UNIVERSITY OF CAMBRIDGE
Faculty of Education

Faculty of Education

MPhil in Education

Thesis

Novel metacognitive problem-solving task for 8- to 11-year-old students

Keywords- metacognitive control, metacognitive monitoring, problem-solving, metacognitive memory, secondary data analysis.

Name: Jwalin Nishith Patel

Route: Educational Research

Supervisor: Dr Michelle Ellefson

College: Fitzwilliam College

Submission Date: 6th July, 2017

Word Count: 19,980 Words (excluding title page, acknowledgements, tables, table notes, figures, equations and reference list as per course guidelines).

Abstract

Metacognition is important for monitoring and regulation of cognitive processes, decision making, problem-solving and learning. Despite the widespread interest in metacognition, measuring metacognition in children poses a significant challenge. Some qualitative and observational measures exist, but they are restricted by the number of components they measure and the sampling size. Some meta-cognition tasks of memory have been developed for children, but these only measure a narrow range of skills involved in metacognition. A novel metacognitive problem-solving task, previously developed by the lab, provides scalable means to holistically measure metacognition. The thesis, developed a new coding scheme, recoded the data and studied the reliability and validity of the novel task by comparing it with demographic variables known to be associated with metacognition and a metamemory task. The results indicate the novel task is reliable and valid. It operationalizes metacognitive measures similarly to a classical metamemory task, suggesting that the new task could be a bridge between existing measures of metacognition in children and adults.

The thesis uses the novel task to then explore other broader questions with 182, 8- to 11-year-old students, pertaining to cognitive levels in low SES, ethnic-minority students, domain-generality/specificity of metacognition and the association between metacognition and executive-functions. The results indicate that low SES, ethnic-minority students have poor cognitive levels and low amounts of cognitive development across the various age groups. The results also suggest metacognitive components to be domain-general in nature and were tapped into by the novel metacognitive problem-solving, metamemory and a complex executive-function tasks. The results provide further evidence for association between metacognition and executive-functions.

Declaration of originality

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. This thesis does not exceed 20,000 words in length.

Signature (hard copy submission only)……………………………………………………….

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110932 to the University of Cambridge. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Acknowledgements

The data presented here were collected by a large team of researchers from the University of Cambridge, Virginia State University, Virginia Commonwealth University and Ashley-Parr, LLC (listed alphabetically by last name): Mariah Adlawan, Annabel Amodia-Bidakowska, Cortney Anderson, Aaron Blount, Lakendra Butler, Parul Chaudhary, Laura Clarke, Jackson Collins, Aiden Cope, Amenah Darab, Asha Earle, Mary Elyiace, Sophie Farthing, Pippa Fielding Smith, Aysha Foster, Kristine Gagarin, Marleny Gaitan, Summer Gamal, Katie Gilligan, Cynthia Gino, Aditi Gupta, Jennifer Hacker, Donita Hay, Rachel Heeds, Joy Jones, Spencer Kearns, Hyunji Kim, Steven Mallis, Dr. Geoff Martin, Alexandria Merritt, Kelsey Richardson, Fran Riga, Tennisha Riley, Kristin Self, Amelia Swafford, Krystal Thomas, Quai Travis, Jorge Vargas, Tony Volley, Elexis White, Sterling Young. These researchers were directed by the PI/Co-PIs: Drs. Michelle Ellefson, Zewelanji Serpell, and Teresa Parr.

Special thanks to Dr Geoff Martin and Dr Michelle Ellefson for programming and developing the overall participant and research interfaces for the Mind Match Chess secured website, to Dr Michelle Ellefson, Dr Zewelanji Serpell and Aysha Foster for the development of the instructions and online presentation of the Disks and Memory Challenges, to Dr Amanda Aldercotte, Dr Michelle Ellefson and Dr Zewelanji Serpell for developing the Zoo-task and to the PIs of the primary project: Dr Michelle Ellefson, Dr Zewelanji Serpell and Dr Teresa Parr to allow me to use the primary data.

I would like to thank my supervisor, Dr Michelle Ellefson, for guiding me through the research and for supporting me to think deeply and critically about the research design and its analysis.

I would also like to thank my family for supporting me in throughout this year and all my EdRes classmates for their constructive feedback and friendship. I would like to particularly thank Darshini Sundar, Layla Patel and Yushan Xie for proofreading the thesis.

Table of contents

Table of figures

Table of tables

Table of abbreviations

AAI:	Absolute Accuracy Index
BI:	Bias Index
EF:	Executive Function
GCA:	General Cognitive Ability
JOK:	Judgement Of Knowledge
MCON:	Metacognitive Control
MEVAL:	Metacognitive Evaluation
MMON:	Metacognitive Monitoring
RCJ:	Retrospective Confidence Judgement
SES:	Socio-economic Status

Chapter 1 introduction

1.1 Background

Metacognition, colloquially understood as 'thinking about thinking', is involved in understanding and regulating one's own thinking (Flavell, 1979). Metacognition refers to a collection of cognitive processes that are responsible for awareness of one's cognitive processing, knowledge of tasks and cognitive strategies, planning, decision-making and effective monitoring of cognitive processes. Metacognition is crucial for problem-solving, memorizing, learning (Schunk & Zimmerman, 1994; Wang, Haertel, & Walberg, 1990) and has been suggested to be the building-block for academic attainment, success at work (Zimmerman & Bandura, 1994) and social competence (Eisenberg, Fabes, Carlo, & Karbon, 1992). Metacognition has also been associated with other cognitive processes like critical thinking (Deanna Kuhn, 2004), self-regulation (Schunk, 1995; Winne & Hadwin, 1998; Zimmerman, 1995) and executive-functions (EF; Bryce, Whitebread, & Szűcs, 2015; Fernandez-Duque, Baird, & Posner, 2000; Roebers, Cimeli, Röthlisberger, & Neuenschwander, 2012).

There is longstanding sustained interest in metacognition research in the educational context since cognitive psychologists (Hart, 1965; Underwood, 1966) studied the relation between test performance and students' perception of knowledge. Meta-analyses have revealed metacognition to be the single most important, at times along with self-regulation, predictor of learning (Wang *et al.*, 1990). Metacognition research in education has been trying to explore how children become independent problem-solvers, self-regulate their learning, adapt behavior and thinking patterns in responses to making errors and continuously monitor their progress on a given task. These skills are generally considered to be domain-general and crucial to all learning processes. There is overwhelming evidence of the importance of metacognition for learning and efficacy of school-based metacognition-targeting interventions (Schunk & Zimmerman, 1994; Veenman & Spaans, 2005).

1.2 Thesis aims and significance

The thesis lies in the interdisciplinary field of education and cognitive-psychology/ experimental-psychology and it studies (a) cognition levels and domain-generality in ethnic-minority, low Socio-Economic Status (SES), 8- to 11- years old students; (b) the reliability and validity of a novel metacognitive problem-solving task exploring associations with demographics data and a commonly used metamemory-task; and (c) association between metacognition and EF, a cognitive skill that is essential for decision making, planning and responding to novel situations (Diamond, 2013; Luria, 1966), using a commonly used complex EF task. The thesis aims to provide metacognition researchers with a novel assessment tool that holistically assesses various metacognitive components in children.

Despite the wide-spread recognition of the importance of metacognition for learning there is very little research on students aged 8- to 11-years, there are many papers available on studies with children in pre-primary, high-school, colleges and in adults. There are very few age-appropriate, holistic metacognition assessment measures for 8- to 11-year-olds however there are a multitude of tools available for young-adults and children aged less than six, although most being observational and non-scalable for the latter. The currently available age-appropriate tools provide limited information on key metacognitive components (control and monitoring). The novel metacognitive problem-solving task is an age-appropriate problem-solving task that provides a scalable means of collecting data on multiple metacognitive components from a single task, which would in turn would allow researchers to conduct large-scale studies and teachers to assess students' metacognition levels.

The current thesis also aims to contribute to better understanding metacognition levels in 8- to 11-year-old, ethnic-minority students and domain-general/specific nature of metacognition. The developmental insights could inform interventions and pedagogy while the exploration of metacognition domain-generality/specificity could inform curricula development in order to maximize metacognitive development and academic attainment.

In the near future, once the task has been found to be reliable and valid, it will be used to study the development of metacognition using secondary data from a three-year long longitudinal project and to study the relations between metacognition, academic attainment and EF.

1.3 Thesis outline

This thesis adopts a post-positivistic approach, correlation-regression analysis methodology, quasi-experimental methods and uses secondary data. It assumes that if there is an objective truth it cannot be known but uses the quantitative methodologies to build a plausible model of cognitive development in low SES students, test the validity of the novel task and explore metacognition-EF association.

The next chapter discusses the broader literature on metacognition, its components, development, associations with demographic variables, task validity and reliability studies and associations with EF. Chapter three discusses the methodology used both in the primary project for data-collection and in the thesis for recoding and analyzing the data. Chapter four and five summarize the findings from the analyses and discuss their meaning in the light of previous research, how it builds onto existing knowledge and limitations of the study. Chapter six concludes the thesis by discussing the broader implications of the study and the researcher's personal learning curve during the course of the study.

Chapter 2 literature review

This chapter begins by introducing and defining metacognition and theories about its components. It further discusses the role of age, gender, general cognitive ability and SES on metacognition development, followed by a review of the existing measurement tools, their limitations and the need for a new task to measure metacognition in children. It then discusses ways of testing (and developing) a novel task and introduces a wider debate of metacognition domain-generality/specificity which affects comparisons across different tasks. It then discusses proposed theories of metacognition-EF association before ending with the key research questions.

2.1 Metacognition and its components

Research in metacognition is a very diverse field with ongoing research in education, cognitive science and experimental psychology. This has a resulted in different conceptualizations and theoretical models of its components, all with minor theoretical variations (Flavell, 1987; Metcalfe & Dunlosky, 2008; Nelson & Narens, 1990; Schraw & Moshman, 1995; Tobias & Everson, 2002; Winne & Hadwin, 1998). However, there is common consensus in the field for the two core components: metacognitive knowledge and metacognitive skills (further discussed in details in following subsections). Metacognitive knowledge is the knowledge about one's own cognitive strategies and tasks and impacts test performance and school careers (e.g. Lockl & Schneider, 2006; OECD, 2005). Metacognitive skills explain individual differences in test performance, even when controlling for GCA (Van der Stel & Veenman, 2008) and can be further sub-classified as: (a) metacognitive control (MCON)- involved in regulation of cognitive skills; (b) metacognitive monitoring (MMON)- awareness of one's cognitive processes; and (c) metacognitive evaluation skills (MEVAL, frequently also referred to as metacognitive accuracy)-

judgement of one's task performance (Flavell, 1987; Jacobs & Paris, 1987; Pintrich, Wolters, & Baxter, 2000). Table 1 outlines the various terms that have been used by different researchers to describe metacognitive components.

Table 1
Summary of conceptualizations of metacognitive components.

Component	Conceptualization in this thesis	Conceptualizations in the literature	Paper
Metacognitive knowledge	Knowledge of one's learning ability	Personal knowledges, declarative knowledge, epistemological understanding	Flavell, 1979, Kuhn 2000, Schraw *et al.* 2006
	Knowledge of the task requirement	Procedural knowledge, Meta-task knowledge	Flavell, 1979, Schraw, 2006, Kuhn 2000
	Knowledge of cognitive strategies	Strategy knowledge, conditional knowledge, meta-strategic knowledge	Flavell, 1979, Schraw *et al.* 2006, Kuhn 2000,
	Metacognitive experiences	Cognitive experiences	Flavell, 1979
Metacognitive skills	MCON, regulates other cognitive skills	Planning, regulation, control	Schraw & Moshman, 1995, Jacobs & Paris 1987, Whitebread 2009
	MMON, monitors cognitive processes	Monitoring,	Schraw & Moshman, 1995, Whitebread 2009
	MEVAL, evaluates the final output	Evaluation,	Schraw & Moshman, 1995, Jacobs & Paris 1987, Whitebread 2009

MCON- metacognitive control, MMON- metacognitive monitoring, MEVAL-metacognitive evaluation

This thesis has adapted the Nelson & Narens (1990) model of metacognition (figure 1), which defines two levels of cognitive processing; object-level and meta-level. The meta-level consists of a model of the object level; corresponding to metacognitive knowledge. The meta-level is updated with bottom up information *via* monitoring processes while the meta-level decisions feed into the object level *via* control processes. Models like Metcalfe & Dunlosky, (2008)'s model conflate MMON and MEVAL into one process as both the processes involve assessment of one's performance either during the task (on-line) or before/after it (off-line), respectively. However, in this thesis the two are considered to be separate as done in many studies (Bryce *et al.*, 2015; Whitebread *et al.*, 2009) because the thesis aims to develop a novel task that holistically measures metacognition.

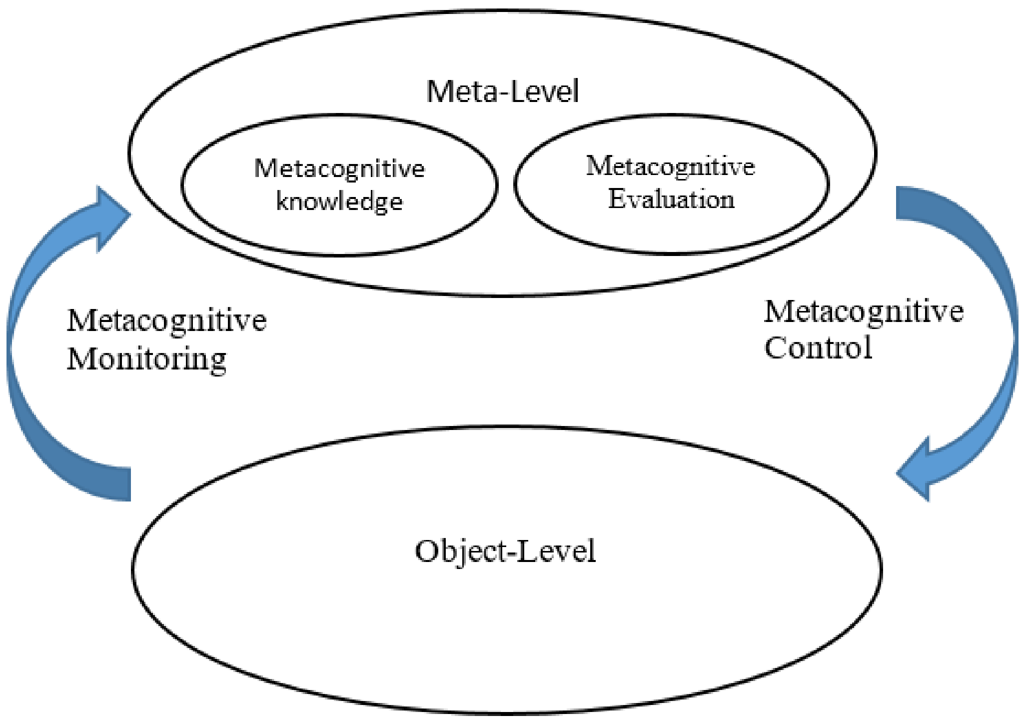

Figure 1: **adapted Nelson and Narens metacognition model**; adapted from (Fernandez-Duque, Baird, & Posner, 2000 page 3, figure 1. MEVAL has been placed in the meta-level as over period of time it could potentially inform 'metacognitive experiences' and influence metacognitive knowledge.

2.1.1 Metacognitive knowledge

Metacognitive knowledge refers to one's understanding of their cognitive process. Flavell, (1979) defines metacognitive knowledge as acquired knowledge of cognitive processes that can influence the control of the skills and categorized metacognitive knowledge as: (a) knowledge of one's learning abilities; (b) knowledge of the task requirement; (c) knowledge of strategies to facilitate task completion. Metacognitive knowledge includes three different types of metacognitive awareness: Declarative knowledge (factual knowledge "about" things), Procedural knowledge (knowledge of "how" to complete processes and tasks) and Conditional knowledge (contextual understanding of "when" and "why") (Deanna Kuhn, 2000; Schraw & Moshman, 1995).

Metacognitive knowledge is generally tested through off-line self-report, survey style questionnaires. Schraw & Dennison, (1994) developed the Metacognitive Assessment Inventory, a reliable and widely used 52-item inventory to measure adults′ metacognitive awareness, which tests both metacognitive knowledge and metacognitive skills.

2.1.2 Metacognitive Control

MCON refers to metacognitive skills that are involved in adapting and changing one's cognition and behavior depending on one's metacognitive knowledge and feedback from MMON. They are involved in planning and self-regulation and allow for changing cognitive strategies upon error detection (informed *via* MMON) or poor performance (informed *via* MEVAL; Jacobs & Paris, 1987).

A variety of measurement tools have been developed including questionnaires (off-line), think-aloud strategies (on-line e.g. Veenman, Wilhelm, & Beishuizen, 2004), experimental protocols (on-line e.g. Bryce & Whitebread, 2012; Bryce, Whitebread, & Szűcs, 2015) and classroom-based interventions.

2.1.3 Metacognitive Monitoring

MMON refers to metacognitive skills that are involved in evaluating on-line progress or performance on a given task. They are considered to be important for staying on-task and self-regulation as the continuous evaluation of the performance feeds into MCON (Bryce & Whitebread, 2012).

Given many studies conflate MMON and MEVAL, off-line MEVAL measurement tools have been interpreted as MMON measures. However, several recent studies have developed on-line MMON measures based on think-aloud and experimental protocols involving observational coding of children's behaviors on problem-solving tasks (Bryce & Whitebread, 2012; Pino-Pasternak, Whitebread, & Tolmie, 2010), allowing for distinction from MEVAL.

2.1.4 Metacognitive Evaluation

MEVAL refers to metacognitive skills that are involved in appraising the final outcome (or predicting one's potential outcome) and involves evaluation of one's performance (Schraw & Moshman, 1995). Several off-line judgement measure have been developed including prospective (predictions of one's performance before starting a task) and retrospective (estimation of one's performance upon completion of a task) judgement measures.

Judgement measures are combined with task performance to create judgement outcome-measures, much better measures of MEVAL. Several commonly used metacognitive judgement and outcome-measures relevant to the thesis are reviewed next.

Judgements. The primary study collected data on two judgements: Judgement of Knowledge (JOK) and Retrospective Confidence Judgements (RCJ), which capture students' judgements of their accuracy before and after the task, respectively (Schraw, 2009). Both (other commonly used judgement measures are summarized in table 2) are considered to be based on different informational sources and share a low to medium correlation (Jacob & Nelson, 1990). JOKs are considered to be based on task familiarity (visual, semantic or affective; Koriat, 1993; Metcalfe, Schwartz, & Joaquim, 1993) while RCJs are based on subject's memory and the nature of the given task (amount of memory rehearsals on a memory-task; Busey, 2000; Yonelinas, 1994).

Table 2
Judgement measures and their descriptions.

Judgement	Definition
Ease Of Learning	Prospective judgement of the difficulty/ease of learning a specific item.
Judgement Of Knowledge	Prospective judgement of performance on a specific item given previous knowledge.
Judgements Of Comprehension	Judgement of understanding of a passage; a retrospective adaptation of Ease Of Learning. Though judgements are made before answering questions.
Retrospective Confidence Judgements	Retrospective judgement of performance on a given test item.
Feeling Of Knowledge	Judgement of likelihood of a participant to give a correct answer to question they have answered incorrectly previously.

Judgement outcome-measures. Judgement and their relation to one's task performance is a better means of operationalizing MEVAL. For continuous data, it is recommended that researchers, depending on the research questions, select between five outcome-measures (Table 3; Nietfeld, Enders, & Schraw, 2006; Wallsten, 1996) as opposed to using signal detection theory, which is better suited for binary data.

Table 3
Judgement outcome-measures and their descriptions.

Judgement outcome-measures	Judgement outcome-measure descriptions
Absolute Accuracy Index	Difference between judgement and performance
Bias Index	Degree of over and under-confidence
Relative Accuracy	Correlation between accuracy judgement and accuracy
Scatter index	The variability of judgements of correct and incorrect responses
Discrimination index	Degree of discrimination between correct and incorrect responses.

To comprehensively compare the two tasks the thesis will use: absolute accuracy index (AAI) and bias index (BI). Relative accuracy was dropped through the study due to evolving aims of the project to compare tasks rather than judgements. Scatter and discrimination indices weren't developed because the novel metacognitive task doesn't produce a binary score.

AAI (equation 1; Schraw, 2009) represents squared differences of the judgements from the performance on task, while BI refers to positive and negative differences of the judgment and performance (equation 2; Schraw, 2009). AAI helps understand task difference and improvements, while BI measures under- and overconfidence (Nietfeld *et al.*, 2006).

$$Absolute\ Acuracy\ Index = \frac{1}{N} \times \sum_{i=1}^{N} (c_i - p_i)^2$$

Equation 1

$$Bias\ Index = \frac{1}{N} \times \sum_{i=1}^{N} (c_i - p_i)$$

Equation 2

Where N= number of items, c_i= confidence on a given item and p_i= performance on a given item

AAI ranges from zero to one, where a score of zero correspond to perfect accuracy of judgement and a score of one corresponds to the lack thereof. BI ranges from negative one to positive one where negative one represents under-confidence and one represents overconfidence.

The novel task studied in the thesis measures MCON, MMON and MEVAL (using JOK-AAI, JOK-BI, RCJ-AAI and RCJ-BI) allowing for holistic measurement of metacognitive regulatory skills.

2.2 Effect of demographic factors on metacognition (development)

Several demographic factors including age, gender, GCA and SES are predictive of metacognition and are reviewed in this section. Ideally these demographic factors would still remain predictive of the performance on the novel metacognition tasks and would need to be controlled for in any across-task comparisons.

2.2.1 Metacognitive development (effect of age on metacognition)

Early studies (Flavell, 1979; Kreutzer, Leonard, Flavell, & Hagen, 1975) concluded that metacognition is a late developing skill. However, more recently, there has been increasing evidence of development starting in four-year olds (Flavell, Miller, & Miller, 1993; Montgomery, 1992) and continuing through till adulthood (A. King, 1991; Deanna Kuhn, 1989; Moore & Frye, 1991).

A differential development of the various metacognitive components has been suggested with MEVAL and theory of mind, a mental picture of one's cognitive strategies (similar to metacognitive knowledge), emerging between ages 3- and 5-year olds and continuing to develop through the lifespan (Cultice, Somerville, & Wellman, 1983; Flavell, 2004; Lockl & Schneider, 2006). MCON and MMON emerge between 8- and 10-year, and develop through adulthood

(Veenman *et al.*, 2004), however, Bryce *et al.*, (2015), and Whitebread *et al.*, (2009), found the emergence of MCON and MMON in children aged 5- and 7-years using observational tasks.

