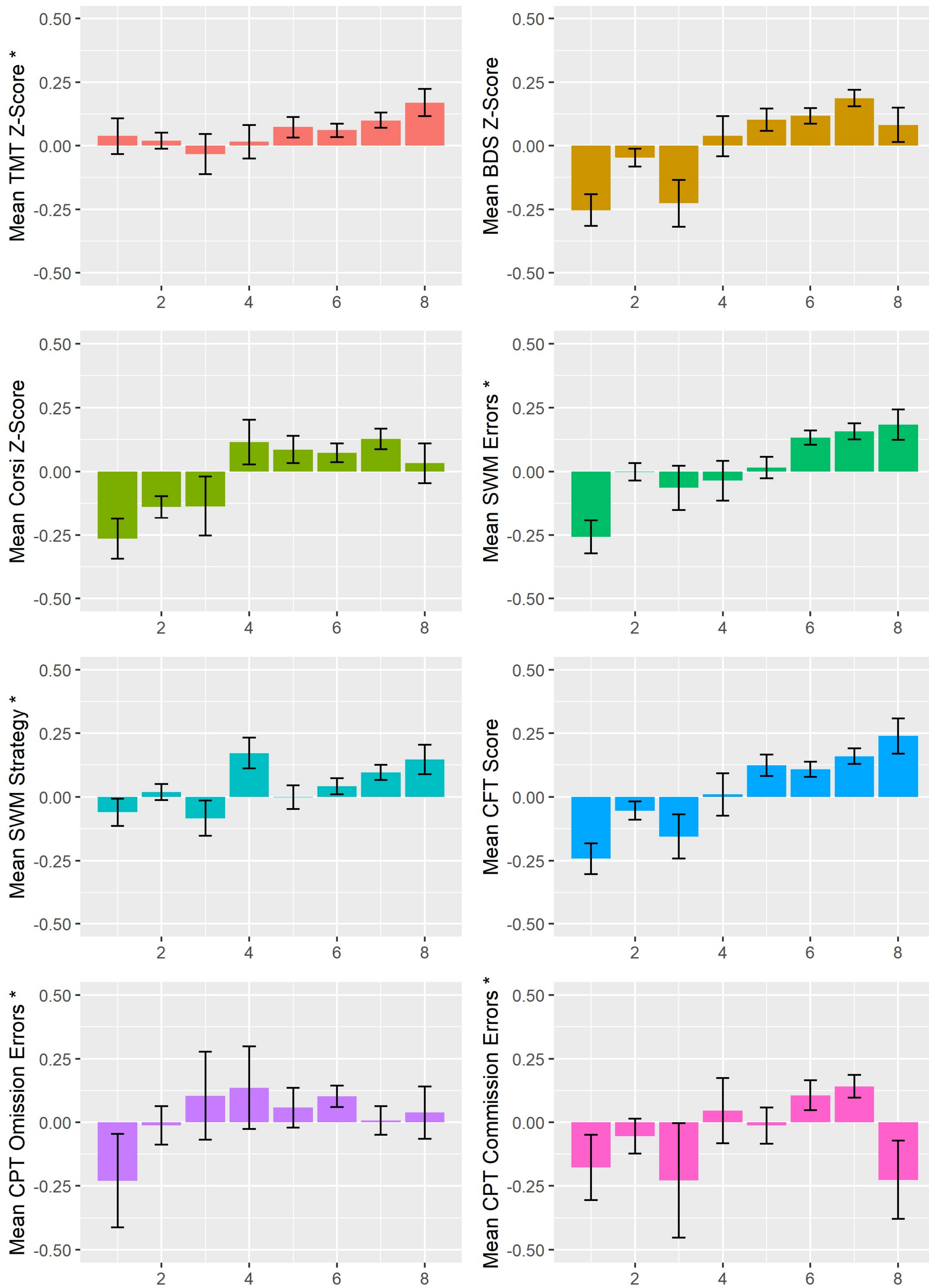
**** Starred tasks have had their scores flipped such that all tasks are shown with higher scores indicating better performance, for easier interpretation of the graphs.***

B: Father Occupation



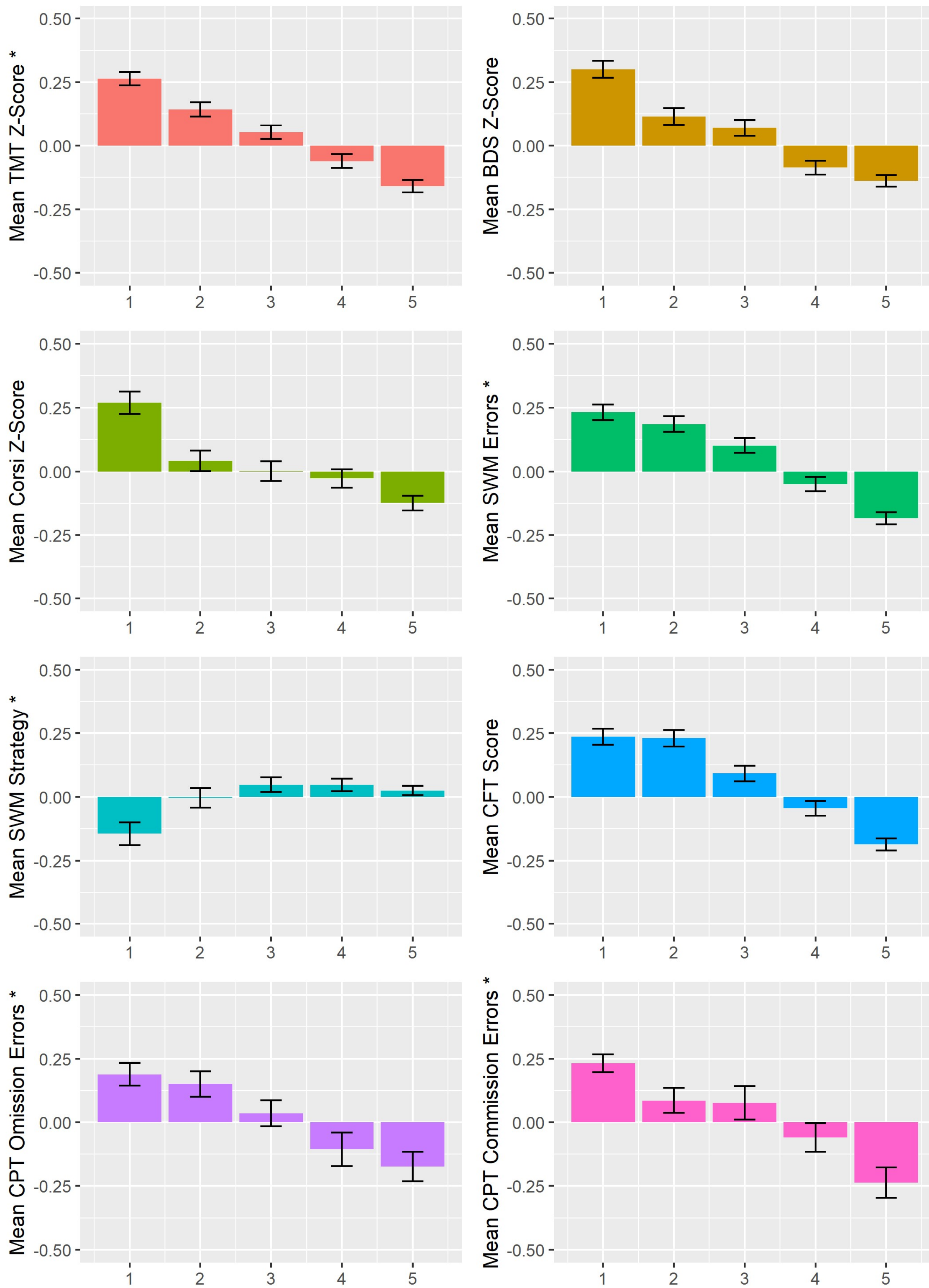
**** Starred tasks have had their scores flipped such that all tasks are shown with higher scores indicating better performance, for easier interpretation of the graphs.***

C: Mother Occupation



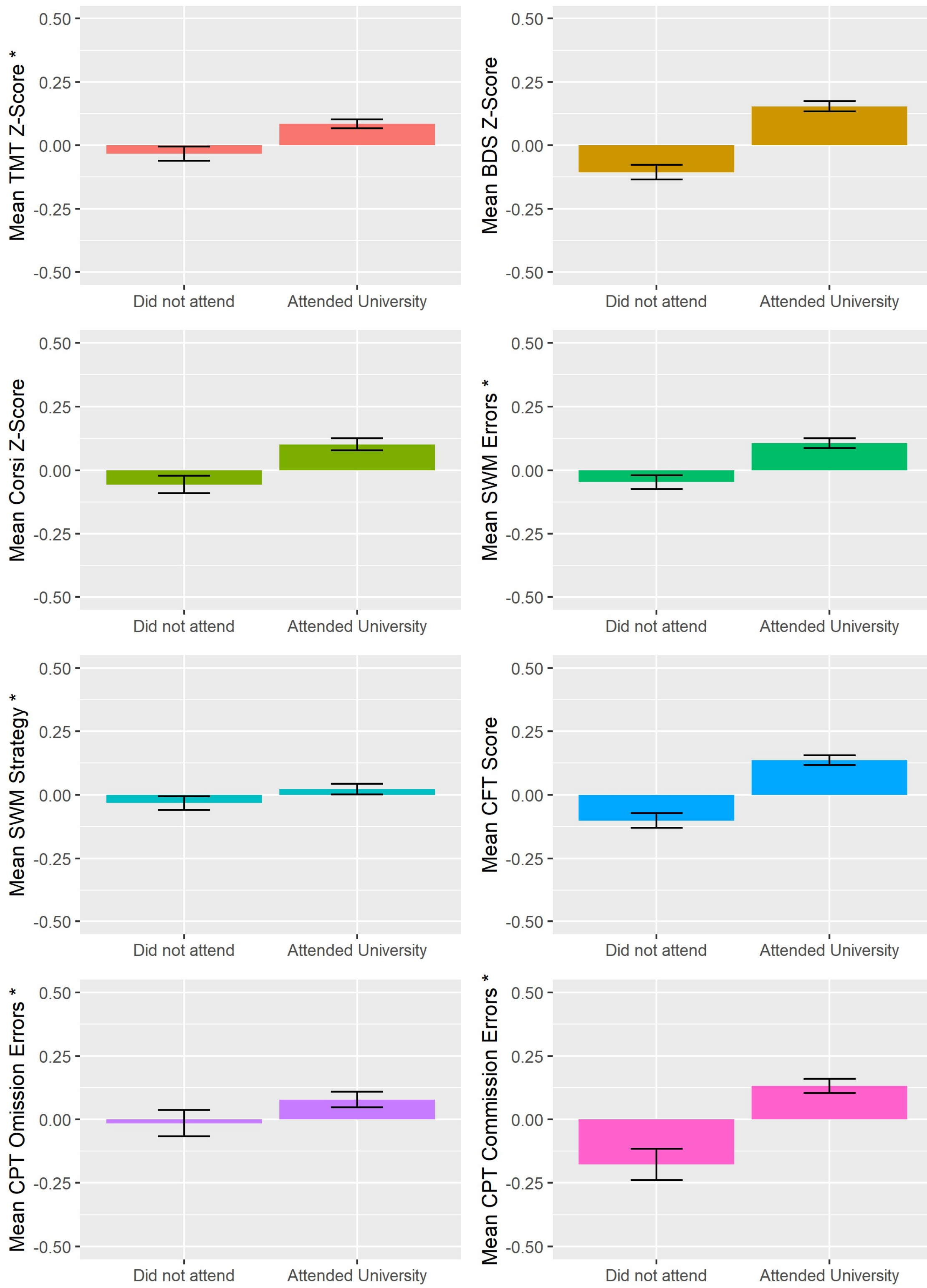
Starred tasks have had their scores flipped such that all tasks are shown with higher scores indicating better performance, for easier interpretation of the graphs.

D: Postcode Deprivation

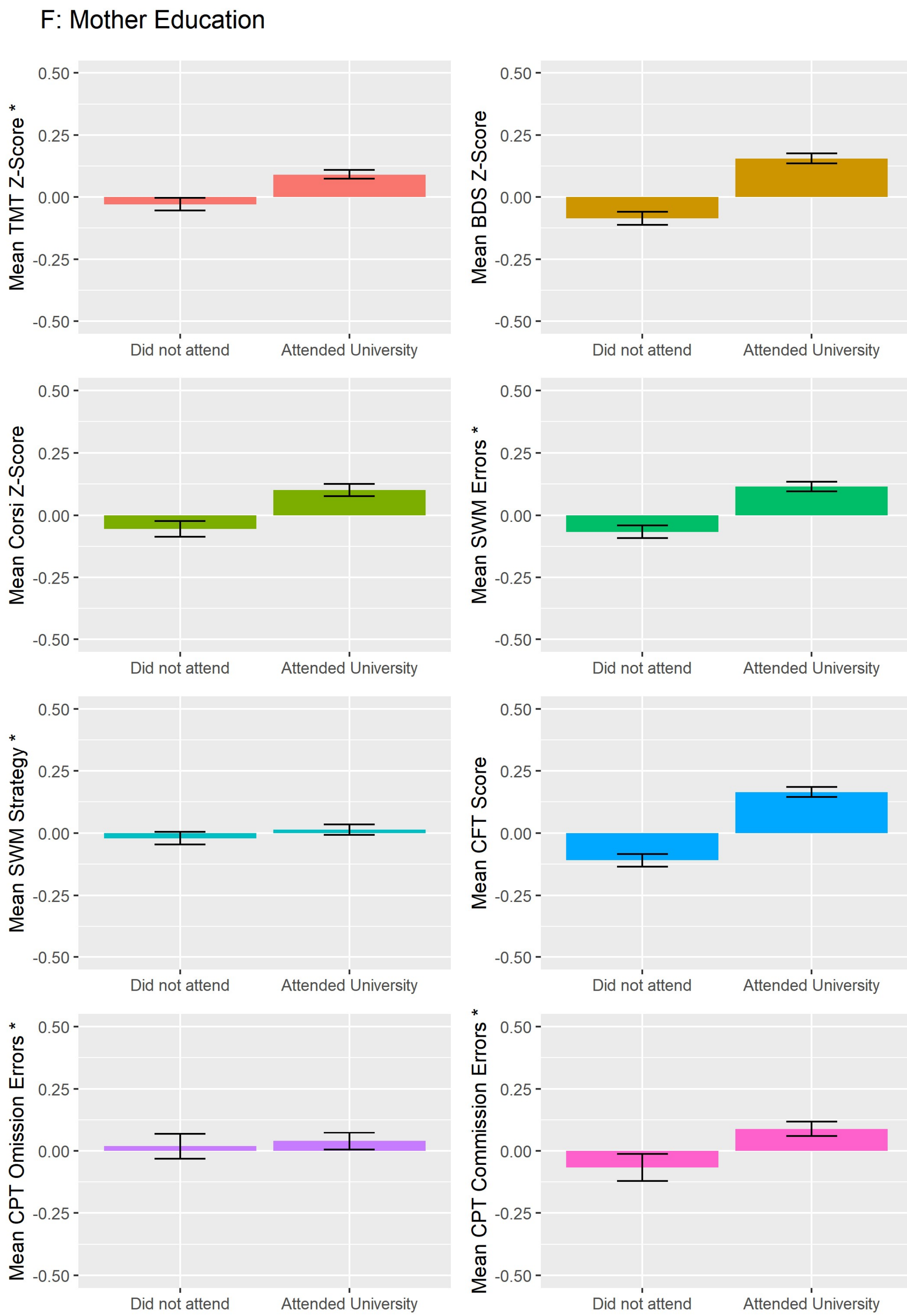


**** Starred tasks have had their scores flipped such that all tasks are shown with higher scores indicating better performance, for easier interpretation of the graphs.***

E: Father Education



**** Starred tasks have had their scores flipped such that all tasks are shown with higher scores indicating better performance, for easier interpretation of the graphs.***



**** Starred tasks have had their scores flipped such that all tasks are shown with higher scores indicating better performance, for easier interpretation of the graphs.***

Table 3.8 Results of multiple regression analyses of association between EF and SES, after accounting for age

	TMT[a]	**BDS**	**SWM Strategy[a]**	**SWM Errors[a]**	**Corsi**	**CFT**	**CPT Omission[a]**	**CPT Commission[a]**
Step 1 (covariates)								
Df	1, 6589	1, 6589	1, 6589	1, 6589	1, 6589	1, 6589	1, 1124	1, 1124
Overall R^2 [b]	<.001	<.001	**.008**	**<.001**	**.002**	**.003**	0.001	**0.017**
	β [b]	**β [b]**	**β [b]**	**β [b]**	**β [b]**	**β [b]**	**β**	**β**
Age	-0.010	**0.002**	**-0.090**	**-0.002**	**0.041**	**4.117**	-0.023	0.005
Step 2 (+ SES measures)								
Df	6, 6583	6, 6583	6, 6583	6, 6583	6, 6583	6, 6583	6, 1118	6, 1118
Overall R^2 [b]	**.04**	**.05**	**.02**	**.04**	**.01**	**.06**	**.03**	**.06**
R^2 change[b]	**.04**	**.05**	**.01**	**.04**	**.008**	**.05**	**.03**	**.06**
	β [b]	**β [b]**	**β [b]**	**β [b]**	**β [b]**	**β [b]**	**B**	**β**
Age	**-0.038**	**0.038**	**-0.073**	**-0.030**	**0.052**	**0.082**	-0.021	-0.049
Father Occupation	**-0.041**	**0.035**	**-0.040**	-0.025	0.037	**0.039**	-0.047	**-0.067**
Mother Occupation	0.012	**0.009**	**-0.044**	-0.030	0.015	0.031	0.012	0.004
School Type	**-0.133**	**0.201**	**0.120**	**-0.132**	**0.055**	**0.126**	**-0.090**	**-0.114**
Postcode Deprivation	**0.068**	**-0.001**	-0.017	**0.060**	-0.004	**-0.075**	**0.090**	**0.081**
Father Education	0.015	**0.017**	-0.022	-0.002	-0.014	0.004	-0.032	**-0.127**
Mother Education	-0.016	**0.025**	-0.013	-0.016	0.008	**0.052**	0.065	0.065

[a] ***Lower scores indicate better performance in these task measures.***

[b] ***Pooled results across imputed datasets are presented for TMT, BDS, SWM, Corsi and CFT; B's are pooled standardised betas.***

[c] ***Standardised betas.***

Bold results are significant at p<.05; Bold Underlined = Remains significant at p < .05 after also covarying for CFT.

3.5 Discussion

This chapter investigated the pattern of relationships between six measures of SES, and seven measures of EF and one measure of fluid intelligence. It also aimed to see whether there are unique associations between SES and EF, over and above associations with fluid intelligence, by controlling for CFT score in analyses. Furthermore, it explored whether there are any significant associations between specific aspects of EF and specific measures of SES. It used a large sample size, representative of the overall London population, and included participants from a wide range of SES backgrounds.

3.5.1 Findings and implications

Results from multivariate MANCOVA analyses showed significant relationships between three specific SES measures (school type, father occupation and postcode deprivation) and the combined EF measures, while accounting for age. When CFT score was added as an extra covariate, these relationships all remained significant, and effect sizes were slightly reduced. In their meta-analysis of SES-EF relationships, Lawson et al. (2018) noted that few prior studies had the power to investigate whether SES-EF relationships remain significant after controlling for general fluid intelligence. These analyses show that even when accounting for CFT, relationships between SES and overall EF remain. The follow-up univariate multiple regression analyses showed the SES measures combined significantly predicted all the individual EF measures and also CFT score, after accounting for age. These analyses also showed that specific measures of SES were associated with some specific EF outcomes.

School type was overall the strongest of the SES predictors considered, and significantly predicted all the EF measures, and also CFT. It was also the SES measure with the largest effect sizes across all models. This is an interesting finding. School type was classified as either State or Independent school (i.e. private and / or selective schools). Effectively, school type acts a kind of proxy measure, as most Independent schools require either fees to be paid, or for pupils to earn a place via a scholarship placement (by achieving exceptionally well in entrance exams). Previous research has shown differences in EF scores between participants in private and public schools (in Mexico), and has suggested that observed differences between these participants may depend on other outside influences, such as parental education level or other factors (Ardila et al., 2005). Of course, the differences in these parental factors endure throughout the schooling period and continue to affect developmental processes, so any effects of school type will be conflated with ongoing differences in parenal factors. Therefore it may be that these ongoing parental factors might be acting to improve cognition in the independent school pupils, rather that these differences being driven by just the differences in the school environment itself.

Often, pupils in Independent schools in the UK complete cognitive tasks similar to our battery at entry, and the schools are able to select or reject pupils based on their ability. So this might mean firstly that these students had more practice at similar types of task, and perhaps that the schools selected pupils based on their ability in these types of task (where State schools are unable to select on this basis). Furthermore, Independent schools might be less likely to have behavioural issues in the classrooms during the assessments, and with smaller group sizes, better computer equipment and greater levels of teacher engagement in research might have meant that students performed better in the tasks in these schools. Our study was carried out in London, which has a particularly high rate of students attending Independent schools relative to other areas of the UK. It would be interesting to see whether this remained a significant predictor of cognitive outcomes in other areas of the UK, or in other countries.

Postcode deprivation was associated with TMT, BDS, SWM errors, both CPT error measures, and also with CFT score. Participants from areas with greater deprivation performed worse on these tasks. The Carstairs postcode deprivation measure is a wide measure of SES, which itself has multiple contributing factors. Given the general findings of previous research that has shown SES to be associated with EF measures, the fact we have found multiple tasks to be associated with postcode deprivation is unsurprising. It is perhaps a little more surprising that school type has a greater effect size than the postcode deprivation measure.

Father occupation was significantly associated with TMT, BDS, and SWM strategy. Mother occupation was associated with BDS and SWM Strategy. Previous research has shown that paternal occupation in a child’s early life is more closely associated with their later life health outcomes than is maternal occupation (Pinilla et al., 2017). Here, we have found that mother and father occupation are both important predictors of strategy use and of verbal working memory capacity, and that father occupation is also predictive of switching ability. This perhaps reflects the fact that mother and father are both likely to work, and in London there are relatively high costs of living which will mean that the incomes of both parents are important in determining overall family SES levels.

Father education was associated with BDS and CPT omission errors, and mother education was associated with BDS and CFT. Previous research has found that mother’s education has been found to be a stronger predictor of cognitive performance than father’s education (Gutman et al., 2003), where other research has found that both parents’ education level are predictors of EF (Ardila et al., 2005). Here we see father and mother education both predicting verbal working memory capacity, and each also predicting one other task measure. Intervention studies have shown that parental education about parenting skills can improve EF outcomes in children, suggesting that this is an important factor in predicting EF (Neville et al., 2013).

Overall we found that some relationships between SES and EF were still present after accounting for CFT, with effect sizes only reduced slightly by including this covariate in the models. This indicates significant relationships between some measures of SES and EF, over and above any associations with fluid intelligence (Gf). In their meta-analysis, (Lawson et al., 2018) suggested that controlling for Gf in future SES-EF research would help to identify whether SES and EF have unique associations, over and above more general intelligence; this has been done in this chapter.

Our task selection is an important factor in terms of identifying which aspects of EF are more closely related to SES. For example, we do not have a specific individual measure of inhibition. Rather, inhibition is an embedded element of task performance across multiple of our tasks. It might be possible to better tap inhibition by considering other task measures, such as the number of non-dot clicks in the spatial working memory or TMT tasks, or random type errors in the CPT task. Our tasks included multiple measures of working memory, and we have indicated that this is significantly associated with SES measures. These relationships extend over and above relationships with general intelligence. We found working memory was more closely related to SES than was switching ability (as measured by the TMT task).

The mechanisms of SES-EF relationships are not clearly understood at present. Cognitive stimulation is an overarching theory that might explain the links between our observed associations between SES and EF. This suggests that children in higher SES households experience greater cognitive stimulation from various routes that drives their improved cognitive development. For example, maternal education level is associated with maternal language complexity, which is in turn asaociated with better language abilities in early childhood, which in turn is associated with improved cognitive outcomes. The improved quality of language exposure is one example of a source of cognitive stimulation which may be an overall driver of our observed SES-EF associations.

Our findings lend weight to the idea that school and neighbourhood environment might have wide impacts on cognitive development, as the school type and Carstairs measures are more strongly associated with cognitive scores than our other SES measures. School type has previously been shown to have associations with academic outcomes, but more limited evidence of association with cognitive outcomes. In terms of mechanisms, independent schools generally have smaller class sizes and better resources, which could be a source of improved cognitive stimulation and may drive improvements in certain cognitive areas as well as academic achievement. However, they also have different cohorts with independent schools having higher performing pupils at the time of entry. Once these differences are accounted for in studies, effects of private schooling are reduced (Ndaji et al., 2016). In our study we do not have any data to be able to account for pupils' abilities at time of entry to the schools, so cannot account for this in our analysis.

Regarding the findings that the local area deprivation is a significant predictor of EF, various mechanisms might be at play here. Previous research has indicated that the increases in pollution exposure associated with living in less well-off areas tend to have higher lead concentrations in the blood, and as this is a neurotoxin, this could impact on children's neurocognitive development and be a factor driving the relationship between Carstairs and EF in our study. Another factor that could underpin this relationship is that lower SES areas tend to also have poorer access to greenspace in the local environment, which a recent paper using our dataset has shown to be associated with overall EF score (Maes et al., 2021).

If we can better understand the links between the composite parts of SES and EF, then this might suggest places to look for specific hypotheses about causal mechanisms relating SES to EF. For example, in this chapter it is clear that school type is the best predictor of all measures EF it may be that a key driving mechanism between SES and EF is to be found in the differences between the independent and state schools in this sample, or their intakes, or some other circumstance surrounding them. Answers to those questions may in turn suggest hypotheses to identify causal mechanisms between SES and EF.

3.5.2 Limitations and suggestions for future research

One issue in this study is that the outcome measures were all significantly correlated with one another. This makes it somewhat difficult to disentangle the effects. Furthermore, we did not correct for multiple comparisons in the follow-up univariate multiple regression analyses, after the MANCOVA test was completed. Previous researchers using omnibus tests with follow-up univariate tests have suggested that the results of the follow-up tests are 'protected' from issues with inflated type 1 error risk by performing the omnibus test first – however this claim is not accepted by all researchers in the field (Field, 2013). Another factor that limits the wider interpretation of our results is that we have a limited task selection. Future research could consider including a wider range of EF tasks, and using a combination of factor analysis methods and MANCOVA or multiple regression analysis to explore whether wider components of EF are related to SES. It would also be useful to explore specific measures of inhibitory control to explore which aspects of EF are more closely related to SES.

The SES as measured by the parental occupations in our sample was reasonably representative, though slightly higher than we might expect for London distributions. This perhaps relates to the fact that only 30 participants were in the lowest category of occupation both parents 'never worked', so these were excluded from analysis as the group was too small. Our group sizes were uneven which might have affected the analysis – but this is to be expected as this is representative of the London population. Our parental education measure was a binary measure – did they attend university – this could be critiqued as we therefore had less variation in this measure than for example if we had assessed the parents' years of schooling. This variable and the parental occupation variable were both reported by the students – previous studies have shown parental occupation assessments to be reasonably reliable when reported by teenagers (Lien et al., 2001), however they might be less able to answer more detailed questions on parental education levels than the binary measure we used.

The MANCOVA was conducted in whole case data only. This may have meant that the poorer performers were excluded from this analysis, for example if those who would have performed worse did not manage to complete all of the tasks analysed. It would be useful in future to explore whether this is the case in our data, and perhaps to carry out the MANCOVA in the imputed dataset to account for this potential source of bias.

It would be useful to consider whether our effects were driven across the whole spectrum of SES, or whether there are specific levels of SES that are driving the significant effects we saw. For example, ANOVA with post-hoc tests could reveal whether the associations between SES occupation measures and EFs are driven across the whole gamut of assessed SES, or whether these effects are driven by differences between only some of the SES categories. A further analysis of interest could be to consider whether occupation related disparities in the EF tasks also occur within State or Independent school samples – i.e. is the relationship between occupation stronger or weaker within participants who attend state or independent schools?

The effect sizes across this study were very small – in the MANCOVA, the partial eta square values were between .004 and .055. This does however fit with previous research – in a meta-analysis, Lawson et al. (2018) concluded that it is possible that SES and EF are only weakly related, with small effect sizes. In their meta-analysis SES accounted for only around 2.6% and 7.8% of the variance in EF measures. Our effect sizes are considerably smaller than this for some of our observed SES and EF associations, but the school type variable does fit into this range.

Results were not corrected for multiple comparisons. If we were to conduct Bonferroni correction to each analysis, i.e. within each multiple regression analysis with a single outcome variable, we would divide the intended p-value of .05 by the number of considered SES and covariates (i.e. .05/7). In this case, almost all of the results would remain significant at this corrected value, suggesting our conclusions would survive Bonferroni corrections. A post-hoc power calculation was carried out for the multiple regression analyses, using the online calculator by Soper (2024). For an effect size of around .05, with .90 power, we would require 373 participants for each multiple regression analysis. Thus we have sufficient power to detect the expected effect sizes based on the literature search, for all analyses carried out. For an effect size of .004 (which was our smallest observed significant effect size) we would have required 4,574 participants for a .90 power – therefore we do have sufficient participants in the imputed dataset, but not sufficient power for the analyses with CPT omission or commission errors at the smallest effect sizes.

3.5.3 Conclusions

This chapter has demonstrated that EF is associated with SES, even after controlling for CFT. Associations between specific aspects of SES and of EF suggest that different measures of SES might have different cognitive impacts on adolescents. Some of these specific associations survive after accounting for fluid intelligence, suggesting that SES and EF relationships are present over and above associations with fluid intelligence. School type was overall the strongest SES predictor of our EF task and CFT scores. The reasons for this are yet to be explored, though the selective nature of Independent schools seems likely to play a part in this effect. The effect sizes were all really quite small here, with partial eta squared values of between .004 and .055 for the relationship between SES measures and overall EF in the MANCOVA analysis. This is encouraging from a wider social point of view, in that SES is not determining most of the variance in EF in adolescents. SES can be said to have a small but significant relationship with EF given our findings here.

Chapter 4.
Developmental Trajectories of Executive Function and Fluid Intelligence

4.1 Abstract

This chapter explores the development of Executive Functions (EFs) and fluid intelligence across late childhood and early adolescence. Participants included in analysis in this section after data cleaning (and exclusions of data where the ages were likely incorrectly reported – see **Chapter 2** for details)were aged *M*= 12.05 (*SD*=0.48, range 10-13 years) at baseline, and at follow-up were aged *M*=14.62 (*SD*=0.52, range 13-16 years). Scores from the cognitive tasks in the main SCAMP battery described in **Chapter 2** are used to assess EF and fluid intelligence.

When considering data from across both assessment points, linear regressions reveal a small but significant effect of age on all but one (CPT omissions) of the task measures considered, with older participants performing better than younger ones. Multi-level modelling shows that these effects remain significant when also including a random intercept per participant in the model, suggesting that even after accounting for the fact data come from the same participants at baseline and follow-up, there are still significant age-related improvements in performance. Further multiple regressions show that task scores at follow-up were better predicted by participants' baseline scores than by their age at baseline or age change. Another set of regression analysis shows no significant association between change in task score between baseline and follow-up and age change or age at baseline. Change in task score is significantly negatively associated with baseline score. Finally, explorations to investigate potential practice effects showed that not all of the age-related improvements in task performance observed in the first set of analyses were related to practice effects. There is little evidence of any age-related development of task scores within either of the assessment periods, however, comparing age-related development within subsamples of participants with task data at only one or both assessment points shows that some but not all of the age-related improvement in task scores observed in analysis Part 1 were due to practice effects. Findings are discussed in relation to previous literature.

4.1 Introduction

This chapter considers the developmental trajectories during early adolescence of the EF and general intelligence cognitive task measures. The general development of EF during adolescence was discussed in the general introduction Chapter 1.

As reviewed in **Chapter 1,** overall, age related improvements in EF task performance can be observed across childhood and adolescence. The three key EF components of inhibition, working memory and switching having somewhat different developmental trajectories across this period. Furthermore, within those broad components, specific tasks and task measures have also been shown to have different patterns of development. Developmental studies have indicated that measures of inhibition performance reach adult-like levels at some point between age 11 and 14, with specific tasks and task measures showing some variation in age that peak performance is reached (Brocki & Bohlin, 2004; Huizinga et al., 2006b; Luna et al., 2004). Working memory capacity is likely to reach a peak in performance around age 13-15 (De Luca et al., 2003; Huizinga et al., 2006b); where more strategic aspects of WM may not reach their peak until early adulthood (De Luca et al., 2003). Improvements in switching ability with age can be observed until mid-adolescence, perhaps reaching a peak in performance by around age 15 (Best et al., 2009; Huizinga et al., 2006b). Some measures of switching ability might begin to level off by around age 8-10 in some measures (De Luca et al., 2003).