Early studies underestimated children's metacognition levels due to methodological difficulties in measuring metacognition in younger children arising due to (a) reliance on children's verbal abilities (Kreutzer *et al.* 1975); (b) task content which children didn't relate with and (c) non age-appropriate task requirements for example working-memory, which doesn't develop early on, (Bryce & Whitebread, 2012; Pressley, Borkowski, & Schneider, 1987). However, studies using more ecologically valid tasks found significant metacognition levels were found in 4- and 5-year olds (Cultice *et al.*, 1983, Bryce *et al.*, 2015). Cultice *et al.* presented children with photographs of people with varying familiarity instead of the commonly used strategy of showing pairs of word in two different languages while Bryce *et al.* used observational problem-solving-task involving train tracks.

The absence of age-appropriate tasks have led to ceiling- and floor-effects preventing studies in students aged 8- to 11-years. Veenman *et al.,* (2004), conducted a significant study (one of the very few studying the current sample's age group) across four age groups (fourth-, sixth-, and eighth-graders, and university students) to study the developmental relation between metacognition and IQ. Metacognition was scored *via* a computerized task on a scale of 0-4 on hypotheses generation, planning, evaluation and generating explanations. However the performance on all criteria was highly correlated and combined into a single metacognitive skillfulness measure and doesn't allow differential analysis of MCON, MMON and MEVAL.

2.2.2 Gender

There is mixed evidence for the role of gender in predicting metacognitive abilities. Two major studies found contradictory findings. Sperling *et al.*, (2002), studying 344 children in grades K-9 in a rural school revealed insignificant gender differences. The extremely large sample sizes would have allowed the study to catch small effect-size correlations but didn't, suggesting an absence of gender differences. Zimmerman & Martinez-Pons, (1990) studying 90, 5-12 year old children's self-regulation ability in mathematics and English found that Girls displayed more monitoring, goal-setting and planning and lower self-efficacy than boys. However, given they didn't have access to academic results, there is no way to match the self-efficacy scores with performance to create true MEVAL measures.

More recent studies have found evidence for the gender differences in metacognition. Liliana & Lavinia, (2011) used the metacognitive awareness inventory to study 91, 8th grade children and found that girls demonstrated metacognitive knowledge, higher planning and monitoring. Bidjerano, (2005) administered questionnaires to 198 university students and found female students reported higher use of metacognition, self-monitoring, goal setting and planning. However, all studies used self-report tools and needs to be interpreted cautiously as the observed gender differences could be the result of stereotypical biases.

2.2.3 General Cognitive Ability

General Cognitive Ability (GCA) has been traditionally conceptualized in different ways including intelligence and IQ. The thesis uses Elshout, (1983)'s conceptualization of GCA, a domain-generic cognitive toolbox responsible for basic cognitive operations and to be developmentally dependent on environmental factors and genetics. GCA is a well-known academic attainment predictor (Roca *et al.*, 2010), and is correlated with metacognition (Veenman *et al.*, 2004). Sternberg, (1990), initially conceptualized metacognition to be a core component of intelligence though, there is a constant ongoing debate on if the two are integrally associated, independent or just associated but separable. The metacognition domain-generality-specificity debate (section 2.5) also informs the GCA-metacognition debate as metacognition cannot be an integral part of GCA if found to be domain-specific (Veenman *et al.*, 2004).

Veenman *et al.*, (1997) proposed three alternative models for the metacognition-GCA relation:

(a) Metacognition as an integral part of GCA. Several studies (including Elshout & Veenman, 1992; Zimmerman & Martinez-Pons, 1990) report either learning performance to be independent of metacognition when IQ levels are controlled for or differences in metacognitive levels between intellectually gifted and average students. However many times the correlations with performance weren't reported (Alexander, Carr, & Schwanenflugel, 1995; Veenman *et al.*, 2004).

(b) Metacognition and GCA as independent constructs. Allon, Gutkin, & Bruning, (1994) and Swanson, Christie, & Rubadeau, (1993), using questionnaires and problem-solving tasks, respectively, reported low correlations and only partial dependence between metacognition and GCA.

(c) A mixed model where the metacognition and GCA are related but separable implying that metacognition will have an additional predictive value on performance once GCA is controlled for. This is the generally accepted model and it has been suggested that "IQ mediates metacognition, but does not explain it" (Berger & Reid, 1989). More than 8 studies by Veenman's lab using simulation labs, text comprehensions and problem-solving have provided evidence for the same (Veenman *et al.*, 2004). This has been further supported by developmental studies; Alexander *et al.*, (1995), conclude that there is a monotonic developmental relation between metacognition and GCA whereby both develop alongside.

Alexander *et al.*, (1995) and Eme, Puustinen, & Coutelet, (2006) using 'gifted' or 'skilled' and 'unskilled' students found that there is potentially a differential association where MCON, and not MMON, correlates with GCA.

2.2.4 Socio-Economic Status

There is a limited understanding of the role of SES on the metacognition developmental process. Blair & Raver, (2016), recently highlighted the need for a better understanding of the developmental process to inform the development of school-, home- and community-based interventions to improve cognitive development and reverse any cognitive developmental delays.

It is generally considered that students from low SES backgrounds demonstrate lower metacognition (Hall, Bartlett, & Hughes, 1988; Pappas, Ginsburg, & Jiang, 2003; Thompson & Williams, 2006), lower EFs (Blair *et al.*, 2011; Mezzacappa, 2004) and poorer academic attainment (Raver, Smith-Donald, Hayes, & Jones, 2005). Developmental science research has corroborated the same with correlational evidence that children with poorer SES have smaller amounts of grey matter in the regions associated with metacognition and EF (prefrontal cortex and parietal lobes) at ages 4 (Hanson *et al.*, 2013), 10 (Luby *et al.*, 2013) and 11 (Lawson, Duda, Avants, Wu, & Farah, 2013). However, these studies were correlation and more research is required for causal understanding.

However, Jordan, (1994), demonstrated that the performance of students with poorer SES performance on non-verbal tasks is similar to that of other groups of children (Jordan, Huttenlocher, & Levine, 1992), implying that the students with poorer SES may have the relevant skills and knowledge and cannot demonstrate them on verbal tasks due to poorer verbal abilities. Further comparative studies using non-verbal tasks are required for a better understanding of the influence of SES on metacognition.

2.2.5 Summary

Age and GCA are generally considered to be related with and affecting metacognition development while there is mixed evidence for the role of gender difference. There is a very limited understanding of the role of SES, although potentially quite none upon using non-verbal tasks.

2.3 Metacognition measurement tools

Methodological limitations have posed to be a significant challenge in the field; Veenman, Hout-Wolters, & Afflerbach, (2006), in a major review emphasized the need for comprehensive, age-appropriate measures. Over the past few years, children specific tasks have been developed however they use very different metrics to assess metacognition as compared to tasks for adults not allowing for any comparative or developmental studies across age-groups.

Veenman (2005) and Sperling, Howard, Miller, & Murphy, (2002), have reviewed various commonly used tasks in literature, to measure different components for different age groups. Each of these methods have several pros and cons, discussed below, and provide basis for the design and development of the novel task.

2.3.1 Questionnaires and interviews (verbal self-report)

Several questionnaires based on Likert-scales have been developed that assess general metacognition, domain-specific metacognition for reading and mathematics (Cross & Paris, 1988; Kramarski & Mevarech, 2003; Schraw & Dennison, 1994) and have the advantage of large-scale administration. However, individual biases in self-report measures, especially in children, limits the validity of the measure. Veenman, (2005), found that scores on questionnaires didn’t correspond with behavioral measures. Additionally children, and even adults, are prone to making errors while reporting on their metacognitive processes due to metacognition’s nonconscious nature (Nisbett, 1977). Interview and verbal self-report have been further criticized by Whitebread *et al.*, (2009), due their over reliance on students' verbal ability and language fluency.

2.3.2 Prospective and retrospective judgements

Prospective and retrospective judgements are widely used to assess MEVAL. The use of the off-line measurements with children has been challenged and are considered less reliable than online measures due to the disruption they cause, attitudinal/individual biases and variance in the scores with remoteness from the task (Sperling *et al.*, 2002; Veenman, 2005; Veenman *et al.*, 2006).

2.3.3 Think-aloud strategies

Schraw & Moshman, (1995), strongly recommend the use of think-aloud strategies, especially for studies in children, as they allow researchers to observe unobservable and nonconscious aspects of the thinking process allowing for a better and deeper understanding of the metacognitive processes. However, think-aloud strategies are once again limited by verbal fluency and in children lead to a significant increase in cognitive load, conflating and distorting task performance (Whitebread *et al.*, 2009).

2.3.4 Observations

Whitebread *et al.*, (2009) and Winne & Perry, (2000) propose that observational checklists can be conducted in naturalistic environment and provide an advantage by studying non-verbal behavior and recording what children do rather than their perception or memory of it. However, the time-intensive nature of the tool limits the scale of the study and statistical rigor of quantitative studies.

2.3.5 Observational problem-solving

Whitebread *et al.*, (2009) and Pino-Pasternak *et al.*, (2010), developed an observational problem-solving task that allowed for on-line coding of metacognitive skills (MCON and MMON). This provides greater advantage as compared to observation checklists, albeit conducted in quasi-lab environments. Kramarski & Mevarech, (2003), using a similar philosophy, developed observational tasks in which students graphically representing their conceptual understanding. However, observational problem-solving involves some subjectivity and restricts the scale of the study due to the requirement of physical presence of the researcher or video recording analyses.

2.3.6 Challenges to metacognitive assessment in children

Metacognition assessment is challenging and it is even more so with children for the following six reasons. First, metacognition is a complex construct made up of multiple components and many a times what is measured isn't a reflection of the construct itself (tasks commonly have poor construct validity). Second, the over-reliance on verbal proficiency leads to under-estimation in children that are still learning the language. Third, measurement tools are confounded by pre-requisites, like literacy and working-memory (required for processing instructions), that are still developing in children and haven't reached adult-levels. Fourth, existing measures, due to the complex metacognition nature, tend to narrow-down and test only a single metacognitive component rather than the whole metacognition construct leading to poor content validity. Fifth, due to the highly interrelated nature of metacognitive components, the different components work in tandem leading to a task impurity problem whereby several components contribute to the measurement of the other components. Sixth, many tasks are conducted in laboratory environments and children can perform very differently in non-naturalistic environments.

Veenman *et al.*, (2006), suggest the use of multi-method designs in order to better understand the disparity of the assessment methods though this pragmatically makes studies far less feasible, especially studies that aim to build theories pertaining to MC's association with other cognitive skills. Alternatively there have been ongoing efforts to develop tasks that holistically measure the multiple metacognitive components.

There are no age-appropriate tasks that holistically measure the various metacognitive components for 8- to 11-year-old; tasks meant for younger children and adults lead to floor- and ceiling-effects, respectively. Self-report and behavior rating scales are highly biased, observational measures limit the study size and commonly used metamemory-tasks only allow for MEVAL measures. Additionally newer observational tasks for children do not use similar metrics for metacognition measurement as adult-version of the tasks do leading to growing divide in the literature. The thesis aims to address this gap in the field by testing the reliability and validity of a novel task that is age-appropriate, scalable, measures multiple metacognitive regulatory skills and similar to tasks used with adults.

2.4 Developing novel tasks

Designing novel psychometric tasks is a challenging process as the instruments quantitatively measure relatively complex, at time unobservable, phenomena. It is essential to establish both, reliability and test validity, of novel tasks before they can be used for further research to ensure that the tasks operationalize the construct that is being measured. Tasks that aren't reliable and valid can lead to erroneous results and a typical 'garbage in, garbage out' situation (Field, 2013). Test/task reliability refers to the consistency of the results produced by a task while test validity measures the extent to which the construct operationalization reflects the concept that the construct is testing for. *Standards for educational and psychological testing*, (1999) define test validity as "the degree to which evidence and theory support the interpretations of test scores".

2.4.1 Reliability

All or a combination of the following three reliability are generally recommended to be carried out depending on the task. First, internal consistency- tests if the various items on a given scale or (sub)task measure the same underlying construct. Second, inter-rater reliability- for tasks that have some subjectivity in their coding like observational or interview, scores from two different raters need to be compared to see if different people rate the performance similarly. Third, test-retest reliability- to test if the same people get similar scores when tested on more than one occasion (Field, 2013). Cronbach's alpha (Cronbach, 1951) is used to test the correlation between performance on various items, scores from different raters or over time. A Cronbach's alpha greater than a critical value of 0.7 suggests acceptable reliability (Kline, 1999).

2.4.2 Validity

There are three types of test validity and it is recommended that, where possible, all three be conducted.

Construct validity tests the extent to which the operationalization of a given construct tests the construct. This is usually done using either convergent validity or divergent validity by comparing the association of the outcome-measure with parameters that are theoretically known to be associated with it and those that don't, respectively (Cronbach & Meehl, 1955). This thesis tests convergent validity using demographic variables.

Content validity is a non-statistical type of validity that tests the extent to which items on the task represent the complete construct being measured (Anastasi, 1997). Foxcroft, Paterson, LeRoux, & Herbst, (2004) recommend that content validity can be established by experts reviewing the task and its various items. Due to the complex nature of metacognitive tasks very few tasks attempt to provide a complete reflection of the whole construct (metacognitive knowledge and skills).

Criterion validity compares the extent to which the outcome-measures from the novel task matches those from a previously established task. Criterion validity can be subdivided into concurrent and predictive validity testing associations between the outcomes of two tasks at a given time-point and or predictive relations across multiple time-points, respectively (Anastasi, 1997). This thesis studies concurrent validity by comparing the novel problem-solving task with a metamemory-task.

2.5 Metacognition Domain-generality/specificity

The comparison of metacognition across two different tasks, for concurrent validity analysis, draws in a wider debate of domain-generality/specificity of metacognition. There is mixed evidence; Tobias & Everson, (2002), Schraw, Dunkle, Bendixen, & Roedel, (1995), Schraw & Nietfeld, (1998) and Van der Stel & Veenman, (2008), found metacognition to be domain-general while Glaser, Schauble, Raghavan, & Zeitz, (1992), Kelemen, Frost, & Weaver, (2000), Scott & Berman, (2013) and Veenman & Spaans, (2005), found evidence for domain-specificity. The following subsections outline evidence for domain-generality/specificity of metacognitive components.

2.5.1 Metacognitive Control and Metacognitive Monitoring Domain-generality/ specificity

Veenman *et al.*, (1997) and Veenman & Verheij, (2003) studied university students using think-aloud strategies and found MCON and MMON to be domain-general. Veenman *et al.*, 2004, studying 9- to 12-year olds using discovery–learning non-academic tasks found evidence for domain-generality of MCON. Though the study was not corroborated by Veenman & Spaans, (2005), in 12- and 15-year olds. They used academic problem-solving tasks and found domain-specificity in 12-year olds and domain-generality in 15-year olds. The differences in the findings could be explained by the later onset of metacognition development in academic settings versus non-academic settings (Alexander *et al.*, 1995; D. Kuhn, 1999). Van der Stel & Veenman, (2008) using think-aloud strategies in 12- year olds found a two-factor solution supporting the presence of both, a major domain-general and a minor domain-specific factor.

2.5.2 Metacognitive Evaluation Domain-generality/ specificity

Schraw *et al.*, (1995) tested university students in eight non-academic domains (general knowledge, comprehension, spatial and problem-solving) and found metacognitive knowledge and MEVAL were domain-general once the variability in task difficulty was controlled for.

Contradictorily, Kelemen *et al.*, (2000), found evidence of domain-specificity of MEVAL in university students in non-academic domains (Swahili-English word pairs, general knowledge and comprehension). They found an 8% correlation between tasks for MEVAL (studied by Ease Of Learning, JOK, Judgement Of Learning and Judgement of Comprehension) suggesting MEVAL is domain-specific. Scott & Berman, (2013) used a 30 item metacognitive questionnaire on 644 university students and found evidence of domain-generality of metacognitive knowledge and MCON and domain-septicity of MEVAL.

2.5.3 Summary

Findings have been inconclusive due to differences in conceptualization, measurement-tasks (observational, think-aloud, questionnaires), sampling and controls used. Metacognitive Knowledge, MCON and MMON develop in a domain-specific manner (information encapsulation theory) before they become more transferable (Veenman & Spaans, 2005) while MEVAL may potentially follow a similar trend or remain domain-specific. The domain-generalization transition seem to happen at different times for academic and non-academic tasks. This suggests that the non-academic nature of the two task used in this study would allow for validity analysis.

2.6 Metacognition-Executive function associations

The thesis also aims to use the novel task to study the association between metacognition and EF. Metacognition and EF are considered to be synergetic with overlapping functions, albeit separable. Additionally, several studies have suggested EF to be precursors for metacognitive development (Bryce & Whitebread, 2012) though methodological limitations have limited most of the studies to adults.

McCloskey & Perkins, (2012), identified 32 EF subtypes though there is consensus for three core EFs: (a) working-memory, involved in holding, processing and manipulating information; (b) inhibitory-control, involved in self-control and selective attention; and (c) cognitive-flexibility, involved in task switching, adapting to novel situations and creative thinking (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Miyake *et al.*, 2000).

Fernandez-Duque *et al.* (2000) explored synergy between the Nelson and Narens', (1990), metacognition framework and Norman & Shallice's, (1986), EF framework, whereby both cognitive functions have a reciprocal relation of control and feedback between the object and the meta-levels. Following this several other papers have found metacognition-EF associations.

Garner, (2009), using metacognition self-report questionnaires with undergraduate students found that EF predicted metacognition. However, the paper has been criticized for the use of self-report questionnaires and the use of single measures per cognitive ability. Roebers *et al*., (2012), *via* a longitudinal study using novel spelling-based task to measure MCON and MMON and three EF tasks found EF (inhibitory-control and cognitive-flexibility) predicted latent MCON and not latent MMON. However, most of the tasks (three of the four) relied on literacy levels. Bryce *et al.* (2015), using observational metacognition problem-solving, metacognition metamemory-tasks and EF tasks found that at age 5 (and not at age 7) working-memory predicted MCON and inhibitory-control predicted MMON processes. However, their observational metacognition task limited the study size and didn't allow for regression analysis.

The various studies have concluded that metacognition and EF share multiple processes and are associated, although separable.

2.7 Research questions

This thesis uses secondary data from seven schools in Eastern USA of 182 economically disadvantaged, ethnic-minority, 8- to 11-years-old students. It aims to contribute to the development of an age-appropriate, scalable metacognitive problem-solving task by studying the four key research questions.

What are the cognitive levels and trends in the given sample? What are the levels of GCA, metacognition and EF in the sample and how do the cognitive levels develop as children age? Several studies have suggested delayed cognitive development in low SES students (Hall *et al*., 1988; Mezzacappa, 2004; Thompson & Williams, 2006); however very little is known about cognitive development in the novel sampling context (Blair & Raver, 2016) this exploratory question doesn't have any specific hypothesis. Across task comparisons will allow testing for the domain-generality/specificity of metacognitive skills operationalized by the novel task, despite the mixed evidence (Tobias & Everson, 2002; Van der Stel & Veenman, 2008; Veenman & Spaans, 2005) the thesis hypothesis metacognition to be domain-general.

Is the novel metacognitive problem-solving task internally consistent and reliable? Do the various levels of the task test the same problem-solving and metacognitive skills? The thesis hypothesizes high internal consistency across the various levels as theoretically the different levels should be testing the same skill and high inter-rater reliability because of the objective nature of the coding scheme (Field, 2013).

Is the novel task valid (convergent construct validity and concurrent criterion validity)? The novel task data were compared with demographic and classical metamemory-task data.

Demographic data- how does the novel task relate with the known demographic variables? The literature provides evidence for association of metacognition with age (Veenman *et al.,* 2004; Whitebread *et al*., 2009), GCA (Alexander *et al*., 1995; Veenman *et al*., 2004) and mixed evidence for association with gender (Sperling *et al.,* 2002; Zimmerman & Martinez-Pons, 1990). The thesis hypothesizes the novel task to be related with age and GCA however doesn't make a specific hypothesis for association with gender.

Classical metamemory-task- do the problem-solving and metamemory-tasks tap into the same metacognitive components and how do the components behave across the tasks? The thesis, like Veenman & Spaans, (2005), hypothesizes that at the current age MCON, MMON (and MEVAL) to be domain-general skills for non-academic domains that would be tapped into by both the tasks.

What is the nature of the relation between EF and MC? Does a complex EF task, which taps into multiple EF subtypes, predict the various metacognitive measures? The comparative study of the two tasks will provide a basis for the future use of the task to study metacognition-EF association. Bryce *et al.,* 2015; Garner, 2009 and Roebers *et al.,* 2012 have found EF to predict metacognitive components including MCON and MMON.

Chapter 3 Methods

The chapter initially discusses the philosophical perspectives and its influences on the methodology before discussing the ethical and contextual considerations. It then elaborates on data-collection, processing and analyses procedures used.

3.1 Philosophical and methodological perspectives

The thesis is a basic research that aims to develop a new tool for comprehensive assessment of metacognitive skills in children aged 8- to 11-year-old. It adopts an objectivistic epistemology, post-positivist theoretical perspective, correlation-regression research design and secondary data analysis (quantitative) methodology.

3.1.1 Theoretical perspectives

Tashakkori & Teddlie, (2010), suggest that the philosophical perspectives (and research design) should be determined by the research aims. As the thesis aims to test a novel task it adopts a post-positivist approach which allows for quantitative analysis and generalizability of the findings. Furthermore, the researcher had the unique opportunity to use a large-scale quantitative database with more than 120,000,000 data-points, which has been essential for an in-depth analysis and allowed for controlling for various demographic variables though it limited the researcher to a post-positivistic theoretical perspective.

Post-positivism perspective lies between positivism and interpretivism and aims to use quantitative tools to test and build theories about the world. As opposed to the positivism perspective, it adopts the world view that if there is a single objective truth it cannot be found and only aims to develop possible models of a given phenomenon using evidence and rational consideration while still trying to be generalizable (Creswell, 2013).

The quantitative findings (cognitive levels in the sample, novel task reliability and validity, metacognition-EF associations) are a model of the possible truth and do not represent the absolute truth however the study aims to be generalizable in similar sampling contexts.

3.1.2 Research design

The thesis adopts a quasi-experimental, correlation-regression research design. Correlation and regression research designs, embedded in post-positivistic perspectives, use quantitative analyses to study interactions between variables (Creswell, 2013). Although, unlike true experimental designs, they cannot study causal links, they have high ecological validity (essential for theory generation) due to researchers' limited influence on the dependent variables (Field, 2013). The quasi-experimental nature of the study arises from experimental manipulation within a given psychological task resulting in a higher internal validity than pure non-experimental, correlational methods, which rely purely on observations and do not manipulate any independent variables (Field, 2013).

The regression design allows the thesis to study predictive relations between the novel task measures and (a) demographics; (b) classical-metacognition task measures; and (c) EF measures. The quasi-experimental nature of the tasks have allowed for sub-dividing the measures into factors that test different aspects of a cognitive process.

3.1.3 Methodology

The thesis adopts secondary data analyses methodology to analyze pre-collected data and has the significant advantage of availability of large amounts of high quality data collected by trained professionals using standardized protocols resulting in high reliability and validity (Boslaugh, 2007). However, secondary data analysis methodology limits: (a) the analyses to pre-collected data; (b) the depth of understanding of the social phenomena due to reduction of complex data into numeric forms; and (c) the flexibility in terms of how constructs are measured (Smith, 2008).

Smith (2008), recommends making two considerations while using secondary data. First, using predesigned research design, research questions and research hypothesis in order to avoid data mining. Second, aligning the secondary study's philosophical perspectives with those of the primary study due to their strong influences on research design and data analysis protocols. Both, the primary project and this thesis adopt a post-positivistic theoretical perspective with similar aims i.e. to study the development of cognitive skills and relations between cognitive skills. The primary project was longitudinal project that aimed to evaluate a chess intervention to bring about cognitive development. The thesis uses data only from the initial time-point and in the current context could be considered quasi-experimental in nature.

3.2 Ethics

The primary project for this given time-point received ethical approvals from multiple institutions: University of Cambridge's Psychology Research Ethics Committee, its Institutional Review Board (2011.39), Virginia State University's Institutional Review Board (1011-37) and Virginia Commonwealth University's Institutional Review Board (HM20000017). The project was registered under the name of "Malleability of Executive Control" in 2011 and thereafter received repeated yearly renewals over the course of the longitudinal project.

The project received informed consent from both the children and the parents by having them fill an opt-in form, which summarized the project, data confidentiality, use of the data and parents' and children's right to withdraw from the study at any point of time without any consequences. The primary project ensured data confidentiality and ensured that no data could be traced back to the students. The database uses code numbers instead of children names and doesn't name the research assistants involved in data collection. Most of the data were collected using a secure online website and the website doesn't store personal details while any paper based assessments were stored in locked cabinets in the PI/co-PIs offices and shredded once entered.

The primary researchers have set authorship and data use guidelines that hold the secondary researcher to the same ethical standards (the completed data request form is attached in Appendix D). Additionally the guidelines also require any data to be presented in an aggregate form to avoid it being traced without any indication of any schools that the data has been collected from.

3.3 Sample and context

The primary project collected data on 8- to 11-year-old, ethnic-minority students in eastern USA. The data were collected during or after school hours from seven schools in high poverty urban areas from a single school district. The schools were selected using convenience sampling based on the co-PIs' locations and the presence of urban high-poverty ethnic-minority students. A total of 182 children were sampled with a mean age of 9.30 years (SD 0.78 years). A majority (90%) of the sampled students received free or reduced-cost lunch, indicating lower SES. 86% of the students were African-American, 8% Latin-American students, 1% Asian, 3% white and 2% had mixed ethnic backgrounds.

The primary project selected the current context to further understand the cognitive development process given the gap in the literature and potential of maximizing social change with smaller interventions. Due to this sampling bias care needs to be drawing broader implications to the larger student population. However, the sample doesn't pose a threat but rather provides an advantage to the primary goal of the current study because if the task is found to be age-appropriate in these contexts they could also be used in other contexts for the same age groups and even lower age groups as the analyses prevent ceiling-effects.

The sampling context also had further implications for the methods used in the study; the various tasks and tests had to be designed such that they wouldn't be limited by verbal abilities and would test non-verbal abilities of the children. Additionally, the sampling bias doesn't allow enough statistical power to study the effect of SES or ethnicity on metacognition.

3.4 Data collection methods

The primary-project researchers collected data on a multitude of parameters including metacognition (two tasks), EF (six tasks), GCA (one task) and academic attainment (three assessments) either *via* a secure site (thinking game; http://instructlab.educ.cam.ac.uk/thinkinggames/) to allow for large-scale testing or *via* paper-based tasks. Three of the eight cognitive tasks, zoo, memory and disc, have been used for the thesis to focus on the primary goal of the study; to test a novel task and to provide basis for future studies on metacognition-EF association.

The selected cognitive tasks involved a practice-trial with students receiving assistance from research assistants, which were followed by unassisted test-trials. Students were instructed to try and solve problems as quickly as possible while being accurate. The outcome-measures generated by the zoo and memory-task included accuracy, real-time and metacognitive measures while the disc-task generated accuracy and real-time data.

Cognitive tasks suffer from task impurity problems whereby measures of a single cognitive subtype/component are influenced by levels of other subtypes/components. A latent variable (multi-method design) approach, entailing the administration of atleast three tasks per component, is generally recommended in order to partially alleviate task impurity (Miyake *et al.*, 2000; Veenman *et al.*, 2006). However, due to the scale of the primary project, limited amount of time with students and in-depth measurements of other parameters including academic attainment and GCA, additional metacognition tasks weren't feasible.

3.4.1 Novel metacognitive problem-solving task

The primary researchers developed a novel task that requires planning, algorithmic thinking and measures problem-solving efficiency, MCON, MMON and MEVAL. This novel metacognitive problem-solving task is an adaptation of errands task, designed for adults. The errands task is a shopping task that involves finding the best route to shop for multiple items while completing necessary pre-requisites like withdrawing money and filling petrol before going shopping (HayesRoth & HayesRoth, 1979). Radziszewska & Rogoff, (1988, 1991), adapted the task for children supported with adult guidance and Evans, Chua, McKenna, & Wilson, (1997) further adapted the task for use without adult assistance; removed prerequisites, item availability from multiple locations and turned it to a more 'attractive setting' of a zoo.

The current task could potentially allow for bridging metacognition research in children and adults as it is very similar to the errands task, operationalizes problem-solving efficiency similarly, uses judgement outcome-measures and prevents ceiling-effects (section 3.5.4).

The task was administered as a paper-pencil task and students were instructed to draw a map of the shortest route to 'feed' a given set of animals in a zoo (figure 2 provides a sample of the zoo map used). They were instructed to try and solve the task as quickly as possible. The task has three test-trials with increasing difficulty, based on the number of animals that needed to be fed (4, 6 or 8 animals). Before and after every trial prospective and retrospective judgements of their performance on the given trial were recorded on a 5-point Likert-scale of "definitely no" to "definitely yes" in response to the questions "will you be able to find the best way through the zoo?" and "did you find the best way through the zoo?", respectively.

The primary study generated JOK, RCJ, accuracy (based on instances of metacognition) and real-time outcome-measures. An alternative coding scheme has been developed for the thesis and it included three additional outcome-measures: (a) MCON- measures instances of seeking strategies, changes in strategies (corresponding with persistence of error) and efficiency of the order of animals visited; (b) MMON- measures instances of error detection and following given instructions (starting at "start" and ending at "finish", visiting all required animals, sticking to paths between cages and not backtracking); and (c) accuracy measure- compares students' absolute path-length with the shortest path while factoring in the number of animals visited and their relative weightage of the animals on the basis of their distance from the start point.

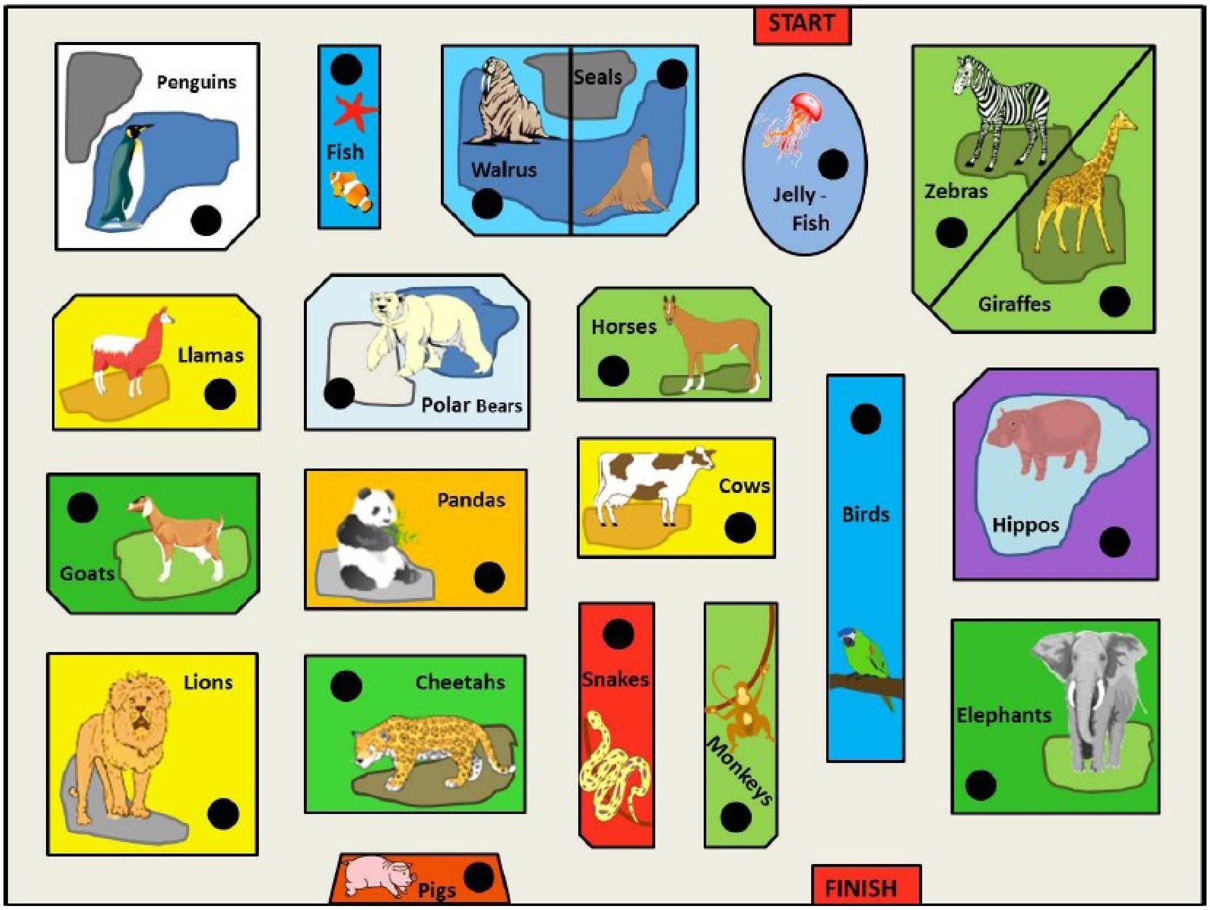

Figure 2: sample zoo map. *©2013 University of Cambridge (Michelle Ellefson, Zewelanji Serpell, Teresa Parr), used with permission from the Mind Match Chess Project. Zoo entrances are marked by a black dot. Students are required to start at "start", visit a set number of animal cage entrances and finish at "finish".*

3.4.2 General Cognitive Ability task

Raven's Standard Progressive Matrix has been used to test GCA. Raven's task tests non-verbal intelligence and is considered fairer to students with language and learning difficulty. It tests for one's ability think clearly, understand complexity, store and reproduce information. Students are administered 60 questions split into 5 sets and instructed to find missing pieces that fit geometric designs. Raven’s task has a high test-retest reliability, content validity and construct validity (Raven, 1938).

The GCA task results have been used to study the zoo task’s criterion validity.

3.4.3 Metamemory-task

The standard metamemory-task measures students’ ability to recall information and their prospective and retrospective judgements of their memory. Previously, the task has widely been used to study MEVAL and delineate predictive relations between MEVAL and academic attainment (Everson, Smodlaka, & Tobias, 1994; Tobias & Everson, 2000). Several version of the tasks are used depending on the age of the participants including showing pairs of words, words in different languages, faces and pictures.

The task used pairs of pictures that children commonly encounter like furniture items, animals, vegetables and cutlery. The task was administered *via* http://instructlab.educ.cam.ac.uk/thinkinggames/. Various pairs of pictures were auto-generated using a random combination of available set of pictures (figure 3 depicts some of the pairs of images). The task was splits into three phases: learn, break and recall. In the learn phase they were shown 24 pairs of pictures for 5 seconds each. For each pair of images, students made prospective judgements of their likelihood to remember the pair of images on a 5-point Likert-scale of “definitely not” to “definitely yes”. This was followed by a 3-minute break phase. In the recall phase they were shown 24 pairs of pictures and were asked to note if they had seen the exact pair before in the learn phase and their retrospective judgement of if they had correctly answered the question on a 5-point Likert-scale of “definitely not” to “definitely yes”. The recall phase has three task-sets (referred to as set through the thesis) of pairs of pictures: (set 1) 12 pairs of pictures that were matched in the learn phase; (set 2) 8 pairs of pictures with pictures that weren't shown in the learn phase; and (set 3) 4 pairs of pictures with one new picture and one picture that they have seen before. Another set (set 4), although not used in this task due to its age-inappropriateness, shows pictures that have been previously seen but weren't paired together.

The memory task is used to study the novel task's concurrent validity and to compare the metacognition operationalization across the two tasks.

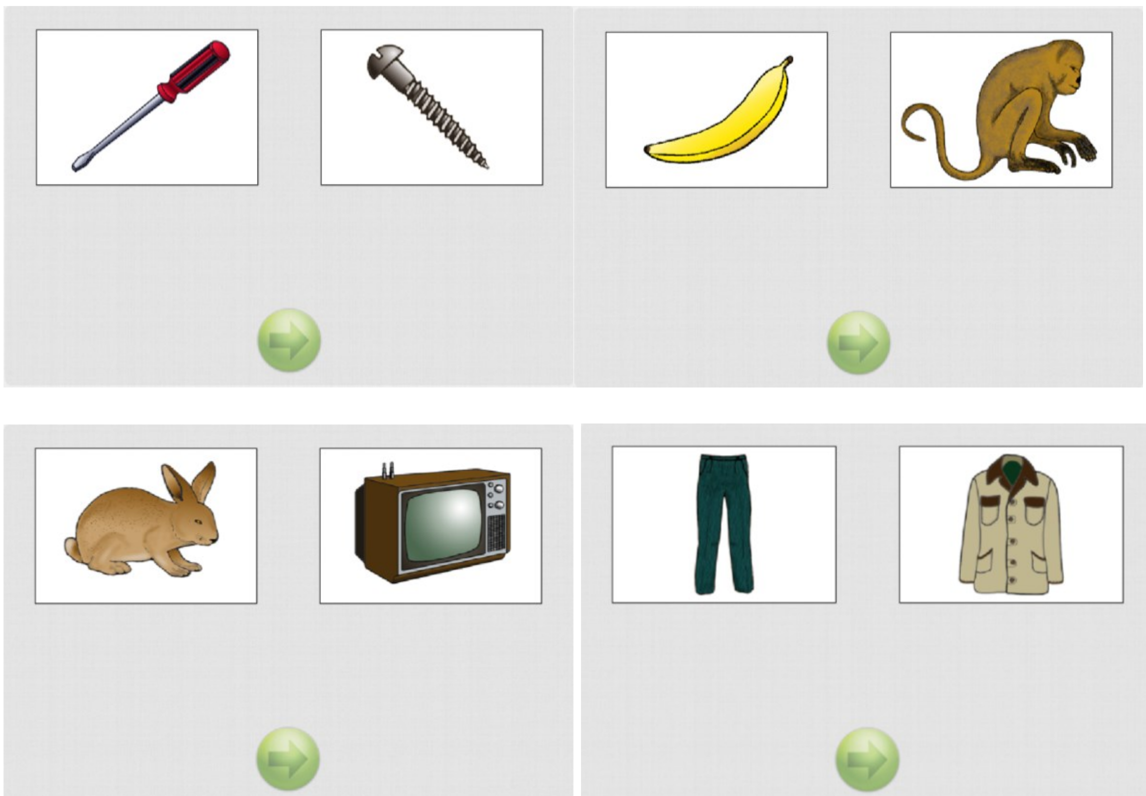

Figure 3: sample memory-task image representation. *©2012 University of Cambridge (Michelle Ellefson, Zewelanji Serpell, Teresa Parr), used with permission from the Mind Match Chess Project; pictures taken from http://instructlab.educ.cam.ac.uk/thinkinggames/. Pairs of image like the above were presented during the memory-task.*

3.4.4 Complex Executive function-problem-solving task

The discs challenge is a commonly used problem-solving task and an adaptation of Tower of Hanoii (Welsh, 1991). It measures the planning and problem-solving abilities and requires multiple EF subtypes including working-memory, cognitive-flexibility (Huizinga, Dolan, & van der Molen, 2006) and perhaps inhibitory-control (Bull, Espy, & Senn, 2004 provided evidence for the same in a similar task called Tower of London). This task has been widely used in cognitive psychology to study EF in cognitively impaired, lesion-studies and school-going children (Goel & Grafman, 1995; Huizinga *et al.*, 2006).

The task was administered *via* http://instructlab.educ.cam.ac.uk/thinkinggames/. The students were shown two stacks of discs and were instructed to move the bottom stack of discs to match the stacking pattern in the top stack (figure 4 depicts the eight trials used). They were instructed to plan their attempt and accomplish the goal in the least possible moves (shifting of any discs) without placing any larger disc on smaller discs as quickly as possible. Post a practice trial the students were administered eight test-trials: 3-ring problems, with 2-, 3-, 4-, 5-, 6-, and 7-move solutions and 4-ring problems, with 7-, and 15-move solutions. In order to progress through to the next levels students needed to solve a given level twice on consecutive attempts in the minimum number of moves. The task would end if the students couldn't progress onto the next level on 6 consecutive trials ensuring students reach the top of their ability. Any illegal moves (placing a larger disc on a smaller disc) were disallowed with the disc being replaced in the original location, NOT counted as a move and reminder message, "sorry- you cannot place a big block on top of a smaller one", was displayed for 2,000 milliseconds.

The disc's task is used to study metacognition-EF association in the sampled children.

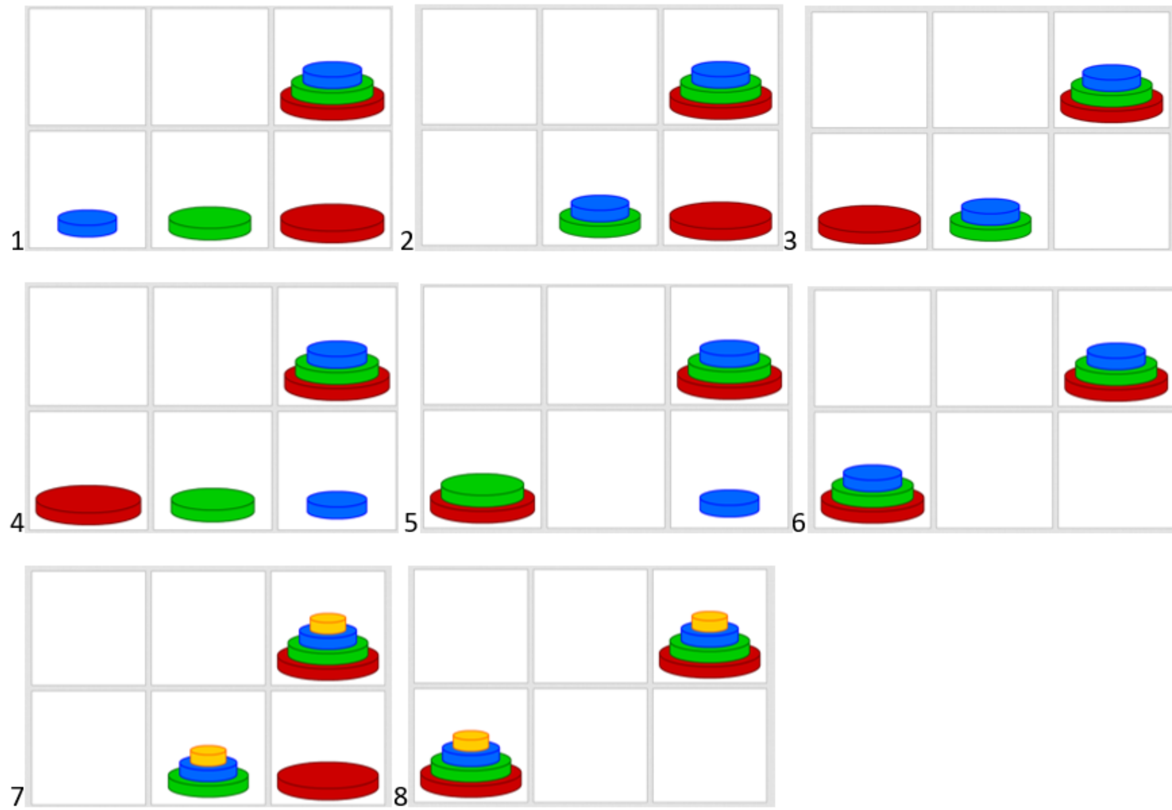

Figure 4: 8 disc-tasks. *©2012 University of Cambridge (Michelle Ellefson, Zewelanji Serpell, Teresa Parr), used with permission from the Mind Match Chess Project; pictures taken from http://instructlab.educ.cam.ac.uk/thinkinggames/. Each level has two rows with multiple discs. The bottom row represents the discs that the students rearrange in order to match the stack of discs in top row.*

Table 4 summarizes the constructs measured, data-collection methods used and outcome variables generated.

Table 4
Data collection methods' summary table.

Variables	Construct measured	Method/ Task name	Resulting outcome variables	Mode of collection
Meta-cognition	Metacognitive problem-solving	Zoo-task	Accuracy, Real-time, Efficiency, Judgements (pre and post), monitoring and control measures	Paper
	Metamemory	Memory Challenge	Accuracy, Real-time, Efficiency, Judgements (pre and post)	Secure site
Executive Functions	Planning; Complex EF task	Discs challenge	Accuracy, Real-time, Efficiency	Secure site
General Cognitive Ability	Nonverbal General Cognitive Ability	Raven's Progressive Matrices	Accuracy and standardized scores	Paper

3.5 Data-management and processing

The primary study has more than 120,000,000 data points of raw-data and the thesis uses a part of it for a cross-sectional analysis. The requested data included trial-by-trial for the aforementioned tasks. The zoo-task required to be recoded (section 3.5.3). The relatively large amount of chosen data required data-processing i.e. processing the raw-data till they are ready for analysis (Bourque & Clark, 1992).

Three statistical software were used to manage and analyze the selected data: (a) MS Excel for Zoo-task data entry, (b) JMP for most of the data management and analysis and (c) SPSS for factor analysis and dimension-reduction.