Specific EF tasks might show specific age-related developmental trends during childhood and adolescence. The tasks considered in this chapter are: Trail making task (TMT), Backward digit span (BDS), Spatial working memory (SWM), Corsi block task, Continuous performance task (CPT), and Cattell's culture fair task (CFT). The tasks are described in detail in **Chapter 2.** After data cleaning and exclusions (See **Chapter 2.7** for details) participants in the SCAMP cohort range in age from 10 to 16 years, with *N* = 6,591 participants assessed at baseline, aged between 10.4 and 13.5 years (*M* = 12.05 ; *SD* = 0.48), and *N* = 5,116 participants assessed at follow-up, aged between 13.1 and 16 years (*M* = 14.26; *SD* = 0.52). A summary of previous research that has investigated development in the same or similar tasks within similar age groups follows.

4.1.1 Development of specific task performance during late childhood and early adolescence

Trail Making Task (TMT) and Switching

Previous findings regarding the age at which development of switching ability reaches a peak have been somewhat conflicting. For example, in a study that considered the development in scores across three different switching-type tasks, some measures of switching show levelling off by around age 6 (Flanker task), where for the Picture-Symbol task the oldest participants (14-15 years) performed better than the younger groups (K. Lee et al., 2013). The Simon task switch-cost measure was shown to be levelling out by around age 8, with only small improvements observable beyond this age in this study. This illustrates the fact it is difficult to directly tap EF concepts, and the use of different embedding tasks can result in differing estimates of EF performance. The observed results perhaps reflect the idea that specific tasks are tapping somewhat different aspects of switching ability that have unique developmental trajectories, or perhaps that the switching cost measures are embedded in tasks of varying difficulty. From individual tasks it is difficult to tell whether the embedding task difficulty is conflated with switching ability.

The results of Lee et al's (2013) study are somewhat ambiguous, as some inhibitory tasks are still showing development in the age groups we are considering, but others appearing to have reached a plateau in performance. The picture-symbol task is probably most similar to the task we are using, and the switch-cost measure they used probably a more pure measure of switching than the other two tasks analysed in Lee et al.'s research. Therefore it is more likely that we will see results similar to their results for this task than others, with development extending up to age 14-15. This suggests we could expect to see age-related development on our switch cost measures of TMT.

A previous study has investigated development of TMT scores in this age group, though a different task outcome measure was used. Lehto et al. (2003) investigated TMT development in a sample of 173 girls and 196 boys aged 8-13 years. Time for completion of Part A and B was used as their task measures. No significant effect of age was observed in an ANOVA analysis looking at potential effects of their five age group categories on task scores. This would suggest that scores have reached a developmental plateau by the age of 8 in this study, and it is therefore possible that there will be little development in scores associated with age within our data. However, they used the overall completion times of each part rather than a more 'pure' measure of the switching component such as the proportion (B-A)/A we are using, so it does not rule out that we might see development in this 'purer' EF measure. It is also possible that there will be continued development beyond age 13 that we might be able to observe in our sample, which this study was unable to observe due to the age range considered. Finally, their sample sizes were much smaller than ours; perhaps they did not have sufficient statistical power to detect relatively small effect sizes that might be detectable in our sample.

Backward digit span (BDS)

Prencipe et al. (2011) found improvement in BDS scores across a sample from late childhood (8-9) to early adolescence (14-15) in an ANOVA analysis. They found the significant effect of age was driven by differences between the 8-9 year group and the 10-11 and 14-15 groups, but interestingly, the 12-13 group did not perform significantly better than the 8-9 group, and the 12-13 group had slightly lower absolute scores than the 10-11 group. The task measure used for BDS was total correct, in the traditional progressive presentation of the task. Brocki and Bohlin (2004) also used BDS in a developmental study of children from 6-13 years. Again they used a progressive presentation, eight trials / levels, they had two attempts at same trial regardless of success on first attempt, one point per time each trial was correct, average points across the trials (so average of 0, 1, 2 which was scored for each level). Performance in BDS improved linearly with age across this age band. Another developmental study using BDS found no significant development with age in BDS in a sample of 138 participants aged 11-17, using total correct in a progressive presentation as their measure. There were also no significant differences in performance between any of their individual age groups, i.e. single years between 11 and 15 and a final 15+ group (Anderson et al., 2001).

Looking to an overall prediction for our study based on these findings, given that some previous studies have found age-related development in BDS performance in similar age groups to our sample, that our youngest participants will perform worse than our oldest participants. It is therefore possible that we will see an overall improvement in BDS scores with age across our sampled age groups. Furthermore it is possible that there will be a non-linear effect of age, as for example Prencipe et al. (2011) found a small dip in performance in participants aged 12-13 years compared with participants slightly older and younger than this, and the study where linear effects of age were found did not cover older participants than 13 years.

Spatial Working Memory (SWM) – Phones task

Lehto et al. (2003) investigated age related development in children aged 8-13 in tests from the Cambridge Neuropsychological Test Automated Battery (CANTAB) battery, including the original ‘boxes’ version of the Spatial Working Memory task that our SWM ‘phones’ task was based upon. They found no significant effect of age in children aged 8-13 years on the SWM task, in an ANOVA analysis. This was true for three measures of SWM they considered: Within and Between Search errors, and Strategy score. This somewhat conflicts with other prior research which has found that strategy measures of Working memory continue to develop into early adulthood (See also **Chapter 1** for summary of development of Working Memory overall). Previous research has suggested that although working memory accuracy and capacity might have reached adult-like levels by early adolescence, other aspects of WM performance may continue to develop into young adulthood. Functional gains have been observed in an efficiency measure of spatial working memory, between the ages of 15 and 19 years, with further increases until 20–29 years of age (De Luca et al., 2003). De Luca et al. also found that a strategic planning measure of working memory (equivalent to our SWM Strategy measure) peaked in their 20-29 year old age group.

Overall, findings suggest that we might see some improvement with age in our SWM task, in particular we are more likely to see significant development with age in the strategy score measure as this is a higher-level EF and previous research has suggested this kind of measure shows extended developmental period into young adulthood.

Corsi block tapping task

The Corsi task is a second measure of spatial working memory used in our SCAMP battery. Previous research investigating development in childhood in this task has found no significant effect of age on a spatial span measure obtained from a traditional progressive presentation of the Corsi task, in children aged 8-13 years (Lehto et al., 2003). De Luca et al. (2003) also used the Corsi task in their research looking at EF development across the lifespan, from age 8 to 64 years. They found significant effects of age in spatial span capacity, with 15-19 and 20-29 year olds showing the best performance in this task (De Luca et al., 2003).

Given the mixed findings of previous research, it is not clear what we might expect to see in our study. There could be some improvement with age amongst our sample, since spatial span capacity has been found to still be developing until age 15-19 in one study (De Luca et al., 2003), where Lehto et al. (2003) found no significant improvement in Corsi performance between ages 8 and 13. Taken together we might predict that we could see a non-linear effect of age, for example there may be little difference in span capacity between our youngest participants and those in the middle of our range (i.e. roughly between age 10 and 13), but then an improvement may occur between ages 13 and 16. Overall we might see a small linear change effect.

Continuous Performance Task (CPT)

Brocki and Bohlin (2004) used a version of a CPT task in their assessment of EF development among children aged 6-13. Their task used squares with different symbols inside, with a target sequence of a square containing an X followed by square with a vertical line. Measures they considered were omission errors (which they considered to be a direct measure of inattention) and three types of commission errors (following the model of dividing the commission errors in Halperin et al., 1991). In a latent variable analysis, they found that an inhibition component of EF generally improved with age, with the 10-11 year olds better than 8-9 year olds. For the specific task measures of CPT, significant effects of age were found for the inattentive omission error measure and the impulsivity commission error measure, but no significant effect of age was observed on the disinhibited or impulsive commission error measures.

A prediction based on this previous research might suggest that we will see a significant effect of age on our omission error measure, but perhaps not on the commission error measure, as two out of the three types of commission error did not improve with age in this previous study (Brocki & Bohlin, 2004). However it is worth noting that our study is using older participants – they found no significant differences in CPT performance between their oldest participants and the other groups in post-hoc analysis – suggesting that perhaps performance might be beginning to level off by this age and therefore we won't see large differences across our whole age sample.

Cattell's Culture Fair Task (CFT) – IQ proxy measure of General Fluid Intelligence (Gf)

Scores on CFT increase across childhood. Research considering children in school grades 4, 5 and 6 (aged approximately 8-12 years) shows significant increases in CFT score over this age group (Cahan & Cohen, 1989). Pubertal timing has been shown to be related to CFT scores in boys aged between 8 and 12 years (Shangguan & Shi, 2009). In this study, levels of testosterone in saliva samples were used as proxy markers of pubertal development stage. Findings were somewhat complex, with amount of testosterone at different ages predicting being correlated with sometimes improved scores and at other points worse scores in CFT. Overall, there was as significant improvement in CFT scores between the ages of 8 and 12. Overall these studies suggest that we are likely to see some development in CFT task scores in our cohort.

4.1.2 Aims

This study aims to see whether there is any significant development in EF and Gf task scores in a large sample of adolescents, between age 10 and 16 years. We will investigate the association between Age and Task score in our dataset using linear regressions and multi-level modelling (MLM) techniques. We will also explore rates of change and whether these are associated with change in age between baseline and follow-up, and will explore whether the task score at follow-up and change in score are associated with baseline scores. We will also consider the potential impact of practice effects on our results.

4.2 Methods

4.2.1 Task Measures

A summary of the key measures used in this chapter are presented in **Table 4.1**. Details of task presentation methods, measure calculation, and data cleaning processes were covered in **Chapter 2.** We used the proportion score (time taken in (A-B)/A) for the TMT task. This was selected as it is not possible to show overall changes in performance from baseline to follow-up using the residual measure which was used in **Chapter 3** analysis. This is because the residual score is centred on zero and is calculated across the whole sample of participants. Thus, any increase in score between two time points would reflect a relative improvement in switching ability compared to the other participants in the sample. Given that in our study the samples at the two time points are not identical (as many participants are missing data from one or other time point), improvements in a BrA residual score would not be easily interpretable. Therefore we used a proportion score (B-A)/A in this chapter consider scores across both time points.

Table 4.1 Summary of cognitive task measures used in this chapter

Task	Measure	Brief summary of how measure is calculated
Trail making task (TMT)	Proportion score	Difference in response times between switching and letters condition, divided by the response time in letters condition (A-B)/A
Backward digit span (BDS)	Mean pass / fail level	Average of the mean level of correct trials and the mean level of incorrect trials
Spatial working memory (SWM)	Total errors	Total number of errors (within and between search errors) across the task levels with 4, 6, 8 and 10 items
	Strategy score	Total number of excess switches of search start location across levels with 6, 8 and 10 items
Corsi block task	Mean pass / fail level	Average of the mean level of correct trials and the mean level of incorrect trials
Cattell's culture fair task (CFT)	Total correct	Total number of correct trials across odd-one-out and complete-the-pattern subtasks
Continuous performance task (CPT)	Omission errors	Total number of omission errors when A-X target was missed across the task
	Commission errors	Total number of commission errors when response was made to something other than A-X target across the task

4.2.2 Participants

Participants were tested at two time points with the SCAMP computerised assessments as described in **Chapter 2.** Participants were tested during school years 7-8 at baseline, and school years 9-10 at follow-up. Data have been excluded for participants who reported that they were very much older or younger than the target age at each assessment point(see **Section 2.7** for details on exclusions)**.** After these exclusions, a total of n = 6,591 participants had attempted the baseline assessment, aged between 10.4 and 13.5 years (M = 12.05 ; SD = 0.48), and n = 5,116 participants attempted the follow-up assessment, aged between 13.1 and 16 years (M = 14.26; SD = 0.52). Ages are given at time of testing within each assessment point. Age in days was first calculated from the date of testing and the date of birth that was reported by the participant in the assessment questionnaire. Age in days was then divided by 365.25 to give age in years - this measure of age is used in analyses in this chapter. The distribution of participant ages included in the analysed data at each time point were shown in **Chapter 2, Figure 2.10.**

4.2.3 Data Cleaning and Statistical Analysis

Before analysis, the baseline task measures were z-scored, and follow-up task measures were standardised using the means and standard deviations from the baseline data (See **Chapter 2.7**). This was to more easily compare effect sizes across tasks, and to show the progression from baseline to follow-up. For measures where a lower raw score would usually indicate better performance (TMT, SWM total errors and strategy score, CPT omission and commission errors) scores were multiplied by -1 such that a higher score indicates better performance for all measures in the statistical analyses, for easier interpretation of results. Analysis was carried out in R version 4.0.0, with packages dplyr, psych, Hmisc, pastecs, lme4, lmer, QuantPsyc.

4.2.4 Modelling Approach

The analysis consists of four key sections. In part 1, linear regressions and MLM were used to investigate the relationship between overall task score and age in our dataset, across data from both baseline and follow-up assessments. Part 2 investigates the associations between follow-up task scores and both age and baseline score. Part 3 looks at whether the rate of change in task score between baseline and follow-up is associated with the participants' change in age, age at baseline or their baseline score. Finally, Part 4 considers whether any observed age-related development is due entirely to practice effects.

4.2.5 Part 1: Overall associations between task scores and age

In this section of analysis, data from both time points were combined into a single dataset. In Analysis A, separate linear regressions were run with data for each task respectively to see whether age predicted task scores overall, for any of the cognitive tasks. In Analysis B, Multi-level Modelling (MLM) was used to improve these estimates, using task scores as outcomes, age at testing as a predictor, and adding a random effect of participant to account for the fact that data are not independent samples within the two time points.

Analysis A: Linear Regression Analysis

Data from both assessment points were combined into a single dataset for this analysis. This method treated each data point as a separate observation, without considering whether the data from follow-up assessment came from the same participants as the data from the baseline assessments. Separate linear regressions were then run for each task, with respective task score as the outcome, and age at time of testing as a predictor.

This analysis aimed to assess whether there is an overall association between age and cognitive task score, ignoring which assessment point the data came from.

Analysis B: Multi-level Modelling

Multi-level modelling (MLM) is a way of accounting for the similarity of the data gathered from the same participants multiple testing points. MLM is essentially an extension of linear regression methods, where random effects are first included in the model, which are used to account for the similarity of data obtained from a particular participant group while considering fixed effects of any predictors (Ntoumanis, 2014). Multi-level models, also known as linear mixed models, apply a linear regression model to data with a continuous outcome, with fixed effects for predictors (the same as a simple regression model) and adds random effects of factors which are common across time (in this case we apply random effects of participant to the data from baseline and follow-up).

In this case, a random intercept per participant is used. This allows us to model data for each individual, remove variance which is explained by the fact the data at the two time points are coming from the same participant, before a standard linear model (with fixed effects of age on task score as outcome) is applied to the data. Simpler linear regression methods as used in Analysis A assume that all data points are independent – MLM may therefore be considered to be more appropriate when data are obtained from same participants at multiple time points (Ntoumanis, 2014). Multi-level modelling adds random effects to a linear model to account for the fact that scores at baseline and follow-up are not independent, as the same participants were present in both samples. This model uses all of the data from both time points, i.e. includes data for participants who have data at both assessments and at only one time point.

There are two key aims for these analyses:

1. Does age predict task score, when accounting for the fact data at the two time points comes from the same participants?
2. What is the effect size of any observed association between age and task score, whilst accounting for individual intercepts?

In the multi-level models we used a similar approach to the linear regressions in Analysis A, with age entered as a predictor and respective task score as outcome. We also added random intercept effects per participant to account for the fact much of the data came from the same participants across the two time points. It was not possible to include a random slope per participant in this model, as there were only at most two data points per individual and this would have resulted in an overfitted model. The models listed in **Table 4.2** were conducted sequentially. Models were run using the Enter method. At each stage we checked to see if the more complex model added significantly to the previous one by running an ANOVA. If the ANOVA returned a significant result (p<.05) then the more complex model was considered to be an improvement over the previous one.

Table 4.2 Sequential models used for multi-level modelling

Model	Predictors included
1	Age (intercept only)
2	Participant (random intercept)
3	Age (fixed) + Participant (random intercept)

4.2.6 Part 2: Associations of score at follow-up, with age measures and score at baseline

Next we ran multiple regressions to investigate whether score at follow-up was predicted by task score at baseline, age at follow-up, or change in age from baseline to follow-up. Age at baseline testing was calculated in decimal years from the date of birth the participant entered during the assessment battery, and the date on which they completed the battery. An age change measure was calculated by taking age in decimal years at baseline testing from age at follow-up testing. Task scores at baseline were the Z-scored key measures, and the follow-up scores were the transformed task scores (calculated as described in **Chapter 2**). Multiple regressions were performed separately for each task, with respective task score at follow-up as the outcome; predictors were respective task score at baseline and the two age measures: age at baseline and change in age from baseline to follow-up. These predictors were entered into the models simultaneously using the Enter method.

The key questions this analysis aimed to address were:

1. Do baseline scores predict follow-up scores?
2. Does age at baseline predict score at follow-up?
3. Does change in age between baseline and follow-up predict score at follow-up?
4. Are these relationships positive or negative? What are the relative effect sizes between age and score at baseline in predicting follow-up score?

4.2.7 Part 3: Associations of change in task score with change in age, age at baseline and score at baseline

Next we wanted to investigate whether the rate of change in task performance was associated with the rate of change in age. This analysis is also intended to investigate whether development in task score is the same at all ability levels, by using score at baseline as a predictor. For each participant, simple change scores were calculated by taking their score at baseline away from their score at follow-up, using the z-scored and transformed data (as described in **Chapter 2**) as the basis for the task scores. It was therefore only possible to include participants with data at both time points for each task respectively in this analysis. Multiple regressions were then performed with these task change scores as outcome; predictors were respective task score at baseline and the two age measures: age at baseline and change in age from baseline to follow-up. These were entered into the model simultaneously using the Enter method.

The key questions addressed by this analysis were:

1. Does task performance at baseline predict the rate of change in task score? Is the effect seize of any relationship between baseline score and rate of change greater or less than effects of age and/or age change?
2. Do those who are younger at baseline improve more than those who are older at baseline?
3. Is change in task score related to absolute age at baseline or follow-up, independent of their age at baseline testing?
4. Is age change a strong predictor of increase in task score?

4.2.8 Part 4: Exploration of potential practice effects

As many participants completed the same tasks in baseline and follow-up, we wanted to rule out any practice effects which might confound our developmental results. Practice effects in EF tasks have been shown to have durations extending for at least a year in adult populations (Basso et al., 1999). Practice effects in EF tasks can be investigated by running analyses within specified subsamples of participants, to see whether effects remain in these groups or whether effect sizes are decreased. (Salthouse, 2011).

The approach here was two-fold. In Analysis A, we checked whether there were any associations between age and task scores within either of the assessment points. The logic of this is that any significant developmental effects within the time points would not be subject to practice effects, as participants only completed each of the tasks once within that time point. For Analysis A, we ran linear regressions of task score by age within data from each of the two time points, to see whether there are any significant effects of development within either of the time windows covered by the assessments at baseline and follow-up. Any significant developmental effects here would not be associated with practice effects, as participants only completed each task once within a time point (NB – any occasions where individual participants completed repeat attempts of tasks were excluded during the data cleaning steps, as described in **Chapter 2**).

Analysis B checked whether participants who had completed the tasks only once showed differing patterns of age related effects on task score to those who completed the tasks at both time points. Logically, the subsample with data for a given task at only one time point will not display any practice effects, as they only completed the tasks once, where those with data at two time points may have some practice effects in addition to any age-related or developmental change in score. In Analysis B, we ran linear regressions of task score by age for each of the cognitive tasks within two subsamples of participants: The first group consisted of participants with data at only one time point for each task respectively, and the second group were those who had data at both time points for each task respectively.

4.3 Results

4.3.1 Descriptive Statistics

Overall means and standard deviations for the cognitive task data in each assessment point are presented in **Table 4.3**. These are the raw scores for each task measure. Higher scores indicate better performance in CFT, Corsi and BDS. Lower scores indicate better performance for TMT proportion (A-B)/A, SWM errors and strategy errors measures, and the CPT Omission and Commission errors measures. On average, performance is slightly better at follow-up than at baseline for all tasks. The subsample columns are presented to show that participants who have data for a given task at both time points show slightly better performance on average than those who have data for that task at only one time point. This is true for both baseline and follow-up data – showing this effect is not only due to practice effects in the follow-up testing. Rather, it indicates that the participants who have data at only one time point do not score as well in the tests overall as those who were present at school on the days of both testing points.

Table 4.3 Descriptive statistics for the tasks baseline data raw task scores

	Baseline Data								
	All data			**Subsample with baseline data only**			**Subsample with data at baseline and follow-up**		
Task Measure	***N***	***M***	***SD***	***N***	***M***	***SD***	***N***	***M***	***SD***
TMT proportion (A-B)/A	6424	0.84	0.64	2898	0.84	0.67	3526	0.83	0.61
BDS Score	6084	4.09	0.9	2705	3.99	0.89	3379	4.17	0.89
SWM errors	6305	28.99	13.52	3022	30.41	13.86	3283	27.69	13.06
SWM strategy errors	6305	10.99	3.59	3022	10.86	3.1	3283	11.12	3.98
Corsi Score	3791	5	0.78	2533	4.97	0.79	1258	5.06	0.75
CFT Total correct	5808	13.27	3.9	2675	12.86	4.08	3133	13.62	3.69
CPT omission errors	1572	0.54	1.04	1173	0.6	1.1	399	0.37	0.84
CPT commission errors	1572	4.61	5.74	1173	4.8	5.85	399	4.04	5.37

	Follow-up Data								
	All data			**Subsample with follow-up data only**			**Subsample with data at baseline and follow-up**		
Task Measure	***N***	***M***	***SD***	***N***	***M***	***SD***	***N***	***M***	***SD***
TMT proportion (A-B)/A	4918	0.79	0.62	1392	0.78	0.65	3526	0.79	0.61
BDS Score	4864	4.4	0.99	1485	4.19	0.99	3379	4.49	0.97
SWM errors	4627	25.37	13.05	1344	27.63	13.86	3283	24.44	12.6
SWM strategy errors	4627	10.09	2.95	1344	10.26	2.97	3283	10.03	2.95
Corsi Score	2093	5.32	0.89	835	5.18	0.89	1258	5.41	0.88
CFT Total correct	4701	14.57	4.11	1568	13.86	4.57	3133	14.92	3.81
CPT omission errors	912	0.57	1.15	513	0.68	1.3	399	0.43	0.9
CPT commission errors	912	3.14	5.16	513	3.85	5.72	399	2.23	4.16

Note. Means and SDs are presented for the raw data before Z-scoring or transformation applied.

4.3.2 Correlations Between Cognitive Task Measures

Data from both time points were combined into a single dataset. Overall means and standard deviations for each measure, and the results of Pearson's correlations, are presented in **Table 4.4**. Positive correlations are observed between almost all of the task measures, with only CPT omission and SWM strategy having no significant relationship with each other at p<.05. Moderate correlations with R >.3 are present between BDS and SWM errors, Corsi, and CFT; between SWM errors and SWM strategy score and Corsi; between CFT and Corsi; and between CPT omission and commission errors measures. No issues with multicollinearity are present.

Table 4.4 Pearson's Correlation matrix showing relationships between the cognitive task measures. With means and standard deviations for each task measure.

Task Measure	*M*	*SD*	1	2	3	4	5	6	7
1. TMT Proportion	0.03	0.99							
2. BDS	0.16	1.06	.12**						
3. SWM Errors	0.11	0.99	.09**	.33**					
4. SWM Strategy	0.11	0.94	.05**	.15**	.51**				
5. CORSI	0.15	1.07	.15**	.35**	.32**	.19**			
6. CFT	0.15	1.04	.10**	.39**	.32**	.18**	.31**		
7. CPT Omission	-0.01	1.04	.07**	.13**	.12**	.04*	.13**	.11**	
8. CPT Commission	0.09	0.97	.08**	.21**	.20**	.12**	.25**	.20**	.34**

Note. Data were Z-scored at baseline; and transformed according to the mean and SD from baseline at follow-up – these Z-scored / transformed values are presented above; presented Mean and SD figures are from the overall combined dataset i.e. across baseline and follow-up data. Values are presented for the Pearson's coefficient (R) for each pair of task measures. Pair-wise deletion is used. *p<.05; **p<.01.

4.3.3 Part 1: Overall associations between age and task scores

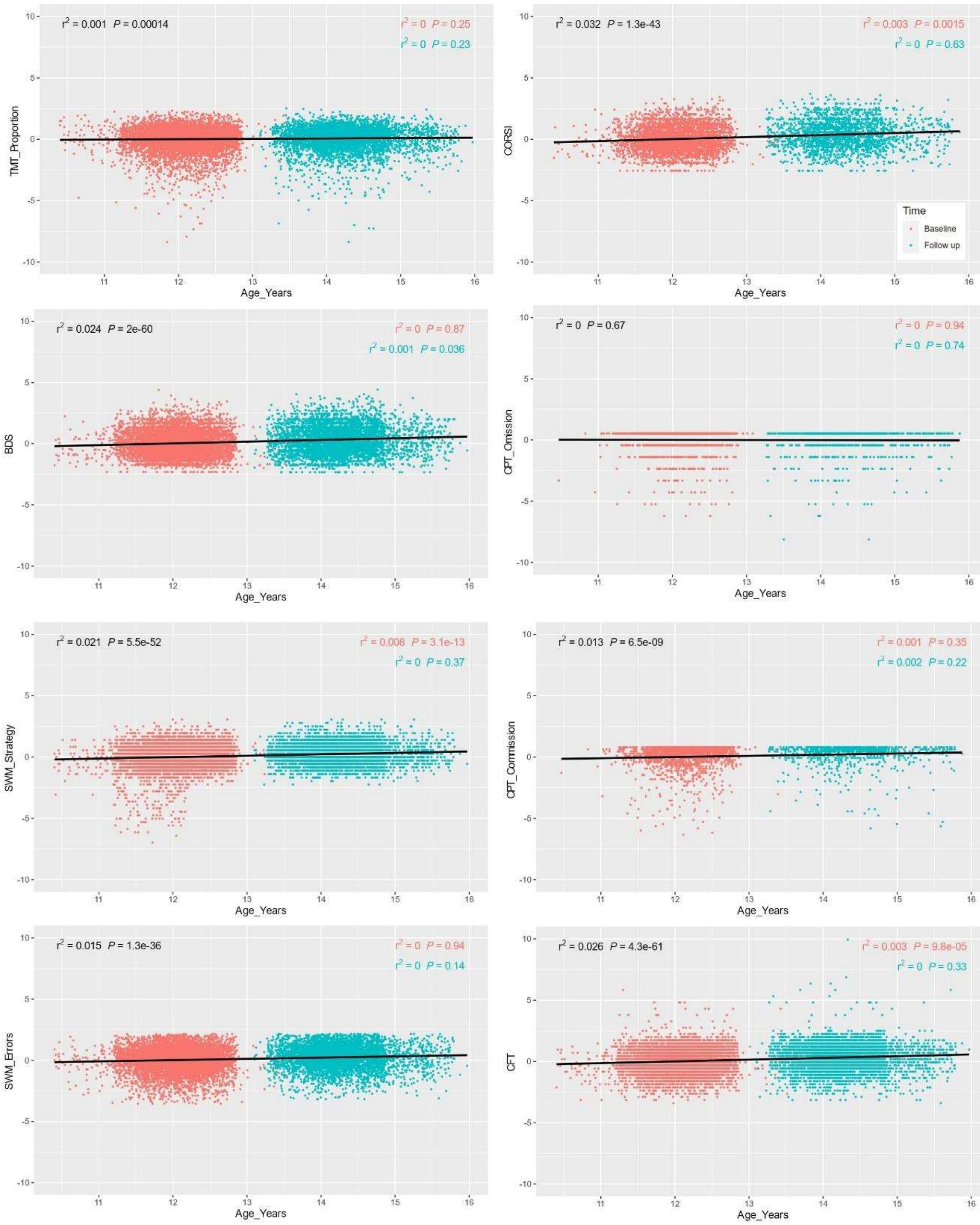
Analysis A: Linear Regression Analysis

The numbers of data points for each task along with the results of the linear regressions are shown in **Table 4.5.** Age was found to be a significant predictor of score for TMT, BDS, both SWM measures of errors and strategy, Corsi, and CPT commission errors, but not CPT omission errors. The largest effect sizes were for the Corsi task, followed by CFT then BDS, SWM strategy and errors measures, then the CPT commission errors and finally TMT. Using Cohen's rules of thumb for effect sizes, the modelled effect size was extremely small for the TMT score (R^2 = .0016), and also very small for all of the for the other significant task measures (R^2 = .014 to .032) (Cohen, 1988). Graphs to illustrate these regressions are presented in **Figure 4.2.** A weakness of this method is that is likely to underestimate error terms, and *p* values, as the assumption of independence of data points in the linear regression is violated. Data from the two assessment points are not truly independent, as many participants completed the tasks at both baseline and follow-up. Analysis B will attempt to address this issue.

Table 4.5 Results of linear regressions with task score as outcome and age as predictor

Outcome	Predictor	*Std. Beta*	*CI*	*p*	R^2 / R^2 adjusted	*N* obs.
TMT	Age Years	.04	.02 – .05	**<.001**	.001 / .001	11342
BDS	Age Years	.16	.14 – .17	**<.001**	.024 / .024	10948
SWM errors	Age Years	.12	.10 – .14	**<.001**	.015 / .014	10932
SWM strategy	Age Years	.14	.13 – .16	**<.001**	.021 / .021	10932
Corsi	Age Years	.18	.15 – .20	**<.001**	.032 / .032	5884
CFT	Age Years	.16	.14 – .18	**<.001**	.026 / .025	10509
CPT omission	Age Years	-.01	-.05 – .03	.669	.000 / -.000	2484
CPT commission	Age Years	.12	.08 – .16	**<.001**	.013 / .013	2484

Figure 4.1 Results of linear regressions showing associations between cognitive task scores and age, across (black) and within (colour) the two assessment time points.



***Note.** Task scores are scaled (Z-scores at baseline; follow-up transformed with baseline mean and SD) and higher scores indicate better performance for all tasks. Black line represents the linear regression result using all data, with age as predictor and task score as outcome. Grey shaded area (very narrow) represents Standard Error of the Mean. R^2 represents the zero-order correlation; p is significance for overall model fit. R^2 and p-values for regression for all data shown in black, and within time points are shown in colour. Regression lines for within time points are not shown as these overlapped with overall regression lines. Scatterplot of individual scores also shown in colour; Red = Baseline data, Turquoise = Follow-up data.*

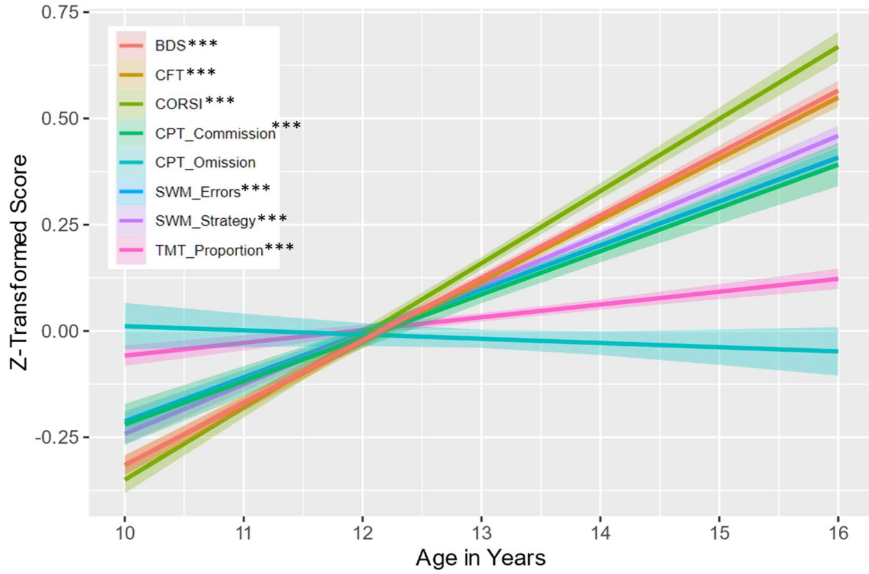
Analysis B: Multi-level Modelling

For each task measure, except CPT omissions, Model 3 (which included fixed effect of Age and Random participant intercepts; see **Table 4.2**) was the best model. For CPT Omission, there was no significant relationship between age and task score, and therefore Model 2 including only a random intercept for Participant was the preferred model. Results for Model 3 for each task are reported in **Table 4.6,** including for CPT Omissions. Method used was Maximum Likelihood for all MLMs.