3.5.1 Inclusion criteria

An inclusion criteria of students completing the zoo-task has been used to fit the thesis's primary objective. Additionally each task uses exclusion criteria for participants based on (a) consent being provided to use the data and (b) the number of attempts at completing the task, irrespective of their success. Fewer attempts on a given task resulted from technical glitches including Wi-Fi connectivity problems, students accidently logging out of the session or computer/program crashes. The memory-task excluded participants if there were less than 12 attempts (out of 24) during the learning phase or less than five attempts during the recall phase. The disc-task excluded participants that quit the task before using all six attempts for level one, or four attempts for levels two or three (of the total 8 difficulty levels). 23 and 18 participants were excluded from the memory and disc-task, respectively.

3.5.2 Missing data

Missing data, commonly found in social science research, when arising because of random causes is usually less problematic (Shaughnessy, Zechmeister, & Zechmeister, 2014). Mean substitution is recommended and has been used for instances where the missing data are less than 10% as (McKnight, McKnight, Sidani, & Figueredo, 2007), over discarding data, which would decrease the statistical power. Mean substitution entails substituting the missing data with the observed mean of other results, however leading to a decrease in the variance in the data, which is primarily compared in regression analyses (Field, 2013).

Mean substitutions has been used for Zoo judgement scores as the missing data arose was quite low due to students either circling multiple options or forgetting to select an option. Though no substitutions were made for missing real-time scores (>25%) for zoo-task. The memory and the disc-tasks, post the exclusion criteria, didn’t have any missing data as the tasks were computer-based.

3.5.3 Re-coding zoo data

The zoo data previously been coded for the following variables on a scale of zero to two: (a) number of required animal that were visited; (b) starting and finishing at the right places; (c) using the instructed dots to visit the animal cages and not walking through the animal cages; (d) avoiding backtracking as had been instructed; (e) order of animals visited and (f) the presence of any strategy including previously marking animals or checking off animals as they have been visited.

MMON coding. MMON variable was recomputed using trial-by-trial data for (a) starting and finishing at the right places; (b) using the instructed dots to visit the animal cages; (c) avoiding backtracking; and (d) order of animals visited. The maximum possible score is 24 points.

MCON coding. MCON variable was recoded using trial-by-trial data for (a) order of animals visited; (b) presence of a strategy; and (c) an additional binary variable corresponding to change in strategy. The maximum possible score is 16 points.

Accuracy measure (further details in Appendix F). In the zoo-task, children drew paths on the map and a new accuracy measure was developed using absolute path-length of the drawn path. An average of 10 minutes was spent coding the paths for every child (the sample contained 182 children with each child attempting 3 trials). The new accuracy measure factors the following: (a) the difference of the student's absolute path-length with the shortest possible path; (b) number of required animals visited; and (c) the weightage of visiting a given animal based on the relative distance from "start" (Equations 3-5).

In order to computerize the calculation of the absolute path-length the zoo map was converted into a grid (Appendix F). Then the paths were physically traced, the list of animal entrances passed by while following the path were used to calculate the absolute path-length. The accuracy measure further processes the absolute path-length using equations 3 to 5. The percentage difference of the path-lengths and accuracy shares an inverse relation hence the percentage difference was subtracted from 100. Subtraction was chosen over using 1/percentage difference in order to avoid an exponential graph. The second half of equation 3 (values range from 0-1) corresponds to the number of required animals visited and their weightage.

Students can achieve the maximum possible score of 100 when their absolute path-length equals the shortest path-length and they visit all the required animals. However, when the path chosen is twice the shortest path-length the first half of the equation results in a negative number. In order to avoid worse accuracy scores with an increase in the number of required animals visited Equation 4 is used.

$$If\ Path\ chosen\ \leq 2 \times Shortest\ Path$$

$$Accuracy = \left[1 - \left(\frac{|Shortest\ Path - Path\ chosen|}{Shortest\ path}\right)\right] \times 100 \times \left[\sum_{i}^{n}(Distance\ of\ cage\ from\ start \times x_i)\right]$$

Equation 3

$$If\ Path\ chosen\ \geq 2 \times Shortest\ Path$$

$$Accuracy = \left[1 - \left(\frac{|Shortest\ Path - Path\ chosen|}{Shortest\ path}\right)\right] \times 100 \times \left[\frac{1}{\sum_{i}^{n}(Distance\ of\ cage\ from\ start \times x_i)}\right]$$

Equation 4

Where i-n represents the required animals the child visits and x_i represents the relative weightage of animal i and can be found from Equation 5.

$$\sum_{i}^{n}(Distance\ of\ cage\ from\ start \times x_i) = 1$$

Equation 5

3.5.4 Computing outcome-measures for analysis

Accuracy tends to be usually non-normally distributed because of ceiling- and floor-effects though efficiency is more normally distributed and is a measure of the accuracy, speed trade-off. Efficiency measures have been computed by dividing the accuracy by real-time data. They have been generated on a trial-by-trial basis before being averaged for a given level (zoo) set (memory) or task (disc).

The standardized judgements along with standardized task accuracy (further information about standardization is provided in the next section) have been used to create AAI and BI for a child using equations 1 and 2 (section 2.1.4; Schraw, 2009).

3.5.5 Testing and correcting data to match statistical assumptions

Normality. Field, (2013), suggests that for regression analyses a normal distribution of the residuals is required rather than that of the data. Multivariate normality has been tested for using the following methods. First, univariate normality was tested as univariate normality implies multivariate normality (Field, 2009; Tabachnick, 2014). Second, B. King & Minium, (2003) Lumley, Diehr, Emerson, & Chen, (2002), theorize that for large sample sizes the samples approximate a normal distribution given the central limit theorem and the relatively large size of the current data allows for approximation of the normal behavior. Third, normality of the variables was tested using Normal-Quantile (also referred to as QQ plots) rather than the Shapiro Wilk's test as the latter is considered to be too stringent for larger data sets (N>100). Variables where most of the data fell in the QQ plot confidence limits were preserved while those that were non-normal due to skew or kurtosis were modified into ranks using the procedure described in the next section.

Additional post-regression analyses tests also found residuals to be normally distributed.

Two-step transformation of non-normal continuous data. Zoo-MCON, Zoo-MMON, Zoo-Efficiency, Zoo Accuracy, Zoo AAI, Memory-RCJ-AAI (factors 1-3) and Memory-RCJ-BI (factor 3) didn't follow normal distribution and have been transformed to create continuous variables. Field, (2009), recommends the use of ranked variables as they preserve the general order of performance although the information on the magnitude of difference between participants is lost. Templeton, (2011), recommends a two-step transformation over Field, (2009)'s, single-step rank-transformation as simple rank transformation leads to high kurtosis. Therefore the non-normal variables were converted into fractional ranks (using SPSS). The uniform distribution of fractional ranks (cumulative frequency) allows for the second step of data transformation into a continuous cumulative frequency density using the inverse DF function with a mean of 0 and standard deviation of 1. The final outcome variable represents a normally distributed rank variable.

Outlier analysis. Outliers are data points that don't follow the pattern of the rest of the data, they are numerically distant and can potentially bias the data. Outliers are frequently found in statistical analysis and especially in education and psychology. Field, (2013), suggests that outliers, if caused by chance or exceptional results, should be adjusted or excluded from the analysis to prevent biases.

Outlier analysis was conducted by analyzing data on box plot and the data outside the whiskers were considered to be outliers and were corrected for using Winsorizing. Winsorizing involves replacing outliers with data-points that aren't outliers (Field, 2013); a mean plus or minus three standard deviation has been used in the thesis. A second confirmatory post-correlation **multivariate outlier analysis** test has been conducted using Mahalanobis distances (Field, 2013) and any data-points with distances greater than 10 would need to be corrected for. Though all distances were less than 7 and the data-points were uniformly distributed suggesting no remaining outliers.

3.5.6 Standardization

Non-transformed variables across the various tasks were converted into standardized z-scores as the different tasks were measured on different scales; the transformed variables already follow a z-distribution. Calculations of the standardized z-scores were done with datasets split by trials in order to maintain the variance across the trial. Standardization prevents within-task mean comparisons, though it allows for between-task regression analyses.

3.5.7 Dimension-reduction and grouping

Exploratory-factor-analysis is commonly used to identify common underlying factors between sets of variables and to decide if the related variables can be combined into a single composite variable while still retaining most of the original variability and information (Field, 2013). Exploratory-factor-analysis was used to test if the following could be combined: three levels in zoo-task, 24 trials on memory-task, eight tasks on disc-task and two judgements for every judgement outcome-measure. Factor analysis has two prerequisites: (a) minimum sample size- Pedhazur & Schmelkin, (1991), recommend a sample size of 100 while the Kaiser-Meyer-Olkin measure allows a better measure for sample size adequacy; and (b) significant correlation- Bartlett's test of sphericity, tests the significance of the multivariate correlation (Field, 2013).

3.6 Main statistical analysis

An initial power analyses (data not included) helped inform decisions on the statistical analyses to use. Power analyses test if the sample size would be large enough to have a recommended power of 80% (Cohen, 1988) and a medium expected effect-size.

3.6.1 Preliminary correlation analysis

A preliminary correlation analysis has been conducted to understand the zoo-task and inform further analyses.

3.6.2 Reliability analysis

Internal consistency. Task reliability was tested using Cronbach's alpha (Cronbach, 1951). Cronbach's alpha explores the properties of items/levels on a task and compares the performance of each item. Values greater than 0.7 suggests acceptable reliability (Kline, 1999). Additionally, it is also suggested that if the removal of any item results in a significant increase in the Cronbach's alpha, then the item/level should be dropped.

Inter-rater reliability. Intra-class correlation coefficient with a cut-off of 'excellent reliability (>0.75; Cicchetti, 1994) has been used rather than kappa in order to study the reliability in continuous data (Hallgren, 2012). Previous researchers have studied reliability of MMON and MCON measures. For the new accuracy measure an independent rater, supported with a code-book (Appendix E), evaluated 10% of the data selected at random using data from every 4th, 14th, 24th… child from a given school. The number four was generated using a random number generator.

3.6.3 Hierarchical-Multiple-Regression analyses

Multiple-regression is commonly used to study predictive relations of multiple independent variables on a single dependent variables (Osborne, 2007). Hierarchical-multiple-regressions involve inserting variables into the regression analyses in a block-wise manner, in a theory-driven predetermined order and allows for controlling for previously known predictors (Field, 2013). Two-step hierarchical-multiple-regressions have been used over other forms of multiple-regression to study the effect of the factors of interest above and beyond the effect of previously well-established factors. Confounding demographic variables (age, GCA and gender) were added in stage one and auto-fit modelling of other predictors (metacognitive measures and other demographic control variables like school and ethnicity) in stage two. School and ethnicity additions were added in stage two due to absence of pre-established relation and purposive-sampling used.

The auto-fit modelling was based on p-value thresholds using mixed modelling which involved addition and removal of predictors with p-values < 0.25 and p-values >0.25, respectively. The auto-fit modelling used the combine rule (for any interaction effects it also adds significant/non-significant main effects involved in the interaction) rather than the restrict rule (prevents addition of interaction effects unless all associated main effects have been added) as there is a possibility of presence of interaction effects in the absence of main effects. The hierarchical-multiple-regressions were followed by power analysis and regression assumption tests (outlined in the next sub-section). The power analyses reveal Cohen's δ effect-sizes, measures of the strength of the relations, as the regression coefficient (unlike correlation coefficients) do not do so. Effect-sizes of .01, .20, .50 and .80 are considered to be very small, small, medium and large, respectively (Cohen, 1988).

Hierarchical-multiple-regression assumptions. Hierarchical-multiple-regression assumes that the data has a normal distribution, linear relation between the independent and dependent variables, lacks outliers, lacks influencers, that there is little multicollinearity and is homoscedastic. The previous data processing steps have corrected the data for violations of normality (*via* two-step rank transformation), outliers (*via* winsorizing) and multicollinearity (*via* Exploratory-factor-analysis). All hierarchical-multiple-regression analyses have been followed by post-tests for: (a) normal distribution of the residuals; (b) linearity of relations using a scatter-plot of residuals by predicted values; (c) absence of multicollinearity using Pearson's r ($r>.8$ imply collinearity; Abu-Bader, 2010) and Variance Influence Factors (VIF/tolerance; VIF>10 imply collinearity; Abu-Bader, 2010); (d) absence of influencers using cook's D plot (influencers normally lie at the extremes of the plot away from the majority of data). And (e) homogeneity of errors using a residual by predicted-value plot.

Construct (convergent) validity; comparing zoo-task variables with demographic variables. Hierarchical-multiple-regression is used to test the predictive nature of the demographic variables on the zoo-task measures. The zoo-task variables were used as dependent variables. While age is expected to positively associated there are no hypothesis for gender and GCA due to the presence of mixed evidence in the literature. There is a possibility of no relation with age if the students hit floor- and ceiling-effects.

Criterion (concurrent) validity; comparing memory-task variables with the zoo-task variables. The predictive nature of the zoo-task on memory-task variables was studied for validity analyses and to better understand the similarities and differences between the two tasks. The memory-task variables were used as dependent variables and the zoo-task variables as independent variables while controlling for demographic variables. Differences could potentially arise due to systematic errors, the nature and complexity of the tasks or the task-specificity of metacognition. However, the thesis hypothesizes metacognition to be domain-generic.

3.6.4 Exploratory hierarchical-multiple-regression exploring metacognition-EF association and problem-solving skills

The zoo-task was compared with a commonly used complex EF, problem-solving task's (Disc-task) to explore the EF-metacognition relation and if the tasks test for the same problem-solving skills (although unlikely given the difference in the nature of the tasks). Hierarchical-multiple-regressions were run using the zoo-task variables as dependent variables and the disc-task efficiencies as independent variables while controlling for demographic variables because previous studies (Bryce *et al.*, 2015; Garner, 2009; Roebers *et al.*, 2012) have proposed and used EF as predictors of metacognition.

3.6.5 Summary of statistical analysis

This study used five different statistical analyses tools to study the research questions: (a) exploratory-factor-analysis for dimension-reduction and production of composite variables and factors; (b) Cronbach's alpha (and exploratory-factor-analysis) to establish zoo-task's internal consistency; (c) intra-class correlation coefficient for inter-rater reliability (d) a preliminary multivariate correlational analysis of the various variables to inform regression analysis and (e) hierarchical-multiple-regression analyses to study the relations between metacognitive skills within a task, across metacognition and EF tasks.

The pre-analysis power analysis showed the study had a large enough sample size to detect medium effect-size associations using the aforementioned analyses tools.

Chapter 4 Results

The chapter outlines the results from the various statistical analyses. Sections ‘data processing results’ and ‘preliminary correlational analyses’ detail preliminary analyses and decisions made that feed into the main analyses targeting the four research questions in sections ‘contextual understanding of the cognitive levels’, ‘reliability studies’, ‘validity studies’ and ‘exploratory analysis of metacognition-EF relation’.

4.1 Data processing results

4.1.1 Data management decisions

Mean substitutions were used for missing Zoo judgement scores (4.9%). However, a significantly large amount of missing zoo real-time data (23.4%) posed a challenge and lead to missing Zoo-Efficiency data. The Zoo-Efficiency data were estimated using Accuracy data due to the presence of a large effect-size correlation ($r = 0.92$, $p<0.0001$). Accuracy is used for within-task analyses (exploratory-factor-analysis and internal consistency) and efficiency is used for between-task comparisons to allow for comparisons with other tasks’ efficiencies.

Normally distributed data once corrected for outliers and standardized, demonstrated low kurtosis and skew. Non-normal variables once corrected by the two-step transformation don’t demonstrate outliers, kurtosis and skewness.

Scatter-plots depicted linearity of most of the variables except Zoo-BI, which was slightly curvilinear. However, it was retained as Tabachnick, (2014) suggest that slight curvilinearity is acceptable and care must be taken while interpreting the results due to a loss in statistical power of models.

4.1.2 Dimension-reduction (*via* exploratory-factor-analysis)

Can the three levels of zoo-task be considered to be one? Exploratory-factor-analysis (using principal-axis factor-analysis) tested zoo-task's internal consistency across the three trials. Post-tests revealed the sample size was adequate as the Kaiser-Meyer-Olkin measure equaled .684 which falls in the mediocre range (.5-.7) (Field, 2013); and significant correlation for factor analysis using Bartlett's test of sphericity ($\chi^2(3)= 109.32$; $p=.00$).

The Eigen values (factor 1=1.94, factor 2=0.53 and factor 3=0.52) and scree plot (Appendix A) reveal a single factor solution explaining 64.7% of the variance with the following factor loadings: Trial 1=.80; Trial 2=.80; and Trial 3=.80.

A new combined variable for the accuracy scores was computed using equation 6. The single-factor loading also allows for the creation of BI and AAI as they are based on a summation of bias and squared biases, respectively, across multiple levels.

$$Zoo\ Efficiency = 0.804 \times Trial\ 1\ Efficiency + 0.802 \times Trial\ 2\ Efficiency + 0.807 \times Trial\ 3\ Efficiency$$

Equation 6

Can the three sets of the memory-task be combined? The memory-task was split into the three sets and accuracy was used to test if the three groups could be combined into a single composite variable. The results revealed insufficient correlation as tested by Bartlett's test of sphericity ($\chi^2(3)= 4.38$; $p=.22$). Therefore the three sets were considered to be separate for the remaining analyses.

Judgement outcome-measures. Exploratory-factor-analysis revealed significant correlations for (a) Zoo-JOK-BI and Zoo-RCJ-BI; and (b) Zoo-JOK-AAI and zoo-RCJ-AAI. And none across any of the memory-set JOK-BI and RCJ-BI or JOK-AAI and RCJ-AAI.

Principal axis factor analysis on the Zoo-BI revealed the sample size to be just adequate as the Kaiser-Meyer-Olkin measure equaled .50 which falls in the mediocre range; and significant correlation for factor analysis using Bartlett's test of sphericity ($\chi^2(1)$= 182.07; p=.00). The Eigen values (factor 1=1.80 and factor 2=0.20) and scree plot (Appendix A) reveal a single factor solution explaining 89.98% of the variance with the following factor loadings: JOK-BI=.95; and RCJ-BI=.95. A new combined variable for the Net BI scores was computed using the equation 7.

$$ZooBI = 0.949 \times ZooJOKBI + 0.949 \times ZooRCJBI$$

Equation 7

Principal axis factor analysis on the zoo absolute accuracy indices revealed the sample size to be adequate as the Kaiser-Meyer-Olkin measure equaled .50 which falls in the mediocre range; and significant correlation for factor analysis using Bartlett's test of sphericity (χ^2 (1)= 182.07; p=.004). The Eigen values (factor 1= 1.21 and factor 2= 0.78) and scree plot (Appendix A) reveal a single factor solution explaining 60.70% of the variance with the following factor loadings: JOK-AAI=.78; and RCJ-AAI=.78. A new combined variable for the Net BI scores was computed using the equation 8.

$$ZooAAI = 0.78 \times ZooJOKAAI + 0.78 \times ZooRCJAAI$$

Equation 8

Can the eight levels on the disc-task be combined? Exploratory-factor-analysis using principal component analysis on the disc-task efficiencies revealed: the sample size to be just adequate as the Kaiser-Meyer-Olkin measure equaled 0.52 which falls in the mediocre range; and significant correlation for factor analysis using Bartlett's test of sphericity ($\chi^2(28)$= 62.70; p=.00). The Eigen values (factor 1=2.53, factor 2=1.48 and factor 3=1.24) and cumulative variance reveal a three-factor solution explaining 65.65% of the variance. Rotated (Varimax) factor loadings are listed in table 5.

Table 5
Rotated factor loading of the eight disc-tasks.

	Disc Factor 1	Disc Factor 2	Disc Factor 3
Task 1	.84	-.19	.23
Task 2	.61	.24	.46
Task 3	.00	.01	.87
Task 4	.87	.18	-.20
Task 5	-.01	.51	.07
Task 6	.19	.54	.56
Task 7	-.04	.85	.15
Task 8	.38	.61	-.27

The factor loading patterns suggest: (a) tasks one, two and four load onto factor 1, (b) tasks five, six, seven and eight load onto factor 2; and (c) tasks three and six load onto factor 3. Combined factor variable were computed using the equations 9-11.

$$Factor\ 1\ Eff = L1\ Eff \times 0.841 + L2\ Eff \times 0.612 + L3\ Eff \times 0.004 + L4\ Eff \times 0.868 + L5\ Eff \times (-0.008) + L6\ Eff \times 0.185 + L7\ Eff \times (-0.043) + L8\ Eff \times 0.376$$

Equation 9

$$Factor\ 2\ Eff = L1\ Eff \times (-0.187) + L2\ Eff \times 0.235 + L3\ Eff \times 0.012 + L4\ Eff \times 0.177 + L5\ Eff \times 0.508 + L6\ Eff \times 0.539 + L7\ Eff \times 0.851 + L8\ Eff \times 0.605$$

Equation 10

$$Factor\ 3\ Eff = L1\ Eff \times 0.234 + L2\ Eff \times 0.463 + L3\ Eff \times 0.866 + L4\ Eff \times (-0.200) + L5\ Eff \times 0.073 + L6\ Eff \times 0.564 + L7\ Eff \times 0.151 + L8\ Eff \times (-0.268)$$

Equation 11

4.1.3 Summary of data processing results

The final outcome variables available post data-processing and dimension-reduction are listed in table 6.

Table 6
Final outcome variable table.

	Zoo-task	Memory-task	Disc
Final Outcome Variables	Accuracy	Accuracy (sets 1, 2 and 3)	Efficiency (Factor 1, Factor 2 and Factor 3)
	Efficiency	Efficiency (sets 1, 2 and 3)	
	MCON	JOK AAI (set 1)	
	MMON	RCJ AAI (sets 1, 2 and 3)	
	BI (net JOK BI and RCJ BI)	JOK-BI (set 1)	
	AAI (net JOK AAI and RCJ AAI)	RCJ-BI (sets 1, 2 and 3)	
Intermediate variables	JOK-BI		
	RCJ-BI		
	JOK-AAI		
	RCJ-AAI		

RCJ-retrospective confidence judgement, JOK-judgement of knowledge, BI-bias index, AAI-absolute accuracy. Since the memory-task has been split into task–sets on the basis of if the exact pair of images was seen before (Set 1), if neither of the two images were seen before (Set 2) and if one image was seen before (Set 3), there are no JOK outcome-measures available for sets 2 and 3.

4.2 Contextual understanding of students' cognitive levels and their development

4.2.1 Sampling context

182 students aged 8- to 11-years fulfilled the inclusion criteria. The students came from seven schools that provided an average of 26 students each (with one school providing a minimum of 15 students). And a majority of the students are ethnic-minority students (African-Americans, Latin-Americans and Asians).

Table 7
Descriptive statistics of demographic variables.