Results of the multi-level modelling (**Figures 4.3 & 4.4; Table 4.6**) showed a significant effect of age on task scores in the following tasks: TMT, BDS, Corsi, SWM strategy and errors measures, CFT and CPT commission errors. No significant association with age is observed for CPT omission errors. The results show that the Corsi task has the greatest age-related development, with an improvement of around 1 SD in score observable from the youngest participants to the oldest. Next greatest are BDS and CFT which show similar amounts of age related development, with around 0.75 SD difference between the oldest and youngest participants. Then SWM Strategy use, SWM Errors and CPT Commission errors with around 0.6 SD difference across the age groups assessed. Finally, TMT shows a very small improvement in score, less than 0.1 SD, across this time period. CPT omission errors shows no significant change over this time period.

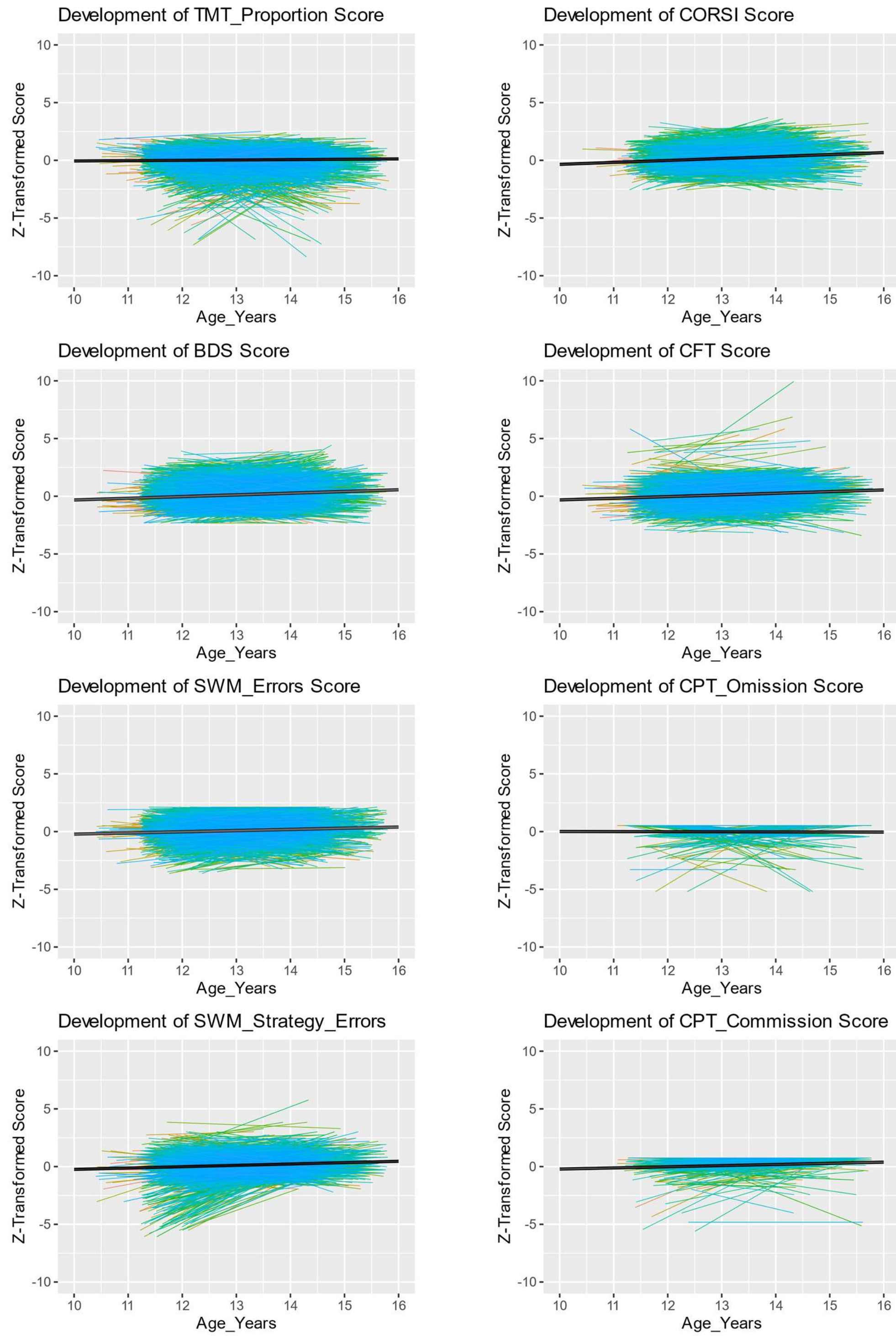
The marginal R^2 illustrates the effects of the fixed effects only (i.e. age in years), where the conditional R^2 is the overall effect size of the model once random effects are also added (i.e. accounting for age and also the random participant intercepts). By comparing the marginal and conditional R^2 values, we can see that the multi-level model including random intercepts per participant accounts for a much greater proportion of the variance in the data than the marginal effect of age alone. This is true for all the tasks. This indicates that the effect of individual differences on task scores is much greater than the effect of age on scores.

Figure 4.2 Results of Multi-level Models Showing Development of Cognitive Task Scores with Age



Note. Task scores are scaled (Z-scores at baseline; follow-up transformed with baseline mean and SD). Higher scores indicate better performance for all tasks. Estimates predicted using multi-level models with respective task score as outcome, fixed effect of age, and random intercept per participant. Shaded bands represent Standard Error of the Mean. *** indicates the effect of age in the model is significant at p<.001. TMT_Proportion=Trail making task proportion score (B-A)/A; BDS=Backward Digit Span; SWM_Errors=Spatial Working Memory errors; CFT=Cattell's Culture Fair Task score; SWM_Strategy= Spatial Working Memory strategy; CPT_Omission=Continuous Performance Task omission errors; CPT_Commission=Continuous Performance Task commission errors.

Figure 4.3 Development of cognitive task scores with age, shown in black, with individual participant trajectories in colour.



Note. Individual participant trajectories are shown in randomly coloured lines, for participants with valid task data at both time points for each task respectively. Task scores are scaled (Z-scores at baseline; follow-up transformed with baseline mean and SD), and higher scores indicate better performance for all tasks. Black lines are predicted estimates from multi-level models with respective task score as outcome, fixed effect of age, and random intercept per participant. Grey shaded areas (very narrow) show standard error of the mean. TMT_Proportion=Trail making task proportion score (B-A)/A; BDS=Backward Digit Span; SWM_Errors=Spatial Working Memory errors; CFT=Cattell's Culture Fair Task score; SWM_Strategy= Spatial Working Memory strategy; CPT_Omission=Continuous Performance Task omission errors; CPT_Commission=Continuous Performance Task commission errors.

Table 4.6 Results of Multi-level modelling, with task score as outcome, fixed effects of age in years, and random intercept per participant.

		TMT			BDS			SWM errors	
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Age in years	0.04	0.02 – 0.05	<0.001	0.16	0.15 – 0.18	<0.001	0.12	0.11 – 0.14	<0.001
σ^2	0.76			0.53			0.6		
$\tau_{00\ ID}$	0.22			0.56			0.38		
ICC	0.22			0.52			0.39		
N_{ID}	7816			7569			7649		
Observations		11342			10948			10932	
Marginal R^2 / **Conditional R^2**		0.001 / *0.226*			0.027 / *0.530*			0.015 / *0.394*	

		SWM strategy			Corsi			CFT	
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Age in years	0.15	0.13 – 0.16	<0.001	0.18	0.16 – 0.21	<0.001	0.16	0.15 – 0.18	<0.001
σ^2	0.75			0.56			0.45		
$\tau_{00\ ID}$	0.11			0.55			0.62		
ICC	0.13			0.5			0.58		
N_{ID}	7649			4626			7376		
Observations		10932			5884			10509	
Marginal R^2 / *Conditional R^2*		0.021 / *0.151*			0.034 / *0.513*			0.026 / *0.588*	

		CPT Omissions			CPT Commissions	
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Age in years	-0.01	-0.05 – 0.03	0.564	0.12	0.09 – 0.16	<0.001
σ^2	0.21			0.37		
$\tau_{00\ ID}$	0.19			0.4		
ICC	2085			2085		
N_{ID}	0.88			0.57		
Observations		2484			2484	
Marginal R^2 / *Conditional R^2*		0.000 / *0.193*			0.015 / *0.406*	

Note. Results for Model 3 (see Table 4.2) are presented for all task outcomes. Model 3 used age as a fixed effect predictor, individual participant as a random intercept, and task score as outcome. Values in non-italics ***relate to the Fixed Effects in the model. Marginal R^2 is the effect of Age alone in the model without accounting for the random effects. Values in*** *Italics* ***(σ^2, $\tau_{00\ ID}$, ICC, N_{ID}) relate to the Random Effects in the model, the random intercept per participant that is included. Conditional R^2 is the overall model R^2 when the Random Effects are included.***

4.3.4 Part 2: Associations between score at follow-up, age, and score at baseline

Next we ran multiple regressions to investigate whether task score at baseline, age at follow-up, or change in age from baseline to follow-up were good predictors of task score at follow-up. Results are shown in **Table 4.7**. For all task measures except CPT omission errors, there is significant association present between score at baseline and score at follow-up (p's<.05). There is a positive relationship between score at baseline and score at follow-up – suggesting that those who scored worse at baseline also scored worse at follow-up. The effect sizes here vary between tasks. For BDS, Corsi and CFT the effect size is large. For TMT, SWM errors and CPT commission, we have medium effect sizes. SWM strategy has a small effect size.

Table 4.7 Results of Multiple Regressions using Score at Follow-up as Outcome, with Score at Baseline, Age Change and Age at Baseline as predictors

	TMT at Follow-up			**BDS at Follow-up**			**SWM at Follow-up**		
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Baseline Score	0.21	0.18 – 0.24	**<.001**	0.51	0.48 – 0.54	**<.001**	0.36	0.32 – 0.39	**<.001**
Age change	0	-0.03 – 0.04	.895	0.02	-0.01 – 0.05	.111	0	-0.03 – 0.03	.935
Baseline Age	-0.01	-0.04 – 0.03	.63	0.01	-0.02 – 0.04	.46	0.02	-0.01 – 0.06	.168
Observations		3526			3379			3283	
R^2 / R^2 adjusted		.044 / .043			.264 / .263			.127 / .126	

	SWM Strategy at Follow-up			**CORSI at Follow-up**			**CFT at Follow-up**		
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Baseline Score	0.16	0.12 – 0.19	**<.001**	0.5	0.45 – 0.54	**<.001**	0.53	0.50 – 0.56	**<.001**
Age change	-0.02	-0.05 – 0.02	.373	0.01	-0.04 – 0.06	.807	0.01	-0.02 – 0.04	.369
Baseline Age	0.02	-0.01 – 0.06	.188	-0.03	-0.08 – 0.02	.207	0	-0.03 – 0.04	.771
Observations		3283			1258			3133	
R^2 / R^2 adjusted		.025 / .024			.247 / .245			.282 / .281	

	CPT Omissions at Follow-up			**CPT Commissions at Follow-up**		
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Baseline Score	0.07	-0.02 – 0.17	.14	0.29	0.20 – 0.39	**<.001**
Age change	0.02	-0.08 – 0.12	.754	0.06	-0.03 – 0.16	.201
Baseline Age	0.08	-0.02 – 0.18	.134	0.09	-0.01 – 0.18	.074
Observations		399			399	
R^2 / R^2 adjusted		.011 / .004			.097 / .090	

Beta = Standardized Betas for each predictor. CI are 95% CI for the Standardized Beta values. P-values <0.05 are highlighted in Bold.

4.3.5 Part 3: Associations between change in task score and change in age, age at baseline and score at baseline

Results of the multiple regressions are shown in **Table 4.8,** and relationships are illustrated in **Figures 4.5 and 4.6.** For all tasks, there is a significant negative association between score at baseline and change in score between baseline and follow-up. This suggests that people who were scoring well at baseline improved less than those who scored poorly at baseline. The strongest predictor of change in score was respective task score at baseline. Further, there is no significant association of change in score with either of the age predictors, i.e. there was no association between participants' age at baseline testing and their change in task scores, or their difference in age between baseline and follow-up (age change) and their change in task scores. This applies to all the cognitive tasks.

Table 4.8 Results of Multiple Regressions using Score Change from Baseline to Follow-up as Outcome, and Score at Baseline, Age Change and Age at Baseline as predictors

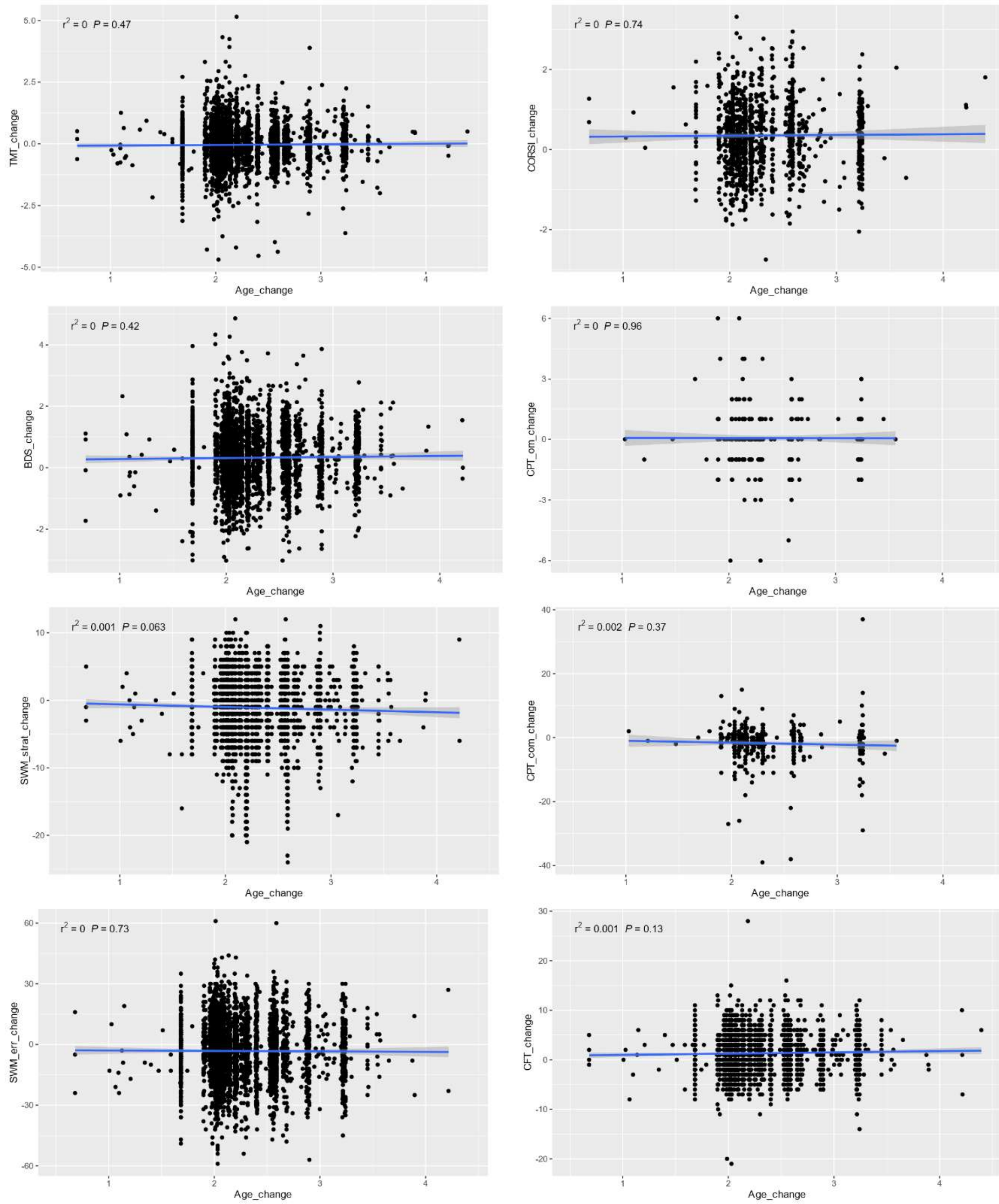
	TMT Change			BDS Change			SWM Change		
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Baseline Score	-.63	-.66 – -.61	**<.001**	-.42	-.45 – -.39	**<.001**	-.59	-.62 – -.56	**<.001**
Age change	0	-.02 – .03	.895	.03	-.01 – .06	.111	0	-.03 – .03	.935
Baseline Age	-.01	-.03 – .02	.63	.01	-.02 – .04	.46	.02	-.01 – .05	.168
Observations	3526			3379			3283		
R^2 / R^2 adjusted	.401 / .400			.180 / .180			.347 / .347		

	SWM Strategy Change			CORSI Change			CFT Change		
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Baseline Score	-.77	-.79 – -.75	**<.001**	-.38	-.44 – -.33	**<.001**	-.46	-.49 – -.43	**<.001**
Age change	-.01	-.03 – .01	.373	.01	-.05 – .06	.807	.01	-.02 – .05	.369
Baseline Age	.02	-.01 – .04	.188	-.03	-.09 – .02	.207	0	-.03 – .04	.771
Observations	3283			1258			3133		
R^2 / R^2 adjusted	.595 / .594			.149 / .147			.212 / .211		

	CPT Omissions Change			CPT Commissions Change		
Predictors	***Beta***	***CI***	***p***	***Beta***	***CI***	***p***
Baseline Score	-0.65	-0.73 – -0.58	**<.001**	-0.72	-0.79 – -0.66	**<.001**
Age change	0.01	-0.06 – 0.09	.754	0.05	-0.02 – 0.12	.201
Baseline Age	0.06	-0.02 – 0.13	.134	0.06	-0.01 – 0.13	.074
Observations	399			399		
R^2 / R^2 adjusted	.428 / .424			.525 / .522		

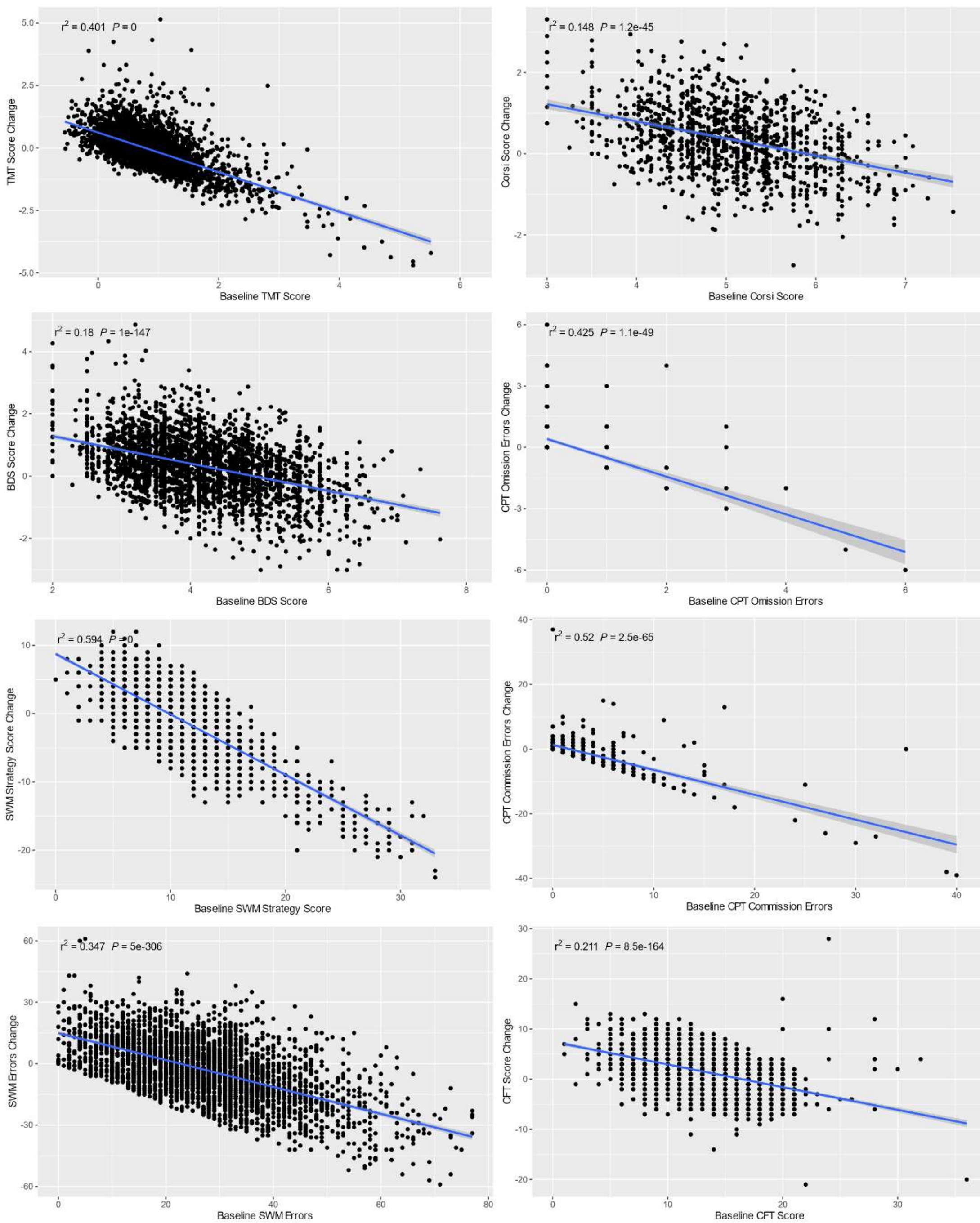
Beta = Standardized Betas for each predictor. CI are 95% CI for the Standardized Beta values. P-values <0.05 are highlighted in Bold.

Figure 4.4 Results of linear regressions with age change as predictor, and change in task score between baseline and follow-up as outcome.



Note. Individual task scores are shown as black dots. The blue line represents the association between change in task score and age change. The Grey shaded area represents the Standard Error of the Mean.

Figure 4.5 Results of linear regressions with score at baseline as predictor, and change in task score between baseline and follow-up as outcome.



Note. Individual task scores are shown as black dots. The blue line represents the association between change in task score and score at baseline in that task. The grey shaded area represents the Standard Error of the Mean. The graphs illustrate the negative relationship between task score at baseline and improvement in that score between baseline and follow-up.

4.3.6 Part 4: Exploration of potential practice effects across time points

Analysis A: is there any association between age and task scores within baseline or follow-up assessment periods?

We ran linear regressions of task score by age within data from each of the two time points. R^2 and p-values for each regression within time points are shown in colour alongside the scatterplots above in **Figure 4.2,** and results are shown in **Table 4.9**. There was some improvement in specific task scores associated with age within the assessment windows: Within baseline, CFT, Corsi and SWM total errors showed significant association with age, and for only the BDS task within follow-up. However, the effect sizes were extremely small for all the associations observed (Adjusted R^2 values between .003 and .008 at baseline, and .001 at follow-up).

Table 4.9 Results of linear regressions, with each task score as outcome, and age in years as predictor in each model. Regressions are run within data from each assessment point separately

Outcome	**Predictor**	***Std. Beta***	***CI***	***p***	**R^2 / R^2 adjusted**	***N* obs.**
			Baseline			
TMT	Age Years	-.01	-.04 – .01	.246	.000 / .000	6424
BDS	Age Years	0	-.02 – .03	.872	.000 / -.000	6084
SWM errors	Age Years	0	-.02 – .03	.942	.000 / -.000	6305
SWM strategy	Age Years	.09	.07 – .12	**<.001**	.008 / .008	6305
Corsi	Age Years	.05	.02 – .08	**.002**	.003 / .002	3791
CFT	Age Years	.05	.03 – .08	**<.001**	.003 / .002	5808
CPT omission	Age Years	0	-.05 – .05	.941	.000 / -.001	1572
CPT commission	Age Years	.02	-.03 – .07	.354	.001 / -.000	1572
			Follow-up			
TMT	Age Years	.02	-.01 – .05	.231	.000 / .000	6424
BDS	Age Years	.03	.00 – .06	**.036**	.001 / .001	6084
SWM errors	Age Years	-.02	-.05 – .01	.143	.000 / .000	6305
SWM strategy	Age Years	.01	-.02 – .04	.37	.000 / -.000	6305
Corsi	Age Years	.01	-.03 – .05	.632	.000 / -.000	3791
CFT	Age Years	.01	-.01 – .04	.334	.000 / -.000	5808
CPT omission	Age Years	.01	-.05 – .08	.735	.000 / -.001	1572
CPT commission	Age Years	-.04	-.11 – .02	.219	.002 / .001	1572

Std. Beta = Standardized Betas for each predictor. CI are 95% CI for the Standardized Beta values. P-values <0.05 are highlighted in Bold.

Analysis B: Are associations between age and task score the same within sub-samples of participants who completed only one assessment point vs those who completed both assessments?

We ran linear regressions of task score by age for each of the cognitive tasks within two subsamples of participants: The first group were participants with data at only one time point for each task respectively, the second were those who had data at both time points for each task respectively.

Comparing the two sections of results in **Table 4.10,** we can see that the observed developmental effects are not entirely related to practice effects. Within the subsample with data at only one data point, there are still significant associations of task scores with age for all the task measures that showed significant development in the whole group analyses presented in Part 1 (i.e. we see significant effects of age for all task measures except CPT omission errors). However, there are notable decreases in age-related effect sizes in the one-time only group compared with those who completed the tasks at both time points, suggesting that at least some of the age-related performance increase effects we observed in the whole group analysis above may be down to practice effects caused by completing the same tasks at both time points.

One caveat to this finding is that the two groups are not identical in terms of task performance: as shown in the descriptive statistics in the Methods section of this Chapter, **Table 4.3**, the participants with data at both time points show slightly higher scores on average than those with data only at one time point, and this difference in scores is present in baseline assessment. It is therefore feasible that development trajectories across this time period may also differ between those who attended school on both assessment days and therefore had data for both time points, and those who completed just one assessment.

Table 4.10 Results of linear regressions, with each task score as outcome, and age in years as predictors. Regressions run with data from both assessment points combined as a single dataset

Outcome	**Predictor**	***Std. Beta***	***CI***	***p***	**R^2 / R^2 adjusted**	***N* obs.**
	Subsample with data at only one time point per task					
TMT	Age Years	.05	0.02 – 0.08	**<.001**	.003 / .003	4290
BDS	Age Years	.12	0.09 – 0.15	**<.001**	.013 / .013	4187
SWM errors	Age Years	.09	0.06 – 0.12	**<.001**	.009 / .008	4363
SWM strategy	Age Years	.10	0.07 – 0.13	**<.001**	.010 / .010	4363
Corsi	Age Years	.12	0.09 – 0.16	**<.001**	.016 / .015	3366
CFT	Age Years	.13	0.10 – 0.16	**<.001**	.017 / .017	4240
CPT omission	Age Years	-.02	-0.07 – 0.03	.414	.000 / .000	1686
CPT commission	Age Years	.07	0.02 – 0.12	**.004**	.005 / .004	1686

Outcome	Predictor	*Std. Beta*	*CI*	*p*	R^2 / R^2 adjusted	*N* obs.
	Subsample with data at both time points per task					
TMT	Age Years	.02	.00 – .05	**.046**	.001 / .000	7052
BDS	Age Years	.16	.14 – .18	**<.001**	.026 / .026	6761
SWM errors	Age Years	.11	.09 – .14	**<.001**	.013 / .012	6569
SWM strategy	Age Years	.17	.14 – .19	**<.001**	.027 / .027	6569
Corsi	Age Years	.19	.15 – .23	**<.001**	.036 / .035	2518
CFT	Age Years	.16	.14 – .19	**<.001**	.027 / .027	6269
CPT omission	Age Years	-.04	-.11 – .03	.213	.002 / .001	798
CPT commission	Age Years	.16	.09 – .23	**<.001**	.026 / .025	798

Std. Beta = Standardized Betas for each predictor. CI are 95% CI for the Standardized Beta values. P-values <.05 are highlighted in Bold.

4.4 Discussion

This chapter used a combination of linear and multiple regressions, and multi-level modelling, to explore the associations between age and cognitive task scores during late childhood and early adolescence, between the ages of 10 and 16 years. In the first set of analyses, we found significant associations between age and task score overall for all task measures except CPT omissions. These effects were not entirely due to practice effects, as shown in part 4 of our analyses. In part 2 of our analysis, score at baseline was shown to be a strong, positive predictor of score at follow-up for all measures except CPT omissions. Part 2 further showed that age at baseline or age change had no significant association with follow-up score, when including score at baseline in the analysis. Part 3 showed that score at baseline was also a strong predictor of change in score between baseline and follow-up for all measures except CPT omissions, but the association was in a negative direction. Part 3 further showed that age change and age at baseline were not significant predictors of score change for any task measure.

4.4.1 Key Findings and Interpretation

Overall, in line with prior research, we found that measures of EF and general fluid intelligence improve with age across the time period of 10 to 16 years. However, the observed effect sizes were quite small, and some but not all of the improvements observed were likely due to practice effects as the same tasks were repeated at two assessment points by most participants.

Part 1: Age is associated with task scores

The results of the fist set of analyses showed that there are small but significant associations between age and task score across the EF task measures for the TMT, BDS, SWM errors and strategy, Corsi, CPT commission errors, and also on CFT. All task measures except the CPT omission errors measure showed significant associations with age. Our observed effect sizes were all very small – suggesting that there are small but still significant improvements with age in EF tasks across this time period. The observed significant effects remained in the MLM analysis, though the effect sizes were decreased for the associations between age and task score when accounting for a random intercept per participant. Descriptive statistics also showed higher scores on average at follow-up than at baseline.

In relation to previous research, these findings are largely in-keeping with previous studies. The finding that a relative measure of switch-cost from the TMT task improves with age between 10 and 16 fits with findings that a switch-cost measure in a Flanker task was still developing until around age 15 (Lee et al., 2013). We found improvement in our proportion switch-cost measure, but only a very small improvement of around .1 of a SD of the mean between baseline and follow-up.

Previous findings were mixed for the BDS task, in that some studies found no significant development in performance during early adolescence (Brocki & Bohlin, 2004), where others did find development across adolescence (Prencipe et al., 2011). We found that BDS capacity did improve with age in our sample, but capacity only increased between baseline and follow-up by around 1 item on average.

In the Corsi task, some previous research suggested we might see some improvement in spatial working memory capacity (Prencipe et al., 2011), where other research suggested that we might not (Lehto et al., 2003). We found significant improvement in spatial working memory span capacity across our sample, with a very small increase in capacity from baseline to follow-up of around .3 items on average.

For the SWM task, we saw significant but reductions in the total numbers of errors associated with age between ages 10 and 16, with a drop of around 3 errors in the task from baseline to follow-up on average. We also saw improvement in strategy use, with around 1 fewer strategy error being made at follow-up than at baseline. Previous research had suggested that we would see development in the strategy score in particular (De Luca et al., 2003) which fits with our findings. SWM Strategy score is the highest level skill we assess here, which involves planning an overall task approach.

A previous study suggested that CPT omission scores improved with age until around age 12 (Brocki & Bohlin, 2004). Here we found no significant improvement in CPT omission scores with age in our sample. However, we used a shortened CPT task, and therefore the absolute numbers of omission errors was relatively low in our sample. This might be one reason that we have no significant effect of age here. We did see small but significant reductions in the numbers of commission errors with age in our sample. Previous research had found that some types of commission errors measures continue to develop with age until age 13 (Brocki & Bohlin, 2004), but not other commission error types. Given the smaller sample size that we had in the CPT task, and the fact we used a shortened version of the CPT task, we thought that we would not have sufficient variance in the data to split apart the types of commission errors within our analysis.

Part 2: Score at follow-up positively associated with score at baseline and age

We found that score at baseline was a strong, positive predictor of score at follow-up. This means those who scored relatively higher at baseline were also scoring relatively higher at follow-up. With score at baseline included in the model, we found no significant effects of age, either age at baseline or age change. It would be possible to investigate this further by including a multi-step multiple regression model instead of the single, Enter version used here, for example, using age at baseline and age change as a first step, and then considering whether score at baseline would add anything to this model.

Part 3: Change in task score negatively associated with baseline score but not change in age

We found that score at baseline was a strong, negative predictor of score change between baseline and follow-up. This means that those who scored higher at baseline did not improve as much as those who had poorer scores. This is interesting because it counteracts some previous work looking at task score change over time – where often gaps between participants present at early assessment points continue to widen over time.

Taken together, the results across Parts 2 and 3 show that participants with a lower score at baseline: a) continued to be lower scorers in follow-up; but also b) their performance improved relatively the higher scorers' performance between baseline and follow-up assessments. Taken together, this suggests that adolescents who have lower scores and therefore poorer EF skills during late childhood or the start of adolescence continue to perform below their peers who had higher EF skills at the start of adolescence, but that they may start to 'catch up' to their higher performing peers, with the gaps between the highest and lowest performers on average narrowing over the course of adolescence.