Variable	Statistic	Descriptive statistics
Sample size	N	182
Age (years)	Mean	9.31
	SD	0.78
GCA	Mean	31.05
	SD	9.24
Gender		
	Male	101
	Female	81
Ethnicity		
	African American	154
	Asian	2
	Latino	14
	White	5
	Mixed	4
	N missing	3

GCA-General Cognitive Ability.

4.2.2 Cognitive levels

Table 7, depicts that students demonstrate a mean GCA of 31.05 and a median (representative of 50% of the population) of 32, out of a maximum possible score of 60, and a good spread of data. Table A1 lists the mean and standard deviations of various cognitive task variables.

The zoo, memory and disc-task measures demonstrate high amounts of variability, except the judgement measures, and would allow for the various regression models. Judgement measures of both the metacognitive tasks reveal that students generally tend to be generally (over)confident of their abilities. Median scores are all above 4.5 out of 5 (except for Memory-JOK). Zoo-MCON measures are measured out of 16 points and they remain quite low as compared with Zoo-MMON scores, which are measured out of 24 points.

4.2.3 Correlational analysis on cognitive development

Correlation analysis of various cognitive measures with age revealed very few significant correlations (summarized in table 8) and **no** significant correlations were observed between: (a) GCA and age; (b) any zoo-task variables and age; and (c) disc efficiency and age. Additionally, for significant memory-task associations, increases in age lead to lower efficiency and MEVAL.

Table 8
Significant bivariate correlations with demographic variables.

	Age
Memory-Efficiency^{Set1}	**-.22****
Memory-JOK-BI^{Set1}	**.31*****
Memory-RCJ-AAI^{Set2}	**-.20***
Memory-RCJ-AAI^{Set3}	**-.14***

****p<0.0001* ***p<0.001* **p<0.05. Correlation coefficients in bold represent statistically significant results.*

4.3 Preliminary correlation analysis

Several correlation analyses were run within- and between-tasks to inform reliability and validity analyses.

4.3.1 Within-task variables

Zoo-task variables. The zoo-tasks variables are correlated with each other. The accuracy and efficiency measures have a large effect-size correlation (r=0.91, p<.0001; r of .10, .30 and .50 are considered to be small, medium and large effect-sizes, respectively; Cohen, 1988), which allows for predicting missing efficiency scores using accuracy scores. Accuracy measures have small to medium effect-size correlations with MMON and MEVAL measures. MCON and MMON share a large effect-size correlation (r=.537, p<.0001). However, there should be no multicollinearity upon addition of the various variables in a single model as all the Pearson correlation coefficients are less than .8.

Table 9
Bivariate correlations within zoo-task variables.

	Zoo-Accuracy	Zoo-Efficiency	Zoo-MCON	Zoo-MMON	Zoo-AAI
Zoo-Accuracy					
Zoo-Efficiency	**.91*****				
Zoo-MCON	0.06	0.09			
Zoo-MMON	**.15***	**.15***	**.54*****		
Zoo-AAI	**-.42*****	**-.46*****	0.05	0.05	
Zoo-BI	**-.66*****	**-.59*****	**-.15***	**-.32*****	-0.05

$^{***}p<0.0001$ $^{**}p<0.001$ $^{*}p<0.05$. MCON-metacognitive control, MMON-metacognitive monitoring, BI-bias index, AAI-absolute accuracy. Correlation coefficients in bold represent statistically significant results.

Memory-task variables. The memory-task performance of sets 2 and 3 demonstrate a medium effect-size correlation (r=.42, p<.0001) while neither correlates with set 1. The judgement outcome-measures (both for AAI and BI) also follow a similar pattern with retrospective AAI (r=.58, p<.0001) and BI (r= 0.53, p<0.0001) being correlated across sets 2 and

3 and not with set 1. The performance on a set is negatively correlated (medium effect-size) with the AAI and BI for that set.

Disc-task variables. The disc-task variables demonstrate a very strong correlation between factors 1 and factors 3 ($r=.813$, $p<.0001$) though not with factor 2. However, the three factors need to be considered to be separate and distinct, due to the dimension-reduction results.

4.3.2 Between-task variables

Correlational analysis revealed only two significant correlation between the zoo-task and the other two task: between Memory-RCJ-BI^{Set1} and Zoo-Efficiency; and between Memory-RCJ-BI^{Set3} and zoo-AAI. There are several correlations that are significant at 10% significance-level for the Zoo-MCON and Zoo-MMON with disc and memory factor variables.

Table 10

Bivariate correlations of between cognitive task variables.

	Zoo-Efficiency	Zoo-MCON	Zoo-MMON	Zoo-AAI	Zoo-BI
Disk Factor 1	-.06	.02	**.14**$^{-}$	.04	-.02
Disk Factor 2	.05	**.14**$^{-}$	**.15**$^{-}$	.04	-.10
Disk Factor 3	-.05	.00	**.15**$^{-}$	.03	-.03
Memory-Efficiency^{Set1}	.06	**-.14**$^{-}$	-.13	-.03	-.02
Memory-Efficiency^{Set2}	-.08	-.05	.03	.10	.10
Memory-Efficiency^{Set3}	-.12	**-.14**$^{-}$	-.00	.11	.08
Memory-JOK-AAI^{Set1}	-.11	-.01	-.01	.11	.02
Memory-RCJ-AAI ^{Set1}	.06	**-.15**$^{-}$	-.06	-.02	-.10
Memory-RCJ-AAI^{Set3}	.10	-.01	-.10	-.05	-.05
Memory-RCJ-AAI^{Set3}	.10	.13	.08	.02	-.04
Memory-JOK-BI^{Set1}	.00	.08	.02	-.12	.08
Memory-RCJ-BI^{Set1}	**-.16***	-.06	.01	-.08	.12
Memory-RCJ-BI^{Set2}	-.01	.04	-.05	-.06	.06
Memory-RCJ-BI^{Set3}	-.03	-.04	-.02	**-.22*****	.06

$^{***}p<0.0001$ $^{**}p<0.001$ $^{*}p<0.05$ $^{-}p<0.1$. MCON-metacognitive control, MMON-metacognitive monitoring, MEVAL- metacognitive evaluation, RCJ- retrospective confidence judgement, JOK- judgement of knowledge, BI- bias index, AAI- absolute accuracy. Correlation coefficients in bold represent statistically significant results.

4.4 Reliability studies

4.4.1 Internal consistency

The exploratory-factor-analysis (section 4.1.2) suggested that the three levels in the zoo-task were correlated, tested the same factor and could be combined into one single variable. The correlation matrix (table 11) reveals medium effect-sizes. Internal consistency analysis using Cronbach's alpha found high correlations across the three trials with the Cronbach's alpha of .73, above the commonly used critical value of .7 (Field, 2013). And removal of any trial lead to a reduction in the alpha (table 12).

Table 11
Correlation matrix of zoo-task performance.

	T1 Accuracy	T2 Accuracy
T1 Accuracy		
T2 Accuracy	**.43*** **	
T3 Accuracy	**.47*** **	**.46*** **

$^{***}p<0.0001$ $^{**}p<0.001$ $^{*}p<0.05$. *Correlation coefficients in bold represent statistically significant results.*

Table 12
Cronbach's alpha table.

Levels	Cronbach's alpha
Levels 1, 2 & 3	.73
Levels 2 & 3 (Level 1 excluded)	.64
Levels 1 & 3 (Level 2 excluded)	.64
Levels 1 & 2 (Level 3 excluded)	.64

4.4.2 Inter-rater validity

The reliability for MCON and MMON codes were coded previously by other researchers while the accuracy inter-rater reliability was coded for this thesis (summarized in table 13).

Table 13
Inter-rater reliability table.

Code	ICC (3,2; N=186)	ICC (2,1; N=20)
Raters	Raters 1 and 2	Raters 2 and 3
Clear Path Drawn	.85	
All animals visited (MMON)	.83	
Visiting start and end points (MMON)	.88	
Visiting dots and not going through animal cages (MMON)	.92	
Backtracking (MMON)	.82	
Order of animals visited (MCON)	.90	
Strategy used (MCON)	.91	
Accuracy		0.95

MCON-metacognitive control, MMON- metacognitive monitoring, ICC- Intra-class coefficients. The complete sample was evaluated for MCON and MMON and an average score was used, while 10% data were evaluated for the accuracy measure and scores from a single rater were finally used.

To summarize, the zoo-task has medium effect-size correlations across each level, a Cronbach's alpha of .73 (greater than the cut-off of .7) and excellent inter-rater reliability (ICC>.75) for all zoo-task variables.

4.5 Validity analysis

4.5.1 Construct (convergent) validity- comparison with demographics' data

Hierarchical-multiple-regressions revealed that age (section 4.2.3) and gender weren't associated with performance on the zoo-task (along with other problem-solving, disc, task) while GCA was. Regression and power analyses revealed that gender and age have very small effect-sizes ranging from 0.02 to 0.11 and that the current sample size can only allow detection of effect-sizes greater than 0.12.

Correlation analyses (table 14) revealed GCA shared small to medium effect-size correlations with various variables in both the problem-solving tasks, zoo (MCON, MMON and MEVAL) and disc (all the factor efficiencies), and didn't share any correlations with the memory-task.

Table 14
Significant bivariate correlations with demographic variables.

	GCA
Zoo-MCON	**.19***
Zoo-MMON	**.43*****
Zoo-AAI	**.20***
Disc Factor 1 Efficiency	**.23***
Disc Factor 2 Efficiency	**.16***
Disc Factor 3 Efficiency	**.27*****

****p<0.0001 **p<0.001 *p<0.05. MCON-metacognitive control, MMON- metacognitive monitoring, AAI- absolute accuracy.*

Simple linear regression analyses reconfirmed the associations with GCA.

A standard deviation increase in GCA leads to 0.20 standard deviations increase in Zoo-MCON. There is a significant regression equation of $F(1,175)=6.91$; $p<.009$ and the model had a small effect-size that predicted 3.8% variance (R^2 of 0.038; R^2 of .02, .13 and .26 are considered to be small, medium and large effect-sizes, respectively; Cohen, 1988) in Zoo-MCON

A standard deviation increase in GCA leads to 0.42 standard deviations increase in Zoo-MMON. There is a significant regression equation of $F(1,175)=37.19$; $p<.0001$ and the model had a medium effect-size that predicted 17.5% variance in Zoo-MMON

A standard deviation increase in GCA leads to 0.20 standard deviations increase in Zoo-AAI. There is a significant regression equation of $F(1,175)=7.26$; $p<.008$, and the model had a small effect-size that predicted 4% variance in Zoo-AAI

To summarize, key metacognition variables on the novel metacognitive problem-solving (zoo) task positively correlate (small to medium effect-sizes) with and are predicted by GCA measures and not with/by other demographic variables like age and gender.

4.5.2 Criterion (convergent) validity- comparison with a standard metamemory-task

In order to test the validity of the novel task an exploratory hierarchical-multiple-regression was run on the Memory-task efficiency which was followed by the main evaluation on the metacognition measures. The hierarchical-multiple-regression tables in the thesis depict only significant variables (not all variables entered selected for auto-fit) though, commonly associated variables age, GCA and gender are depicted for comparisons with model 1.

Preliminary exploratory analysis to understand the performance measures of the two tasks. The initial hierarchical-multiple-regression was conducted to explore skills required to do well on both the tasks. However, these do not provide any evidence for task validity, which would require the metacognition measures on the tasks to be compared (next sub-section).

A two-step hierarchical-multiple-regression was used to test if the Zoo-task variables would predict Memory-Efficiency. Stage 1 involved addition of age, gender and GCA and step 2 involved addition of Zoo-Efficiency, Zoo metacognition measures, their interaction effects, ethnicity and school using auto-fit modelling. Zoo-MCON predicted Memory-Efficiency for sets 1 and 3, Zoo-Efficiency predicted the memory-Efficiency for set 3 and there were no significant zoo-task predictors for set 2 (table 15 and B1).

Memory-Efficiency set 1 (table 15), for a standard deviation increase in Zoo-MCON, demonstrates a 0.20 standard deviation decrease. The model had a medium effect-size that predicted 15.6% variability in memory efficiency.

Table 15

Hierarchical-multiple-regression on Memory-set 1 efficiency with Zoo-task predictors.

	Zoo task variables → Memory-Efficiency^{Set1}				
	Coefficient	SE	P	Effect-size	Power
Model 1					**F (3,150)=3.18; p=.03, R^2=.060**
GCA	-0.08	0.08	.29	.08	.18
Age	-0.29	0.10	.01	.22	.80
Gender	**0.01**	**0.08**	**.90**	**.01**	**.05**
Model 2					**F (8,143)=3.31; p=.002, R^2=.156**
GCA	-0.06	0.08	.47	.06	.11
Age	-0.25	0.11	.02	.18	.65
Gender	0.05	0.08	.57	.04	.09
Ethnicity{B,W&L&AfA &W&W,B-A&B/L}	**-0.82**	**0.30**	**.01**	**.21**	**.79**
School{F&T-S&W&J&H&R}	**-0.21**	**0.10**	**.04**	**.16**	**.53**
Zoo-MCON	**-0.20**	**0.08**	**.01**	**.20**	**.73**

GCA- general cognitive ability, MCON- metacognitive control. *Variables in bold represent statistically significant results. The remaining hierarchical-multiple-regression tables are presented in appendix B and this table serves as an example of how the regression models were built.*

Memory-Efficiency set 3 (table B1), for a standard deviation increase in Zoo-MCON, demonstrates a 0.22 standard deviation decrease. And for a standard deviation increase in Zoo-BI there is a 0.16 standard deviation decrease in Memory-Efficiency. The model had a medium effect-size that predicted 15.8% variability in memory efficiency.

Due to the addition of Zoo-Efficiency along with Zoo metacognition measures and their correlated nature (section 4.3.1), despite VIF tests revealing no collinearity, independent hierarchical-multiple-regressions were also run and they provided similar associations.

Major Analysis to understand the association between the metacognitive measures. The major analysis involved eight, two-step hierarchical-multiple-regressions testing if Zoo metacognition measures predicted memory-tasks' metacognition measures. Memory-task's JOK-AAI (set 1), RCJ-AAI (sets 1-3), JOK-BI (set 1) and RCJ-BI (sets 1-3) were used as dependent variables. Stage 1 involved addition of age, gender and GCA and stage 2 involved addition of Zoo Variables (MMON, MCON, Net-BI, Net-AAI, and their interaction), Ethnicity and school as predictors using auto-fit modelling. No zoo measures significantly predicted Memory-JOK-AAI^{Set1} (model also lacks fit), Memory-RCJ-AAI^{Set1} (model also lacks fit), Memory-RCJ-BI^{Set2} and Memory-RCJ-AAI^{Set2}. While zoo measures predicted JOK-BI^{Set1}, RCJ-BI^{Set1}, RCJ-BI^{Set3} and RCJ-AAI^{Set3}; the models are reported in tables B2–B5 and summarized in table 16.

Memory-JOK-BI^{Set1} (table B2)- for a standard deviation increase in Zoo-MCON there is a 0.17 standard deviation increase in Memory-JOK-BI and for a standard deviation increase in age there is a 0.25 standard deviation increase in Memory-JOK-BI. The model had a large effect-size that predicted 27.0% variability in Memory-JOK-BI.

Memory-RCJ-BI^{Set1} (table B3)- for a standard deviation increase in Zoo-BI there is a 0.17 standard deviation increase in Memory-JOK-BI and for a standard deviation increase in age there is a 0.25 standard deviation increase in Memory-JOK-BI. The model had a small effect-size that predicted 12.8% variability in Memory-JOK-BI.

Memory-RCJ-AAI^{Set3} (table B4)- for a standard deviation increase in Zoo-BI there is a 0.16 standard deviation decrease in Memory-JOK-AAI. The model had a small effect-size that predicted 9.6% variability in Memory-RCJ-AAI.

Memory-RCJ-BI^{Set3} (table B5)- it's predicted by Zoo-AAI a crossover interaction between Zoo-AAI and Zoo-MMON (figure 5). At a poor/low Zoo-MMON level, with a standard deviation increase in Zoo-AAI there is a 0.21 standard deviation decrease in Memory-RCJ-BI though at high MMON levels there is only a 0.02 standard deviation decrease in Memory-RCJ-BI. The model had a medium effect-size that predicted 15.4% variability in Memory-RCJ-BI.

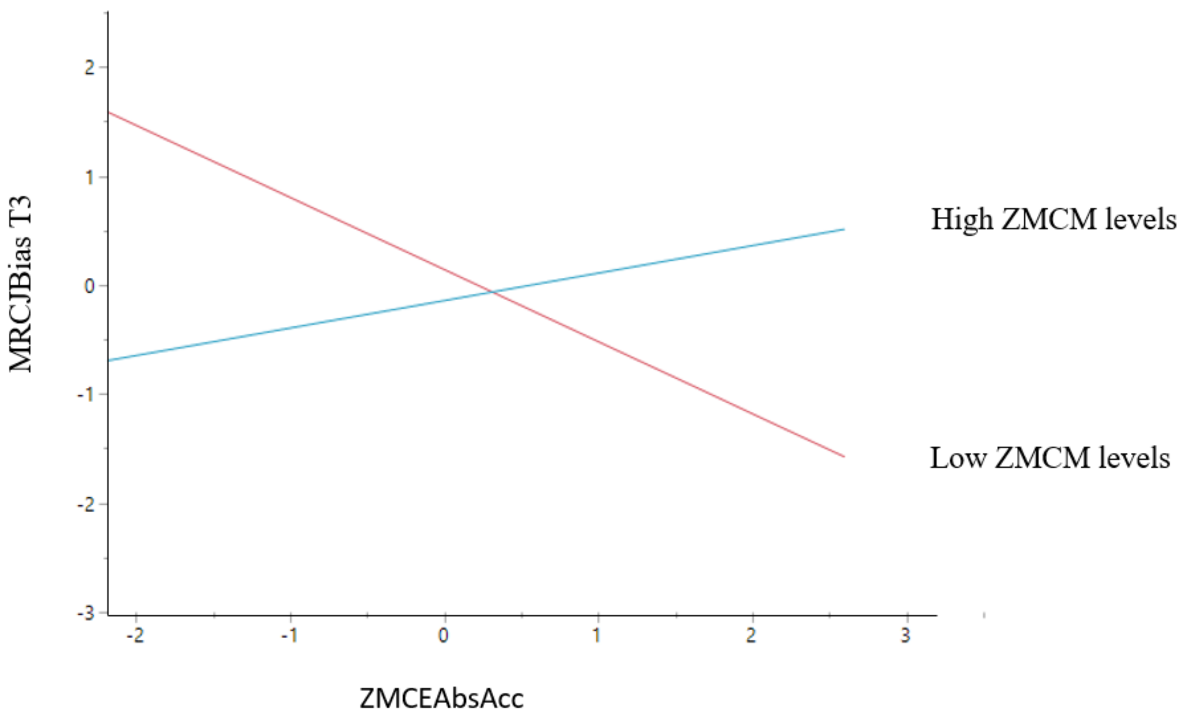

Figure 5: **interaction plot of Zoo-AAI and Zoo-BI predicting Memory-RCJ-BI.** Adapted from JMP output. The red line represents low MMON levels and the blue line represents high MMON. For high MMON levels a positive slope is observed as opposed to a slope of 0 due to insignificant main effects of MMON on Memory-RCJ.

Post-regression analyses tests were run and all suggested that the residuals follow a normal distribution, presence of a linear relation, lack of multicollinearity, absence of influencers and homogeneity of error.

Several memory-task variables at Set 1 and Set 3 are predicted by zoo-task variables of Efficiency, MCON, AAI and BI (summarized in table 16). Age is the only demographic variable that remains a significant predictor of two memory-task variables, prospective and retrospective BI at set 1, after the addition of zoo-task variables. All the predictors in the various models demonstrate effect-sizes between 0.15 and 0.27.

Table 16

Memory-task hierarchical-multiple-regression summary table

Dependent Variable	Memory Task	Independent Variable	Effect-size
Set 1	Efficiency	Zoo-MCON	Medium
	JOK-BI	Zoo-MCON	Large
	JOK Abs Acc	No zoo predictors - Lacks model fit	
	RCJ-BI	Zoo-BI	Small
	RCJ-AAI	Insignificant zoo predictors- Lacks model fit	
Set 2	Efficiency	No zoo predictors - Lacks model fit	
	RCJ-AAI	No zoo predictors	
	RCJ-BI	No zoo predictors	
Set 3	Efficiency	Zoo-MCON, Zoo-Efficiency	Medium
	RCJ-AAI	Zoo-BI	Small
	RCJ-BI	Zoo-AAI, Zoo-MMON*Zoo-AAI	Medium

MCON-metacognitive control, MMON- metacognitive monitoring, MEVAL-metacognitive evaluation, RCJ- retrospective confidence judgement, JOK- judgement of knowledge, BI- bias index, AAI- absolute accuracy. Variables in bold represent statistically significant results.

4.6 Exploratory analysis of metacognition-Executive function association

Five, two-step hierarchical-multiple-regression were used to test if the three disc factor efficiencies predicted zoo-task variables. Stage 1 involved addition of age, gender and GCA and stage 2 involved addition of disc factor efficiencies, their interactions, ethnicity and school using auto-fit modelling. The disc-task efficiencies were predictive of all zoo task variables (table 17 and C1-C4; summarized in table 18).

Zoo-Efficiency- there are no significant disc factor main effects though there is a significant interaction effect of Disc factor 1 and 2 (figure 6). In students with low disc factor 1, for every standard deviation increase in disc factor 2 there is a 0.16 standard deviation increase in Zoo-Efficiency. However, at high disc factor 1, for every standard deviation increase in disc factor 2 students there is a 0.07 standard deviation decrease in Zoo-Efficiency. The model had a small effect-size that predicted 11.9% variability in Zoo-Efficiency.

Table 17

Hierarchical-multiple-regression on zoo-task efficiency with disc-task predictors.

	Disc Factor → Zoo-Efficiency				
	Coefficient	SE	P	Effect-size	Power
Model 1					F (3,172)=0.46; p=.71, R^2=.008
GCA	0.05	0.08	.48	.05	.11
Age	-0.08	0.10	.39	.07	.14
Gender	0.02	0.08	.77	.02	.06
Model 2					**F (10,147)=1.99; p=.04, R^2=.119**
GCA	0.05	0.08	.51	.05	.10
Age	-0.05	0.11	.63	.04	.08
Gender	0.02	0.09	.78	.02	.06
Factor 1 * Factor 2	**-0.23**	**0.09**	**.01**	**.20**	**.71**

GCA- general cognitive ability. Variables in bold represent statistically significant results. The remaining hierarchical-multiple-regression tables are presented in appendix C and this table serves as an example of how the regression models were built.

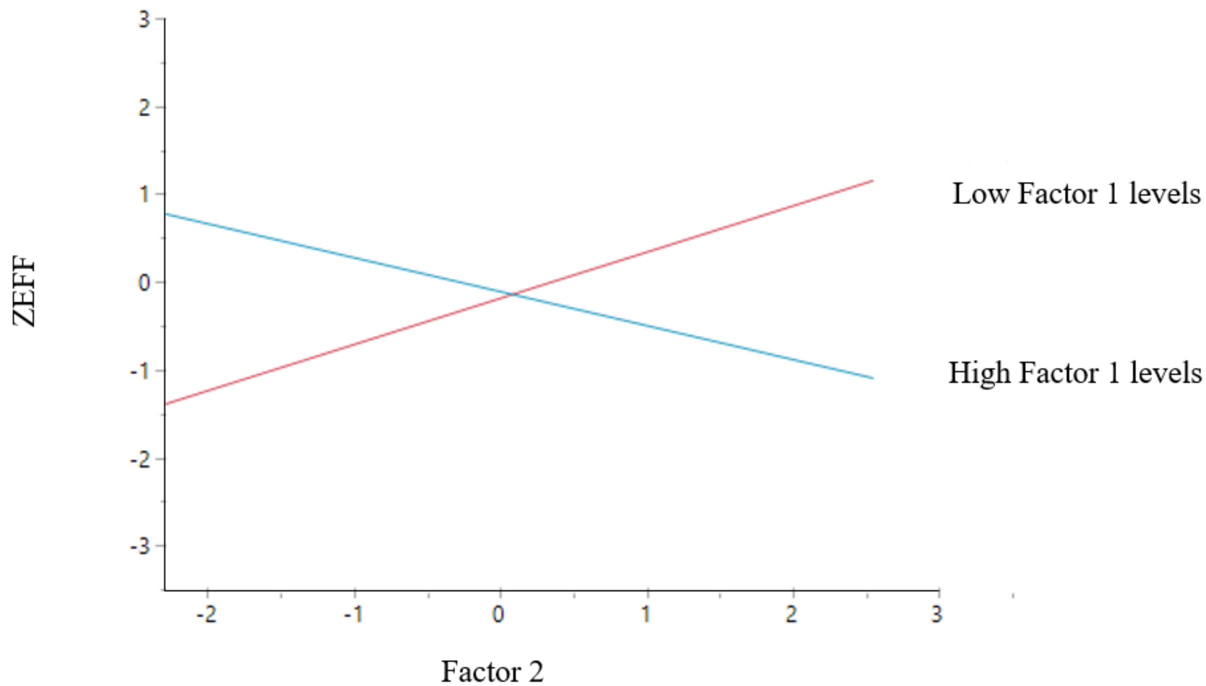

Figure 6: **interaction plot of Factor 1 and Factor 2 predicting Zoo-Efficiency.** Adapted from JMP output. The red line represents low Factor 1 levels and the blue line represents high Factor 1 levels.

Zoo-MCON- for a standard deviation increase in disc factor 2 there is a 0.17 standard deviation increase in Zoo-MCON and for a standard deviation increase GCA there is a 0.17 standard deviation increase in Zoo-MCON. The model had a small effect-size that predicted 11.7% variability in Zoo-MCON.

Zoo-MMON- GCA acts as a main effect along with a significant interaction effect of Disc factor 2 and disc factor 3 (figure 7). For a standard deviation increase GCA, there is a 0.33 standard deviation increase in Zoo-MCON. In students with high disc factor 3, for every standard deviation increase disc factor 2 demonstrate a 0.25 standard deviation increase in Zoo-MMON. In students with low disc factor 3 levels, also for every standard deviation increase in disc factor 2 students also demonstrate an increase, albeit of a smaller magnitude, of 0.07 standard deviation in Zoo-Efficiency. The model had a large effect-size that predicted 31.3% variability in Zoo-MMON.

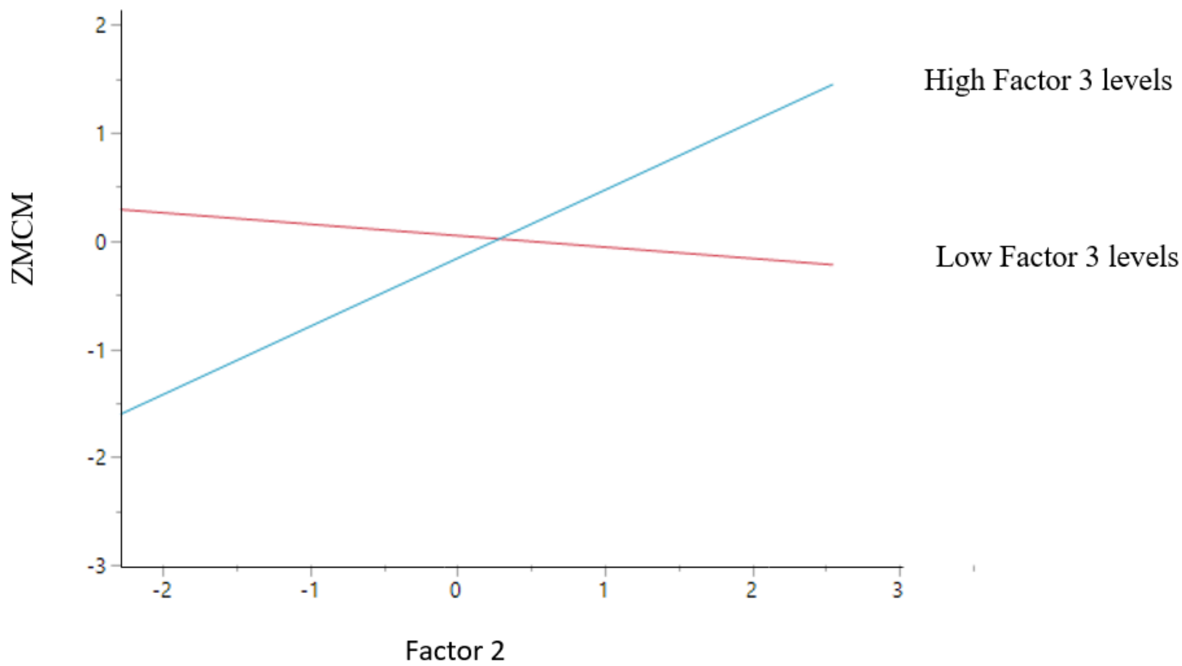

***Figure 7:* interaction plot of Factor 2 and Factor 3 predicting Zoo-MMON.** Adapted from JMP output. The red line represents low Factor 3 levels and the blue line represents high Factor 3 levels. For low Factor 3 levels a negative slope is observed as opposed to a positive slope of 0.16 due to other effects in the model.

Zoo-AAI- there are no significant disc factor main effects though there is a significant interaction effect of Disc factor 1 and disc factor 2 (figure 8). In students with high disc factor 1, for every standard deviation increase disc factor 2 there is a 0.15 standard deviation increase in Zoo-AAI. However, in students with low disc factor 1, a standard deviation increase in disc factor 2 leads to a 0.05 standard deviation decrease in Zoo-AAI. The model had a small effect-size that predicted 11.2% variability in Zoo-AAI.

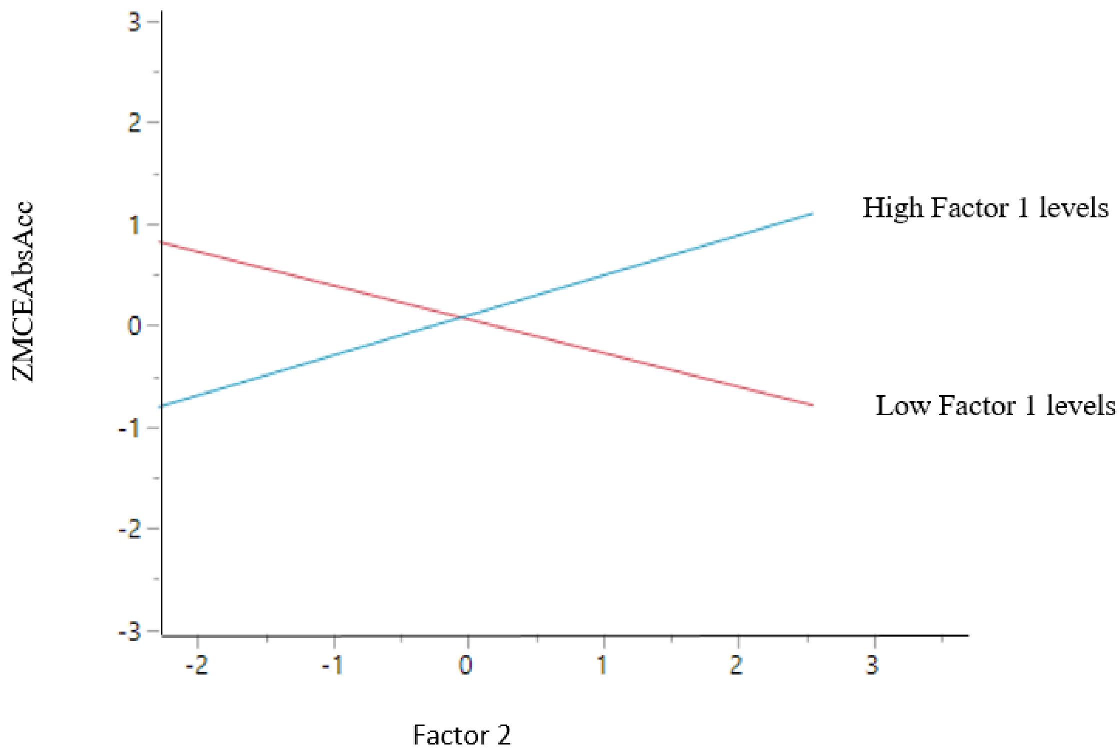

***Figure 8:* interaction plot of Factor 1 and Factor 2 predicting Zoo-AAI.** Adapted from JMP output. The red line represents low Factor 2 levels and the blue line represents high Factor 2 levels. For low Factor 2 levels a negative slope is observed as opposed to horizontal line with a slope of 0 due to Factor 2 effects in the model.

Zoo-BI- a main effect of disc factor 2 and a crossover interaction between disc factors 1 and 3 (figure 9) were observed. For a standard deviation increase in disc factor 2 there is a 0.20 standard deviation decrease in Zoo-BI. At a poor/low disc factor 1 level, with a standard deviation increase in disc factor 3 there is a 0.28 standard deviation increase in Zoo-BI though at high disc factor 1 levels there is only a 0.10 standard deviation increase in Zoo-BI. The model had a medium effect-size that predicted 16.4% variability in Zoo-BI.

VIF=4.21 is less than the suggested cut-off of 10, allowed for the preservation of the correlated factors, factor 1 and 3, in the model.

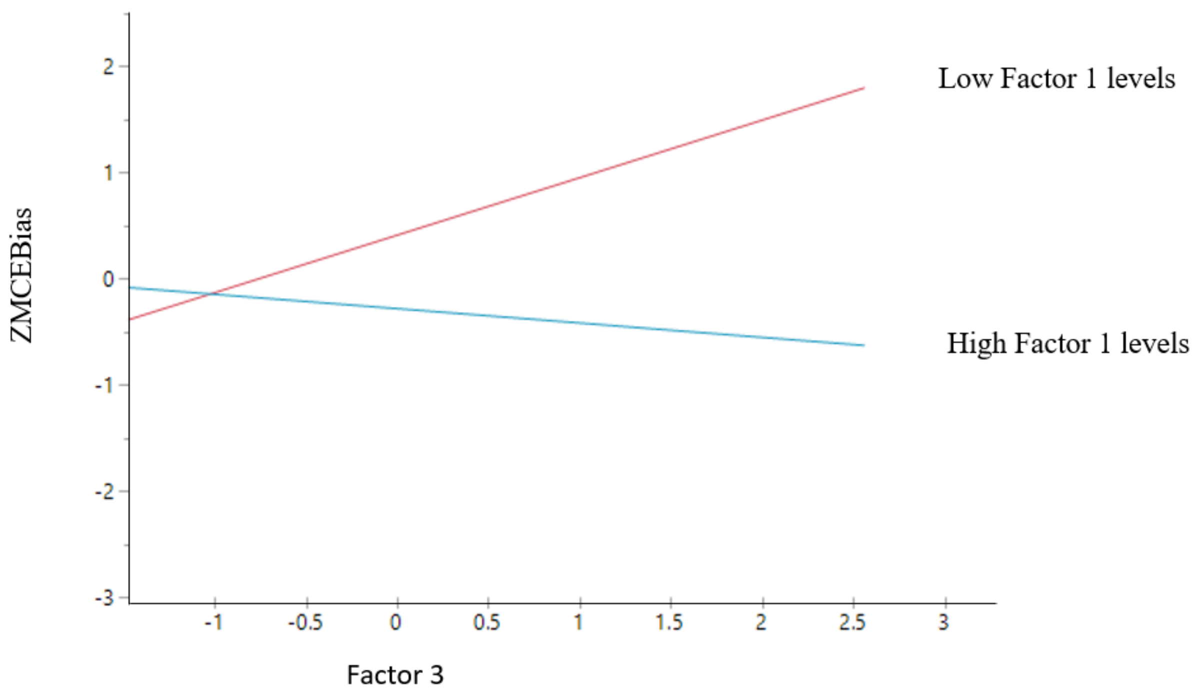

Figure 9: **interaction plot of Factor 1 and Factor 3 predicting Zoo-BI.** Adapted from JMP output. The red line represents low Factor 1 levels and the blue line represents high Factor 1 levels. For high Factor 1 levels a negative slope is observed as opposed to a positive slope of 0.102 due to an insignificant factor 1 effect and other effects in the model.

The post-regression analyses tests were run and all suggested that the residuals follow a normal distribution, presence of a linear relation, lack of multicollinearity, absence of influencers and homogeneity of error.

All zoo-task variables were predicted by different disc-task factors (summarized in table 18). GCA is the only demographic variable that remains a significant predictor of two zoo-task variables, MCON and metacognition, after the addition of disc-task variables. The predictors in the various models demonstrate small to large effect-sizes between 0.16 and 0.31.

Table 18

Zoo-task hierarchical-multiple-regression summary table

Dependent Variables	Significant Predictors	Effect-size of the model
Zoo-Efficiency	Factor 1 * Factor 2	Small
Zoo-MCON	Factor 2	Small
Zoo-MMON	Factor 2 * Factor 3	Large
Zoo-AAI	Factor 1 * Factor 2	Small
Zoo-BI	Factor 2; Factor 1 * Factor 3	Medium

MCON-metacognitive control, MMON- metacognitive monitoring, BI- bias index, AAI-absolute accuracy.

Chapter 5 discussion

The chapter discusses the outcomes of the study in light of previous research. The ‘dimension reduction’ section describes implications of preliminary analyses for further analyses. The following sections, ‘students’ cognitive levels’, ‘zoo-task reliability’, ‘zoo-task validity’ and ‘metacognition-EF associations’ discuss the various findings for each of the four research questions, respectively. The chapter ends with a summary of the key findings and limitations of the study.

5.1 Dimension-reduction

5.1.1 Zoo and memory-task levels

The exploratory-factor-analysis revealed that the zoo-task levels could be combined while memory-sets couldn’t. The medium effect-size correlations of the zoo levels and nearly equal factor loadings for the three levels suggests that the task is internally consistent (further discussed in section 5.3.2).

The memory task is split into three sets, where sets 1, 2 and 3 show children pairs of photos they have seen before, images that they haven’t seen before and a pair with one photo that they have seen before and one that haven’t, respectively. Previous studies (like Everson *et al.*, 1994) create sets based on similar principles, using signal detection theory. The lack of correlation between the three memory-sets suggests that the three sets have to be considered to be separate due to the difference in nature of the sets.

5.1.2 Zoo-task judgement measures

Exploratory-factor-analysis revealed a large effect-size correlation between (a) Zoo-JOK-BI and Zoo-RCJ-BI; and (b) Zoo-JOK-AAI and Zoo-RCJ-AAI implying that the there is little difference between a child’s JOK and RCJ judgements. This finding doesn’t support the literature wherein the two judgements are considered to be testing different aspects of metacognition (Jacobs & Paris, 1987). This apparent similarity could be the result of task administration. The paper-based task had very similar formats for the entry of JOK and RCJ measures. The zoo task deviated in sequence of completing the task, whereby RCJ of a given trial were followed by JOK of the next trial, as opposed to other tasks whereby all JOK measurements are taken together, followed by a break and then the RCJ measurements are taken together. This change in sequence in the zoo-task wouldn't have allowed enough time for reflection on the following trial and could potentially lead to students conflating their answers for post-task judgements with the next task's pre-task judgement. Schneider, Visé, Lockl, & Nelson, (2000), suggested that small methodological changes with the use of prospective and retrospective judgement can potentially lead to significantly different conclusion. This further implies that the zoo judgement outcome-measures would either interact with memory JOK outcome-measures or memory RCJ outcome-measures and not both of them.

5.1.3 Disc-task tasks

Exploratory-Factor-Analysis revealed a three factor solution for the disc-task with the following loading pattern: (a) tasks one, two and four loading onto factor 1, (b) tasks five, six, seven and eight loading onto factor 2; and (c) tasks three and six loading onto factor 3. Figure 4 depicts the problems posed for the eight difficulty tasks. Tasks one, two and four represent a new problem altogether, tasks five, six, seven and eight correspond to significant increase in difficulty tasks and tasks three and six correspond to problems that are relatively similar in nature to the ones solved previously.

Factor 1 (named "new problem-solving factor") corresponds with solving a new problem, a slight increase in task and is hypothesized to tap into working-memory, cognitive-flexibility and basic planning abilities. Factor 2 (named “hard problem-solving”) represents a significant increase in difficulty, involve the use of in-depth analysis of previously used strategies and is hypothesized to tap into working-memory, planning and inhibition of certain previously used strategy. And Factor 3 (named “repeat problem-solving factor”) corresponds to a repeated use of a previously used strategy and is hypothesized to tap into working-memory, inhibitory-control and short term memory.

The disc-task is a complex EF task that is known to tap into a variety of different EFs (Bull *et al*., 2004; Huizinga *et al*., 2006). However, the studies don’t split the task into factors and there is no empirical evidence supporting the hypothesized association of specific factors and EFs. Any EF analysis should be considered to be exploratory in nature and would need to be studied further using EF subtype-specific tasks.

5.2 Students’ cognitive levels and their development

5.2.1 Low cognitive levels

The primary researchers developed age-matched scores and found that the students in the current sample demonstrate poorer GCA levels as compared with the national average (Ellefson, Zewelanji, & Parr, 2016). The thesis also reveals that ethnic-minority, low SES, 8- to 11-year-old students demonstrate poor levels of GCA, poor MCON, poor MEVAL and overconfidence in their judgement (prospective and retrospective) on both the metacognitive tasks suggesting a possibility of under-developed cognitive abilities. The thesis supports previous findings of low SES being associated with poor cognitive development, lower amounts of grey matter in the prefrontal cortex (Luby *et al.*, 2013), poor metacognition (Pappas *et al.*, 2003) and poor EF (Blair *et al.*, 2011). However, age-matched comparisons with other peers is required to confirm cognitive under-development and is beyond the scope of the thesis.

The new coding scheme developed for the thesis catches high variability in students' performance and MMON and not for MCON. The Zoo-MCON generally remains low and could result from either the observed poor cognitive levels or floor-effects of the task, where the task isn't able to measure MCON. Future analyses of the task could build on the current coding scheme to include error preservation, the use of eraser to change paths and converting the order of animals visited from a discrete variable to a continuous variable (could be extracted from the accuracy coding scheme).

The consistently high judgements and negative correlations between MEVAL measures (measured by AAI and BI) and performance on both the tasks suggest late-primary, ethnic-minority, low-SES students are generally overconfident. The overconfidence could be the result of the observed general bias, difficulty of the tasks or nature of the administration of the tasks where children were regularly encouraged. Students at this age have been known to suffer from other biases like confirmation bias (Dunbar & Klahr, 1989; Wason, 1977), i.e. they tend to hold onto real-life conclusions which maybe inconsistent with the empirical data. Alternatively, Lichtenstein, Fischhoff, & Philips, (1982), described hard-easy effect whereby participants demonstrate overconfidence of judgement on hard tasks and other studies (Glenberg, Wilkinson, & Epstein, 1982) have found similar findings in primary-school and university students. This implies that the students either have biases or the tasks were very hard and the children aren't tapping into metacognition.

The bias in the confidence measures makes the judgement outcome-measures, based on differences of judgement and performance, slightly invalid as the students may not be forced to consciously tap into metacognitive skills and the measures may not reflect MEVAL levels. The differences in the judgement outcome-measures could be reflection of performance rather than the metacognitive abilities. This has further implications for validity analyses; comparing MEVAL outcome-measures may not allow for a complete reflection of the validity of the new task. Additionally, the reduced validity of judgement outcome-measures re-emphasizes the need for alternative metacognitive tasks that measure other metacognitive aspects.

5.2.2 Lack of developmental association

The literature has increasing evidence for a relation of cognitive skills with age (Alexander *et al.*, 1995; Best & Miller, 2010; Flavell, 1992, 2004). However, correlational analyses surprisingly revealed that sampled students' age didn't correlate with GCA, zoo-task and disc-task variables. Few (four out of 11) memory variables were weakly associated with age, though three of these associations were negatively associated. A possible explanation could be students hitting ceiling- or floor-effects on the various measures though the descriptive analysis suggests that this isn't the case and students potentially only demonstrate floor-effects on Zoo MCON measures but not on any other measures. Additionally, primary study researchers (Ellefson *et al.*, 2016), reporting on the same sample found that children's age wasn't a significant predictor in any of their models.

The lack of a significant relation between age and the various cognitive measures could be explained by a number of factors like the nature of the sample and confounders. The sampled children may potentially have under-developed cognitive abilities, as discussed in the previous sub-section. And confounding variables like maternal education, availability of resources or learning environments (combining into a SES measure) could mediate the effect of age and GCA (Pappas *et al.*, 2003; Luby *et al.*, 2013).

5.3 Zoo-task reliability

5.3.1 Preliminary zoo-task evaluation

The large effect-size correlation of zoo accuracy and efficiency measures despite speed-accuracy trade-off suggests that the current coding scheme for accuracy measures effectively captures a strong variability in performance.

The medium effect-size correlation between performance (efficiency) and MMON, reflects the need for monitoring skills through the problem-solving task. There is surprisingly an insignificant correlation of performance with MCON which could be the result of students having poor levels of cognitive ability and low variability, as previously discussed; or the measure as operationalized by the task not being able to catch the variability in MCON. Observational problem-solving tasks have found the performance to be related with both MMON and MCON (Bryce *et al*., 2015; Kramarski & Mevarech, 2003; Whitebread *et al*., 2009). The performance is also associated MEVAL, though this observed association may not be the result of an innate relation between the two variables but rather because the students appear to be generally overconfident.