One possible reason that we did not find association between age change and score change is that the absolute change values for the tasks were quite low. Furthermore, the age change variable was found to be non-normal, which might have affected the results. The non-normality was not able to be corrected, as it was caused by the fact that participants within the same school had testing carried out on the same day, or at least within the same week, as each other. Therefore we don't have the full gamut of possible values in age change covered – rather, we have substantial clustering of the age change values. Finally, the change scores were only possible to calculate for participants who had data at both time points. This subsample may not be representative of the overall population. Looking at the descriptive statistics in **Table 4.3,** we can see that those who have data for both time points score higher on average than those with only data at that one time point. This effect could be tested statistically with a t-test or ANOVA to see if the groups are significantly different to one another.

Part 4: Practice effects don't explain all associations of task scores with age

This set of analysis investigated whether practice effects were playing a major role in the findings from Part 1, where we observed a significant improvement in many of the cognitive task scores with age overall in the dataset.

Firstly, we found there was some limited development within the baseline assessment window for three task measures: Corsi, CFT and SWM strategy. There was only one significant association observed between a task score and age within the follow-up window, which was the BDS task. The effect sizes were all extremely small, though slightly larger for the baseline tasks than at follow-up. This indicates that the time windows within our assessment points are likely too small to show clear developmental effects. Developmental improvements in cognitive task performance would not be expected to be particularly large within a 1-2 year time frame, and it is likely that individual differences will be greater than the overall association with age. It is worth noting that there were more effects observed within the baseline assessment point. Many of the tasks assessed would still be expected to be showing improvement with development in early adolescence, the time window covered by our baseline assessment. However, for many of the tasks, we might be expecting to see a slowing in the rate of development towards the middle of adolescence, and for some tasks, we would be expecting to see little development in certain task measures by the age of around 13. The finding observed here where more development is observed within the baseline window than the follow-up window therefore fits with previous findings that EF tasks are developing at a faster rate in early adolescence than later in adolescence.

4.4.2 Limitations of this study

One limitation in this study is that participants within a particular school were usually all assessed on or very close to the same day. This means that the age differences between baseline and follow-up were not normally distributed, as the age change values tended to cluster together and were not spread out across the whole possible range. This means that the age change distribution is not normal, and it was not possible to make any correction for this non-normality.

No correction for multiple comparisons was performed in the multiple regression analysis or the multi-level modelling. However, where results have reached significance at the uncorrected p<.05, these significant values would in fact survive Bonferroni correction across this whole set of analyses, as the p-values are all very low. Therefore the findings would still be applicable. Post-hoc power calculations were conducted using the online calculators by Soper (2024). For our multiple regression analyses, we required a sample of 1,411 for an effect size of .01 and power of .90. This is met for all of our analyses, except for a few of the analyses in Part 4, the within-time point analysis for the CPT tasks only. For the MLM models, this showed we would require 1,255 participants for an effect size of .01 at .90 power – this is met for all of our MLM models.

Within the multi-level modelling used in Part 1 Analysis B, we could not include random slopes per participant as we only had at most two data points per participant and this would have resulted in an overfitted model. If we had three or more data points per participant we could model with random slopes per participant in addition to the random intercepts to better account for the similarity of the data within participants. It would also be possible to extend the multi-level modelling to account for other clustering that might be present in the data, for example, it would be possible to add a random effect for each school that the participants attended. This would be a 'nested' form of MLM with multiple levels of clustering nested within each other. This could be useful to include as the data from individuals within certain schools are likely to be somewhat similar to each other, particularly for example within selective schools vs. state schools.

It is important to consider the possibility of practice effects driving the associations observed between age and score. More than half of the participants with data for each task at time 2 had completed the task previously at time 1. It has been noted in the literature that practice effects can persist over at least two year intervals, and that in longitudinal studies, practice effects may to positively bias age-related improvement trends (Salthouse, 2010). In the Salthouse study this effect was particularly observed in younger adults as opposed to older adults. In another study with adults, Salthouse et al demonstrated that an interval of at least 7 years between test administration was needed to reduce the positive practice effects to 0 (Salthouse et al., 2004).

Another issue here is that the participants who are present at both time points may not be representative of the sample as a whole. Students who miss a greater number of school days are more likely to have been absent on one of the testing days, than those who miss fewer school days. There is also evidence that students who miss fewer school days do better in school, and also may have higher IQs and overall better cognitive performance. This is backed up in the data here – inspecting the scores within each assessment point, we can see that the subsample who have valid data at both time points generally perform better than the subsample who only have data in one time point. This is a potential confounding factor, especially in the regression models which were run on the subsample with both time point data – in these models we see no effect of age change on either score at follow-up (Part 2) or with change in task scores (Part 3) but there is a significant association between score at baseline and score at follow-up (Part 2), with those scoring better at baseline also scoring better at follow-up. There is also an effect in analysis Part 3 where those scoring lower at baseline improved more than those who scored better at baseline– since this model is only run within the participants with data at both points.

One potential improvement to this study is that we could consider regressing out any gender and ethnicity effects. When data checks were carried out, there were some significant differences observed between genders on some of the cognitive tasks, and a few effects of ethnicity group on task score also. The effect sizes were not very large, but these are potential confounders which could be addressed by regressing out effects of these variables by including them as covariates in the models.

4.1.1 Future Research Directions

The onset of puberty happens at different chronological ages. It has been suggested that the changes in pubertal hormone levels might act to induce a sensitive period with increased neural plasticity in association with EF development during adolescence (Laube et al., 2020). This means that it is important to consider not only chronological age, but also pubertal status when considering stages of EF development. During the SCAMP assessments, a sub-set of participants in the Bio-Zone add-on study (see **Chapter 2 & Table 2.1**) had saliva samples collected in order to analyse their pubertal hormone levels at the time of assessment, at both baseline and follow-up. There is potential to re-analyse the data for this sub-set of participants with some indication of their individual pubertal stage. However, the samples were only collected on the two occasions at testing – to better understand individual's pubertal stage it would be necessary to have collected multiple additional samples before and after the testing point to see the *relative change* in the individual's hormone levels, rather than just taking a one-off snapshot sample. We also did not consider any non-linear effects of age, where previous research might have suggested we could see this. Future research could consider exploring whether there are non-linear effects present for the Corsi and BDS tasks in particular.

4.4.3 Conclusions

In Chapter 4 we investigated the development of task scores across the age range of 10 - 16 years. We found small but significant effect of age on all but one of our EF task measures, and also on our measure of Gf. This effect was present over and above the practice effect, i.e. the normally expected improvement associated with some participants having completed the tasks more than once.

Chapter 5.
Structural Models of Executive Functions at Baseline and Follow-up Assessment

5.1 Abstract

This chapter investigates the latent structure of EF in the SCAMP cohort at two time points in early adolescence. Participants were *N* = 6,591 participants assessed at baseline, aged between 10.4 and 13.5 years (*M* = 12.05 ; *SD* = 0.48), and *N* = 5,116 participants assessed at follow-up, aged between 13.1 and 16 years (*M* = 14.26; *SD* = 0.52). Task measures and questionnaire data are taken from the main SCAMP assessment battery. A series of exploratory and subsequent confirmatory factor analysis was carried out on data from baseline and follow-up separately to investigate the structural relationships of EF, fluid intelligence and mental health at two time points across early adolescence.

At baseline, a three-factor model best explained the structure of EF data, with components labelled as combined switching and working memory; planning; and attention. A three-factor model was also best fit at follow-up, with slightly different latent components which were labelled as combined switching and working memory; spatial working memory; and attention. At both time points the EF components were significantly correlated with each other, supporting a unity-yet-diversity model of EF structure.

Next, a fluid intelligence measure was added in addition to EF task scores. At baseline, fluid intelligence fit within an underlying component alongside some of the EF measures, resulting in a three-factor model with components labelled combined EF and general fluid intelligence; spatial working memory; and attention. In the follow-up sample, a single factor explained the data best when fluid intelligence was added, which was labelled general intelligence or G.

Finally, we added scores on the five sub-scales of the Strengths and Difficulties Questionnaire (SDQ; Goodman & Goodman, 2009) as a measure of mental health. Results showed that SDQ formed two additional components at baseline, resulting in a five-factor model: three EF components; a problems SDQ component; and a strengths SDQ component. At follow-up, a single factor structure was found containing all the cognitive tasks and the SDQ measures, which was labelled general intelligence or G. Results are discussed in the context of the wider literature.

5.2 Introduction

As covered in **Chapter 1**, adolescence marks a period of significant development in terms of EF components. Measures of inhibition, working memory and switching ability all show significant improvements during adolescence. The structure of EF in unity-yet-diversity models of EF during adulthood has been covered in **Chapter 1.** Here, I will summarise findings relating to the underlying structure of EF in childhood and during adolescence, with a focus on the structure of EF within the age range covered by the SCAMP sample.

5.2.1 EF Structure in Children

Studies looking at structure of EF components in young children have generally found less differentiated EF structural models than in adulthood. One influential example of a factor analysis study looking at EF structure in children aged between 2-6 years, with a mean age of 3.9 years, found a single-factor, unitary model best described EF in this age group (Wiebe et al., 2008). This study was conducted using data from a battery of 10 EF tasks, and used CFA methods to investigate whether children's EF is best modelled by a single factor, two factors or a full three-factor model similar to Miyake's model in adults. Their findings suggest that EFs in young children are less differentiated than in adults, as a single factor was the best model here. This suggests that the neural and cognitive processes that underpin EF task performance undergo some reorganisation during the period between childhood and adulthood, i.e. during adolescence, and that there might be a process of moving from less-differentiated EFs to more differentiated ones over the period of late childhood to adolescence.

5.2.2 EF Structure in Adolescence

Studies considering EF structures in late childhood and adolescence have more mixed findings. Evidence from cross-sectional studies where children within a similar age range as the participants in the present study are presented here. In the studies mentioned in this section, children are grouped into a single group across multiple ages for structural analyses.

In a cross-sectional study of EF structure in children aged 8-13, mean age 10.5, Lehto et al. (2003) found that the best model of EF was a three factor model with inhibition, switching and working memory components, similar to the model found in adults by Miyake et al. (2000). Initially this three-factor solution was identified using EFA. Using follow-up CFA analysis in the same data, they compared the identified three-factor model to one- and two-factor models. They found the three-factor model was the best fit for the data. They also found that allowing the factors to correlate with each other produced better fit metrics than forcing the factors to be orthogonal to one another.

Alfonso and Lonigan (2021) found a two-factor model of EF in early adolescence, in a sample of 174 middle school students in Florida (Mean age of 12.8 years, range between 10 and 15). The two factors were labelled as working memory and a combined shifting and inhibitory control factor. They had theorised a model similar to Miyake and they used three tasks per expected component – yet still found a two-factor model.

In a sample that included a clinical population of adolescents with ADHD and controls, aged 12-19, Barkley et al., (2001) found a three factor model of EF using a battery of cognitive tasks. They described these factors as CPT inattention, working memory, and CPT inhibition. CPT omission errors, CPT correct response speed standard error, and variability of the standard error of CPT response speed measures loaded on CPT inattention; CPT commission errors and CPT correct response speed loaded on the inhibition factor; backward digit span, Simon task, verbal fluency, object usage task and fluency task loaded on a working memory factor. Their clinical population of teens with ADHD showed significantly greater inattention than controls without any diagnosed conditions. No differences between controls and these clinical participants were observed for the working memory or inhibition components.

Malagoli & Usai (2015) conducted a cross-sectional analysis of WM and inhibition measures in a sample of participants aged 14-19. They found a two-factor model was the best fit for their data, with their four inhibition tasks clustered on one factor, and WM measures forming a separate factor. They find 2 factor model of EF is a better fit than a one factor model. They only used tasks of WM and inhibition so it was not really possible for them to find any more factors as they did not assess switching for example. They do also found that the WM and inhibition tasks are reasonably well related to each other – so a unity-yet-diversity hypothesis of EF in adolescence is supported here. Inhibition measures were all on one factor here – this contrasts to Friedman & Miyake (2004) who found inhibition forms multiple factors in adults.

In slightly younger participants aged 6-13, Brocki & Bohlin (2004) found a three-factor model of EF. They described their components as Disinhibition, Speed/arousal, and Working memory/Fluency. EFA was used so factors could be any structure – this is similar to our analysis approach, as they were not looking for any pre-defined structure. They found low unity among the factors identified, with only correlations of -.12 to -.15, which contrasts with most other EF factor analysis findings. Anderson et al. (2001) used PCA methods, and found a five-factor model was the best explanation for their EF data in participants aged 11-17, who were treated as a single group.

Adolescence is a period of rapid development with multiple biological, behavioural, cognitive and societal phases. As such collapsing across time points during adolescence means it is possible that we are collapsing across developmental phases within adolescence. In the cross-sectional studies reviewed above, participants were often grouped into widely ranging age groups, with a single model being used to describe the data across all age groups overall. This means that it is not possible for this research to highlight any developmental changes in EF structure across the tested age groups. This could be problematic as other research has indicated that structures of EF components change and develop significantly across adolescence. Better evidence for the ways in which structure of EF changes during adolescence comes from studies where the same tasks are applied to participants of different ages. These studies may be either cross-sectional in approach, with data from participants in different age groups are structurally analysed separately, or longitudinal, where the same participants are assessed at different times. In studies providing the best evidence for the changing structure of EFs over adolescence, factor analysis is carried out within narrow age groups, perhaps as small as single year of age. A strength of this type of study is by using the same tasks across different age groups, variance due to differences in task requirements is removed, and therefore gives us a purer estimate of the developmental related differences in EF structure than comparing results across studies which have used different sets of tasks. A further strength of longitudinal studies, i.e. of using follow-up assessments using the same tasks with the same participants, is that variance due to both task requirement and individual participant characteristics is reduced.

Huizinga et al. (2006) used a cross-sectional design, performing latent class modelling with children in different age groups (7, 11, 13 and 15 year olds) and young adults (21 year olds). They used tasks designed to tap working memory, inhibition and shifting, and two complex EF tasks (WCST and TOL). They found that a two-factor model best explained their data at all ages considered. The two factors were working memory and switching. Interestingly they found that inhibition did not lie on a single factor within their sample, rather was best explained with individual tasks as manifest variables in the final model. Xu et al. (2013) also carried out cross-sectional structural analysis of EF, using groups of 7-9, 10-12, and 13-15 year olds. They found a single EF factor was the best model in their youngest and middle groups, then a three-factor model with working memory, inhibition, and shifting in oldest group.

Lee et al. (2013) used structural analysis on data from a battery of EF tasks in children aged between 6 and 15, considering models within single year age groups. They used CFA methods to check for the most appropriate structure at each age, using Miyake et al.’s three factors as a basis for their model structures. Using CFA they compared a full three-factor model with all possible two-factor models and a unitary model. For participants aged 6-13 years, they found a two-factor model was the best fit for the data. The two factors were updating working memory, and a combined inhibition-switching factor. Although a three-factor model also had good fit metrics for participants aged 6-12, the three-factor model was inadmissible as the model parameters required a correlation of greater than 1 between the inhibition and switching components. Within their age 11 and 12 participants, they found that the three-factor model was admissible, but there was a strong correlation between the inhibition and switching components (r =.86), and the fit metrics were similar between the two- and three- factor models in this age group, so the two-factor model was deemed to be more parsimonious and therefore was accepted for this group. For participants age 13 and 14, they selected the three-factor model, though the authors noted reservations regarding high correlations between latent variables of switching and inhibition in the 13 year old group, and some issues with similarity of fit metrics in the 14 year old group. In the age 15 group, they found more convincing evidence for a three-factor model, where a structure with working memory, inhibition and switching as separate components had improved fit metrics over any two- or one- factor model, and also the correlation between the any of the latent variables was not considered to be overly high.

A good feature of the Lee et al. (2013) study is that they followed participants longitudinally over 4 years, reducing variance due to individual participant characteristics over time. Results for the latent variable modelling at each age point use all the data from all participants across that age group, i.e. the data were combined from participants who had previously completed the tasks at previous time points, and from participants who were completing the tasks for the first time. This is somewhat similar in approach to the method used in this paper – in that some participants in the SCAMP follow-up assessment had completed the assessment at baseline and some were not present at baseline assessment, so were completing the tasks for the first time.

5.2.3 Summary of Findings from Adolescent Studies

Findings relating to the development of structure of EF during adolescence have been rather mixed. Three-factor structures of EF have been observed in participants aged 12-19 (Barkley et al., 2001), and at ages 13-15 (Lee et al., 2013; Xu et al., 2013). Among younger participants, a one-factor model has been observed in participants aged 7-9 and 10-12 (Xu et al., 2013), where in other studies a two-factor model has been found in participants aged 6-12 (Lee et al., 2013). Taken together, these findings could suggest that some point around age 13 marks the transition to a more differentiated structure of EFs. However, other research has also found three-factor models in younger participants aged between 6 and 13 (Brocki & Bohlin, 2004; Lehto et al., 2003), and even a five-factor model has been observed in older teenage participants aged 11-17 (Anderson et al., 2001).

Furthermore, the nature of the factors identified in these studies varies significantly, with studies using exploratory factor analysis methods (such as EFA and PCA) frequently finding structures other than the widely used model of working memory, inhibition and switching. The fact that factors are not always found to match this structure might depend on the specific tasks and measures used in studies; or it may be that there is no consistent structure to EF during adolescence, rather structures are in a state of flux across this period.

One reason for the lack of consistency across studies might be the use of varying tasks in the factor analysis research. As discussed in **Chapter 1,** EF tasks by their nature are embedded within complex tasks which require multiple cognitive processes to work together to successfully complete them. Labelling of factors identified by EFA is therefore complex. In EFA, we do not have a specific expected structure to the data. When considering the labels we apply to the components identified in EFA, we need to carefully consider both the main EF ability that is needed for task performance across the contributing tasks, any other EFs which are required, and any other important cognitive abilities that may also be required for performance across the tasks. An example illustrating this complexity comes from Lehto et al. (2003). They labelled an EF component which included Tower of London performance as ‘inhibition’. However this task is often considered as a more complex EF task, rather than purely relating to inhibitory processes, and the component might therefore be better labelled as something like ‘planning’ or ‘general EF’.

Task selection is very important in factor analysis research, and changing the specific tasks used to assess EF can greatly influence the latent variable structures that are identified when modelling. This is of particular issue in CFA analysis, where tasks are chosen carefully to target particular hypothesised EF components. Miyake et al. (2000) make clear that they only put in tasks thought to fit with their three hypothesised components; they also explain that including other types of EF tasks would be theoretically justifiable and might well change the latent structures a factor analysis model would produce. Alfonso and Lonigan (2021) found a two-factor model of EF in early adolescence, in a sample of 174 middle school students in Florida (M age=12.78, SD =1.18). The two factors were labelled as working memory and a combined shifting and inhibitory control factor. They had theorised a model similar to Miyake and they used three tasks per expected component – yet still found a two-factor model. Use of EFA is also complex, in that labelling factors within the models can be difficult, as it may not be clear exactly what cognitive process is underlying a given factor, or what process the task measures have in common with each other.

Many different tasks have been used across the studies cited here, and where the same tasks have been used, specific task measures have also differed. Given that specific EF tasks and even individual task measures have been shown to have different developmental trajectories, it is likely that this is a factor in different researchers finding slightly different results in EF structural research. This makes it difficult to reconcile from the existing literature overall what the pattern of development of EF structure is during childhood and adolescence. Studies use differing sample sizes, clinical or non-clinical populations, and participants of slightly differing ages, making overall conclusions more difficult (Zelazo et al., 2016). Furthermore, in adolescence there is a complicating factor of puberty to take into account. Within any particular age or school year sample we will have people of different biological stages of puberty. Girls and boys also begin puberty at different ages, on average. The onset of puberty happens at different chronological ages. It has been suggested that the changes in pubertal hormone levels might act to induce a sensitive period with increased neural plasticity in association with EF development during adolescence (Laube et al., 2020). This means that it is important to consider not only chronological age, but also pubertal status when considering EF development in adolescence.

Based on previous work, it is rather difficult to predict what exact structure of EF is likely to come out of our research. However, one possibility is that we will find a less-differentiated structure in the younger participants, and a more differentiated factor structure in the older participants – this broad pattern has been found in studies that have considered the development of EF factor structure in adolescence (Lee et al., 2013; Xu et al., 2013).

5.2.4 Rationale for investigating EF structure in the SCAMP cohort

The above summary of previous research in this area sets the scene for our study – in younger children, research has generally shown a less differentiated, perhaps single factor or two-factor model to be a good fit for EF data. In adults, a three-factor model of EF is largely accepted in the literature, as described in **Chapter 1.** Adolescence marks a period of development in terms of improving performance across broad EF components (as described in **Chapter 1**) and of specific EF tasks (see Introduction to **Chapter 4** for a review of previous research, and the findings from **Chapter 4** relating to our specific sample). Structural underpinnings of EF are also believed to be in a period of transition, from the less differentiated models found in early childhood, to the more differentiated models observed in later adolescence and adulthood. Findings from research considering the structure of EFs during adolescence generally indicates that adolescents gradually move from a less-differentiated structure present in childhood, to a more differentiated three-factor structure by the mid-teens (K. Lee et al., 2013). This study will explore the factor structure present in EF at two time points during adolescence, firstly within the baseline assessment period, with participants aged between 10 and 13 years, and secondly within the follow-up assessment period, where participants were aged between 13 and 16 years.

5.2.5 Strengths and Difficulties Questionnaire

The SDQ is a well-validated indicator of child mental health and behavioural problems. Higher total difficulties scores reflect an increased chance of psychopathology diagnosis (A. Goodman & Goodman, 2009). The questionnaire covers five subscales, over four domains of difficulties: emotional symptoms, conduct problems, hyperactivity problems and peer problems; and one strength domain: prosocial behaviours. The difficulty subscales can be grouped into internalising problems (measured by the emotional problems and peer problems scales) and externalising problems (conduct problems and hyperactivity scales).

EF performance has been found to be negatively related to concurrent mental health in adolescence. Participants aged 11-16 with more errors in the WCST were more likely to have higher scores of depression symptoms. EF performance was also predictive of later anxiety symptoms two years later (Han et al., 2016). EF predicted later anxiety even after accounting for IQ, suggesting a specific association between adolescent EF and anxiety symptoms in later adolescence. WM capacity may act as a mediator of the relationship between trait anxiety and academic achievement (Alfonso & Lonigan, 2021). Anxiety and depression are examples of internalising disorders, where the person has mental health symptoms which are related to their internal cognitive, affective and behavioural problems. Although the SDQ does not measure specific symptoms of anxiety or depression disorders directly, more general internalising problems are assessed by the peer and emotional problem scales.

A recent study has shows that aspects of SDQ and EF are related in adolescence (Donati et al., 2021). Within early adolescence (age 10-13 years), higher internalising and externalising problem scores on the SDQ were associated with poorer working memory and inhibition task performance. In later adolescence (age 14-19 years), significant association was observed between externalising problems and working memory scores, but there was no concurrent association between internalising problem scores and working memory within their later adolescence time point. Using a cross-lagged model, they further showed that internalising and externalising problems as measured by the SDQ predicted worse later working memory scores, but that later inhibitory task performance was not predicted by earlier SDQ scores. Earlier executive function measures did not predict later SDQ score. These findings suggest a developmental link from earlier internalising and externalising problems to a later reduction in EF scores. In the context of the present study, the findings of Donati, Meaburn and Dumontheil (2021) would suggest that we can expect negative correlations between the SDQ problem subscales and working memory and inhibition within our baseline sample, and that we are less likely to observe such correlations between SDQ and EF within the follow-up sample.

In a study investigating group differences between children identified as hyperactive by their teachers and controls, the hyperactivity subscale of the SDQ has been found to be associated with literacy measures, tasks of inhibition and executive function, but not with verbal working memory measures (Adams & Snowling, 2001). This suggests that certain aspects of the SDQ may associate with certain aspects of EF.

Given that the hyperactivity subscale of SDQ has previously been associated with inhibition and executive functions (Adams & Snowling, 2001), we might expect to see that this subscale clusters with some of our EF measures, and may be most likely to cluster with the TMT measure. Adams and Snowling did not however find association between hyperactivity and verbal working memory, leading to a prediction that we may not see this SDQ subscale clustering with our BDS measure.

5.2.6 Factor Analysis Methods

Exploratory factor analysis (EFA) is statistical method by which data from multiple source measures is reduced to a number of factors (also called components, or latent variables) without prior specification of the number of factors to be included in the model (Beaujean, 2014). EFA is somewhat similar in this regard to principal component analysis (PCA), which is another statistical method that extracts components from data without a pre-defined structure in mind. This is in contrast to confirmatory factor analysis (CFA), where a specific pre-planned model structure of latent variables is tested to see whether the pre-planned model structure is a good fit to the dataset. In both EFA and CFA, the methods allow the factors extracted to be either orthogonal to each other, or to be correlated with one another. In previous EF structural research, factors have been found to be correlated with one another (Miyake et al., 2000; Zelazo et al., 2008). This consideration will determine the factor rotation method that will be applied in this study – we have selected an oblique factor rotation method (oblimin), as recommended by Osborne and Costello (2009). Another consideration when constructing factor analysis models is the selection of which model is the 'best' fit for the data. When conducting EFA, multiple possible models could theoretically be used to describe the underlying data patterns, and the researcher must choose between the possible structures. To this end, goodness-of-fit metrics are used. Hu and Bentler (1999) suggest that comparative fit index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) can be used to estimate the performance of a model. Multiple other metrics are also used in the literature. Each fit metric assesses a particular aspect of goodness-of-fit. When judging an EFA model, these metrics can either be chosen a-priori, or it is possible to apply many metrics to the data then see which model produces the greatest number of significant results in order to choose the most appropriate model, this is referred to as a method agreement procedure (Makowski, 2018). We will apply the method agreement procedure to select the best models here.

5.2.7 Aims

This study has three major aims. Firstly, we are interested in seeing what structure of EFs can be observed during adolescence at our two assessment points, in school years 7-8 and school years 9-10 respectively, and whether there are any differences between the EF structures between these age groups. The tripartite model as described by Miyake et al. (2000) in adults provides a working framework for investigation. Using exploratory factor analysis, we would like to investigate what structure EFs have in adolescence; whether this structure is a similar three-factor organisation to that which has been noted in adulthood; and whether we see both unity and diversity of the latent variable structures in adolescence. There is evidence that EF structure changes across adolescence, with research prior research generally indicating that EFs become more differentiated across development (e.g. Lee et al., 2013). Factor analysis will therefore also be used to investigate whether similar EF structures are found at both our assessment points, or whether there are differences in structures at our two assessment points.

Secondly, we want to see whether fluid intelligence sits alongside EF measures within any of the identified EF latent variables, or whether it forms its own distinct factor from the EF tasks. This will be investigated at both assessment points.

Finally, prior research has indicated that mental health issues such as internalising and externalising problems as measured by the SDQ are associated with EF in adolescence ; Donati et al., 2021; Han et al., 2016). Our last aim is to see whether any of the subscales from the SDQ form latent variable factors alongside either EFs or fluid intelligence, and also whether the SDQ subscales sit together in the same factor as each other. Again, this will be investigated in both assessment points.

This study will extend the available literature considering structure of EF in adolescence. A potential strength of this study is that we are using the same participants at both time points; this will enable us to see the development of structures in the same sample of participants, and should reduce variance that might occur by using different participants in the separate analyses. However it should be noted that practice effects might come into play and influence task performance at follow-up (see **Chapter 4**). This study will offer a structural analysis of EF at two time points across adolescence. This should help to identify whether the structure of EF in adolescence is differentiated into separate though related components, or whether a single factor is a better explanation for adolescent EF structure. As we will not be collapsing across wide age bands, rather conducting separate analysis of participants at younger and older time points, we should be able to see whether the structure of EF changes from early to later adolescence. By adding CFT and SDQ measures into the structural analysis, we should be able to see whether EFs are structurally separate from fluid intelligence, and whether mental health measures cluster with EFs.

5.3 Methods

5.3.1 Task and Questionnaire Measures

The EF task measures were: TMT Proportion score, BDS score, SWM errors, SWM strategy, CORSI score, CPT omission errors, and CPT commission errors. For fluid intelligence, we used the total score on the CFT task. For all the cognitive tasks, we used Z-scored data at baseline and scaled data at follow-up. See **Chapter 2** for details on measure selection, calculation and scaling methods. For the SDQ measures, sum scores out of 10 for each of the five subscales (emotional symptoms, conduct problems, hyperactivity problems, peer problems, and prosocial behaviours) were used. For details of the SDQ scoring methods and the items in the SDQ, see **Chapter 2 and Appendix A** respectively.

5.3.2 EFA and CFA Modelling Procedures

Three sets of models were investigated within each testing point, i.e. within baseline and follow-up data. Models 1-3 were carried out using baseline data. After exclusions based on incorrect ages were carried out (see **Section 2.7** for details on exclusions and data cleaning), *N* = 6,591 participants were assessed at baseline, aged between 10.4 and 13.5 years (*M* = 12.05 ; *SD* = 0.48).Models 4-6 used the follow-up data. Participants at follow-up were *N* = 5,116, aged between 13.1 and 16 years (*M* = 14.26; *SD* = 0.52). In the first stage of modelling, Models 1 and 4 used the EF task scores: TMT Proportion score, BDS score, SWM errors, SWM strategy, CORSI score, CPT omission errors, and CPT commission errors. In the second set of models (Models 2 and 5), CFT score was added, to see if the resulting factor structure would match with that from Models 1 and 4 respectively, and to see whether fluid intelligence would appear as a separate factor from the EF measures. Finally, Models 3 and 6 added scores for the five subscales of the strengths and difficulties questionnaire (SDQ) to see whether these were related to any of the identified EF or cognitive factors from the previous models, or whether the SDQ formed factor(s) separate to the cognitive tests.

For each model, the datasets were first inspected to ensure they were suitable for factor analysis using the Kaiser, Meyer, Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity to check for sufficient correlation in the data to support factor analysis. EFA was then applied to the dataset to explore the factor structure. EFA was conducted as we have no clear idea of what structures we are expecting to observe a-priori: as we have only a limited number of EF measures, and the tasks were not specially selected to tap specific aspects of EF, rather to act as general EF ability indicators, and also we don't have multiple tasks per EF, the data were not suitable to pre-plan how we might expect the tasks to line up with particular EF components in e.g. Miyake's model, so CFA was deemed unsuitable in the first instance. As there is no consensus in the literature on what is the "best" goodness-of-fit statistic to use when determining the success of a structural model, a method agreement procedure was used (Makowski, 2018). This procedure runs multiple factor selection routines, producing multiple goodness-of-fit statistics for all the potential model structures, then selects the factor structure with the highest rate of consensus across the different methods. This procedure was implemented via the psycho package in R for each model.

We next wanted to check whether the optimal model identified from the EFA analysis was better than other possible models. EFA was applied to the data in order to identify the optimal factor structure. Optimal structure was identified using the method agreement procedure (Makowski, 2018). CFA was then applied to the testing set using the optimal structure obtained from the training set. We then statistically compared the optimal model identified in the EFA phase with the structure with the second-highest consensus rating obtained from the method agreement procedure. Goodness-of-fit of the two compared models was assessed using comparative fit index (CFI) statistics to decide which of the models to accept. Finally, CFA using the model structure obtained from the EFA analysis was applied to the whole dataset to get the overall fit metrics and factor weightings.

R was used to carry out the factor analysis, using packages psycho, lavaan and performance; images illustrating the structural models were produced using R packages ggplot2 and tidySEM. Oblimin rotation method was used for all factor analyses. This is an oblique rotation method which allows the latent variables to correlate with each other. This was chosen as correlations between the underlying factors would be expected based on the literature review (e.g. Lehto et al., 2003). Missing data were removed listwise for each model. The method of estimation used was maximum likelihood. Model fit statistics are provided for chi-squared (χ^2), comparative fit index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Rules of thumb were applied to determine whether the models were a good fit within the different model fit metrics: For χ^2, significant p-values of below .05 indicate good model fit (Beaujean, 2014); for CFI and TLI, values greater than 0.90 indicate adequate fit, and greater than 0.95 indicate good model fit; for RMSEA, values of below .06 indicate good fit; for SRMR, values of below 0.08 indicate good fit (Hu & Bentler, 1999).

5.4 Results

5.4.1 Descriptive Statistics

Means (*M*) and standard deviations (*SD*) for the age of the participants at testing, and variables entered into the structural analysis for the cognitive tasks and the SDQ subtasks within each assessment point are given in **Table 5.1**. The means and SDs presented below are the Z-scored variables at baseline, and the follow-up scores that have been transformed using the mean and SD from baseline results. This means at baseline, the mean scores for the cognitive tasks are set to 0; the SDs to 1. At follow-up, the scores reflect the average increase in performance compared with baseline. The process of creating the Z-scored and transformed variables is described in **Chapter 2, Section 2.7.6**. The Z-scored and transformed variables for the cognitive tasks were entered into each of the structural models. The SDQ measures were not transformed, the raw scores presented below were entered into models.

Table 5.1 Descriptive statistics for the Baseline Z-scored and Follow-up Transformed measures used in Factor Analysis

	Baseline			**Follow-up**		
Measure	*N*	*M*	*SD*	*N*	*M*	*SD*
Age at testing (years)	6672	12.07	0.47	5138	14.26	0.51
TMT proportion (B-A)/A	6424	0	1	4918	0.08	0.97
BDS score	6090	0	0.99	4864	0.35	1.1
SWM total errors	6313	0	1	4627	0.27	0.97
SWM strategy errors	6313	0	1	4627	0.25	0.82
Corsi score	3795	0.01	1	2093	0.41	1.14
CPT omission errors	1573	0	1	912	-0.03	1.11
CPT commission errors	1573	0	1	912	0.25	0.9
CFT total score	5816	0	1	4701	0.33	1.05
SDQ emotional problems	5456	2.79	2.22	3967	3.18	2.38
SDQ conduct problems	5456	2.75	1.45	3967	2.71	1.48
SDQ hyperactivity	5456	4.68	1.57	3967	4.61	1.61
SDQ peer problems	5456	4.31	1.28	3967	4.21	1.24
SDQ prosocial behaviours	5456	7.5	1.9	3967	7.12	2.01

Note: Data are presented for the Z-scored data at Baseline and Transformed data at Follow-up, as entered into the factor analysis models in this chapter.

Correlations between all the task and questionnaire measures used in factor analysis are given in **Tables 5.2 and 5.3**. At both baseline and follow-up, significant positive correlations at $p<.05$ are present between all the cognitive task measures, except between CPT omission errors and SWM strategy score. The effect sizes are all small, with the highest R^2 value at baseline being .12 for the correlation between CFT and BDS, and .15 in follow-up for the same pair of tasks. The correlations between the EF and fluid intelligence tasks at follow-up were all somewhat higher than those at baseline, suggesting a general increase in the relatedness between all the tasks (this may also be interpreted as a reduction in the differentiation of EF components across this time period).

At baseline, significant ($p<.05$) correlations are present between all the subscales of the SDQ except between SDQ prosocial and SDQ emotions. Interestingly, the direction of these associations is positive between all the subscales, i.e. amongst all the difficulty domains (emotion, conduct, hyperactivity and peer problems) and the strength scale (prosocial behaviours), except for between prosocial behaviours and conduct problems, which are negatively correlated. Significant ($p<.05$) correlations are present between all the SDQ subscales at follow-up. The direction of these associations are the same as at baseline: all positive except for the correlation between prosocial behaviours and conduct problems.

Finally, there are some significant correlations between the cognitive tasks and SDQ subscales. At baseline, SWM Strategy is not correlated with any of the SDQ measures (p's all >.05). Emotional problems are correlated negatively with BDS, Corsi, CPT commission and CFT. Conduct problems are correlated negatively with all task measures except SWM strategy. Hyperactivity problems are only associated with CPT commission errors. Peer problems are negatively associated with BDS, SWM errors, Corsi and CFT. Prosocial behaviours are positively correlated with BDS, SWM errors, Corsi, CPT omission errors and CFT (all except TMT and CPT commission errors). At follow-up, we see no significant associations between TMT and any of the SDQ measures (*p*'s all >.05). Positive correlation is present between emotional problems and SWM errors. Conduct problems are negatively associated with BDS, SWM errors, SWM strategy, Corsi, and CFT. Hyperactivity and peer problems are not associated with any cognitive tasks at follow-up. Prosocial behaviours are positively correlated with all the cognitive tasks except for TMT. All of the correlations observed are .45 or below, therefore there are no major issues with multicollinearity.

Table 5.2 Correlations between factor analysis variables at baseline

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. TMT												
2. BDS	.10**											
3. SWM Errors	.09**	.28**										
4. SWM Strat	.03*	.07**	.45**									
5. Corsi	.12**	.31**	.28**	.15**								
6. CPT Omission	.05*	.13**	.14**	.05	.14**							
7. CPT Comm	.07**	.18**	.19**	.10**	.22**	.33**						
8. CFT	.09**	.35**	.29**	.14**	.28**	.13**	.18**					
9. SDQ Emotion	-.02	-.06**	-.02	.01	-.04*	-.03	-.05*	-.09**				
10. SDQ Cond	-.03*	-.13**	-.12**	.01	-.11**	-.01	-.10**	-.14**	.34**			
11. SDQ Hyper	.01	-.00	-.00	-.01	.01	.01	-.06*	-.02	.27**	.26**		
12. SDQ Peer	-.03	-.09**	-.04**	.01	-.04*	-.00	-.02	-.08**	.29**	.32**	.24**	
13. SDQ Pro	-.02	.11**	.10**	.01	.09**	.06*	.04	.12**	.01	-.09**	.15**	.20**

Note. Values represent Pearson's correlation R. * indicates p < .05. ** indicates p < .01. White background are correlations amongst cognitive tasks ; light grey are correlations between SDQ measures and cognitive tasks; darker grey are correlations amongst SDQ measures.

Table 5.3 Correlations between factor analysis variables at follow-up

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. TMT												
2. BDS	.12**											
3. SWM Errors	.09**	.36**										
4. SWM Strat	.07**	.23**	.60**									
5. Corsi	.18**	.37**	.33**	.20**								
6. CPT Omission	.10**	.15**	.10**	.04	.13**							
7. CPT Comm	.08*	.22**	.20**	.13**	.24**	.37**						
8. CFT	.09**	.39**	.32**	.19**	.32**	.08*	.20**					
9. SDQ Emotion	.00	-.02	.05**	-.00	.00	-.01	.01	-.02				
10. SDQ Cond	-.03	-.11**	-.09**	-.07**	-.10**	-.04	-.05	-.12**	.29**			
11. SDQ Hyper	.00	.02	.01	.02	.01	-.00	-.05	.01	.26**	.23**		
12. SDQ Peer	.01	.00	.02	.00	.00	.02	.04	-.03	.25**	.31**	.22**	
13. SDQ Pro	.02	.16**	.11**	.04*	.10**	.14**	.12**	.11**	.09**	-.09**	.15**	.22**

Note. Values represent Pearson's correlation R. * indicates p < .05. ** indicates p < .01. The area with a white background shows correlations amongst the cognitive tasks; light grey area shows correlations between SDQ measures and cognitive tasks; darker grey area shows correlations amongst the SDQ measures.

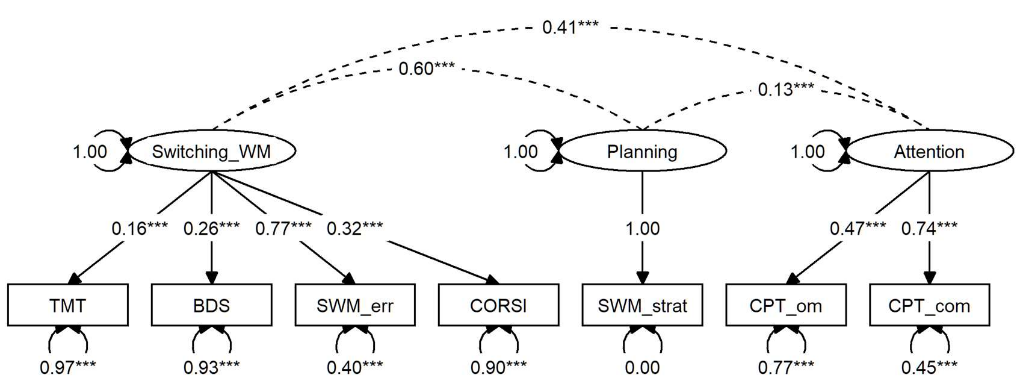
5.4.2 Model 1: Baseline Structure of Executive Functions

EFA analysis suggested a model with three factors was optimal. Using the method agreement procedure, the choice of three dimensions is supported by 10 (43.48%) methods out of 23 (Bentler, CNG, Optimal coordinates, Parallel analysis, Kaiser criterion, VSS complexity 1, BIC, BIC (adjusted), CRMS, BIC). The three latent variables were labelled as: Switching and Working Memory, consisting of TMT, BDS, SWM errors and Corsi scores; Planning, consisting of SWM strategy; and Attention consisting of both CPT measures. In the EFA model, the three latent factors accounted for 40.23% of the total variance of the original data (Switching and Working Memory = 16.35%, Planning = 12.73%, Attention = 11.15%).

The structure identified in the EFA was then applied to CFA, and the three-factor model was compared with a one-factor model as this was the next best in terms of method agreement consensus. The three-factor model gave improved fit statistics compared with the one-factor. Fit metrics indicate the final model is not a very good fit for the data:SRMR (.06) indicates an adequate fit, and the other fit metrics indicate a poor fit of the model to the original data with χ^2(185.20, 12, p<.001), CFI (.83), TLI (.70), RMESA (.10). Factor loadings of each task on the latent variables, and the covariances between the latent factors, are shown in **Figure 5.1.** Goodness-of-fit statistics are presented in **Table 5.4**.

Factor loadings were all significant at *p*<.001, however the loading for the TMT task on the Switching_WM component was very low (0.16), and had a very high and significant error term (0.97), which indicates variance in this task is not particularly well explained by the Switching_WM component. The BDS task also did not load strongly with the Switching_WM component (0.26 with an error term of 0.93). Correlations between the identified components were all significant at *p*<.001, with R values between Switching_WM and Planning .060, Switching_WM and attention .41. Although significant at *p*<.001, the association between planning and attention was only R = 0.13.

Figure 5.1 Results of Model 1: Three-factor latent variable model of executive functions at baseline



Ellipses represent latent variables. Rectangles represent manifest variables. Double headed circular arrows represent error terms. Arrow labels are standardised parameter estimates. Dashed lines represent covariances between latent variables.* indicates p < .05. ** indicates p < .01. *** indicates p < .001. Switching_WM = Switching and Working Memory; TMT = Trail making task; BDS = Backward Digit Span; SWM_err = Spatial Working Memory errors; SWM_Strat= Spatial Working Memory strategy; CPT_om = Continuous Performance Task omission errors; CPT_com = Continuous Performance Task commission errors.

Table 5.4 Latent variable structures, variance explained and goodness-of-fit metrics for Models 1-6

	Baseline models			Follow-up models		
Metric	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Latent factor structure	3 factors: Switching + WM; Planning; Attention	3 factors: EF_G; SWM; Attention	5 factors: EF_G; Planning; Attention; SDQ Probs; SDQ Prosocial	3 factors: Switching + WM; SWM; Attention	1 factor: G	1 factor: G
Total variance accounted for in EFA	40.23 %	36.44 %	41.65 %	26.32 %	23.66 %	15.08 %
CFA Fit Metrics						
χ^2	185.20	43.90	390.27	10.20	293.54	676.21
Df	12.00	17.00	57.00	11.00	20.00	65.00
P	<.001	<.001	<.001	0.51**	<.001	<.001
CFI	0.83	0.98**	0.80	1.00**	0.72	0.52
TLI	0.70	0.96**	0.73	1.00**	0.60	0.43
RMSEA	0.10	0.03**	0.06**	0.00**	0.13	0.11
SRMR	0.06**	0.02**	0.05**	0.02**	0.09	0.09

Models 1 and 4 included the executive function (EF) measures: trail making task, backward digit span, spatial working memory errors and strategy score, continuous performance task omission and commission errors. Models 3 and 5 included EF measures and Cattell's culture fair task (CFT) score. Models 3 and 6 included EF and CFT measures, and scores in the five subscales of the strengths and difficulties questionnaire (SDQ). Latent factors: WM = Working memory; EF_G = Executive Function and General fluid intelligence; SDQ probs = Problems scales of SDQ; SWM = Spatial Working Memory; G = General fluid intelligence. Fit metrics: χ^2 = Chi Squared; Df = Degrees of freedom; p = significance of Chi Squared test; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean residual. ** indicates this metric shows the model is considered to be a good fit to the data.

5.4.3 Model 2: Baseline Structure of Executive Functions and General Intelligence

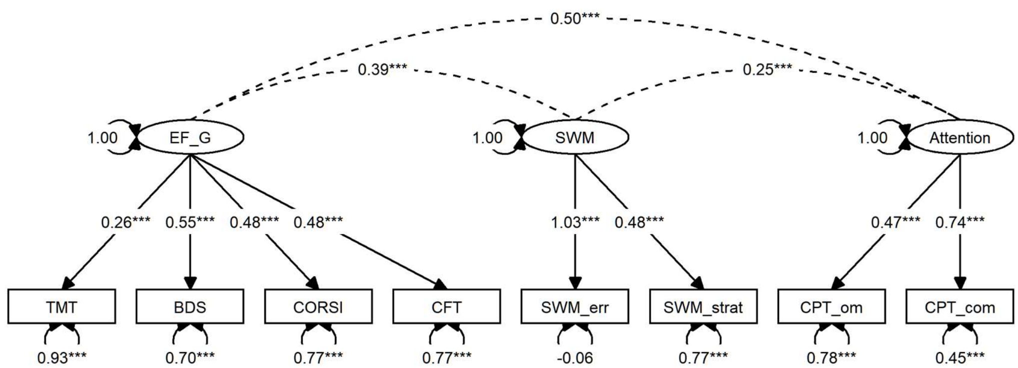
In this model we added CFT into the model in addition to the EF measures, to see if this loads with the EF skills or whether it forms its own separate factor. EFA analysis suggested a model with three factors was optimal. Using the method agreement procedure, the choice of three dimensions was supported by 7 (36.84%) methods out of 19 (CNG, Parallel analysis, Kaiser criterion, VSS complexity 1, VSS complexity 2, BIC, BIC (adjusted)).

The three latent variables were labelled as: executive function and general fluid intelligence, consisting of TMT, BDS, CFT, and Corsi scores; Spatial Working Memory, consisting of SWM errors and SWM strategy; and Attention consisting of both CPT measures. In the EFA model, the three latent factors accounted for 36.44% of the total variance of the original data (Executive functions and general fluid intelligence = 14.40, Spatial Working Memory = 12.40%, Attention = 9.65%).

The structure identified in the EFA was then applied to CFA, and the three-factor model was compared with a one-factor model as this was the next best in terms of method agreement consensus. The three-factor model gave improved fit statistics compared with the one-factor. Fit metrics indicate the final model is a good fit for the data: $\chi^2(43.90, 17)$, CFI (.98), TLI (.96), RMESA (.03) and SRMR (.02) all indicate a good fit of the model to the original data.

Factor loadings of each task on the latent variables, and the covariances between the latent factors, are shown in **Figure 5.2.** Goodness-of-fit statistics are presented in **Table 5.4**. Factor loadings of all measures were all significant at $p<.001$, however the loading for the TMT task on the EF_G component was very low (0.26), and had a very large (0.93, $p<.001$) error term, which indicates variance in this task is not particularly well explained by the EF_G component. The factors were significantly associated with each other, with correlations of R=.50 between Attention and EF_G, and lower correlations between SWM and EF_G (R=.39) and SWM and Attention (R=.25).

Figure 5.2 Results of Model 2: Three-factor latent variable model of executive functions and fluid intelligence at baseline



CFA results. Ellipses represent latent variables. Rectangles represent manifest variables. Double headed circular arrows represent error terms. Arrow labels are standardised parameter estimates. Dashed lines represent covariances between latent variables.* indicates p < .05. ** indicates p < .01. *** indicates p < .001. EF_G=Executive functions and general fluid intelligence; TMT=Trail making task; BDS=Backward Digit Span; SWM_err=Spatial Working Memory errors; CFT=Cattell's Culture Fair Task; SWM_Strat= Spatial Working Memory strategy; CPT_om=Continuous Performance Task omission errors; CPT_com=Continuous Performance Task commission errors.

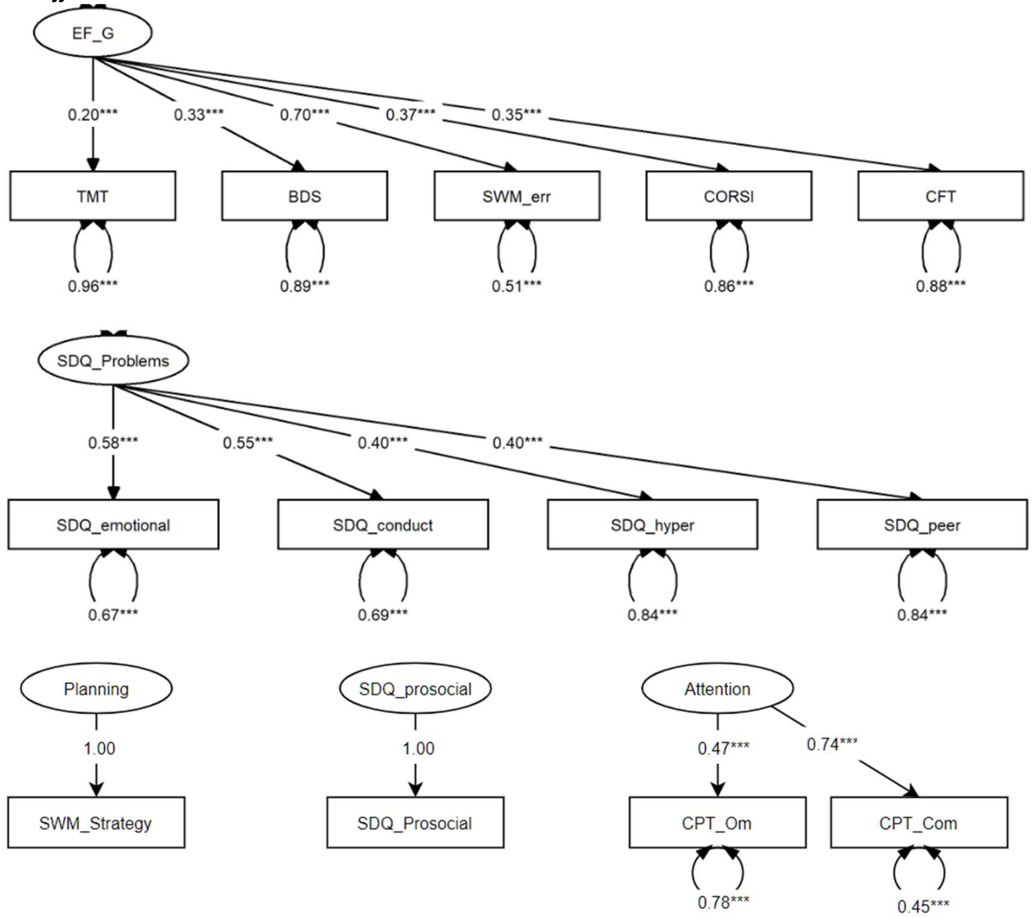
5.4.4 Model 3: Baseline Structure of Executive Functions, General Fluid Intelligence and Strengths and Difficulties Questionnaire

In this model we added the scores on the five SDQ subscales into the analysis in addition to the EF and CFT measures, to see if these load with any of the cognitive tasks. EFA analysis suggested a model with five factors was optimal. Using the method agreement procedure, the choice of three dimensions was supported by 5 (26.32%) methods out of 19 (Parallel analysis, Kaiser criterion, VSS complexity 1, BIC, BIC (adjusted)).The five latent variables were labelled as: Executive function and G general fluid intelligence, consisting of TMT, BDS, SWM errors, CFT, and Corsi scores; SDQ Problems, consisting of the four SDQ problem subscales emotional, conduct, peer relationship and hyperactivity problems; Planning, consisting of the SWM strategy score; and Attention, consisting of both CPT measures. The 5 latent factors accounted for 41.65% of the total variance of the original data (Executive function and G general fluid intelligence = 9.29%, SDQ problems = 9.12%, Strategy = 8.45%, SDQ prosocial = 8.16%, Attention = 6.63%).

The structure identified in the EFA was then applied to CFA, and the five-factor model was compared with a one-factor model as this was the next best in terms of method agreement consensus. The five-factor model gave improved fit statistics compared with the one-factor. Goodness-of-fit statistics were presented in **Table 5.4**. The fit metrics indicate the final model is a fairly good fit for the data: SRMR (.05) indicate good fit, and RMSEA (.06) indicates adequate fit, although the other reported fit metrics (χ^2 = 390.27, 57, CFI = .80 and TLI = .73) indicate a poor fit of the model to the original data.

Factor loadings of each task on the latent variables are shown in **Figure 5.3.** Factor loadings of all measures were all significant at $p<.001$, however the loading for the TMT task on the EF_G component was very low (0.20), and had a very large (0.96, $p<.001$) error term, which indicates variance in this task is not particularly well explained by the EF_G component. **Table 5.5** shows the covariances between all five latent factors. Covariances between the factors indicate significant interrelations amongst the three cognitive task components (EF_G, Planning and Attention), and interrelations between the two SDQ components and Attention and EF_G components, but no association between Planning and the SDQ components.

Figure 5.3 Results of Model 3: Five-factor latent variable model of executive functions, fluid intelligence and strengths and difficulties at baseline



Ellipses represent latent variables. Rectangles represent manifest variables. Double headed circular arrows represent error terms. Arrow labels are standardised parameter estimates. * indicates p < .05. ** indicates p < .01. *** indicates p < .001. EF_G=Executive functions and General Fluid Intelligence; TMT=Trail making task; BDS=Backward Digit Span; SWM_err=Spatial Working Memory errors; SWM_Strat=Spatial Working Memory strategy; CFT=Cattell's Culture Fair Task; SDQ=Strengths and Difficulties Questionnaire; SDQ_hyper=SDQ hyperactivity subscale; CPT_om=Continuous Performance Task omission errors; CPT_com=Continuous Performance Task commission errors.

Table 5.5 Interrelations between the five latent variables found in Model 3

Latent variable pairs	Covariance
EF_G and SDQ_Problems	-0.09*
EF_G and SDQ_Prosocial	0.09**
EF_G and Planning	0.57***
EF_G and Attention	0.49***
Planning and Attention	0.13***
SDQ_Problems and SDQ_Prosocial	0.03
SDQ_Problems and Planning	0.06
SDQ_Problems and Attention	-0.11*
SDQ_Prosocial and Planning	0.02
SDQ_Prosocial and Attention	0.08*

**** indicates p < .05. ** indicates p < .01. *** indicates p < .001.***

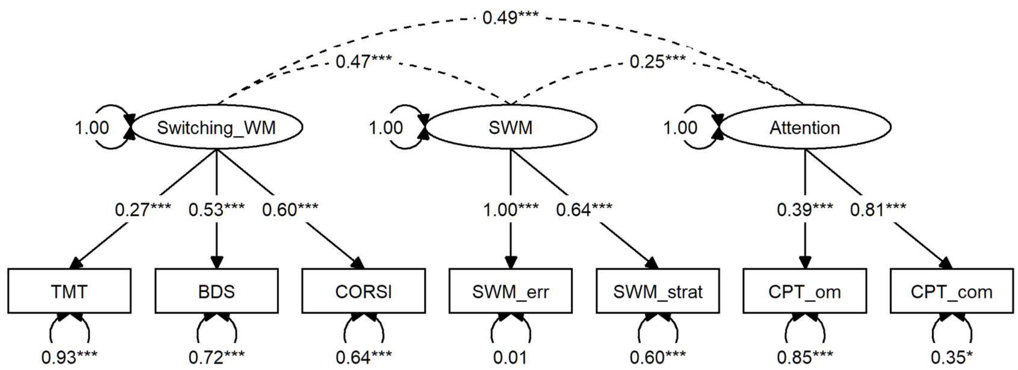
5.4.5 Model 4: Follow-up Structure of Executive Functions

Using the method agreement procedure with the EFA analysis, there was poor agreement on which model would be the best, with models with 1, 3, 4 and 6 factors having agreement in 3 methods each. Further investigation was carried out within the training portion of the data. Here, the choice of 3 dimensions was supported by 6 (31.58%) methods out of 19 (CNG, Parallel analysis, Kaiser criterion, VSS complexity 2, BIC (adjusted), BIC). The three latent variables were labelled as: Switching and Working Memory, consisting of TMT, BDS, and Corsi scores; Spatial Working Memory, consisting of SWM errors and strategy scores; and Attention consisting of both CPT measures. In the EFA model, the three latent factors accounted for 47.85% of the total variance of the original data (Spatial Working Memory = 19.09%, Attention = 16.02%, Switching and Working Memory = 12.74%).

The structure identified in the EFA on the training set was then applied to CFA, and the three-factor model was compared with a one-factor model as this was the next best in terms of method agreement consensus. The three-factor model gave improved fit statistics compared with the one-factor model. Goodness-of-fit statistics are presented in **Table 5.4**. The fit metrics indicate the final model is overall a good fit for the data: χ^2 (10.20, 11), SRMR (.02), RMSEA (<.001), CFI (>.99) and TLI (>.99) all indicate good fit.

Factor loadings of each task on the latent variables, and latent variable covariances, are shown in Figure 5.4. Factor loadings of all measures were all significant at p<.001, however the loading for the TMT task on the Switching_WM component was very low (0.27), and had a very large (0.93, p<.001) error term, which indicates variance in this task is not particularly well explained by the Switching_WM component. Significant correlations are present between all three components at p<.001; with R values of .46 between Switching_WM and SWM, .49 between Switching_WM and Attention, and a small correlation of .25 between SWM and Attention.

Figure 5.4 Results of Model 4: Three-factor latent variable model of executive functions at follow-up



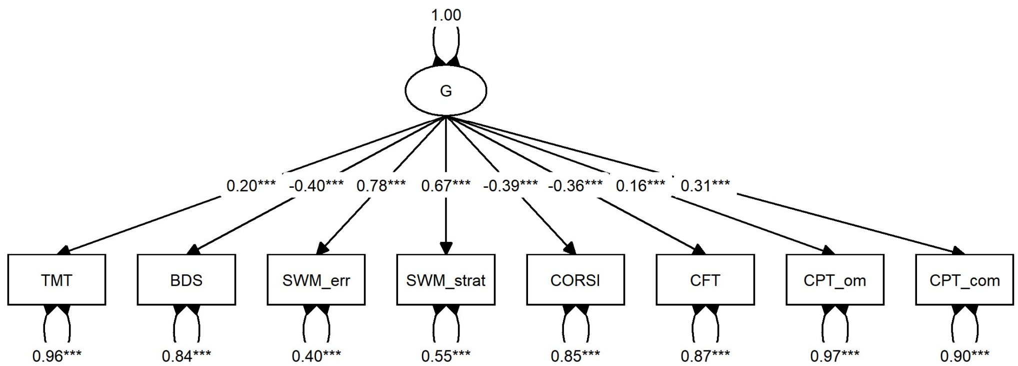
Ellipses represent latent variables. Rectangles represent manifest variables. Double headed circular arrows represent error terms. Arrow labels are standardised parameter estimates. Dashed lines represent covariances between latent variables. * indicates p < .05. ** indicates p < .01. *** indicates p < .001. Switching_WM=Switching and Working Memory; SWM=Spatial Working Memory; TMT=Trail making task; BDS=Backward Digit Span; SWM_err=Spatial Working Memory errors; SWM_Strat= Spatial Working Memory strategy; CPT_om=Continuous Performance Task omission errors; CPT_com=Continuous Performance Task commission errors.

5.4.6 Model 5: Follow-up Structure of Executive Functions and General Intelligence

Using the method agreement procedure with the EFA analysis, the choice of 1 dimensions was supported by 6 (31.58%) methods out of 19 (t, p, Acceleration factor, Scree (SE), Scree (R2), Velicer's MAP). The unique latent variable was labelled as G or general fluid intelligence. In the EFA model, the unique latent factor accounted for 23.66% of the total variance of the original data. The single factor structure was then applied to CFA. Fit metrics indicate the final model is a relatively poor fit for the data: χ^2 (293.54, 20), SRMR (.09), RMSEA (.13), CFI (.72) and TLI (.60) all indicate poor fit of the model to the original data.Goodness-of-fit statistics are presented in **Table 5.4**.

Factor loadings of all measures were all significant at p<.001, however the loading for the TMT task on the G component was very low (0.20), and had a very large (0.96, p<.001) error term, which indicates variance in this task is not particularly well explained by the single G factor. Factor loadings of each task on the latent variable in the final model are shown in **Figure 5.5.**

Figure 5.5 Results of Model 5: One-factor latent variable model of executive functions and fluid intelligence questionnaire at follow-up



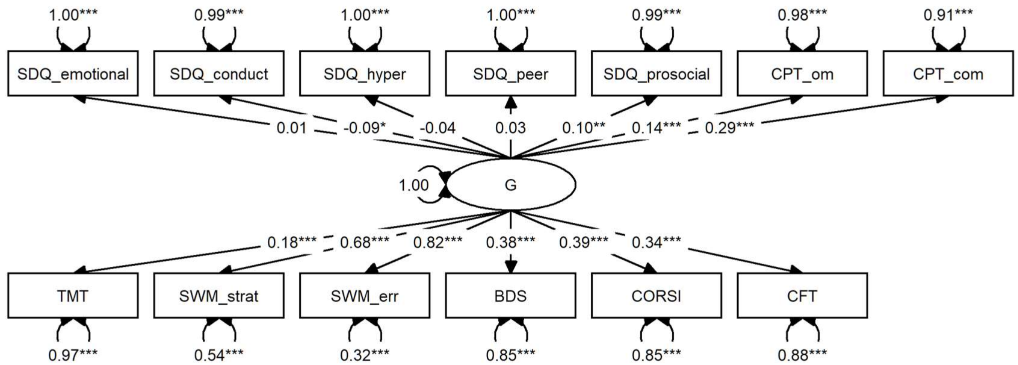
Ellipses represent latent variables. Rectangles represent manifest contributing variables. Double headed circular arrows represent error terms. Arrow labels are standardised parameter estimates. Dashed lines represent covariances between latent variables.* indicates p < .05. ** indicates p < .01. *** indicates p < .001. G=General fluid intelligence; TMT=Trail making task; BDS=Backward Digit Span; SWM_err=Spatial Working Memory errors; SWM_Strat= Spatial Working Memory strategy; CFT=Cattell's Culture Fair Task; CPT_om=Continuous Performance Task omission errors; CPT_com=Continuous Performance Task commission errors.

5.4.7 Model 6: Follow-up Structure of Executive Functions, General intelligence and Strengths and Difficulties

Using the method agreement procedure with the EFA analysis, the choice of a single dimensions is supported by 4 (21.05%) methods out of 19 (t, p, Scree (R2), Velicer's MAP). The unique latent variable was labelled as G or general fluid intelligence. In the EFA model, the unique latent factor accounted for 15.08% of the total variance of the original data. The single factor structure was then applied to CFA. Fit metrics indicate the final model is a relatively poor fit for the data: χ^2, SRMR (.09), RMSEA (.11), CFI (0.52) and TLI (0.43) all indicate poor fit of the model to the original data. Goodness-of-fit statistics are presented in **Table 5.4**.

Factor loadings of each task on the latent variables in the final model are shown in **Figure 5.6.**Factor loadings of all the cognitive task measures were all significant at p<.001, however the loading for the CPT omissions and commission errors measures and the TMT task measure on the G factor were very low (<0.30), and all had very large (>0.9, p<.001) error terms, which indicates variance in the CPT and TMT tasks is not particularly well explained by the single G factor. Furthermore, the SDQ measures did not load well with the G factor, with very low weightings and very high error terms, suggesting the SDQ measures do not fit well with the cognitive task measures and that the single factor model is not a particularly effective model to explain variance in the SDQ and cognitive tasks together.

Figure 5.6 Results of Model 6: One-factor latent variable model of executive functions, fluid intelligence and the strengths and difficulties questionnaire at follow-up



Ellipses represent Latent Variables. Rectangles represent manifest contributing variables. Double headed circular arrows represent error terms. . Arrow labels are standardised parameter estimates. Dashed lines represent covariances between latent variables. * indicates p < .05. ** indicates p < .01. *** indicates p < .001. G=General Fluid Intelligence; SDQ=Strengths and Difficulties Questionnaire; TMT=Trail making task; BDS=Backward Digit Span; SWM_err=Spatial Working Memory errors; SWM_Strat= Spatial Working Memory strategy; CPT_om=Continuous Performance Task omission errors; CPT_com=Continuous Performance Task commission errors.