5.3.2 Internal consistency

Both exploratory-factor-analysis and Cronbach’s alpha depict that the three levels of the zoo-task test the same skill. The performance on any level is significantly correlated with the performance on any other levels i.e. a student, as compared to other students, performing better on any level will also perform better than other students on any other levels. The removal of any levels leads to poorer internal consistency therefore the performance measures for the three different levels can be combined into a single outcome variable (Field, 2013).

5.3.3 Inter-rater reliability

Excellent inter-rater reliability was found for all zoo-task variables suggesting the coding scheme is objective in nature and different raters score students similarly on the various constructs.

5.4 Zoo-task validity

5.4.1 Construct (convergent) validity- comparison with demographics' data

Neither age, as discussed in section 5.2.2, nor gender predict zoo-task variables, though GCA predicts various zoo-task metacognition constructs. Age and gender may be weakly associated though the current study's sample size doesn't allow for the detection of the very weak association. Additionally, the lack of relation is limited to the current sample and cannot be extended to all children. Previous metacognition literature has provided mixed evidence for MC's relation with gender (Sperling *et al.*, 2002; Zimmerman & Martinez-Pons, 1990) however there is increasing evidence for its association with GCA (Veenman *et al.*, 2004).

GCA was found to be correlated with and predicted the novel task metacognition measures of MMON, MCON and MEVAL (AAI but not BI) variables providing evidence for construct (convergent) validity. Additionally, hierarchical-multiple-regression analyses on zoo-memory variables revealed that metacognition has a predictive effect post controlling for GCA, suggesting that it has effect over and above the effect of GCA and supporting Veenman *et al.,* (2004)'s, mixed model of GCA-metacognition relation whereby metacognition and GCA are associated with each other but separable.

The relation between the metacognition measures and GCA suggests that the task possesses construct validity and its operationalized metacognition constructs follow the observed trends in literature.

5.4.2 Criterion (concurrent) validity- comparison with metamemory-task

Preliminary zoo-memory-task correlation analysis. There are few correlations between the zoo-task and memory that have $\alpha < .10$. Although not significant, these potential correlations provide further basis to explore the relations using hierarchical-multiple-regression while controlling for demographic variables, to test for mediators and interaction effects which won't be reflected in a simple correlational analysis.

Initial exploration of predictive relations with metacognition performance. MCON as operationalized in the zoo-task, a measure of strategy development and shifting between strategies, was found to be influencing the memory efficiency *via* the time that students take to answer rather than their accuracies (tested *via* a follow-up hierarchical-multiple-regression on memory accuracy). Improvements in MCON and Zoo-Efficiency lead to a decrease in speed with which students make memory decisions and don't affect the accuracy.

The exploratory analyses revealed a differential association with sets 1 and 3 and not with set 2. The differential association of the zoo-task across various memory-sets (associations with sets 1 and 3 and not with set 2) arises because of the nature of the memory-task. It could further be explained either by the three sets of memory-tasks requiring different amounts of MC or set three could involve some amount of problem-solving.

Main analysis studying association between metacognitive measures. Comparisons of the operationalization of metacognitive constructs across the two tasks revealed similarities. On basis of the preliminary analysis it was hypothesized that the Zoo metacognition measures would potentially predict memory-task for sets 1 and 3 metacognition measures. The following predictive relations were found:

Higher levels of MCON lead to increased prospective overconfidence in the sample. Higher MCON levels, theoretically, could lead to increased prospective overconfidence on tasks that the students are seeing for the first time. The lack of relation between memory-task JOK judgement outcome-measures and the zoo-task judgement outcome-measures implies that the net judgement outcome-measures in the zoo-task acts more like RCJ measures and not JOK measures (also discussed previously in section 5.1.2).

Students that are over/under-confident in the zoo-task are also retrospectively over/under confident in the memory-task and the observed overconfidence isn't task-specific.

Higher confidence on the zoo task is associated with better memory retrospective MEVAL suggesting that more confident students are better able to evaluate their performance while those with poorer confidence levels aren't able to do so. However, this relation could be a result of the general bias observed in the sample and students constantly inputting high judgement scores making the scores not reflective of MEVAL but rather of the performance on the two tasks.

Better MEVAL on the zoo task is associated with memory retrospective confidence and the relation is compounded by poor/low MMON levels. Higher MMON tends to buffer the effect of the MEVAL on students' confidence levels.

The zoo-task is significantly and consistently predictive of the memory-task's metacognitive judgement outcome-measures (MEVAL measures) with small, medium and large effect-sizes for different dependent variables. This suggests that the zoo task possesses criterion validity and operationalizes various metacognitive measures in a similar manner as done by the classical metamemory task at sets 1 and 3 though not for set 2. The findings also suggest that MEVAL is a domain-general skill that transfers across the two tasks.

The findings of differential association at the three sets could also explain the contradictory domain-generality/specificity findings of Schraw *et al.* (1995), Kelemen *et al.* (2010), Scott & Berman (2013) and Fitzgerald *et al.,* (2017). Schraw *et al.,* (1995), found MEVAL to be domain-generic by testing students in non-academic domains that the students could potentially be more uncertain about. Scott & Berman (2013) found MEVAL to be domain-specific by testing students on their exam performance in academic domains where the students may not be forced to make conscious judgement decisions and possibly not involving metacognition. Kelemen *et al.* using non-academic domains found MEVAL to be domain-specific though the task wasn't split into sets as done in this thesis and results on each set was combined into a single factor.

The results support Schraw *et al.*, (1995)'s findings and suggest that MEVAL is domain-general as measured in students aged 8- to 11-years on task that are harder, requiring higher levels of decision-making and where metacognitive judgements aren't overweighed by biases, like in set 3. However the small to medium effect sizes (despite the large sample sizes) also suggests the possibility for the presence of both domain-generality and domain-specificity as suggested by Van der Stel & Veenman, (2008). These findings are limited to non-academic tasks and cannot necessarily be extended to academic settings at the 8- to 11-year-old age group due to the later onset of metacognition development for academic settings (Alexander *et al.*, 1995; D. Kuhn, 1999).

5.5 metacognition-Executive function associations

5.5.1 Preliminary disc-task analysis

Disc Factors 1 and 3 demonstrate very strong correlation, even post factor analysis. This suggest that both the new and repeat problem-solving factors are related probably because children could have potentially encountered similar tasks (adaptations are commonly found at school and in market-places) in the past.

Correlational analysis revealed that MCON and MMON correlate with various disc-task variables at a 10% significance-level providing further basis for hierarchical-multiple-regression analysis. The hierarchical-multiple-regression allowed for controlling demographic variables and testing for interaction effects.

5.5.2 Comparing Zoo-task with Disc-task

The hierarchical-multiple-regression revealed that the disc-task was predictive of the zoo-task efficiency and metacognition measures, listed below, despite apparent differences in the tasks; increased requirement of visuospatial thinking in the zoo task, the number of available attempts and the provision of feedback. The zoo-task only allows single attempts per level and doesn't provide any form of feedback or re-trials.

The hard problem-solving factor is associated with zoo-efficiency and the relation is moderated by new problem-solving factor. At low levels of the new problem-solving factor increases in the hard problem-solving factor leads to increases in Zoo-Efficiency and *vice versa* at higher levels of new problem-solving factor. This implies that new problem-solving interacts and inhibits hard problem-solving factors. The new problem-solving factor has been conceptualized to involve trying out multiple strategies and could lead to a shift in cognitive strategy before a plan is completely explored.

This also implies that the zoo-task is a problem-solving task though the observed interaction effects suggest that it may not be tapping into the same problem-solving skills required in the disc-task, as also observed by Scholnick & Friedman, (1993) and Shallice & Burgress, (1991).

The hard problem-solving factor taps into MCON, MMON and MEVAL as operationalized in the zoo task.

Hard problem-solving factor predicts Zoo-MCON. Zoo-MCON is based on strategy selection, error perseverance and ability to change strategy. These skills would also theoretically be important for the harder disc-tasks allowing students to plan and improve their strategies over various attempts to come up with the most efficient way to solve the disc problem.

The relation between hard problem-solving factor and MMON is compounded by higher levels of repeat problem-solving factor. Students that do better on both the factors demonstrate much higher MMON levels. Zoo-MMON is based on children's ability to follow instructions and detect error which would be important for both repeat problem-solving and hard problem-solving.

The relation between hard problem-solving factor and Zoo-MEVAL is moderated by new problem-solving factor. At low new problem-solving levels, improvements in hard problem-solving leads to better MEVAL, and *vice-versa* at high new-problem solving levels. This implies that improved ability to explore a given plan is associated with better judgement of their final products. However, this isn't true for students with better abilities to try new strategies as their ability to quickly shift between strategies doesn't allow an in-depth exploration of the final outcomes.

Repeat problem-solving abilities predict zoo-overconfidence though the overconfidence is downregulated by new and hard problem-solving ability. This implies that the ability to solve problems similar to those previously solved leads to boosts over-confidence however this confidence is downregulated by abilities to explore new strategies or explore a given strategy in-depth.

The predictive nature of disc-task and zoo-task measures suggests that both the task are problem-solving tasks potentially testing different problem-solving skills. The results also suggests that both the tasks tap into metacognitive constructs operationalized in the zoo-task. The various models have small, medium and large effect-sizes.

Previously, Tower of Hanoii and errands task (the adult version of the zoo task) have both been suggested to tap into metacognitive abilities by Brand, Reimer, & Opwis, (2003) and Davis, (2012), respectively. Therefore the results also provide further evidence of construct validity of the novel zoo task.

The disc task is also a complex EF task that taps into multiple EF subtypes including working-memory, cognitive-flexibility (Huizinga *et al*., 2006) and possibly inhibitory-control (Bull *et al*., 2004). The following associations between the disc factors and EFs were hypothesized: (a) new problem-solving factor with working-memory, cognitive-flexibility and planning; (b) hard problem-solving factor with working-memory, inhibitory-control and planning; and (c) repeat problem-solving factor with working-memory, inhibitory-control and planning. The aforementioned results suggests that improvements in Zoo-MCON, Zoo-MMON, Zoo-Efficiency and Zoo-MEVAL are associated with hard problem-solving factor (working-memory and inhibitory-control). And new problem-solving factor (working-memory and cognitive-flexibility) leads to interaction effects while predicting MEVAL and Efficiency, potentially due to the influence of cognitive-flexibility that doesn't necessarily allow students to completely explore or evaluate their given plan.

The thesis reveals a broader relation between metacognition and EF supporting the findings of Bryce *et al.*, (2015), Fernandez-Duque *et al.*, (2000) and Garner, (2009). Bryce *et al.* found MCON to be correlated with working-memory, MMON to be correlated with inhibitory-control and no associations with cognitive-flexibility. The thesis makes very similar observations where MCON and MMON correlate with factors that tap into working-memory and inhibitory-control and not into cognitive-flexibility.

However, given that the disc-task is a complex EF, it isn't possible to have EF subtype-specific factor association and the currently hypothesized association are purely theoretical in nature. The analysis needs be considered to be exploratory in nature, however it provides basis for exploration of metacognition-EF association using EF subtype-specific tasks from the primary project database.

5.6 Summary of findings

The lower-primary ethnic-minority low-SES students demonstrate poor cognitive abilities; low levels of GCA, MCON and MEVAL. The cognitive levels are developmentally stable and are independent of age. Across task analyses demonstrates that metacognition is a domain-general skills that readily transfer across the three tasks.

The results suggest that the novel problem-solving task has high internal consistency, is internally reliable and it tests for the same cognitive constructs across the whole task. The inter-rater reliability suggests an appropriate development of the coding scheme.

GCA was found to be a significant predictor of the novel task's metacognitive measures supporting construct validity of the novel task. And the across task analyses revealed significant associations between performance, EF and metacognition over and above the GCA effects, providing further evidence for Veenman *et al.* (2004)'s mixed model of GCA-metacognition association.

Comparison with a classical metamemory-task suggest that the novel metacognitive problem-solving task has criterion validity. The zoo-task metacognitive measures are predictive of the memory-task metacognitive measures for sets 1 and 3; four predictive relations were found. First, set 1 prospective judgement overconfidence is predicted by MCON skill. Second, set 1 retrospective bias (confidence levels) is predicted by zoo-task's confidence levels. Third, set 3 MEVAL is predicted by zoo-confidence levels. Fourth, set 3 confidence is predicted by zoo-MEVAL measure and the effect is buffered by MMON whereby MMON levels prevent over and under-confidence. The study also demonstrates that metacognition is a domain-general skills that appears to readily transfer across the two tasks.

Comparative analyses with the disc-task revealed that the zoo-task is a problem-solving task, albeit slightly different in nature. Metacognition measures as operationalized by the zoo-task are predicted by the disc-task factors; four predictive relations were found. First, MCON is predicted by hard problem-solving factor (hypothesized to tap into working-memory and inhibitory-control). Second, MMON is predicted by hard problem-solving and the effect compounded by repeat problem-solving factor (hypothesized to tap into working-memory and inhibitory-control) further suggesting MMON is important for doing well on both hard problems and repeat problems. Third, MEVAL is predicted by hard problem-solving factor however the new problem-solving factor (hypothesized to tap into working-memory and cognitive-flexibility) moderates the association. Fourth, overconfidence (another MEVAL measure) is predicted by repeat problem-solving factor and down-regulated by new and hard problem-solving factors. These findings provide supporting evidence for Bryce *et al.*, 2015's metacognition-EF association model whereby MCON and MMON were found to associate with working-memory and inhibitory-control, respectively.

5.7 Limitations and next steps

Several methodological limitations have previously been mentioned, however this section discusses them in further detail. Additionally special care must be taken while drawing conclusion and implications pertaining to the cognitive levels of the sample. The current study cannot be used as a replacement for any psychological evaluations and cannot support generalizations to all cognitive skills as it only tests certain aspects of cognition using single measures (non-comprehensively).

5.7.1 Methodological limitations

The computerized nature of the most of the tasks limits the applicability of the findings as these do not consider other factors like motivation, student (un)familiarity, peer-interaction and comfort (Hacker, Bol, & Keener, 2008). Hacker *et al.* further suggests that the environment in which metacognition takes place affects the levels of metacognition. Additionally several other researchers (Burgress, 1997) suggest that non-naturalistic tools are further limited by test-retest reliability, as theoretically, tasks are only able to pick up cognitive deficits in the first trial and the novelty is lost through the following trials (Chan, Shum, Toulopoulou, & Chen, 2008). However, the greater level of control and larger sample sizes allowed by computerization, as opposed to naturalistic tools, like classroom observations, allows for in-depth correlation and regression analyses for validity analyses. To minimize the effects of non-naturalistic environment the tasks were administered in a classroom environment and the task content and designs were made children-friendly.

The current task is designed as a scalable, child-friendly, nonverbal-task that is independent of language proficiency needs. However, theoretically, it places high demands on working-memory. The task, for future use, should be adapted to include constant written cues, verbal instructions and reminders to lower the working-memory load. Additionally the current MCON coding scheme needs to be extended to add a few more measures of MCON including error preservation, the use of eraser to change paths and converting the order of animals visited from a discrete variable to a continuous variable.

The results suggest that the novel problem-solving task is reliable and valid, though a latent variable approach, using atleast three tasks per metacognitive component, would facilitate holistic task validity study as metacognition is a complex unobservable process. However, there aren't enough holistic metacognitive measures currently available to allow for the latent variable study (Veenman *et al.*, 2006). Additionally, given the scale of the primary project and the breadth of parameters measured including academic attainment and GCA it wasn't practically feasible to have additional tasks. However, a comparison with an observational problem-solving task is recommended as a next step.

5.7.2 Analytical limitations

Violations of normality and the need for two-step rank-transformation suggest that the findings should be interpreted with caution (Field, 2013). The conversion into ranks lead to a loss of information of the magnitude of differences between two participants. However, the consistent presence of significant effect-sizes suggests the presence of the observed and discussed patterns.

Due to the sampling bias (study's focus on low SES, ethnic-minority students), poor cognitive levels and the general overconfidence in the sample, the task validity and metacognition domain-generality/specificity results need to be viewed as exploratory rather than confirmatory. The task needs to be further evaluated and the following three studies are recommended. First, comparisons across other time-points to allow for a developmental understanding. Second, in a different sample of children to allow for generalization to other population although the task should be generalizable to other populations due to Jordan (1994)'s findings that suggested that students with poorer SES perform similarly to other groups of children on non-verbal tasks. Third, using a qualitative study to better understand the nature of the operationalization of the metacognition constructs by the novel task and if they match with observed behavioral traits, like Veenman, (2005)'s, study that tested validity of questionnaires.

The research group is currently working towards a publication about the task whereby the task reliability and validity would be studied across the multiple time-points also allowing for test-retest reliability analysis, predictive criterion validity analyses and to test developmental sensitivity of the task. Following the large-scale validation of the task and it can be used to study the metacognition-EF-academic attainment associations *via* SEM to develop substantive theory pertaining the relation of the three.

The initial metacognition-EF exploration uses a complex EF task that taps into multiple EF subtypes. However, subtype-specific tasks or a latent variable approach need to be used for a better understanding of the metacognition-EF association (as used by Roebers, Schmid, & Roderer, 2009). Additionally the correlation regression design of the thesis doesn't allow for establishment of causal relation between the two and a natural-experimental or true-experimental research-design is required to establish developmental causality. In the near future, data from EF subtype-specific tasks and across various time-points can be used to further understand the relation, however the thesis aimed to primarily test the reliability and validity of the novel task and to test if the task could be used to explore EF-metacognition relations rather than to build substantive theory about the relation.

Chapter 6 conclusion

Ethnic-minority, low SES, 8- to 11-year-old students demonstrated poor levels of cognitive abilities (general cognitive abilities, metacognitive control and metacognitive evaluation). Metacognition, including metacognitive evaluation, was observed to have domain-general characteristics and transferred between memory and problem-solving tasks. A mixed model of metacognition and general cognitive ability association was found whereby both are associated, albeit separable.

The novel problem-solving metacognitive task, an adaptation of a commonly used adult task, is reliable (internally consistent and has high inter-rater reliability) and valid (construct and criterion). It operationalizes metacognitive measures similarly to a classical metamemory-task and the operationalized measures are tapped into by another commonly used problem-solving task. The commonly used problem-solving task, relying on several executive-functions, predicted the novel metacognitive problem-solving task's metacognitive measures suggesting a broader metacognition-EF association.

6.1 Implications of the study

The thesis makes two conceptual contributions to the field. It found poor levels of general cognitive abilities and metacognition and a lack of developmental relation with age in ethnic-minority and low socio-economic status students, suggesting a strong need for interventions that bring about cognitive development or reverse cognitive developmental delays. Additionally it found metacognition to be domain-general and has implications for teaching pedagogy, curricula and interventions whereby metacognition doesn't need to be taught in separate domains or environments; although less malleable or 'teachable' it is easily transferable.

The thesis makes a methodological contribution to the field and found the novel pre-developed problem-solving metacognitive task to be reliable and valid. The task provides metacognition researchers with a much needed age-appropriate measure that could be used with children aged 8- to 11-year-old. As opposed to commonly used tasks, including scalable metamemory-tasks that only measure metacognitive evaluation, observational problem-solving tasks that limit the sample size and highly subjective self-report questionnaires, the novel task will allow researchers to holistically measure multiple metacognitive skill components on a large-scale basis.

The thesis also provides evidence for metacognition-executive association providing basis for substantive-theory developing studies using executive-function subtype-specific task data from the longitudinal project. The developmental understanding between the two cognitive skills can inform pedagogy, curricula and interventions to maximize cognitive development.

6.2 Personal professional development

The thesis till now presents an objective account of the study, in this subsection I would like to reflect on my personal and professional development over the course of the study.

Over the past seven years, I have been working as an educational charity administrator, social reformer and practitioner and have developed a pragmatic attitude (along with a strong sense of idealism). I have always focused on bringing about small amounts of improvements in multiple aspects of any given system/community and have never targeted helping the community reach ‘perfection’ in a single aspect. Over time, I seem to have ingrained a similar attitude on a personal level though through this course I have been able to consciously work towards narrowing down my own focus and targeting perfection.

I have also come to appreciate the importance of in-depth understanding of nuanced educational phenomena for the development of effective teaching practices, curricula and interventions. The process of large-scale data handling and processing gave me an opportunity to learn how to work with repeated measures, multiple tasks, multiple variables and understand the quasi-experimental nature of psychological tasks, not often experienced due to parsimony of time and money for a masters' thesis. The training, in terms of data-handling and understanding psychological constructs and tasks, will help me with my PhD and for devising evaluation tools for the charity.

Chapter 7 List of references

Abu-Bader, S. (2010). *Advanced & Multivariate Statistical Methods for Social Science Research*. Chicago, Ill: Lyceum Books.

Alexander, J. M., Carr, M., & Schwanenflugel, P. J. (1995). Development of Metacognition in Gifted Children: Directions for Future Research. *Developmental Review*, *15*(1), 1–37. https://doi.org/10.1006/drev.1995.1001

Allon, M., Gutkin, T. B., & Bruning, R. (1994). The relationship between metacognition and intelligence in normal adolescents: Some tentative but surprising findings. *Psychology in the Schools*, *31*(2), 93–97. https://doi.org/10.1002/1520-6807(199404)31:2<93::AID-PITS2310310202>3.0.CO;2-X

Anastasi, A. (1997). *Psychological testing / Anne Anastasi and Susana Urbina.* (7th ed. / Anne Anastasi, Susana Urbina.). London: London : Prentice-Hall International, c1997.

Berger, R. S., & Reid, D. K. (1989). Differences That Make a Difference: Comparisons of Metacomponential Functioning and Knowledge Base Among Groups of High and Low IQ Learning Disabled, Mildly Mentally Retarded, and Normally Achieving Adults. *Journal of Learning Disabilities*, *22*(7), 422–429. https://doi.org/10.1177/002221948902200706

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

Bidjerano, T. (2005, October). *Gender Differences in Self-Regulated Learning*. Online Submission presented at the Annual Meeting of the Northeastern Educational Research Association. Retrieved from https://eric.ed.gov/?id=ED490777

Blair, C., Granger, D. A., Willoughby, M., Mills-Koonce, R., Cox, M., Greenberg, M. T., … FLP Investigators. (2011). Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. *Child Development*, *82*(6), 1970–1984. https://doi.org/10.1111/j.1467-8624.2011.01643.x

Blair, C., & Raver, C. C. (2016). Poverty, Stress, and Brain Development: New Directions for Prevention and Intervention. *Academic Pediatrics*, *16*(3 Suppl), S30-36. https://doi.org/10.1016/j.acap.2016.01.010

Boslaugh, S. (2007). *Secondary Data Sources for Public Health: A Practical Guide* (1 edition). Cambridge ; New York: Cambridge University Press.