5.5 Discussion

5.5.1 EF Structure

In terms of the analysis where we considered only EF measures, we found a three factor structure of EF at both time points. This is somewhat similar to what has been observed in literature (e.g. Lehto et al., 2003). However, we did not find evidence for the idea that EFs are differentiated into fewer factors in early adolescence and become more differentiated into a greater number of separate factors in later adolescence, which was one broad prediction that we made based on previous research looking at development of EF structures across adolescence (e.g. Lee et al., 2013; Xu et al., 2013). Rather, we found similar numbers of factorsin EF structure in the younger (age 10-13) and older (age 13-16) participants. In the follow-up data, we found that Model 4 (with three factors) was overall a better fit to the data than was the equivalent three-factor for the baseline data, as the fit metrics indicated a better fit. Furthermore, looking at tables 5.2 and 5.3 we can see that there are consistently slightly higher zero-order correlations between all the EF tasks ijn the follow-up data. It could be argued this indicates greater differentiation into separate factors - in that we have closer correlation between the tasks within the factors identified, with similar correlations across factors as were observed in Model 1 for the baseline data, so the three factors are more differentiated from each other in the follow-up data. However we did not see a greater number of factors at follow-up, which might have been expected from the literature.

Factors were significantly interrelated with each other at both time points. These interrelations had similar effect sizes to those observed in previous literature (e.g. Lehto et al., 2003). However, as we used EFA methods and our tasks were not specifically selected in order to match the working memory, inhibition and switching model proposed by Miyake (Miyake et al., 2000), we did not find clear support for this precise structure of EF. We can however say that we found broad support for a unity-yet-diversity hypothesis of EF in adolescents, with three separable but interrelated factors at age 10-13 and at age 13-16. However this conclusion should be tempered by the fact we have a somewhat limited task selection, which was not designed to replicate previous EF structural modelling research. We have a relatively small number of tasks to conduct factor analysis, and these tasks are somewhat imbalanced in that we have three tasks targeting aspects of working memory, one for switching, and one task measure that might be used to target inhibition, though is not a particularly pure inhibition based task.

Similar structures for EF alone were observed at baseline and follow-up, though the specific task measures that contributed to the three factors differed across our time points, which led to us labelling the factors differently at the two time points. This finding suggests some developmental changes in EF structure occurred between our two assessment points.

Interestingly, we observed a combined switching and working memory latent variable in our baseline data. The fact that WM measures fit into a factor alongside our switching measure appears to be a somewhat unusual finding. In studies where two-factor models are the best explanation for EF data in children, WM is often the EF that is separable from the other two components of EF. For example, (Lee et al., 2013) found a two-factor model of WM and combined switching/inhibition, in participants aged between 6 and 12 years. Alfonso & Lonigan (2021) also found something similar, with a two-factor model with WM and combined switching/inhibition factors in participants aged 10-15. However, we did not have a great number of tasks or measures that tapped inhibition in our analyses, therefore the specific tasks and measures used might have influenced this finding.

One interesting finding is that the working memory tasks (SWM, Corsi, BDS) did not all group together except in the follow up models where the single G factor was the best structure. The Corsi and SWM tasks in particular are ostensibly measuring very similar cognitive processes (in that they are both assessing spatial working memory), but don’t group together in many of our models. One potential avenue of exploring this is to look at what are the differences between the tasks – in the Corsi task, participants repeat an observed order, with little room for any planning or creative problem solving to improve task performance, whereas in the SWM task participants can creatively come up with strategy to solve SWM. Therefore the SWM task is not purely measuring working memory per se, even in the overall performance measure, as the ability to plan and use a helpful strategy will result in improved performance here as compared to the Corsi task, or BDS task.

Another interesting finding is that the TMT task loaded with other task measures across all six models, but it had consistently low weightings values with high error terms. This suggests that there is some shared variance between the TMT task and some of the other tasks, however, there is also some considerable difference between the TMT task and the others that it loaded with. One explanation of this could be that the TMT is the only task to address the switching component of EF – therefore it does not share a huge amount of its variance with the other EF tasks, and that it shares some common EF component with other EF tasks. This finding in itself lends support to the unity-yet-diversity model, in that the TMT did significantly load onto a factor with other EF tasks in each model (suggesting some unity of EF) but its weighting values on these factors were consistently low (suggesting some diversity of switching from the other EFs). However our findings don't lend weight to any specific previous models; rather this research is somewhat unique in its task selection and structures observed. CPT measures formed a separate factor together, which we labelled as attention, across all the baseline models 1-3. One previous study using CPT in an EF structural model also found that measures from a CPT task did not group with other EF measures, however, they found that two CPT measures formed their own individual factors in their model, where ours fit on the same factor (Barkley et al., 2001). One possible reason for our finding is that it could be suggested that the CPT has quite different task requirements to the others in the battery, as it takes considerably longer to complete than the other tasks, and requires continuous extended attention, where the other tasks could mostly be completed successfully even if a person has short breaks in attention. The CPT measures may be more influenced by classroom distractions than other measures, as it is a longer and more boring task requiring more extended focus and attention. Classroom noise has been shown to affect children's performance on cognitive testing (Massonnié et al., 2022). The idea that level of classroom distraction could influence CPT performance in particular could be explored in future analyses by using the data recorded by experimenters on the level of talking that was present in the classroom environment during the assessments. This could be used in for example an ANOVA analysis to see whether reported distraction level is associated with scores in the cognitive tasks, and in particular whether the distraction level is more strongly related with CPT measures than the other tasks.

5.5.2 SDQ and EF

We found no division of the SDQ problems into two factors within this analysis – all four problem subscales clustered on the same factor at both baseline and follow-up. This is different to some conceptions and previous uses of the SDQ, where the problems have been categorised into internalising (combining the emotional problems and peer problems scales) and externalising (combining the conduct problems and hyperactivity scales). Using the five SDQ subscales this study, we found no evidence of the separation of the internalising and externalising scales, in terms which factor they loaded with. Indeed we found no separation of any of the problem subscales from each other – all four difficulty scales clustered on the same factor in both time points. We did find separation of the four problem scales from the prosocial strength scale within the baseline sample, but not within follow-up. The fact that we see the four problem scales clustering together in our sample is somewhat in contrast to previous research which has found independent relationships of internalising and externalising SDQ scales with both concurrent and later EF (Donati et al., 2021). However, as our analysis did not consider the factor structure of SDQ alone, or use the individual questionnaire item responses, we cannot categorically say whether our findings support or refute the use of other structures for SDQ. Future research to investigate the underlying factor structure could be conducted using our dataset, for example carrying out factor analysis on the individual questionnaire items to see what structure is present in the questionnaire subscales. This might be a useful addition to previous research as we have a relatively large sample size of participants who have completed the SDQ, across a representative sample of high school children in London.

Previous research has suggested that SDQ measures are related to EF in a similar age group sample as ours (Donati et al., 2021). We found that SDQ only fit with our EF measures in the follow-up sample, but not at baseline, where it formed its own factors. Even in the follow-up model, we found that the weightings of the SDQ measures on the single G factor were very low, and for most of the SDQ measures, we did not see any significant weighting at $p<.05$ on the G factor. This suggests that even in the follow-up dataset we did not find convincing evidence that the SDQ measures load onto the same factor as EF or fluid intelligence task measures.

The SDQ conduct problems scale had a significant negative correlation with 5 of our 8 cognitive task measures, with the highest association being around $R=.14$. The Prosocial scale had positive correlation with 7 of 8 cognitive measures. Emotional problems, hyperactivity and peer problems scales were not correlated with any cognitive task measures. These findings suggest different patterns of association of the different aspects of SDQ with cognition, with only the conduct problems and prosocial scales being broadly associated with cognition. That the conduct scale is negatively associated with EF tasks is an interesting finding: the conduct scale could perhaps be considered to address some kind of ‘hot EF’, i.e. an ability to control one’s behaviour in emotional contexts appropriately.

5.5.3 Fluid intelligence and EF

Adding CFT into the analysis in the baseline data (Model 2) resulted in a three factor model, labelled executive function and general fluid intelligence; Spatial Working Memory; and Attention. These were significantly interrelated with each other. Adding CFT into the analysis in the follow-up data (Model 5) resulted in a unique factor structure. Although this single factor structure was the best model available according to the fit statistics, the single factor was actually a very poor fit for the data with all of our reported fit statistics showing poor fit. The weightings of individual tasks on this unique G factor were all reduced compared with the weightings to the latent factors in the model for EF alone (Model 4), i.e. the tasks were more closely related to the three components of EF in Model 4 than to G in Model 5. Model 5 with a single factor including the CFT scores alongside EF explained less of the variance of the original data than did the 3-factor solution in Model 4 with only the EF data (23.66% in Model 5 vs 47.85% in Model 4).

When adding CFT, at baseline it groups with one of the EF factors, and the three factor structure was preserved. However at follow-up, including CFT alongside EF led to a single factor model being selected. This is somewhat contrary to the general consensus in the literature, in terms of EF development and increasing differentiation across development during adolescence, as we have found greater separation into distinct EF categories at baseline than at follow-up. However, it is important to note that the models for our follow-up data

SDQ also groups to a single factor G at follow-up. In Baseline, adding SDQ in Model 3 didn't change the structure compared with Model 2 which included CFT and EF measures. SDQ forms two separate clusters: problems and prosocial at baseline.

It is not clear from the literature as yet what the variance that is shared across different EF measures means in terms of cognitive processes. Related to this, a potential explanation for the observed structure of a single factor in Model 5 is that CFT taps the cognitive component that reflects the shared variance between the latent EF structures that were present in Model 4. This would mean that fluid intelligence reflects the common EF or shared variance amongst EF components. To investigate this proposition further, analysis could be conducted where CFT score is instead used as a covariate, i.e. variance associated with CFT could be regressed out of all the other task scores to remove the variance associated with fluid intelligence, then a model similar to Model 4 could be run to investigate the EF latent structure in the remaining data. If we found that the interrelations between the latent EF factors were reduced by this process, it would indicate that the common EF or shared variance amongst latent EFs reflects general fluid intelligence.

An alternative proposition is that common EF could represent processing speed (Rose et al., 2011). This could be investigated in our data by using processing speed as a covariate before modelling the latent structure of EFs, and see whether the covariances between latent EF variables remain the same magnitude. This could be investigated in our dataset by creating a processing speed measure (perhaps by combining measures such as reaction time in the one and two item trials on the enumeration task, speed of reading and response times in some of the longer questions in the main battery questionnaires, or reaction times in the 'dots' portion of the trail making task) then regressing this out of the EF task scores, then running the same structural model on the remaining data. If covariances between the latent EF components were reduced, it would indicate that processing speed reflects common EF.

A final suggestion in previous literature is that common EF represents inhibition processes, rather than inhibition being a similar level of factor to other EF processes. This could be tough to investigate within our data, as we have no single task primarily measuring inhibition. However it would be possible to produce some measures of inhibition from our data – perhaps by considering the commission errors in the CPT task as indicating inhibition failures, or the number of non-dot clicks in the spatial working memory task and the trail making task. Taken together, these measures could form an indicator of inhibitory control. These measures could then be included in a SEM model, to better investigate whether inhibition is a separate factor within our data.

5.5.4 Strengths and Limitations

In the SCAMP battery, we had a selection of five different EF tasks. In terms of structural analysis, this is quite a limited number of tasks. Furthermore, we simplified our analysis to use only a single key measure for most of the tasks, and only two key measures where these were intended to address different specific aspects of EF performance. This was done for ease of interpretation. We could have considered including more measures for each of the tasks, in order to better tap different components of EF and cognitive processing. For example we could have included overall response times in tasks in order to consider processing speed as a factor, or we could have better tapped inhibition as a factor by including measures such as number of non-dot clicks made in the SWM and TMT tasks, or number of random type errors (a subcategory of commission errors) in the CPT task. Previous studies have used multiple measures of CPT, dividing commission errors into three sub-types, and also including reaction times, in structural EF analysis (Brocki & Bohlin, 2004). In the task selection that we did have, we probably did not have sufficient measures to exactly match a three-factor models of EF with working memory, switching and inhibition which has been found in previous structural papers in older adolescents and adults (e.g. Lehto et al., 2003; Miyake et al., 2000).

Our tasks were not selected specially to tap specific EFs, which is usually what is done in CFA research that has a particular hypothesis for EF structure in mind (for example Miyake et al., 2000, specifically chose three tasks for each putative EF). We were therefore unable to apply CFA to test for the specific three-factor Miyake et al. (2000) model in our sample due to the fact that the tasks we used were not intended to exactly match this approach. Rather, the approach to task selection was to cover EF in general, with easy to administer tasks that were well-validated in this age group, and were likely to show development during the SCAMP assessment windows. This meant that underlying structures were identified using EFA instead, then CFA applied to the structures identified in EFA.

EFA has advantages in that it is not confined to a single theoretical approach – however it does mean that there is the complex task of naming the latent variables identified. This process was simple for some areas (for example where the CPT tasks grouped together and were labelled as 'attention' in Model 2) but more difficult in others as tasks grouped in perhaps unexpected ways (for example in Model 4 where our switching measure grouped with some of our WM measures, but not the measures of the SWM task). In our study, we carried out EFA and then follow-up CFA analysis in the same data. This is perhaps not the best method to have selected, as there is a possibility of an inflated Type 1 error when carrying out the follow-up CFA in this kind of method. Previous studies have used this kind of EFA-to-CFA method, such as Lehto et al., (2003). However they did justify their use of this method as they had insufficient data to be able to partition the dataset into a training and test set and still retain sufficient power to explore the structures in their data. We would likely have had sufficient power to split our data into training and test sets, and this could be a critique of the methods in this study.

We used the sum totals for each of the five subscales of SDQ as our SDQ measures. Previous factor analysis research has validated the five-factor model within the SDQ, i.e. has supported the idea that each subscale is different to the others, and that in factor analysis, the question items that contribute to each subscale cluster on a factor for that subscale (Thompson et al., 2021). However, there has also been support in the literature for a two-factor model of internalising and externalising scales, or other structures, especially in non-English versions of the questionnaire (Kóbor et al., 2013); Ruchkin et al., 2007). Further research could be done to investigate the factor structure of the SDQ within our baseline and follow-up samples, to see whether the five-factor model or other structures of SDQ can be validated within this sample. When considering the final goodness-of-fit of the models to the original data, we have found that different fit metrics give differing results. Only in Model 2 (EF tasks plus CFT in baseline data) did we find that all the reported fit metrics agreed with each other. This therefore calls into question how good our models are in their fit to the data, and the conclusions made should be considered in this context. In the literature there is no broadly agreed upon route of selection of which goodness-of-fit statistics should be selected to determine the quality of the model in terms of how it explains the original data. Here we have reported the fit metrics following rules of thumb for model adequacy proposed by Hu and Bentler (1999), and used the same metrics that were reported in the factor analysis paper by Alfonso and Lonigan (2021).

However, where fit metrics disagree with each other, there is only limited advice in the literature about how to go about choosing which fit metric to use, and therefore on what to base our assessment of the quality of a factor analysis model. Lai and Green (2016) discuss the fact that different goodness-of-fit statistics approach the data from different perspectives, and that rule of thumb cut-off values of model adequacy that are generally used in the literature are essentially arbitrary rather than statistically justified. When selecting the best structural model in the EFA phase of the analysis, the method agreement procedure was implemented to attempt to deal with the issue of differing fit metrics yielding differing results (Makowski, 2018). However, even using this method, the model selection process for Model 4 (EF within follow-up data) proved a particular issue. No model structure was better than all the others in terms of a greater number of goodness-of-fit statistics. To resolve this, a 70% training set was used to define the EFA model structure for this model, yielding a three-factor model with highest agreement across fit statistics. Goodness-of-fit statistics in factor analysis each have their own strengths and weaknesses, and it is therefore possible or even likely that they will disagree with each other in some cases. In future, it would be useful to the research field for methodologists to explore which fit metric to use in cases where different metrics conflict, and perhaps a more statistically justified approach to model selection in factor analysis could then be adopted than the somewhat ad-hoc method that was employed here to overcome the particular selection problem we experienced.

One major critique of this analysis is the low number of tasks overall. Some studies have carried out factor analysis on a similar number of tasks, for example, Zanini et al. (2021). Here they used only 6 tasks to investigate EF structure, although it should be noted that they selected specific tasks using two tasks to tap each of the three components of EF, so their design here is still superior to ours. In general however it is better to have a greater number of tasks or task measures when carrying out factor analysis. One issue with using a small number of tasks and measures is the possibility of identifying spurious factors, a kind of type I error, where there is no true underlying grouping such as we have observed, rather we have observed the structures we have due to statistical artefacts in the data. Some of our findings are pretty consistent across the models: that we found similar three-factor structures for EF in both baseline and follow-up – these findings are less likely to be spurious results as we have similar results in two different samples in our analysis. However in each of our models, we have had to name some of the factors differently to the others as different tasks appear to group within different factors. We therefore do not have strong evidence for a single specific structure of EF overall, and our research offers somewhat limited evidence to either support or refute a three-factor structure of EF in this age group.

We had more tasks assessing working memory than other EF components, which means our findings are somewhat limited in terms of attempting to replicate or find similar structures to much previous CFA research investigating EF structure. We also took multiple measures from two of the tasks (the SWM task produced an overall spatial working memory span measure and a strategy use measure, and the CPT produced commission and omission errors measures). These imbalances in design could have caused some of the grouping we observed in the structural models. The CPT task measures in particular grouped together in all the models, suggesting that perhaps the fact we had two measures from this task has influenced those results.

We also have issues with low factor loadings and high error terms in the structural models, for example, the TMT does not load well with its respective factors across the six models. All possible models were considered using the EFA method, and we found that the best model for each data set did include TMT with other tasks, but that this task did not load very well with the other tasks. Interpreting this, TMT could be said to share something in common with the other EF tasks, but that it also has a large amount of variance that is not accounted for by any EF factor in common with our other tasks. It also suggests that the factors identified in the models are not a particularly good explanation of the variance in the TMT.

A post-hoc power calculation was carried out to check the sample size was adequate for the analyses conducted. This was done using an online power calculator for CFA analysis, using the RMSEA fit statistic (Preacher & Coffman, 2006). This calculator applies the rules laid out in MacCallum et al. (1996) and accounts for the sample size N, the model structure applied and the targeted level of significance (Kyriazos, 2018) . For the most complex models we conducted, with inclusion of the EF, CFT and SDQ variables, we had a sample size of N=1,478 and N=823 for the baseline and follow-up models respectively. This analysis had df=57, p=.05, and a good fit for RMSEA set at 0.06 (Hu & Bentler, 1999). For a desired detection power of .99, we required 282 participants – we had far more than this in all analyses. We also saw no issues with convergence or with improper solutions. Therefore we have sufficient power for our CFA analyses for all models considered.

Another statistical issue is that of correcting for multiple comparisons. Here, we have carried out six separate analyses, with our desired alpha level set at $p=.05$. We have not corrected the results for multiple comparisons within the results section above. If we were to carry out Bonferroni correction, the most stringent type of multiple comparison correction, we divide .05 by the six analyses, resulting in $p=.0083$ required for significance for each model. This would not result in any changes to the significance of any of the models considered, i.e. all the relationships that were significant at $p<.05$ remain significant at the Bonferroni corrected level. This is true for both the overall model structures and for the individual factor weightings. Some of the interrelations between factors within the models would disappear at this corrected level, suggesting that there is somewhat weaker evidence for the unity of EFs in our sample than there is for the diversity or differentiation of EFs.

The width of our age bands means that it is hard to identify if there are any differences in EF structure between our younger and older participants within the assessment points. It might be possible to explore this by either school year or year of age in our dataset. Previous work considering the development of EF structure across adolescence suggests that a three-factor structure to EF becomes viable around age 13 (Lee et al., 2013; Xu et al., 2013). As we had participants of around age 13 in both groups, this might have been the reason we found a three-factor model in the baseline assessment as well as the follow-up assessment data. If we considered the year groups separately we might find an undifferentiated or less differentiated model is more supported in the data from the younger participants. We would need to consider whether we have sufficient data to explore structure within single year of age or school year group, as the numbers of datapoints in each analysis would be reduced.

The analyses here grouped participants in quite wide age bands of around 3 years. Developmental research has shown that there is are fairly rapid changes in EF abilities across the period of early adolescence (review in Diamond, 2013). As there is uncertainty in the literature around the exact point at which EFs become more differentiated during this period, and given our sample sizes, it would be possible to re-analyse the data in smaller age categories, such as age in years. However as the testing was carried out during school years, it may be that there are fewer participants in some age categories (for example either testing point could have featured participants who were aged 13 – but only the very oldest participants who were tested at the very end of the baseline testing point would be 13 at baseline, and only the very youngest who were tested at the beginning of the testing period would be 13 at follow-up. Therefore there are fewer age 13 participants than say age 14 at follow-up). A possible option could be to collapse the data across testing points, however, care must be taken to account for the fact that some participants would have completed the testing twice and others only once within the overall sample, and practice effects should be accounted for.

5.5.5 Future Research Directions

Previous studies of EF structure in adolescence has largely been done in non-clinical populations. Our study does include some participants with a variety of specific mental and physical conditions, as it covers the population of students in non-specialist education establishments in London. One previous study used a group of participants with ADHD alongside control participants, finding significant differences between these groups in terms of working memory performance. However they did not explore whether the EF structures were the same across the two groups – rather they were combined for their structural EF analysis (Ardila et al., 2005). It would be interesting to explore whether clinical populations within the SCAMP sample display any differences in terms of their ability across any of our identified components, and perhaps (if numbers of clinical populations are sufficient) comparing whether there are any differences in terms of EF structure between clinical and non-clinical populations.

In the SCAMP dataset, some participants have given consent to link to health data. Therefore future work might be able to explore whether EF structures vary with participants' mental or physical health. An alternative approach to investigate whether EF structures vary with mental health would be to divide the sample in terms of their scores on the SDQ. Although the SDQ measure does not measure anxiety or depression symptoms or any mental health diagnosis directly, it is known to correlate with mental health diagnoses. By dividing the population and analysing EF structures within the high and low scorers on the SDQ problems scales, it would be possible to see whether there are any differences in EF structure between participants with poorer and better mental health.

Another issue at hand is that of chronological age vs. pubertal stage. Pubertal timing has been shown to be related to cognitive scores in early adolescence (Shangguan & Shi, 2009), and pubertal hormone changes might be responsible for initiating a period of neural plasticity that enables the significant changes in EF abilities observed during adolescence (Laube et al., 2020). Future work with the SCAMP dataset could use information on pubertal hormone levels in saliva and urine samples which were gathered as part of the BioZone add-on study to consider whether there is a greater difference in EF structure between those who have and have not started puberty, rather than those of different chronological ages as analysed here.

Our task and measure selection somewhat limited our ability to investigate the structure of EF, in particular, we had no task that mainly assessed inhibition, rather this was only an aspect of cognition required to complete some of our tasks, in particular the commission errors in the CPT task. One possible approach to resolve this could be to use additional task measures to attempt to target a potential inhibition component more directly. For example, in the TMT and SWM phones tasks, the number of non-target clicks could be said to reflect inhibition (with greater numbers of non-target clicks reflecting poorer inhibition ability). In the CPT task, random commission errors (i.e. button presses in response to stimuli other than A or X) could be considered to reflect inhibition, with higher numbers of random errors reflecting poorer inhibition. Including these alternative measures in structural models could better draw out any potential inhibition component of EF. Future work could also consider using a more structured task selection to enable CFA of a three-factor model similar to that proposed by Miyake et al. (2000) within the ages we tested, to better explore whether this particular organisation of EF is supported across early adolescence, and whether there are changes to the structure over this developmental period.

5.5.6 Conclusions

Overall we found some evidence for the unity and diversity of EF components at two time points across early adolescence. At both time points, a three-factor model of EF was supported, with slightly different tasks contributing to the components in the two time points. These findings suggest a three-factor model of EF is likely to be present during adolescence. However we did not find clear cut evidence for the specific three-factor model of inhibition, working memory and switching as proposed by Miyake et al. (2000). This is likely related to the fact we did not select the tasks specifically to target these aspects, and we were in particular lacking a particularly good or pure measure for inhibition. At both time points we observed significant associations between the latent EF components, supporting the idea that there is both unity and diversity of EF components during adolescence. We also observed that the zero-order correlations between tasks were slightly greater in the follow-up data than at baseline, with similar interrelations observed between the identified latent factors. This suggests that we find some limited evidence of increased differentiation of EF into these factors with time across adolescence, while noting that we do not find an increased differentiation into a greater number of factors in the older participants.

CFT was found to fit with one of the three EF components at baseline (age 10-13), where at follow-up (ages 13-16), including CFT meant that a single-factor model was a better explanation for the data. This single factor model however was not a particularly good fit for the data – suggesting the results here are to be taken with caution. The inclusion of SDQ measures at baseline resulted in a five-factor model with three EF / CFT components and two separate SDQ components. SDQ was not related to any of the EF or Gf task measures at baseline, rather it formed two separate factors with all the problem scales in one factor and the prosocial scale separately. At follow-up, the SDQ grouped within the single-factor model along with all the EF measures and CFT. Again, this model was not a very good fit for the data, so this finding should be taken with caution.

Chapter 6. Discussion

6.1 Summary of Findings

The key aims of this thesis were to investigate the development and underlying structure of EF in a longitudinal sample of participants across the period of early adolescence (between ages 10 and 16 years), and to explore the associations between SES and EF in early adolescence.

Both SES and EF are independently predictors of later life outcomes, such as academic achievement (SES: Blair & Raver, 2015;Devine et al., 2016; Sirin, 2005; EF: Diamond, 2013; Purpura et al., 2017). The nature of the relationship between SES and EF during adolescence is currently somewhat unclear – however previous research has shown significant associations between these concepts, and given that both are 'umbrella' concepts which have multiple facets, it would be useful to better understand the detailed relationships between specific aspects of SES and EF. This could potentially allow focused interventions to help improve children's future achievement chances.

Chapter 3 looked at the associations between SES and EF in our baseline data sample, where participants were aged between 10 and 13 at the time of testing. Here, we used MANCOVA analyses to show that SES is associated with overall EF, with higher SES levels being associated with better EF outcomes. This supports previous findings in the literature which have demonstrated many times that EF and SES are related to each other (e.g. Noble et al., 2007**;** Sarsour et al., 2011). A multiple regression analysis showed that CFT, a proxy measure for general fluid intelligence (Gf), was also associated with SES.

In an effort to check whether EFs are particularly associated with SES, over and above any associations of SES with Gf, a second MANCOVA analysis was conducted. This used CFT as an additional covariate. Here we showed that association between overall SES and EF remained significant, with effect sizes slightly reduced. This finding is important because it supports the idea that SES and EF have significant associations over and above associations with Gf. Few previous studies of the relationship between SES and EF have been able to demonstrate this – often because studies have lacked power to also control for CFT, or the measure was not collected (Lawson et al., 2018).

A set of follow-up analyses was conducted to investigate which, if any, individual measures of SES and EF were associated with each other. Here we used Multiple regression methods. These analyses showed that specific aspects of SES and EF were differentially associated with each other. In terms of predictors, we found that the strongest predictor out of the six SES measures we used was school type (independent or stats school). It is possible that school type acts as a kind of proxy measure for multiple SES strands, including for example family income and involvement of parents. On the other hand, it could be that independent schools had quieter environments and better equipment at the time of testing, due to factors such as lower class sizes and more teacher involvement in the research, hence the participants were able to perform better in the tasks than those in state schools, on average. Independent schools are also able to select pupils to attend based on prior attainment or entrance exam results – so the populations may not be the same across state and independent schools. The next strongest SES predictor was Carstairs postcode deprivation measure (O. Morgan & Baker, 2006). This is a combined measure that includes unemployment rates, overcrowding and lack of car ownership and occupation type in the local area. This measure therefore covers a range of SES related measures in itself – this might be why this measure came out as a stronger predictor than the parental occupations or educations.

In terms of the tasks which were associated with SES, the most notable example is that the BDS task was associated with all of our SES measures. BDS is a measure of verbal working memory span (Richardson, 2007). Previous research has also showed significant links between SES and verbal working memory, such as a paper that found that children from a disadvantaged background performed worse in a digit span working memory task (Globerson, 1983). Previous research has also linked parental interactions with the child as predictors of their later life working memory ability – perhaps the links between BDS and SES are related to this.

Chapter 4 looked at the developmental trajectories of EF and Gf tasks in the SCAMP battery. Participants were aged M= 12.05 (SD=0.48, range 10-13 years) at baseline assessment, and at follow-up were aged M=14.62 (SD=0.52, range 13-16 years). Using multiple regression and MLM analysis methods in the overall dataset, i.e. ignoring the assessment point the data were taken from, we found significant associations of age on EF task measures of switching (TMT proportion switch cost), verbal working memory (BDS span), spatial working memory (SWM errors and Corsi span), strategy use in a spatial working memory task (SWM strategy), attention/inhibition (CPT commission errors) and also on a fluid intelligence measure (Gf – measured by CFT score). We did not find any significant association with age in the CPT omission errors measure. Older participants scored better on all of the tests where significant associations with age were observed. The effect sizes were quite small for all of the significant results, suggesting that there is only a small amount of age-related development within the age group we tested.

Previous research has indicated that some aspects of EF are still developing in the period of early adolescence, where other measures of EF are likely to have reached a plateau in performance around this age. For example, inhibition is likely at a plateau in performance by around 11-12 years old (Brocki & Bohlin, 2004; Huizinga et al., 2006b). Previous research has suggested that working memory accuracy and capacity may have reached a peak in performance by early adolescence (De Luca et al., 2003) - however we did see age-related improvements in performance in SWM errors and Corsi capacity. Higher-level aspects of working memory, such as planning or the SWM strategy use measure we used, have been shown to continue to develop until early adulthood (De Luca et al., 2003), which we did see in our data. Switching was also expected to improve over this period (K. Lee et al., 2013), and we also saw this effect. We also expected to see age-related developments in Gf (Cahan & Cohen, 1989), which we did find in our study.

We checked whether these effects were largely down to practice effects in a series of sensitivity analyses. We found that significant development of the same task scores with age was present in the group of participants who only participated in one of the assessment sessions, though effect sizes were reduced – suggesting that some but not all of the developmental effects we observed were down to practice effects.

We also investigated whether score at follow-up was best predicted by age or by score at baseline. We found a strong, significant association between score at baseline and follow-up, where age was not related to follow-up score when baseline score was accounted for. This was true for all of our task measures, except for CPT omission errors. We also considered whether the rate of change in score between baseline and follow-up was predictable by age, or by task score at baseline. We found that previous task score was a strong significant predictor of change in task score from baseline to follow-up for all of our tasks, with a negative relationship – meaning those scoring worse at baseline made more progress than those at follow-up. These findings taken together suggest that the gaps between the worse and best performers on average were decreasing with time across the age groups we analysed.

Chapter 5 investigated the latent variable structure of EF, Gf, and mental health in our two assessment time points, at baseline where participants were aged between 10 and 13, and at follow-up where participants were aged between 13 and 16. We found three-factor models of EF in both time points, with the factors being significantly interrelated with each other, and slightly different tasks contributing to the specific components at the two time points. This lends support to the idea that EFs display both unity and diversity during adolescence, similar to the ideas proposed in adults by Miyake et al. (2000). We failed to find support for the differentiation of EF into the widely discussed components of inhibition, working memory and switching as have been observed in adults, however, as out study was using EFA methods rather than CFA designed to look specifically for this model, this is not conclusive evidence against this structure. It should be noted that our number of tasks overall was quite limited for this kind of factor analysis modelling. Also, our task selection was biased towards working memory, with three of our five EF tasks targeting this component. Therefore our support for this kind of three-factor model is somewhat tentative.

Research looking at EF structural development during adolescence has suggested a broad change from less differentiated models observed in early childhood to more differentiated components observed in adulthood. Research has not been conclusive regarding exactly when during adolescence EF structures become more differentiated, but some studies point to a three-factor model of EF becoming stable around age 13-15 (Lee et al., 2013; Xu et al., 2013). We did not find evidence of EF becoming fractionated into a greater number of factors between our two time points, as both baseline and follow-up data supported a three-factor model of EF fractionation. We did find some evidence that the factors themselves are more differentiated from each other in our older participant group – we note that the zero-order correlations between all the EF tasks are slightly greater in the follow-up data (**Tables 5.2 and 5.3**), where the interrelations between factors are similar in the structural models at baseline and follow-up. However, since our age bands both included participants around age 13, and given that previous structural developmental research suggests this may be the point around which three-factor models begin to become more viable, it may be that our age bands were too wide to be able to spot any possible changes in EF structure that might be occurring around this age.

In the structural models including CFT, we found that Gf fit with one of the EF components in baseline, but that in follow-up, including CFT in the models resulted in a single overall factor including all the tasks being the best model. This is somewhat surprising given that previous research has indicated that cognition broadly becomes more differentiated across adolescence, we did not support this in our findings. However, it is important to note that the single factor structure was not a very good fit for the data based on the fit metrics. We found that mental health measures did not group with EF or CFT measures in the baseline data, rather that they formed two separate factors relating to the Strengths (prosocial scale) and Difficulties (all four problem scales) components of the SDQ items. At follow-up, we found that mental health also measures grouped into the same single factor model that we had identified with CFT and EF measures. Again, this single factor model was not a particularly good fit for the data according to the fit metrics.

6.2 Limitations

The nature of the classroom environment in which participants were assessed results in a reduced amount of control of outside variables, as compared with individual lab-based assessments. For example, there was significant variation in the level of noise between the different testing rooms, which varied at both school and individual classroom level. Research has shown that levels of noise, especially noise of other people talking or 'background babble', can be particularly distracting to some people. For example, those with lower working memory capacity perform worse on mathematics problems in situations with high levels of background babble (Massonnié et al., 2022). It would be possible to investigate whether the noise level during the assessment (as noted by the experimenters at the time) associates with the outcome of task performance across the battery, to rule out or identify the potential magnitude of this potentially confounding variable on the data.

On the other hand, the use of classroom assessments meant that large amounts of data could be collected. This has enabled us to run complex models to investigate executive function structure, development and the associations with SES. It has also enabled us to detect small effect sizes which may be present in the data – more on this shortly. A further advantage of the classroom assessment is that it allowed some level of control over the environment, as opposed to say online testing where people's environments and levels of focus may differ significantly.

Computerized testing was used in the main battery in order to efficiently gather data on the large sample of participants. This reduced the requirement for trained test administrators compared with more traditional presentations of tasks. It also meant that participants were able to get through significant volumes of testing in a relatively short period of time, as there was little or no 'down time' between the individual task assessments or questionnaire items. A downside of using computerized testing is that results of this type of very controlled experimental tasks may have poor construct validity and potentially bear limited resemblance to real-life EF abilities, as discussed by Chan et al. (2008b).

One issue of particular relevance to the developmental trajectories chapter (**Chapter 4**) is that the same tasks were completed at both time points, potentially resulting in practice effects. However it was considered that the advantage of having longitudinal data across the two time points was greater than the downside here. In the analysis in this thesis we did consider the potential impacts of practice effects by conducting a set of sensitivity analyses – we found that the associations with age that we had observed in the overall sample (including those who had repeated the tasks at baseline and follow-up, and those who had only completed one of the two assessment points) were still present in the sub-sample who had data at only one time point, but that effect sizes were reduced. This should be borne in mind when interpreting the results in this section.

In many of the schools, participants did not have the intended full hour to complete the testing, due to time constraints in the school's timetabling. This meant a significant proportion of our participants did not manage to finish all of the assessment battery. Participants were able to work through the assessment at their own pace. A downside of this is that potentially, the participants with quicker response times or reading speed are likely to have completed more of the testing battery. This is of particular concern as it is likely to have affected the sample who completed the final task in the battery, namely the CPT. This task has a much lower sample size than the other cognitive tasks. It is possible that the participants who did manage to complete this task are by their nature the participants who would score more highly on the cognitive tests. This relates to the idea of the 'positive manifold' – in this case, the speed of completion of the tasks and questionnaires would likely positively correlate with performance in any given task, as any psychological task is in general positively associated with any other. It would be possible to analyze statistically, perhaps using logistic regressions, whether those who scored more highly on earlier tasks, or have a quicker reaction time, are more likely to complete the CPT. This could give an estimate of how 'skewed' the data for this task are.

A further complication for the interpretation of our results is that we have not been able to include or account for any measures of pubertal age, as opposed to chronological age. The beginning of puberty is defined by the onset of changes in sex hormones, rather than by a specific chronological age. The hormone changes are thought to be potentially causative of the increase in neural plasticity observed during puberty, and therefore of the rapid changes in cognitive performance also observed at this time. Research in boys aged 8 to 12 years old has shown that levels of salivary pubertal hormone (testosterone) are correlated with performance in Cattell's CFT, after accounting for age and BMI (Shangguan & Shi, 2009). We are likely to have data from some participants who have begun puberty before the first assessment point, others who will have begun puberty at some point between the assessment points, and potentially some who were tested early in the follow-up testing period who may not have entered puberty even by the time of their follow-up assessment. The differences in pubertal stages between participants could explain some of the differences we have observed due to age (in **Chapters 4 & 5**), as research indicates higher levels of puberty hormones are associated with improvements in cognitive tasks. Our overall conclusions are limited by the specific task measure selection we have used. For example, it is difficult to draw wide conclusions about the aspects of EF as described by (Miyake et al., 2000) , as our task selection was not intended to match the components of working memory, switching and inhibition specifically. In particular, we have not included a specific measure of inhibition in our analyses, limiting our conclusions to other EF aspects of switching and working memory, and more general EF where we have shown effects across most or all of our EF tasks.

6.3 Future research directions

Chapter 3 explored the nature of associations between SES and EF. Previous research has suggested that EF acts as a mediator of SES related disparities in other life outcomes, such as academic achievement and numeracy (e.g. Devine et al., 2016; Ellefson et al., 2020). It would be interesting to consider whether this is the case in the SCAMP cohort – future work is planned to collect academic achievement data as the participants have now completed their schooling, and this could be an interesting line of future research to pursue. Furthermore, we only analysed data from the first testing point in this thesis in this chapter, as follow-up data collection had not yet been completed when the analysis was carried out. An area that could be considered in future would be to see whether the association between SES and EF remains similar in the follow-up data as we have seen at baseline. Previous research has come to differing conclusions about whether the associations between SES and EF increases or remains similar over time.

It would also be possible to use the saliva samples collected from the BioZone add-on study to estimate pubertal hormone levels. We could then firstly replicate the work of Shangguan & Shi (2009), who found that pubertal hormone levels in boys aged 10-12 correlated with CFT scores after accounting for age and BMI. We could also extend this research to investigate associations between other cognitive tasks scores and pubertal hormones, include estimates for girls, and extend the age groups up through to around 16 using data from our follow-up sample. We would be able to see whether the improvements in task performance with age that we observed in **Chapter 4** are better explained by pubertal hormone concentrations or chronological age.

The data collected during the SCAMP project included various measures of video game use. Previous research has indicated that people who play video games are more likely to have better subitisation range, and shorter reaction times (Boot et al., 2011; Green & Bavelier, 2006). Subitisation range was measured by our enumeration task, and various measures of reaction times could be calculated from our task data. This could be investigated in relation to participants' responses to the questions regarding video gaming to attempt to replicate these previous findings in a large sample of adolescents in the UK.

A question of import is whether, and if so how, do mobile phone and other technology use influence trajectories of cognitive development. Given that EFs have a protracted period of development, this makes them a likely area of cognition that might be affected by use of mobiles and technology during development (Best et al., 2009). As part of the SCAMP assessment, many measures of mobile phone use amongst our cohort have been collected. This includes questionnaire responses in the main battery and add-on questionnaires, and also a subsample of participants downloaded an app that tracked type and time of mobile usage at points during the data collection period, and a further subsample carried a mobile exposimeter to assess their exposure to RF-EMF in their daily lives. Previous research has indicated that mobile phones do not cause cognitive impairments, but may be associated with poorer mental health (Roser et al., 2016). However the research field is lacking any overall conclusions about the effects of mobile use in a longitudinal study. SCAMP offers the opportunity to investigate this in a large, representative sample of adolescents in London, and might be able to better elucidate causality than previous correlational work in this area due to its longitudinal study design. The work completed as part of this thesis, considering EF structure and development over the SCAMP study period, could be extended to investigate mobile use. For example, the EF latent variables identified in **Chapter 5** could be used in a cross-lagged SEM model along with latent variables describing mobile phone or other technology use, to see whether there is a predictive relationship between these variables, and which direction this relationship runs (i.e. does mobile phone use at baseline predict later EF at follow-up, or the other way around).

6.4 Conclusions

This thesis has investigated the associations between EF and SES in early adolescence, and the structure and development of EF from early to mid adolescence. Small but significant associations between SES and EF were demonstrated in **Chapter 3.** The associations between SES and EF measures were present over and above associations with general fluid intelligence (Gf). This has not been widely demonstrated previously. We used a large sample of participants whose SES was representative of the wider London population, which extends the field by illustrating effects of SES across the whole range of SES, rather than only comparing limited groups of participants. By using multiple measures of both SES and EF, we found significant associations between specific measures of SES and EF. Overall we found that school type was the SES measure with the largest association with EF measures (around 5% of variation in EF scores) with SES measures generally accounting for around 8% of the variation in EF scores, and that the BDS task was most strongly related to multiple measures of SES.

In **Chapter 4** we investigated the development of task scores across the age range of 10 - 16 years. We found small but significant effect of age on all but one of our EF task measures, and also on our measure of Gf. This effect was present over and above the practice effect, i.e. the normally expected improvement associated with some participants having completed the tasks more than once.

In **Chapter 5** we found a three-factor model of EF at both baseline (age 10-13) and follow-up (age 13 -16). Factors were all significantly interrelated, supporting a unity-yet-diversity model of EF in both age groups. At baseline, factors were labelled as switching and working memory; planning; and attention. At follow-up, factors were labelled as switching and working memory; spatial working memory; and attention. We found that Gf fit with the switching and working memory EF component at baseline. At follow-up, including Gf in the model resulted in a unique factor structure which was associated with the variance in all EF and Gf tasks. Mental health measures did not group with EF or Gf task measures at baseline, rather the inclusion of mental health measures resulted in a five-factor model consisting of three EF and Gf related factors (switching and working memory; planning; attention), and two mental health factors (SDQ strengths; SDQ difficulties). At follow-up, mental health measures grouped with the single factor model which explained variance in all of the EF, Gf and mental health measures.

References

Adams, J. W., & Snowling, M. J. (2001). Executive function and reading impairments in children reported by their teachers as 'hyperactive'. *British Journal of Developmental Psychology*, *19*(2), 293–306. https://doi.org/10.1348/026151001166083

Ahlbom, A., Bergqvist, U., Bernhardt, J. H., Cesarini, J. P., Court, L. A., Grandolfo, M., Hietanen, M., McKinlay, A. F., Repacholi, M. H., Sliney, D. H., Stolwijk, J. a. J., Swicord, M. L., Szabo, L. D., Taki, M., Tenforde, T. S., Jammet, H. P., & Matthes, R. (1998). Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health Physics*, *74*(4), 494–521.

Alfonso, S. V., & Lonigan, C. J. (2021). Trait anxiety and adolescent's academic achievement: The role of executive function. *Learning and Individual Differences*, *85*, 101941. https://doi.org/10.1016/j.lindif.2020.101941

Allen, D. N., Thaler, N. S., Ringdahl, E. N., Barney, S. J., & Mayfield, J. (2012). Comprehensive Trail Making Test performance in children and adolescents with traumatic brain injury. *Psychological Assessment*, *24*(3), 556–564. https://doi.org/10.1037/a0026263

Alloway, T. P., & Alloway, R. G. (2013). Working memory across the lifespan: A cross-sectional approach. *Journal of Cognitive Psychology*, *25*(1), 84–93. https://doi.org/10.1080/20445911.2012.748027

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, *28*(7), 725–734. https://doi.org/10.1080/01443410802243828

Alves, A. F., Gomes, C. M. A., Martins, A., & Almeida, L. da S. (2017). Cognitive performance and academic achievement: How do family and school converge? *European Journal of Education and Psychology*, *10*(2), 49–56. https://doi.org/10.1016/j.ejeps.2017.07.001

Anderson, P. (2002). Assessment and Development of Executive Function (EF) During Childhood. *Child Neuropsychology*, *8*(2), 71–82. https://doi.org/10.1076/chin.8.2.71.8724

Anderson, P. J., & Reidy, N. (2012). Assessing executive function in preschoolers. *Neuropsychology Review*, *22*(4), 345–360. https://doi.org/10.1007/s11065-012-9220-3

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of Executive Functions Through Late Childhood and Adolescence in an Australian Sample. *Developmental Neuropsychology*, *20*(1), 385–406. https://doi.org/10.1207/S15326942DN2001_5

Andrews, K., Atkinson, L., Harris, M., & Gonzalez, A. (2020). Examining the effects of household chaos on child executive functions: A meta-analysis. *Psychological Bulletin*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/bul0000311

Ardila, A., Rosselli, M., Matute, E., & Guajardo, S. (2005). The Influence of the Parents' Educational Level on the Development of Executive Functions. *Developmental Neuropsychology*, *28*(1), 539–560. https://doi.org/10.1207/s15326942dn2801_5

Army Individual Test Battery. (1944). *Manual of directions and scoring*. Washington, DC: War Department, Adjutant General's Office.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*(1), 40–49. https://doi.org/10.1002/mpr.329

Baddeley, A. (1996a). Exploring the Central Executive. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 5–28. https://doi.org/10.1080/713755608

Baddeley, A. (1996b). The fractionation of working memory. *Proceedings of the National Academy of Sciences*, *93*(24), 13468–13472.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. http://www.sciencedirect.com/science/article/pii/S0079742108604521

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*(5), 193–200. https://doi.org/10.1016/j.tics.2008.02.004

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews. Neuroscience*, *10*(9), 659–669. https://doi.org/10.1038/nrn2667

Barkley, R. A., Edwards, G., Laneri, M., Fletcher, K., & Metevia, L. (2001). Executive Functioning, Temporal Discounting, and Sense of Time in Adolescents with Attention Deficit Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder (ODD). *Journal of Abnormal Child Psychology*, *29*(6), 541–556. https://doi.org/10.1023/A:1012233310098

Barth, A., Winker, R., Ponocny-Seliger, E., Mayrhofer, W., Ponocny, I., Sauter, C., & Vana, N. (2008). A meta-analysis for neurobehavioural effects due to electromagnetic field exposure emitted by GSM mobile phones. *Occupational and Environmental Medicine*, *65*(5), 342–346. https://doi.org/10.1136/oem.2006.031450

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, *13*(3), 283–292. https://doi.org/10.1076/clin.13.3.283.1743

Beaujean, A. A. (2014). *Latent Variable Modeling Using R: A Step-by-Step Guide*. Routledge. https://doi.org/10.4324/9781315869780

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive Functions after Age 5: Changes and Correlates. *Developmental Review: DR*, *29*(3), 180–200. https://doi.org/10.1016/j.dr.2009.05.002

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between Executive Function and Academic Achievement from Ages 5 to 17 in a Large, Representative National Sample. *Learning and Individual Differences*, *21*(4), 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Bignardi, G., Dalmaijer, E., Anwyl-Irvine, A., & Astle, D. (2020). *Collecting Big Data with Small Screens: Group-Tests of Children's Cognition with Touchscreen Tablets are Reliable and Valid*. OSF Preprints. https://doi.org/10.31219/osf.io/aw6c5

Blair, C., & Raver, C. C. (2015). School Readiness and Self-Regulation: A Developmental Psychobiological Approach. *Annual Review of Psychology*, *66*(1), 711–731. https://doi.org/10.1146/annurev-psych-010814-015221

Blettner, M., & Berg, G. (2000). Are mobile phones harmful? *Acta Oncologica (Stockholm, Sweden)*, *39*(8), 927–930.

Boot, W. R., Blakely, D. P., & Simons, D. J. (2011). Do Action Video Games Improve Perception and Cognition? *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00226

Bos, I. F. den, & Weijer-Bergsma, E. van de. (2020). Classroom versus individual working memory assessment: Predicting academic achievement and the role of attention and response inhibition. *Memory*, *28*(1), 70–82. https://doi.org/10.1080/09658211.2019.1682170

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic Status and Child Development. *Annual Review of Psychology*, *53*(1), 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Brocki, K. C., & Bohlin, G. (2004). Executive Functions in Children Aged 6 to 13: A Dimensional and Developmental Study. *Developmental Neuropsychology*, *26*(2), 571–593. https://doi.org/10.1207/s15326942dn2602_3

Brocki, K. C., Tillman, C. M., & Bohlin, G. (2010). CPT performance, motor activity, and continuous relations to ADHD symptom domains: A developmental study. *European Journal of Developmental Psychology*, *7*(2), 178–197. https://doi.org/10.1080/17405620801937764

Brown, R. E. (2016). Hebb and Cattell: The Genesis of the Theory of Fluid and Crystallized Intelligence. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00606

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Immature frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron*, *33*(2), 301–311.

Burgess, P. W., Dumontheil, I., & Gilbert, S. J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Sciences*, *11*(7), 290–298. https://doi.org/10.1016/j.tics.2007.05.004

Buuren, S. van, & Groothuis-Oudshoorn, C. G. M. (2010). MICE: Multivariate Imputation by Chained Equations in R. *Journal of statistical software*, *45*(3). https://research.utwente.nl/en/publications/mice-multivariate-imputation-by-chained-equations-in-r

Cacciamani, F., Salvadori, N., Eusebi, P., Lisetti, V., Luchetti, E., Calabresi, P., & Parnetti, L. (2018). Evidence of practice effect in CANTAB spatial working memory test in a cohort of patients with mild cognitive impairment. *Applied Neuropsychology: Adult*, *25*(3), 237–248. https://doi.org/10.1080/23279095.2017.1286346

Cahan, S., & Cohen, N. (1989). Age versus Schooling Effects on Intelligence Development. *Child Development*, *60*(5), 1239–1249. https://doi.org/10.2307/1130797

Camilleri, J. A., Müller, V. I., Fox, P., Laird, A. R., Hoffstaedter, F., Kalenscher, T., & Eickhoff, S. B. (2018). Definition and characterization of an extended Multiple-Demand Network. *NeuroImage*, *165*, 138–147. https://doi.org/10.1016/j.neuroimage.2017.10.020

Cattell, R. B., & Cattell, A. (1960a). *Measuring intelligence with the culture fair tests*. Institute for Personality and Ability Testing.

Cattell, R. B., & Cattell, A. K. S. (1960b). *Cattell culture fair intelligence test: Scale 2, parts I and II, a measure of ‘G’*. Bobbs-Merrill Co.; /z-wcorg/.

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008a). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216. https://doi.org/10.1016/j.acn.2007.08.010

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008b). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216. https://doi.org/10.1016/j.acn.2007.08.010

Chetty, R., Hendren, N., & Katz, L. F. (2016). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, *106*(4), 855–902. https://doi.org/10.1257/aer.20150572

Christensen, D. L., Schieve, L. A., Devine, O., & Drews-Botsch, C. (2014). Socioeconomic status, child enrichment factors, and cognitive performance among preschool-age children: Results from the Follow-Up of Growth and Development Experiences study. *Research in Developmental Disabilities*, *35*(7), 1789–1801. https://doi.org/10.1016/j.ridd.2014.02.003

Ciesielski, K. T., Lesnik, P. G., Savoy, R. L., Grant, E. P., & Ahlfors, S. P. (2006). Developmental neural networks in children performing a Categorical N-Back Task. *NeuroImage*, *33*(3), 980–990. https://doi.org/10.1016/j.neuroimage.2006.07.028

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. http://digitool.library.mcgill.ca/R/?func=dbin-jump-full&object_id=93903&local_base=GEN01-MCG02

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience*, *13*(9), 636–650. https://doi.org/10.1038/nrn3313

Crone, E. A., & Steinbeis, N. (2017). Neural Perspectives on Cognitive Control Development during Childhood and Adolescence. *Trends in Cognitive Sciences*, *21*(3), 205–215. https://doi.org/10.1016/j.tics.2017.01.003

Crone, E. A., Wendelken, C., Donohue, S., van Leijenhorst, L., & Bunge, S. A. (2006). Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(24), 9315–9320. https://doi.org/10.1073/pnas.0510088103

Crosnoe, R., Johnson, M. K., & Elder, G. H. (2004a). Intergenerational Bonding in School: The Behavioral and Contextual Correlates of Student-Teacher Relationships. *Sociology of Education*, *77*(1), 60–81. https://doi.org/10.1177/003804070407700103

Crosnoe, R., Johnson, M. K., & Elder, G. H. (2004b). School Size and the Interpersonal Side of Education: An Examination of Race/Ethnicity and Organizational Context*. *Social Science Quarterly*, *85*(5), 1259–1274. https://doi.org/10.1111/j.0038-4941.2004.00275.x

Dahmen, J., Cook, D., Fellows, R., & Schmitter-Edgecombe, M. (2017). An Analysis of a Digital Variant of the Trail Making Test Using Machine Learning Techniques. *Technology and Health Care : Official Journal of the European Society for Engineering and Medicine*, *25*(2), 251–264. https://doi.org/10.3233/THC-161274

Daneri, M. P., Blair, C., & Kuhn, L. J. (2019). Maternal Language and Child Vocabulary Mediate Relations Between Socioeconomic Status and Executive Function During Early Childhood. *Child Development*, *90*(6), 2001–2018. https://doi.org/10.1111/cdev.13065

D'angiulli, A., Lipina, S. J., & Olesinska, A. (2012). Explicit and implicit issues in the developmental cognitive neuroscience of social inequality. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00254

D'angiulli, A., Van Roon, P. M., Weinberg, J., Oberlander, T., Grunau, R., Hertzman, C., & Maggi, S. (2012). Frontal EEG/ERP correlates of attentional processes, cortisol and motivational states in adolescents from lower and higher socioeconomic status. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00306

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

De Luca, C. R., Wood, S. J., Anderson, V., Buchanan, J.-A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative Data From the Cantab. I: Development of Executive Function Over the Lifespan. *Journal of Clinical and Experimental Neuropsychology*, *25*(2), 242–254. https://doi.org/10.1076/jcen.25.2.242.13639

Decker, A. L., Duncan, K., Finn, A. S., & Mabbott, D. J. (2020). Children's family income is associated with cognitive function and volume of anterior not posterior hippocampus. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-17854-6

Deloitte. (2017). *Discover the latest trends via our Mobile Consumer Survey*. Deloitte United Kingdom. https://www2.deloitte.com/uk/en/pages/technology-media-and-telecommunications/articles/mobile-consumer-survey.html

Devine, R. T., Bignardi, G., & Hughes, C. (2016). Executive Function Mediates the Relations between Parental Behaviors and Children's Early Academic Ability. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01902

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, *38*(3), 352–362.

Donati, G., Meaburn, E., & Dumontheil, I. (2021). Internalising and externalising in early adolescence predict later executive function, not the other way around: A cross-lagged panel analysis. *Cognition and Emotion*, *35*(5), 986–998. https://doi.org/10.1080/02699931.2021.1918644

Dumontheil, I. (2016). Adolescent brain development. *Current Opinion in Behavioral Sciences*, *10*, 39–44. https://doi.org/10.1016/j.cobeha.2016.04.012

Dumontheil, I., & Klingberg, T. (2012). Brain Activity during a Visuospatial Working Memory Task Predicts Arithmetical Performance 2 Years Later. *Cerebral Cortex*, *22*(5), 1078–1085. https://doi.org/10.1093/cercor/bhr175

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, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Duncan, J. (2013). The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron*, *80*(1), 35–50. https://doi.org/10.1016/j.neuron.2013.09.015

Duncan, J., Assem, M., & Shashidhara, S. (2020). Integrated Intelligence from Distributed Brain Activity. *Trends in Cognitive Sciences*, *24*(10), 838–852. https://doi.org/10.1016/j.tics.2020.06.012

Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), 5295–5299. https://doi.org/10.1073/pnas.1621147114

Duncan Roger Johnson Michaela Swale, J. (1997). Frontal Lobe Deficits after Head Injury: Unity and Diversity of Function. *Cognitive Neuropsychology*, *14*(5), 713–741. https://doi.org/10.1080/026432997381420

Eamon, M. K. (2005). Social-Demographic, School, Neighborhood, and Parenting Influences on the Academic Achievement of Latino Young Adolescents. *Journal of Youth and Adolescence*, *34*(2), 163–174. https://doi.org/10.1007/s10964-005-3214-x

Ellefson, M. R., Zachariou, A., Ng, F. F.-Y., Wang, Q., & Hughes, C. (2020). Do executive functions mediate the link between socioeconomic status and numeracy skills? A cross-site comparison of Hong Kong and the United Kingdom. *Journal of Experimental Child Psychology*, *194*, 104734. https://doi.org/10.1016/j.jecp.2019.104734

Engelhardt, L. E., Harden, K. P., Tucker-Drob, E. M., & Church, J. A. (2019). The Neural Architecture of Executive Functions Is Established by Middle Childhood. *NeuroImage*, *185*, 479–489. https://doi.org/10.1016/j.neuroimage.2018.10.024

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

Evans, G. W. (2004). The Environment of Childhood Poverty. *American Psychologist*, *59*(2), 77–92. https://doi.org/10.1037/0003-066X.59.2.77

Fecht, D., Fischer, P., Fortunato, L., Hoek, G., de Hoogh, K., Marra, M., Kruize, H., Vienneau, D., Beelen, R., & Hansell, A. (2015). Associations between air pollution and socioeconomic characteristics, ethnicity and age profile of neighbourhoods in England and the Netherlands. *Environmental Pollution*, *198*, 201–210. https://doi.org/10.1016/j.envpol.2014.12.014

Fellows, R. P., Dahmen, J., Cook, D., & Schmitter-Edgecombe, M. (2017). Multicomponent analysis of a digital Trail Making Test. *The Clinical Neuropsychologist*, *31*(1), 154–167. https://doi.org/10.1080/13854046.2016.1238510

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics (Fourth Edition)*. Sage London. https://us.sagepub.com/en-us/nam/discovering-statistics-using-ibm-spss-statistics/book238032%20

Filippi, R., Ceccolini, A., Booth, E., Shen, C., Thomas, M. S. C., Toledano, M. B., & Dumontheil, I. (2022). Modulatory effects of SES and multilinguistic experience on cognitive development: A longitudinal data analysis of multilingual and monolingual adolescents from the SCAMP cohort. *International Journal of Bilingual Education and Bilingualism*, *25*(9), 3489–3506. https://doi.org/10.1080/13670050.2022.2064191

Fitzpatrick, C., McKinnon, R. D., Blair, C. B., & Willoughby, M. T. (2014). Do preschool executive function skills explain the school readiness gap between advantaged and disadvantaged children? *Learning and Instruction*, *30*, 25–31. https://doi.org/10.1016/j.learninstruc.2013.11.003

Foulkes, L., & Blakemore, S.-J. (2018). Studying individual differences in human adolescent brain development. *Nature Neuroscience*, *21*(3), 315–323. https://doi.org/10.1038/s41593-018-0078-4

Fournier-Vicente, S., Larigauderie, P., & Gaonac'h, D. (2008). More dissociations and interactions within central executive functioning: A comprehensive latent-variable analysis. *Acta Psychologica*, *129*(1), 32–48. https://doi.org/10.1016/j.actpsy.2008.04.004

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., & Hewitt, J. K. (2007). Greater Attention Problems During Childhood Predict Poorer Executive Functioning in Late Adolescence. *Psychological Science*, *18*(10), 893–900. https://doi.org/10.1111/j.1467-9280.2007.01997.x

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), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179.

Galobardes, B., Shaw, M., Lawlor, D. A., & Lynch, J. W. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health*, *60*(1), 7–12. https://doi.org/10.1136/jech.2004.023531

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

Gibbons, S., & Silva, O. (2011). Faith Primary Schools: Better Schools or Better Pupils? *Journal of Labor Economics*, *29*(3), 589–635. https://doi.org/10.1086/659344

Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2005). Involvement of rostral prefrontal cortex in selection between stimulus-oriented and stimulus-independent thought. *The European Journal of Neuroscience*, *21*(5), 1423–1431. https://doi.org/10.1111/j.1460-9568.2005.03981.x

Ginkel, J. R. van. (2019). Significance Tests and Estimates for R2 for Multiple Regression in Multiply Imputed Datasets: A Cautionary Note on Earlier Findings, and Alternative Solutions. *Multivariate Behavioral Research*, *54*(4), 514–529. https://doi.org/10.1080/00273171.2018.1540967

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Behavior Rating Inventory of Executive Function. *Child Neuropsychology*, *6*(3), 235–238. https://doi.org/10.1076/chin.6.3.235.3152

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory Factor Analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a Clinical Sample. *Child Neuropsychology*, *8*(4), 249–257. https://doi.org/10.1076/chin.8.4.249.13513

Globerson, T. (1983). Mental capacity and cognitive functioning: Developmental and social class differences. *Developmental Psychology*, *19*(2), 225–230. https://doi.org/10.1037/0012-1649.19.2.225

Goedhart, G., Kromhout, H., Wiart, J., & Vermeulen, R. (2015). Validating self-reported mobile phone use in adults using a newly developed smartphone application. *Occupational and Environmental Medicine*, *72*(11), 812–818. https://doi.org/10.1136/oemed-2015-102808

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

Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire as a Dimensional Measure of Child Mental Health. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(4), 400–403. https://doi.org/10.1097/CHI.0b013e3181985068

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, *177*(6), 534–539. https://doi.org/10.1192/bjp.177.6.534

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*(1), 217–245. https://doi.org/10.1016/j.cognition.2005.10.004

Gregoire, J., & Van Der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition*, *4*(2), 140–149. https://doi.org/10.1080/13825589708256642

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic growth curve trajectories from 1st grade to 12th grade: Effects of multiple social risk factors and preschool child factors. *Developmental Psychology*, *39*(4), 777–790. https://doi.org/10.1037/0012-1649.39.4.777

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13*(2), 65–73. https://doi.org/10.1016/j.tics.2008.11.003

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: Mechanistic insights from human and animal research. *Nature Reviews. Neuroscience*, *11*(9), 651–659. https://doi.org/10.1038/nrn2897

Halperin, J. M., Wolf, L., Greenblatt, E. R., & Young, G. (1991). Subtype analysis of commission errors on the continuous performance test in children. *Developmental Neuropsychology*, *7*(2), 207–217. https://doi.org/10.1080/87565649109540488

Han, G., Helm, J., Iucha, C., Zahn-Waxler, C., Hastings, P. D., & Klimes-Dougan, B. (2016). Are Executive Functioning Deficits Concurrently and Predictively Associated with Depressive and Anxiety Symptoms in Adolescents? *Journal of Clinical Child & Adolescent Psychology*, *45*(1), 44–58. https://doi.org/10.1080/15374416.2015.1041592

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), Article 7302. https://doi.org/10.1038/466029a

Herrmann, D., & Guadagno, M. A. (1997). Memory Performance and Socio-Economic Status. *Applied Cognitive Psychology*, *11*(2), 113–120. https://doi.org/10.1002/(SICI)1099-0720(199704)11:2<113::AID-ACP424>3.0.CO;2-F

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hommel, B., Ridderinkhof, R. K., & Theeuwes, J. (2002). Cognitive control of attention and action: Issues and trends. *Psychological Research*, *66*(4), 215–219. https://doi.org/10.1007/s00426-002-0096-3

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. https://doi.org/10.1037/h0023816

Hostinar, C. E., Stellern, S. A., Schaefer, C., Carlson, S. M., & Gunnar, M. R. (2012). Associations between early life adversity and executive function in children adopted internationally from orphanages. *Proceedings of the National Academy of Sciences*, *109*(Supplement 2), 17208–17212. https://doi.org/10.1073/pnas.1121246109

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hugdahl, K., Raichle, M. E., Mitra, A., & Specht, K. (2015). On the existence of a generalized non-specific task-dependent network. *Frontiers in Human Neuroscience*, *9*. https://www.frontiersin.org/articles/10.3389/fnhum.2015.00430

Huizinga, M., Baeyens, D., & Burack, J. A. (2018). Editorial: Executive Function and Education. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01357

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006a). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006b). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Hwang, K., & Luna, B. (2013). The Development of Brain Connectivity Supporting Prefrontal Cortical Functions. In D. T. Stuss & R. T. Knight (Eds.), *Principles of Frontal Lobe Function*.

Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation--Recommendations to the National Center for Education Statistics. (2012). National Center for Education Statistics. https://eric.ed.gov/?id=ED542101

Iselin, A.-M. R., & DeCoster, J. (2009). Reactive and proactive control in incarcerated and community adolescents and young adults. *Cognitive Development*, *24*(2), 192–206. https://doi.org/10.1016/j.cogdev.2008.07.001

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology*, *26*(1), 403–422. https://doi.org/10.1207/s15326942dn2601_3

Jednoróg, K., Altarelli, I., Monzalvo, K., Fluss, J., Dubois, J., Billard, C., Dehaene-Lambertz, G., & Ramus, F. (2012). The influence of socioeconomic status on children's brain structure. *PloS One*, *7*(8), e42486. https://doi.org/10.1371/journal.pone.0042486

Jensen, S. K. G., Dickie, E. W., Schwartz, D. H., Evans, C. J., Dumontheil, I., Paus, T., & Barker, E. D. (2015). Effect of Early Adversity and Childhood Internalizing Symptoms on Brain Structure in Young Men. *JAMA Pediatrics*, *169*(10), 938. https://doi.org/10.1001/jamapediatrics.2015.1486

Jeynes, W. H. (2002). Examining the Effects of Parental Absence on the Academic Achievement of Adolescents: The Challenge of Controlling for Family Income. *Journal of Family and Economic Issues*, *23*(2), 189–210. https://doi.org/10.1023/A:1015790701554

Jolles, D. D., van Buchem, M. A., Rombouts, S. A. R. B., & Crone, E. A. (2012). Practice effects in the developing brain: A pilot study. *Developmental Cognitive Neuroscience*, *2*, S180–S191. https://doi.org/10.1016/j.dcn.2011.09.001

Joseph, T. N., Hughes, R. W., Sörqvist, P., & Marsh, J. E. (2018). Differences in Auditory Distraction between Adults and Children: A Duplex-mechanism Approach. *Journal of Cognition*, *1*(1), Article 1. https://doi.org/10.5334/joc.15

Judd, N., Sauce, B., Wiedenhoeft, J., Tromp, J., Chaarani, B., Schliep, A., Noort, B. van, Penttilä, J., Grimmer, Y., Insensee, C., Becker, A., Banaschewski, T., Bokde, A. L. W., Quinlan, E. B., Desrivières, S., Flor, H., Grigis, A., Gowland, P., Heinz, A., … Klingberg, T. (2020). *Cognitive* *and brain development is independently influenced by socioeconomic status and polygenic scores for educational attainment* (p. 866624). bioRxiv. https://doi.org/10.1101/866624

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review*, *17*(3), 213–233. https://doi.org/10.1007/s11065-007-9040-z

Kanerva, K., Kiistala, I., Kalakoski, V., Hirvonen, R., Ahonen, T., & Kiuru, N. (2019). The feasibility of working memory tablet tasks in predicting scholastic skills in classroom settings. *Applied Cognitive Psychology*, *33*(6), 1224–1237. https://doi.org/10.1002/acp.3569

Kaplan, G. A., Turrell, G., Lynch, J. W., Everson, S. A., Helkala, E.-L., & Salonen, J. T. (2001). Childhood socioeconomic position and cognitive function in adulthood. *International Journal of Epidemiology*, *30*(2), 256–263. https://doi.org/10.1093/ije/30.2.256

Kessels, R. P. C., van den Berg, E., Ruis, C., & Brands, A. M. A. (2008). The Backward Span of the Corsi Block-Tapping Task and Its Association With the WAIS-III Digit Span. *Assessment*, *15*(4), 426–434. https://doi.org/10.1177/1073191108315611

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience*, *14*(1), 1–10.

Kóbor, A., Takács, Á., & Urbán, R. (2013). The bifactor model of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment*, *29*, 299–307. https://doi.org/10.1027/1015-5759/a000160

Kostyrka-Allchorne, K., Cooper, N. R., & Simpson, A. (2017). The relationship between television exposure and children’s cognition and behaviour: A systematic review. *Developmental Review*, *44*, 19–58. https://doi.org/10.1016/j.dr.2016.12.002

Kowalczyk, A., McDonald, S., Cranney, J., & McMahon, M. (2001). Cognitive Flexibility in the Normal Elderly and in Persons with Dementia as Measured by the Written and Oral Trail Making Tests. *Brain Impairment*, *2*(1), 11–21. https://doi.org/10.1375/brim.2.1.11

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology*, *9*(8), Article 8. https://doi.org/10.4236/psych.2018.98126

Lai, K., & Green, S. B. (2016). The Problem with Having Two Watches: Assessment of Fit When RMSEA and CFI Disagree. *Multivariate Behavioral Research*, *51*(2–3), 220–239. https://doi.org/10.1080/00273171.2015.1134306

Lakes, K. D. (2013). Measuring self-regulation in a physically active context: Psychometric analyses of scores derived from an observer-rated measure of self-regulation. *Mental Health and Physical Activity*, *8*(3), 189–196. https://doi.org/10.1016/j.mhpa.2013.09.003

Lambert, H. K., King, K. M., Monahan, K. C., & McLaughlin, K. A. (2017). Differential associations of threat and deprivation with emotion regulation and cognitive control in adolescence. *Development and Psychopathology*, *29*(3), 929–940. https://doi.org/10.1017/S0954579416000584

Lamm, C., Zelazo, P. D., & Lewis, M. D. (2006). Neural correlates of cognitive control in childhood and adolescence: Disentangling the contributions of age and executive function. *Neuropsychologia*, *44*(11), 2139–2148. https://doi.org/10.1016/j.neuropsychologia.2005.10.013

Laube, C., van den Bos, W., & Fandakova, Y. (2020). The relationship between pubertal hormones and brain plasticity: Implications for cognitive training in adolescence. *Developmental Cognitive Neuroscience*, *42*. https://doi.org/10.1016/j.dcn.2020.100753

Lawson, G. M., & Farah, M. J. (2017). Executive function as a mediator between SES and academic achievement throughout childhood. *International Journal of Behavioral Development*, *41*(1), 94–104. https://doi.org/10.1177/0165025415603489

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

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development*, *84*(6), 1933–1953. https://doi.org/10.1111/cdev.12096

Lee, N. R., Wallace, G. L., Raznahan, A., Clasen, L. S., & Giedd, J. N. (2014). Trail making test performance in youth varies as a function of anatomical coupling between the prefrontal cortex and distributed cortical regions. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00496

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

Lien, N., Friestad, C., & Klepp, K.-I. (2001). Adolescents' proxy reports of parents' socioeconomic status: How valid are they? *Journal of Epidemiology & Community Health*, *55*(10), 731–737. https://doi.org/10.1136/jech.55.10.731

Lipina, S. J., Martelli, M. I., Vuelta, B., & Colombo, J. A. (2005). Performance on the A-not-B task of Argentinean infants from unsatisfied and satisfied basic needs homes. *Interamerican Journal of Psychology*, *39*(1), 49–60.

Lipina, S. J., & Posner, M. I. (2012). The impact of poverty on the development of brain networks. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00238

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, *36*(3), 273–293. https://doi.org/10.1016/S0028-3932(97)00109-7

Luciana, M., & Nelson, C. A. (2002). Assessment of Neuropsychological Function Through Use of the Cambridge Neuropsychological Testing Automated Battery: Performance in 4- to 12-Year-

Old Children. *Developmental Neuropsychology*, *22*(3), 595–624. https://doi.org/10.1207/S15326942DN2203_3

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372. https://doi.org/10.1111/j.1467-8624.2004.00745.x

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Maes, M. J. A., Pirani, M., Booth, E. R., Shen, C., Milligan, B., Jones, K. E., & Toledano, M. B. (2021). Benefit of woodland and other natural environments for adolescents' cognition and mental health. *Nature Sustainability*, *4*(10), Article 10. https://doi.org/10.1038/s41893-021-00751-1

Makowski, D. (2018). The psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Malagoli, C., & Usai, M. C. (2015, September 25). Inhibition and Working Memory: A latent variable analysis during adolescence. *ResearchGate*. EuroAsianPacific Joint Conference on Cognitive Science 2015. https://www.researchgate.net/publication/297717732_Inhibition_and_Working_Memory_A_latent_variable_analysis_during_adolescence

Massonnié, J., Mareschal, D., & Kirkham, N. Z. (2022). Individual Differences in Dealing With Classroom Noise Disturbances. *Mind, Brain, and Education*, *16*(3), 252–262. https://doi.org/10.1111/mbe.12322

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child Development*, *88*(1), 156–166. https://doi.org/10.1111/cdev.12671

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: Developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development*, *75*(5), 1373–1386. https://doi.org/10.1111/j.1467-8624.2004.00746.x

Mills, K. L. (2014). Effects of Internet use on the adolescent brain: Despite popular claims, experimental evidence remains scarce. *Trends in Cognitive Sciences*, *18*(8), 385–387. https://doi.org/10.1016/j.tics.2014.04.011

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

Monette, S., Bigras, M., & Guay, M.-C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology*, *109*(2), 158–173. https://doi.org/10.1016/j.jecp.2011.01.008

Monnier, M., Jaunin, L., Bickle Graz, M., Borradori Tolsa, C., Hüppi, P., Sancho Rossignol, A., Barisnikov, K., & Forcada Guex, M. (2014). [Detection of executive function disorders with a standard neurodevelopmental follow-up of premature children]. *Archives De Pediatrie: Organe Officiel De La Societe Francaise De Pediatrie*, *21*(9), 944–952. https://doi.org/10.1016/j.arcped.2014.04.032

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.

Morandini, H. A. E., Silk, T. J., Griffiths, K., Rao, P., Hood, S. D., & Zepf, F. D. (2020). Meta-analysis of the neural correlates of vigilant attention in children and adolescents. *Cortex*, *132*, 374–385. https://doi.org/10.1016/j.cortex.2020.08.008

Morgan, O., & Baker, A. (2006). Measuring deprivation in England and Wales using 2001 Carstairs scores. *Health Stat Q*, *31*(2).

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2009). Risk Factors for Learning-Related Behavior Problems at 24 Months of Age: Population-Based Estimates. *Journal of Abnormal Child Psychology*, *37*(3), 401–413. https://doi.org/10.1007/s10802-008-9279-8

Morris, R., & Carstairs, V. (1991). Which deprivation? A comparison of selected deprivation indexes. *Journal of Public Health*, *13*(4), 318–326. https://doi.org/10.1093/oxfordjournals.pubmed.a042650

Müller, U., Kerns, K. A., & Konkin, K. (2012). Test-retest reliability and practice effects of executive function tasks in preschool children. *The Clinical Neuropsychologist*, *26*(2), 271–287. https://doi.org/10.1080/13854046.2011.645558

Ndaji, F., Little, J., & Coe, R. (2016). A comparison of academic achievement in independent and state schools. *Durham University: Centre for Evaluation and Monitoring, Durham University*.

Nenty, H. J., & Dinero, T. E. (1981). A Cross-Cultural Analysis of the Fairness of the Cattell Culture Fair Intelligence Test Using the Rasch Model: *Applied Psychological Measurement*. https://doi.org/10.1177/014662168100500309

Neville, H. J., Stevens, C., Pakulak, E., Bell, T. A., Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences*, *110*(29), 12138–12143. https://doi.org/10.1073/pnas.1304437110

Nghiem, H. S., Nguyen, H. T., Khanam, R., & Connelly, L. B. (2015). Does school type affect cognitive and non-cognitive development in children? Evidence from Australian primary schools. *Labour Economics*, *33*, 55–65. https://doi.org/10.1016/j.labeco.2015.02.009

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective & Behavioral Neuroscience*, *12*(2), 241–268. https://doi.org/10.3758/s13415-011-0083-5

Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., Akshoomoff, N., Amaral, D. G., Bloss, C. S., Libiger, O., Schork, N. J., Murray, S. S., Casey, B. J., Chang, L., Ernst, T. M., Frazier, J. A., Gruen, J. R., Kennedy, D. N., Zijl, P. V., … Sowell, E. R. (2015). Family Income, Parental Education and Brain Structure in Children and Adolescents. *Nature Neuroscience*, *18*(5), 773–778. https://doi.org/10.1038/nn.3983

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, *10*(4), 464–480. https://doi.org/10.1111/j.1467-7687.2007.00600.x

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

Norman, D. A., & Shallice, T. (1986). Attention to Action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 1–18). Springer US. https://doi.org/10.1007/978-1-4757-0629-1_1

Ntoumanis, N. (2014). Analysing longitudinal data with multilevel modelling. *The Eurpoean Health Psychologist*, *16*(2), 40–45.

Nyhus, E., & Barceló, F. (2009). The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update. *Brain and Cognition*, *71*(3), 437–451. https://doi.org/10.1016/j.bandc.2009.03.005

Nyongesa, M. K., Ssewanyana, D., Mutua, A. M., Chongwo, E., Scerif, G., Newton, C. R. J. C., & Abubakar, A. (2019). Assessing Executive Function in Adolescence: A Scoping Review of Existing Measures and Their Psychometric Robustness. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00311

Ofcom. (2016, September 30). *The Communications Market Report: United Kingdom*. Ofcom. https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr15/uk

Ofcom. (2019). *Children and parents: Media use and attitudes report 2019*. 36.

Office for National Statistics. (n.d.). *The National Statistics Socio-economic classification (NS-SEC)*. Retrieved 17 September 2018, from https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010#deriving-the-ns-sec-full-reduced-and-simplified-methods

Office for National Statistics. (2016). *Internet access – households and individuals, Great Britain: 2016*. https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2016

Osborne, J. W., & Costello, A. B. (2009). Best Practices In Exploratory Factor Analysis: Four Recommendations To Get The Most From Your Analysis. *Pan Pacific Management Review*, *12*(2), 131–146.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, *28*(10), 1021–1034. https://doi.org/10.1016/0028-3932(90)90137-D

Payne, J. (2006). *Scottish Index of Multiple Deprivation (SMID).*

Peters, M., & Battista, C. (2008). Peters, M. & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a Mental Rotation Stimulus Library. Brain and Cognition, 66,260-264. *Brain and Cognition*, *66*. https://doi.org/10.1016/j.bandc.2007.09.003

Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, *35*, 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Pinilla, J., Lopez-Valcarcel, B. G., & Urbanos-Garrido, R. M. (2017). Estimating direct effects of parental occupation on Spaniards' health by birth cohort. *BMC Public Health*, *17*. https://doi.org/10.1186/s12889-016-3997-6

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Preacher, K., & Coffman, D. (2006, May). *Computing power and minimum sample size for RMSEA [Computer software]*. Quantpsy.Org. https://quantpsy.org/rmsea/rmsea.htm

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011a). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental Child Psychology*, *108*(3), 621–637. https://doi.org/10.1016/j.jecp.2010.09.008

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011b). Development of hot and cool executive function during the transition to adolescence. *Journal of Experimental Child Psychology*, *108*(3), 621–637. https://doi.org/10.1016/j.jecp.2010.09.008

Przybylski, A. K., & Weinstein, N. (2017). A Large-Scale Test of the Goldilocks Hypothesis: Quantifying the Relations Between Digital-Screen Use and the Mental Well-Being of Adolescents. *Psychological Science*, *28*(2), 204–215. https://doi.org/10.1177/0956797616678438

Purpura, D. J., Schmitt, S. A., & Ganley, C. M. (2017). Foundations of mathematics and literacy: The role of executive functioning components. *Journal of Experimental Child Psychology*, *153*, 15–34. https://doi.org/10.1016/j.jecp.2016.08.010

Raiford, S. E., Coalson, D. L., Saklofske, D. H., & Weiss, L. G. (2010). CHAPTER 2—Practical Issues in WAIS-IV Administration and Scoring. In L. G. Weiss, D. H. Saklofske, D. L. Coalson, & S. E. Raiford (Eds.), *WAIS-IV Clinical Use and Interpretation* (pp. 25–59). Academic Press. https://doi.org/10.1016/B978-0-12-375035-8.10002-3

Reilly, E. B., Stallworthy, I. C., Mliner, S. B., Troy, M. F., Elison, J. T., & Gunnar, M. R. (2021). Infants' abilities to respond to cues for joint attention vary by family socioeconomic status. *Infancy*, *26*(2), 204–222. https://doi.org/10.1111/infa.12380

Rice, T. R. (2017). Commentary: Executive functioning—a key construct for understanding developmental psychopathology or a ‘catch-all’ term in need of some rethinking? *Frontiers in Neuroscience*, *11*. https://doi.org/10.3389/fnins.2017.00130

Richardson, J. T. E. (2007). Measures of Short-Term Memory: A Historical Review. *Cortex*, *43*(5), 635–650. https://doi.org/10.1016/S0010-9452(08)70493-3

Richmond, S., Johnson, K. A., Seal, M. L., Allen, N. B., & Whittle, S. (2016). Development of brain networks and relevance of environmental and genetic factors: A systematic review. *Neuroscience & Biobehavioral Reviews*, *71*, 215–239. https://doi.org/10.1016/j.neubiorev.2016.08.024

Richter, M., Leppin, A., & Nic Gabhainn, S. (2006). The relationship between parental socio-economic status and episodes of drunkenness among adolescents: Findings from a cross-national survey. *BMC Public Health*, *6*, 289. https://doi.org/10.1186/1471-2458-6-289

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., Mcinnes, L., & Rabbitt, P. M. A. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society*, *4*(05). https://doi.org/10.1017/S1355617798455073

Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/bul0000227

Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., Manes, F., & Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain*, *133*(1), 234–247. https://doi.org/10.1093/brain/awp269

Rose, D., & Pevalin, D. (2003). *A Researcher’s Guide to the National Statistics Socio-economic Classification*. SAGE Publications, Ltd. https://doi.org/10.4135/9780857024725

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2011). Modeling a cascade of effects: The role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science*, *14*(5), 1161–1175. https://doi.org/10.1111/j.1467-7687.2011.01068.x

Rosen, M. L., Hagen, M. P., Lurie, L. A., Miles, Z. E., Sheridan, M. A., Meltzoff, A. N., & McLaughlin, K. A. (2020). Cognitive stimulation as a mechanism linking socioeconomic status with executive function: A longitudinal investigation. *Child Development*, *91*(4), e762–e779. https://doi.org/10.1111/cdev.13315

Rosen, M. L., Meltzoff, A. N., Sheridan, M. A., & McLaughlin, K. A. (2019). Distinct aspects of the early environment contribute to associative memory, cued attention, and memory-guided attention: Implications for academic achievement. *Developmental Cognitive Neuroscience*, *40*. https://doi.org/10.1016/j.dcn.2019.100731

Roser, K., Schoeni, A., & Röösli, M. (2016). Mobile phone use, behavioural problems and concentration capacity in adolescents: A prospective study. *International Journal of Hygiene and Environmental Health*, *219*(8), 759–769. https://doi.org/10.1016/j.ijheh.2016.08.007

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Ruchkin, V., Koposov, R., & Schwab-Stone, M. (2007). The Strength and Difficulties Questionnaire: Scale validation with Russian adolescents. *Journal of Clinical Psychology*, *63*(9), 861–869. https://doi.org/10.1002/jclp.20401

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029–1040. https://doi.org/10.1016/j.neuropsychologia.2003.12.012

Salthouse, T. A. (2011). What cognitive abilities are involved in trail-making performance? *Intelligence*, *39*(4), 222–232. https://doi.org/10.1016/j.intell.2011.03.001

Sánchez-Cubillo, I., Periáñez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Ríos-Lago, M., Tirapu, J., & Barceló, F. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, *15*(3), 438–450. https://doi.org/10.1017/S1355617709090626

Sanders, T., Liu, Y., Buchner, V., & Tchounwou, P. B. (2009). Neurotoxic Effects and Biomarkers of Lead Exposure: A Review. *Reviews on Environmental Health*, *24*(1), 15–45.

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: The roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society: JINS*, *17*(1), 120–132. https://doi.org/10.1017/S1355617710001335

Sasser, T. R., Bierman, K. L., & Heinrichs, B. (2015). Executive Functioning and School Adjustment: The Mediational Role of Pre-kindergarten Learning-related Behaviors. *Early Childhood Research Quarterly*, *30*(Pt A), 70–79. https://doi.org/10.1016/j.ecresq.2014.09.001

Schoeni, A., Roser, K., & Röösli, M. (2017). Symptoms and the use of wireless communication devices: A prospective cohort study in Swiss adolescents. *Environmental Research*, *154*, 275–283. https://doi.org/10.1016/j.envres.2017.01.004

Selzam, S., Krapohl, E., von Stumm, S., O'Reilly, P. F., Rimfeld, K., Kovas, Y., Dale, P. S., Lee, J. J., & Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular Psychiatry*, *22*(2), Article 2. https://doi.org/10.1038/mp.2016.107

Shallice, T., & Burgess, P. W. (1991). Deficits in Strategy Application Following Frontal Lobe Damage in Man. *Brain*, *114*(2), 727–741. https://doi.org/10.1093/brain/114.2.727

Shangguan, F., & Shi, J. (2009). Puberty timing and fluid intelligence: A study of correlations between testosterone and intelligence in 8- to 12-year-old Chinese boys. *Psychoneuroendocrinology*, *34*(7), 983–988. https://doi.org/10.1016/j.psyneuen.2009.01.012

Sheridan, M. A., Peverill, M., Finn, A. S., & McLaughlin, K. A. (2017). Dimensions of childhood adversity have distinct associations with neural systems underlying executive functioning.

Development and Psychopathology, *29*(5), 1777–1794. https://doi.org/10.1017/S0954579417001390

Sheridan, M. A., Sarsour, K., Jutte, D., D'Esposito, M., & Boyce, W. T. (2012). The Impact of Social Disparity on Prefrontal Function in Childhood. *PLOS ONE*, *7*(4), e35744. https://doi.org/10.1371/journal.pone.0035744

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453. https://doi.org/10.3102/00346543075003417

Soper, D. S. (2024a). *A-priori Sample Size Calculator for Hierarchical Multiple Regression* [Computer software]. https://www.danielsoper.com/statcalc

Soper, D. S. (2024b). *A-priori Sample Size Calculator for Multiple Regression* [Computer software]. https://www.danielsoper.com/statcalc

Spearman, C. (1904). 'General Intelligence,' Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Strickland, P., & Dent, J. (2017). *Online harassment and cyber bullying*. https://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7967

Stuss, D. T. (2011). Functions of the Frontal Lobes: Relation to Executive Functions. *Journal of the International Neuropsychological Society*, *17*(05), 759–765. https://doi.org/10.1017/S1355617711000695

Stuss, D. T., & Levine, B. (2002). Adult Clinical Neuropsychology: Lessons from Studies of the Frontal Lobes. *Annual Review of Psychology*, *53*(1), 401–433. https://doi.org/10.1146/annurev.psych.53.100901.135220

Suchy, Y., Niermeyer, M. A., & Ziemnik, R. E. (2017). Chapter 9—Assessment of Executive Functions in Research. In E. Goldberg (Ed.), *Executive Functions in Health and Disease* (pp. 197–216). Academic Press. https://doi.org/10.1016/B978-0-12-803676-1.00009-X

Sun, J., Zhang, L., Chen, E., Lau, C., & Rao, N. (2018). Preschool Attendance and Executive Function Mediate Early Academic Achievement Gaps in East Asia and the Pacific. *Early Education and Development*, *29*(8), 1039–1060. https://doi.org/10.1080/10409289.2018.1498999

Tamnes, C. K., Ostby, Y., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2010). Brain maturation in adolescence and young adulthood: Regional age-related changes in cortical thickness and white matter volume and microstructure. *Cerebral Cortex (New York, N.Y.: 1991)*, *20*(3), 534–548. https://doi.org/10.1093/cercor/bhp118

Thompson, A., & Steinbeis, N. (2020). Sensitive periods in executive function development. *Current Opinion in Behavioral Sciences*, *36*, 98–105. https://doi.org/10.1016/j.cobeha.2020.08.001

Thompson, J. M. D., Slykerman, R. F., Wall, C. R., Murphy, R., Mitchell, E. A., & Waldie, K. E. (2021). Factor structure of the SDQ and longitudinal associations from pre-school to pre-teen in New Zealand. *PLoS ONE*, *16*(3), e0247932. https://doi.org/10.1371/journal.pone.0247932

Thompson, R., Smith, R. B., Karim, Y. B., Shen, C., Drummond, K., Teng, C., & Toledano, M. B. (2023). Air pollution and human cognition: A systematic review and meta-analysis. *Science of The Total Environment*, *859*, 160234. https://doi.org/10.1016/j.scitotenv.2022.160234

Tiego, J., Testa, R., Bellgrove, M. A., Pantelis, C., & Whittle, S. (2018). A Hierarchical Model of Inhibitory Control. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01339

Toledano, M. B., Mutz, J., Röösli, M., Thomas, M. S. C., Dumontheil, I., & Elliott, P. (2018). Cohort Profile: The Study of Cognition, Adolescents and Mobile Phones (SCAMP). *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyy192

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and Brain Sciences*, *28*(5), 675–691; discussion 691-735. https://doi.org/10.1017/S0140525X05000129

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*(2), 203–214. https://doi.org/10.1016/S0887-6177(03)00039-8

Topor, D. R., Keane, S. P., Shelton, T. L., & Calkins, S. D. (2010). Parent Involvement and Student Academic Performance: A Multiple Mediational Analysis. *Journal of Prevention & Intervention in the Community*, *38*(3), 183–197. https://doi.org/10.1080/10852352.2010.486297

Tsukayama, E., Duckworth, A. L., & Kim, B. (2013). Domain-Specific Impulsivity in School-Age Children. *Developmental Science*, *16*(6), 879. https://doi.org/10.1111/desc.12067

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*(6), 623–628. https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x

Urben, S., Van der Linden, M., & Barisnikov, K. (2011). Development of the ability to inhibit a prepotent response: Influence of working memory and processing speed. *The British Journal of Developmental Psychology*, *29*(Pt 4), 981–998. https://doi.org/10.1111/j.2044-835X.2011.02037.x

Ursache, A., & Noble, K. G. (2016). Socioeconomic status, white matter, and executive function in children. *Brain and Behavior*, *6*(10). https://doi.org/10.1002/brb3.531

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), 575–587. https://doi.org/10.1037/0012-1649.44.2.575

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Hink, R. F., & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology*, *33*(1), 101–111. https://doi.org/10.1080/13803395.2010.493149

World Health Organization. (2010). *WHO 2010 Research Agenda for Radio Frequency Fields*. WHO. http://apps.who.int/iris/bitstream/10665/44396/1/9789241599948_eng.pdf

Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PloS One*, *8*(10), e77770. https://doi.org/10.1371/journal.pone.0077770

Zanini, G. A. V., Miranda, M. C., Cogo-Moreira, H., Nouri, A., Fernández, A. L., & Pompéia, S. (2021). An Adaptable, Open-Access Test Battery to Study the Fractionation of Executive-Functions in Diverse Populations. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.627219

Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). Executive Function: Implications for Education. In *National Center for Education Research*. National Center for Education Research. https://eric.ed.gov/?id=ED570880

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). The development of executive function in childhood. In *Handbook of developmental cognitive neuroscience, 2nd ed* (pp. 553–574). MIT Press.

Appendix A
Strengths and Difficulties Questionnaire, 11-17 Version

Please give your answers on the basis of how things have been for you OVER THE LAST SIX MONTHS.

Options buttons: NOT TRUE; SOMEWHAT TRUE; CERTAINLY TRUE.

1. I try to be nice to other people. I care about their feelings
2. I am restless, I cannot stay still for long
3. I get a lot of headaches, stomach-aches or sickness
4. I usually share with others (food, games, pens etc.)
5. I get very angry and often lose my temper
6. I am usually on my own. I generally play alone or keep to myself
7. I usually do as I am told
8. I worry a lot
9. I am helpful if someone is hurt, upset or feeling ill
10. I am constantly fidgeting or squirming
11. I have one good friend or more
12. I fight a lot. I can make other people do what I want
13. I am often unhappy, down-hearted or tearful
14. Other people my age generally like me
15. I am easily distracted, I find it difficult to concentrate
16. I am nervous in new situations. I easily lose confidence
17. I am kind to younger children
18. I am often accused of lying or cheating
19. Other children or young people pick on me or bully me
20. I often volunteer to help others (parents, teachers, children)
21. I think before I do things
22. I take things that are not mine from home, school or elsewhere
23. I get on better with adults than with people my own age
24. I have many fears, I am easily scared
25. I finish the work I'm doing. My attention is good

Coding of items

Emotional problems: 3, 8, 13, 16, 24

Conduct problems: 5, 7, 12, 18, 22

Hyperactivity problems: 2, 10, 15, 21, 25

Peer problems: 6, 11, 14, 19, 23

Prosocial: 1, 4, 9, 17, 20

Appendix B
Instructions presented for BDS task in the SCAMP battery

BDS

BDS_intro In this task you will see series of numbers, like 2, 5, or 3, 9, 1, appearing on the screen one by one.¬You need to remember the numbers and indicate what they were clicking on a number pad like this.¬ ¬ ¬ ¬ ¬ ¬ The trick is that you have to click on the numbers in the REVERSE ORDER of what they were shown in.¬So if you see 1 then 9 you would need to press number 9 first and then number 1.¬Let's try the practice.

BDS_practice_fail2 Oops you didn't get that one, let's try again, remember click on the numbers you saw IN REVERSE ORDER.¬Press space to try another

BDS_pre_main Well Done! Press space to do the rest of them now - you won't be told if you get them right or wrong any more.

BDS_close That was great, WELL DONE!!¬ ¬Press space to try another task.

BDS_reverseRemind Remember you have to REVERSE the numbers in your head!¬So if you see 1 then 9 you would need to press number 9 first and then number 1.¬Press space to try another

BDS_no_idea_text I DON'T KNOW

BDS_close_early THANKS! Let's try another task. ¬ ¬ Press space to continue

BDS_practice_fail3 You got this last one wrong, remember you have to repeat the numbers backwards, so if you see 1, 2, 3 you should click 3, 2, 1.¬Press space to try another

Appendix C Results for Chapter 3 (SES)

MANCOVA results

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	0.057	16.881[b]	5.000	1401.000	0.000	0.057
School_Type	0.055	16.329[b]	5.000	1401.000	0.000	0.055
SES_F	0.042	1.714	35.000	7025.000	0.006	0.008
SES_M	0.031	1.266	35.000	7025.000	0.135	0.006
cars11_q	0.025	1.759	20.000	5616.000	0.019	0.006
Par_Ed_F	0.004	1.070[b]	5.000	1401.000	0.375	0.004
Par_Ed_M	0.005	1.373[b]	5.000	1401.000	0.232	0.005
Age	0.027	7.735[b]	5.000	1401.000	0.000	0.027

a. Design: MANCOVA test, Intercept + School_Type + SES_F + SES_M + cars11_q + Par_Ed_F + Par_Ed_M + Age

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	0.057	16.840[b]	5.000	1400.000	0.000	0.057
School_Type	0.046	13.579[b]	5.000	1400.000	0.000	0.046
SES_F	0.041	1.649	35.000	7020.000	0.009	0.008
SES_M	0.029	1.184	35.000	7020.000	0.211	0.006
cars11_q	0.025	1.733	20.000	5612.000	0.022	0.006
Par_Ed_F	0.004	1.109[b]	5.000	1400.000	0.354	0.004
Par_Ed_M	0.004	1.085[b]	5.000	1400.000	0.367	0.004
Age	0.019	5.550[b]	5.000	1400.000	0.000	0.019
CFT	0.102	31.848[b]	5.000	1400.000	0.000	0.102

a. Design: MANCOVA test Intercept + School_Type + SES_F + SES_M + cars11_q + Par_Ed_F + Par_Ed_M + Age + CFT

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Appendix D

Results for Chapter 4 (Developmental Trajectories)

Complete Model information for linear regressions, with each task score as outcome, and age in years as predictor in each model.

Outcome	**Predictor**	***b***	***b* 95% CI**	***beta***	***beta* 95% CI**	***r***	**Model Fit R^2, 95% CI**
TMT	Intercept	-0.36**	[-0.56, -0.16]				R^2 = .002**
	Age Years	0.03**	[0.01, 0.05]	0.04	[0.02, 0.05]	.04**	[.00,.00]
BDS	Intercept	-1.67**	[-1.89, -1.45]				R^2 = .024**
	Age Years	0.14**	[0.12, 0.16]	0.16	[0.14, 0.17]	.16**	[.02,.03]
SWM errors	Intercept	-1.21**	[-1.41, -1.00]				R^2 = .014**
	Age Years	0.10**	[0.09, 0.12]	0.12	[0.10, 0.14]	.12**	[.01,.02]
SWM strategy	Intercept	-1.40**	[-1.59, -1.20]				R^2 = .021**
	Age Years	0.12**	[0.10, 0.13]	0.14	[0.13, 0.16]	.14**	[.02,.03]
Corsi	Intercept	-1.97**	[-2.27, -1.67]				R^2 = .032**
	Age Years	0.16**	[0.14, 0.19]	0.18	[0.15, 0.20]	.18**	[.02,.04]
CPT omission	Intercept	0.06	[-0.39, 0.51]				R^2 = .000
	Age Years	-0.01	[-0.04, 0.03]	-0.01	[-0.05, 0.03]	-.01	[.00,.00]
CPT commission	Intercept	-1.15**	[-1.56, -0.74]				R^2 = .014**
	Age Years	0.10**	[0.06, 0.13]	0.12	[0.08, 0.16]	.12**	[.01,.02]
CFT	(Intercept)	-1.70**	[-1.91, -1.48]				R^2 = .026**
	Age_Years	0.14**	[0.12, 0.16]	0.16	[0.14, 0.18]	.16**	95% CI[.02,.03]

Note. A significant b indicates the standardised beta is also significant. b represents unstandardized regression weights. beta indicates standardized regression weights. r represents the zero-order correlation. Numbers in square brackets indicate lower and upper limits of a confidence interval, respectively [LL, UL]. * indicates p < .05. ** indicates p < .01.