Brand, S., Reimer, T., & Opwis, K. (2003). Effects of metacognitive thinking and knowledge acquisition in dyads on individual problem solving and transfer performance. *Swiss Journal of Psychology*, *Vol. 62*, *Nr. 4*(HASH(0x72f2848)), S. 251-261.

Bryce, D., & Whitebread, D. (2012). The development of metacognitive skills: evidence from observational analysis of young children's behavior during problem-solving. *Metacognition and Learning*, *7*(3), 197–217. https://doi.org/10.1007/s11409-012-9091-2

Bryce, D., Whitebread, D., & Szűcs, D. (2015). The relationships among executive functions, metacognitive skills and educational achievement in 5 and 7 year-old children. *Metacognition and Learning*, *10*(2), 181–198. https://doi.org/10.1007/s11409-014-9120-4

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry*, *45*(4), 743–754. https://doi.org/10.1111/j.1469-7610.2004.00268.x

Burgress, P. (1997). Theory and methodology in executive function research. In P. Rabbit (Ed.), *Methodology of frontal and executive function.* (pp. 81–116). Hove, UK: Psychology Press.

Busey, T. A. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, *7*(1), 26–49. https://doi.org/10.3758/BF03210724

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). 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

Cicchetti, D. V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 edition). Hillsdale, N.J: Routledge.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4 edition). Thousand Oaks: SAGE Publications, Inc.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Cross, D. R., & Paris, S. G. (1988). Developmental and Instructional Analyses of Children's Metacognition and Reading Comprehension. *Journal of Educational Psychology*, *80*(2), 131–142. https://doi.org/10.1037/0022-0663.80.2.131

Cultice, J. C., Somerville, S. C., & Wellman, H. M. (1983). Preschoolers' Memory Monitoring: Feeling-of-Knowing Judgments. *Child Development*, *54*(6), 1480–1486. https://doi.org/10.2307/1129810

Davis, A. S. (2012). *Psychopathology of Childhood and Adolescence: A Neuropsychological Approach*. Springer Publishing Company.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Eisenberg, N., Fabes, R., Carlo, G., & Karbon, M. (1992). Emotional responsivity to others. *New Directions for Child Development*, (55), 57–73.

Ellefson, M., Zewelanji, S., & Parr, T. (2016). *Exploring the Malleability of Executive Control* (Grant Report- Executive Summary No. R305A110932).

Elshout, J. J. (1983). Is measuring intelligence still useful? In S. B. Anderson & J. S. Helmick (Eds.), *On educational testing* (pp. 45–56). San Francisco: Jossey-Bass.

Elshout, J. J., & Veenman, M. V. J. (1992). Relation between Intellectual Ability and Working Method as Predictors of Learning. *Journal of Educational Research*, *85*(3), 134–43. https://doi.org/10.1080/00220671.1992.9944429

Eme, E., Puustinen, M., & Coutelet, B. (2006). Individual and developmental differences in reading monitoring: When and how do children evaluate their comprehension? *European Journal of Psychology of Education*, *21*(1), 91–115. https://doi.org/10.1007/BF03173571

Evans, J. J., Chua, S. E., McKenna, P. J., & Wilson, B. A. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine*, *27*(3), 635–646.

Everson, H. T., Smodlaka, I., & Tobias, S. (1994). Exploring the relationship of test anxiety and metacognition on reading test performance: A cognitive analysis. *Anxiety, Stress, & Coping*, *7*(1), 85–96. https://doi.org/10.1080/10615809408248395

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive Attention and Metacognitive Regulation. *Consciousness and Cognition*, *9*(2), 288–307. https://doi.org/10.1006/ccog.2000.0447

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd edition). Los Angeles: SAGE Publications.

Field, A. (2013). *DISCOVERING STATISTICS USING IBM SPSS STATISTICS* (4th edition). Sage Publications.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, Motivation and Understanding* (pp. 21–29). Hillside, New Jersey: Lawrence Erlbaum Associates.

Flavell, J. H. (1992). Perspectives on perspective taking. In H. Beilin & Pufall (Eds.), *Piaget's theory: prospects and possibilities* (pp. 107–141). Hillsdale, NJ: Erlbaum.

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly-Journal Of Developmental Psychology*, *50*(3), 274–290. https://doi.org/10.1353/mpq.2004.0018

Flavell, J. H., Miller, P. H., & Miller, S. A. (1993). *Cognitive Development* (3rd edition). Prentice Hall, Englewood Cliffs, NJ.

Foxcroft, C., Paterson, H., LeRoux, N., & Herbst, D. (2004). *Psychological assessment in South Africa: a needs analysis: the test use patterns and needs of psychological assessment practitioners*. Retrieved from http://www.hsrc.ac.za/en/research-data/view/1642

Garner, J. K. (2009). Conceptualizing the Relations Between Executive Functions and Self-Regulated Learning. *The Journal of Psychology*, *143*(4), 405–426. https://doi.org/10.3200/JRLP.143.4.405-426

Glaser, R., Schauble, L., Raghavan, K., & Zeitz, C. (1992). Scientific Reasoning Across Different Domains. In *Computer-Based Learning Environments and Problem Solving* (pp. 345–371). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-77228-3_16

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*(6), 597–602. https://doi.org/10.3758/BF03202442

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in “planning” functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, *33*(5), 623–642. https://doi.org/10.1016/0028-3932(95)90866-P

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of Memory and Metamemory* (pp. 429–455). Mahwah, New Jersey: Lawrence Erlbaum.

Hall, W. S., Bartlett, E., & Hughes, A. T. (1988). Patterns of information requests. *New Directions for Child Development*, *Winter*(42), 43–58.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.

Hanson, J. L., Hair, N., Shen, D. G., Shi, F., Gilmore, J. H., Wolfe, B. L., & Pollak, S. D. (2013). Family Poverty Affects the Rate of Human Infant Brain Growth. *PLOS ONE*, *8*(12), e80954. https://doi.org/10.1371/journal.pone.0080954

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216. https://doi.org/10.1037/h0022263

HayesRoth, B., & HayesRoth, F. (1979). A Cognitive Model of Planning. *Cognitive Science*, *3*(4), 275–310.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). 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

Jacob, R., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 464–470. https://doi.org/10.1037/0278-7393.16.3.464

Jacobs, J. E., & Paris, S. G. (1987). Children's Metacognition About Reading: issues in Definition, Measurement, and Instruction. *Educational Psychologist*, *22*(3–4), 255–278. https://doi.org/10.1080/00461520.1987.9653052

Jordan, N. C. (1994). Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children. *Learning and Individual Differences*, *6*(4), 413–433. https://doi.org/10.1016/1041-6080(94)90003-5

Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1992). Differential Calculation Abilities in Young Children From Middle- and Low-Income Families. *Developmental Psychology*, *28*(4), 644–653.

Kelemen, W. L., Frost, P. J., & Weaver, I. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory and Cognition*, *28*(1), 92–107.

King, A. (1991). Effects of Training in Strategic Questioning on Children's Problem-Solving Performance. *Journal of Educational Psychology*, *83*(3), 307–17.

King, B. M., & Minium, E. W. (2003). *Statistical Reasoning in Psychology and Education*. Wiley.

Kline, P. (1999). *Handbook of Psychological Testing, Second Edition* (2 edition). London ; New York: Routledge.

Koriat, A. (1993). How Do We Know That We Know? The Accessibility Model of the Feeling of Knowing. *Psychological Review*, *100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing Mathematical Reasoning in the Classroom: The Effects of Cooperative Learning and Metacognitive Training. *American Educational Research Journal*, *40*(1), 281–310. https://doi.org/10.3102/00028312040001281

Kreutzer, M. A., Leonard, C., Flavell, J. H., & Hagen, J. W. (1975). An Interview Study of Children's Knowledge about Memory. *Monographs of the Society for Research in Child Development*, *40*(1), 1–60. https://doi.org/10.2307/1165955

Kuhn, D. (1999). Metacognitive Development. *Child Psychology. A Handbook of Contemporary Issues*, 259–286.

Kuhn, Deanna. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*(4), 674–690. https://doi.org/10.1037/0033-295X.96.4.674

Kuhn, Deanna. (2000). Metacognitive Development. *Current Directions in Psychological Science*, *9*(5), 178–181. https://doi.org/10.1111/1467-8721.00088

Kuhn, Deanna. (2004). Metacognition: A Bridge Between Cognitive Psychology and Educational Practice. *Theory Into Practice*, *43*(4), 268–274.

Lawson, G. M., Duda, J. T., Avants, B. B., Wu, J., & Farah, M. J. (2013). Associations between children's socioeconomic status and prefrontal cortical thickness. *Developmental Science*, *16*(5), 641–652. https://doi.org/10.1111/desc.12096

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

Lichtenstein, S., Fischhoff, B., & Philips, L. D. (1982). S .• , B. & Phillips. L. D. Calibration of probabilities The state of the art to 1980. (pp. 306-334). New York: Cambridge University Press. In P. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (p. 306–334'). New York: Cambridge University Press.

Liliana, C., & Lavinia, H. (2011). Gender Differences in Metacognitive Skills. A Study of the 8th Grade Pupils in Romania. *Procedia - Social and Behavioral Sciences*, *29*, 396–401. https://doi.org/10.1016/j.sbspro.2011.11.255

Lockl, K., & Schneider, W. (2006). Precursors of metamemory in young children: the role of theory of mind and metacognitive vocabulary. *Metacognition and Learning*, *1*(1), 15–31. https://doi.org/10.1007/s11409-006-6585-9

Luby, J., Belden, A., Botteron, K., Marrus, N., Harms, M. P., Babb, C., … Barch, D. (2013). The effects of poverty on childhood brain development: the mediating effect of caregiving and stressful life events. *JAMA Pediatrics*, *167*(12), 1135–1142. https://doi.org/10.1001/jamapediatrics.2013.3139

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, *23*, 151–169. https://doi.org/10.1146/annurev.publhealth.23.100901.140546

Luria, A. R. (1966). *Human brain and psychological processes*. New York: Harper & Row.

McCloskey, G., & Perkins, L. A. (2012). *Essentials of Executive Functions Assessment*. John Wiley & Sons.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing Data: A Gentle Introduction*. New York: Guilford Press.

Metcalfe, J., & Dunlosky, J. (2008). Metamemory. In H. L. Roediger III (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 349–362). Oxford: Elsevier.

Metcalfe, J., Schwartz, B., & Joaquim, S. G. (1993). The Cue-Familiarity Heuristic in Metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 851–861. https://doi.org/10.1037/0278-7393.19.4.851

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

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

Montgomery, D. E. (1992). Young children's theory of knowing: The development of a folk epistemology. *Developmental Review*, *12*(4), 410–430. https://doi.org/10.1016/0273-2297(92)90016-U

Moore, C., & Frye, D. (1991). The acquisition and utility of theories of mind. In C. Moore & D. Frye (Eds.), *Children's Theories of Mind: Mental States and Social Understanding,* (pp. 1–14). Erlbaum, Hillsdale, NJ,.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 26). New York: New York: Academic Press.

Nietfeld, J. L., Enders, C. K., & Schraw, G. (2006). A Monte Carlo comparison of measures of relative and absolute monitoring accuracy. *Educational and Psychological Measurement*, *66*(2), 258–271. https://doi.org/10.1177/0013164404273945

Nisbett, R. E. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–260. https://doi.org/10.1037/0033-295X.84.3.231

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation: Advances in Research and Theory.* Plenum Press.

OECD. (2005). *PISA 2003 studies*. Paris, France: OECD.

Osborne, J. W. (2007). *Best Practices in Quantitative Methods*. Thousand Oaks, Calif: SAGE Publications, Inc.

Pappas, S., Ginsburg, H. P., & Jiang, M. (2003). SES differences in young children's metacognition in the context of mathematical problem solving. *Cognitive Development*, *18*(3), 431–450. https://doi.org/10.1016/S0885-2014(03)00043-1

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach* (Student edition). Hillsdale, N.J: Psychology Press.

Pino-Pasternak, D., Whitebread, D., & Tolmie, A. (2010). A Multidimensional Analysis of Parent–Child Interactions During Academic Tasks and Their Relationships With Children's Self-Regulated Learning. *Cognition and Instruction*, *28*(3), 219–272. https://doi.org/10.1080/07370008.2010.490494

Pintrich, P., Wolters, C., & Baxter, G. (2000). Assessing Metacognition and Self-Regulated Learning. *Issues in the Measurement of Metacognition*. Retrieved from http://digitalcommons.unl.edu/burosmetacognition/3

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategy users coordinate metacognition and knowledge. In R. Vasta & G. Whitehurst (Eds.), *Annals of Child Development* (Vol. 5, pp. 89–129). Greenwich, CT,: JAI Press.

Radziszewska, B., & Rogoff, B. (1988). Influence of adult and peer collaborators on children's planning skills. *Developmental Psychology*, *24*(6), 840–848. https://doi.org/10.1037/0012-1649.24.6.840

Radziszewska, B., & Rogoff, B. (1991). Children's guided participation in planning imaginary errands with skilled adult or peer partners. *Developmental Psychology*, *27*(3), 381–389. https://doi.org/10.1037/0012-1649.27.3.381

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence.* London: H.K. Lewis.

Raver, C. C., Smith-Donald, R., Hayes, T., & Jones, S. M. (2005). Self-regulation across differing risk and sociocultural contexts: Preliminary findings from the Chicago School Readiness Project. Presented at the biennial meeting of the Society for Research in Child Development, Atlanta, GA.

Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., … Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain: A Journal of Neurology*, *133*(Pt 1), 234–247. https://doi.org/10.1093/brain/awp269

Roebers, C. M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: an explorative study on their interrelations and their role for school achievement. *Metacognition and Learning*, *7*(3), 151–173. https://doi.org/10.1007/s11409-012-9089-9

Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *The British Journal of Educational Psychology*, *79*(Pt 4), 749–767. https://doi.org/10.1348/978185409X429842

Schneider, W., Visé, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task. *Cognitive Development*, *15*(2), 115–134. https://doi.org/10.1016/S0885-2014(00)00024-1

Scholnick, E. K., & Friedman, S. L. (1993). Planning in Context: Developmental and Situational Considerations. *International Journal of Behavioral Development*, *16*(2), 145–67.

Schraw, G. (2009). A Conceptual Analysis of Five Measures of Metacognitive Monitoring. *Metacognition and Learning*, *4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a General Monitoring Skill Exist? *Journal of Educational Psychology*, *87*(3), 433–444.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, *7*(4), 351–371. https://doi.org/10.1007/BF02212307

Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology*, *90*(2), 236–248.

Schunk, D. H. (1995). Inherent details of self-regulated learning include student perceptions. *Educational Psychologist*, *30*(4), 213–216. https://doi.org/10.1207/s15326985ep3004_7

Schunk, D. H., & Zimmerman, B. J. (1994). *Self-regulation of learning and performance : issues and educational applications / edited by Dale H. Schunk, Barry J. Zimmerman.* Hillsdale, N.J.: Hillsdale, N.J. : L. Erlbaum Associates, 1994.

Scott, B. M., & Berman, A. F. (2013). Examining the Domain-Specificity of Metacognition Using Academic Domains and Task-Specific Individual Differences. *Australian Journal of Educational & Developmental Psychology*, *13*, 28–43.

SHALLICE, T., & BURGESS, P. (1991). Higher-Order Cognitive Impairments and Frontal-Lobe Lesions in Man. In H. LEVIN, H. EISENBERG, & A. BENTON (Eds.), *In: LEVIN, HS and EISENBERG, HM and BENTON, AL, (eds.) UNSPECIFIED (125 - 138). OXFORD UNIV PRESS (1991)* (pp. 125–138). OXFORD UNIV PRESS. Retrieved from http://discovery.ucl.ac.uk/129658/

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2014). *Research Methods in Psychology* (10 edition). Dubuque: McGraw-Hill Education.

Smith, E. (2008). *Using Secondary Data in Educational and Social Research* (1 edition). Maidenhead; New York, NY: Open University Press.

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of Children's Knowledge and Regulation of Cognition. *Contemporary Educational Psychology*, *27*(1), 51–79. https://doi.org/10.1006/ceps.2001.1091

Standards for educational and psychological testing. (1999). Washington DC: American Educational Research Association, American Psychological Association, & National Council of Measurement in Education.

Sternberg, R. J. (1990). *Metaphors of Mind: Conceptions of the Nature of Intelligence*. Cambridge University Press.

Swanson, H. L., Christie, L., & Rubadeau, R. J. (1993). The Relationship between Metacognition and Analogical Reasoning in Mentally Retarded, Learning Disabled, Average, and Gifted Children. *Learning Disabilities Research and Practice*, *8*(2), 70–81.

Tabachnick, B. G. (2014). *Using multivariate statistics / Barbara G. Tabachnick, Linda S. Fidell.* (Sixth edition, Pearson New International Edition.). Harlow, Essex : Pearson, [2014].

Tashakkori, A. M., & Teddlie, C. B. (2010). *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (2 edition). Los Angeles: SAGE Publications, Inc.

Templeton, G. F. (2011). A two-step approach for transforming continuous variables to normal: implications and recommendations for IS research. *Communications of the Association for Information Systems*, *28*(4), 41–58.

Thompson, R. B., & Williams, D. (2006). Diversity Among Low SES Families: An Exploration of Predictive Variables for Mothers' Metacognitive Questions to Their Children. *Research in Human Development*, *3*(4), 191–209. https://doi.org/10.1207/s15427617rhd0304_2

Tobias, S., & Everson, H. (2000). Assessing Metacognitive Knowledge Monitoring. In G. Schraw & C. I. James (Eds.), *Issues in the Measurement of Metacognition*. Lincoln, NE: Buros Institute of Mental Measurements.

Tobias, S., & Everson, H. T. (2002). *Tobias, S., & Everson, H. T. (2002). Knowing what you know and what you don't: Further research on metacognitive knowledge monitoring.* (No. 3). New York: College Entrance Examination Board.

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, *71*(5), 673–679. https://doi.org/10.1037/h0023107

Van der Stel, M., & Veenman, M. V. J. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences*, *18*(1), 128–134. https://doi.org/10.1016/j.lindif.2007.08.003

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs. In C. Artlet & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis.* Berlin: Waxmann.

Veenman, M. V. J., Elshout, J. J., & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction*, *7*(2), 187–209. https://doi.org/10.1016/S0959-4752(96)00025-4

Veenman, M. V. J., Hout-Wolters, B., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences*, *15*(2), 159–176. https://doi.org/10.1016/j.lindif.2004.12.001

Veenman, M. V. J., & Verheij, J. (2003). Technical students' metacognitive skills: relating general vs. specific metacognitive skills to study success. *Learning and Individual Differences*, *13*(3), 259–272. https://doi.org/10.1016/S1041-6080(02)00094-8

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, *14*(1), 89–109. https://doi.org/10.1016/j.learninstruc.2003.10.004

Wallsten, T. S. (1996). An Analysis of Judgment Research Analyses. *Organizational Behavior and Human Decision Processes*, *65*(3), 220–226. https://doi.org/10.1006/obhd.1996.0022

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What Influences Learning? A Content Analysis of Review Literature. *The Journal of Educational Research*, *84*(1), 30–43. https://doi.org/10.1080/00220671.1990.10885988

Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the tower of hanoi disk-transfer task. *Cognitive Development*, *6*(1), 59–76. https://doi.org/10.1016/0885-2014(91)90006-Y

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., … Demetriou, D. (2009). The Development of Two Observational Tools for Assessing Metacognition and Self-Regulated Learning in Young Children. *Metacognition and Learning*, *4*(1), 63–85. https://doi.org/10.1007/s11409-008-9033-1

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, New Jersey: Lawrence Erlbaum.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In P. Pintrich, M. Boekaerts, & M. Zeidner (Eds.), *Handbook of self-regulation.* Orlando FL: Academic Press.

Yonelinas, A. P. (1994). Receiver-Operating Characteristics in Recognition Memory: Evidence for a Dual-Process Model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341–1354. https://doi.org/10.1037/0278-7393.20.6.1341

Zimmerman, B. J. (1995). Self-regulation involves more than metacognition: A social cognitive perspective. *Educational Psychologist*, *30*(4), 217–221. https://doi.org/10.1207/s15326985ep3004_8

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student Differences in Self-Regulated Learning: Relating Grade, Sex, and Giftedness to Self-Efficacy and Strategy Use. *Journal of Educational Psychology*, *82*(1), 51–59. https://doi.org/10.1037/0022-0663.82.1.51

Chapter 8 Appendices

8.1 Appendix A- data processing and preliminary analyses results

8.1.1 Scree plots from factor analysis

This section provides the scree plots generated by SPSS during dimension reduction.

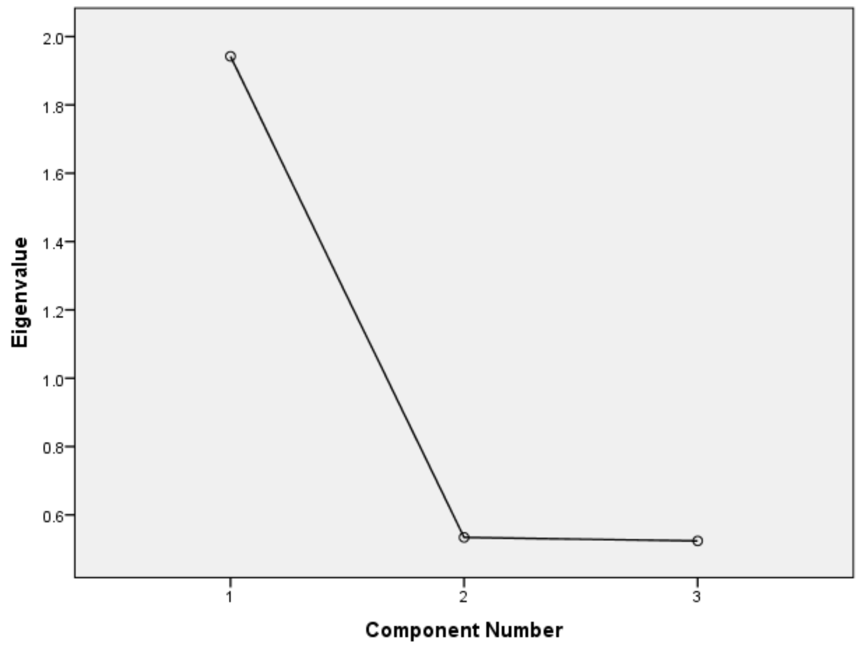

Figure A1: **Zoo-task performance dimension-reduction.** Generated in SPSS.

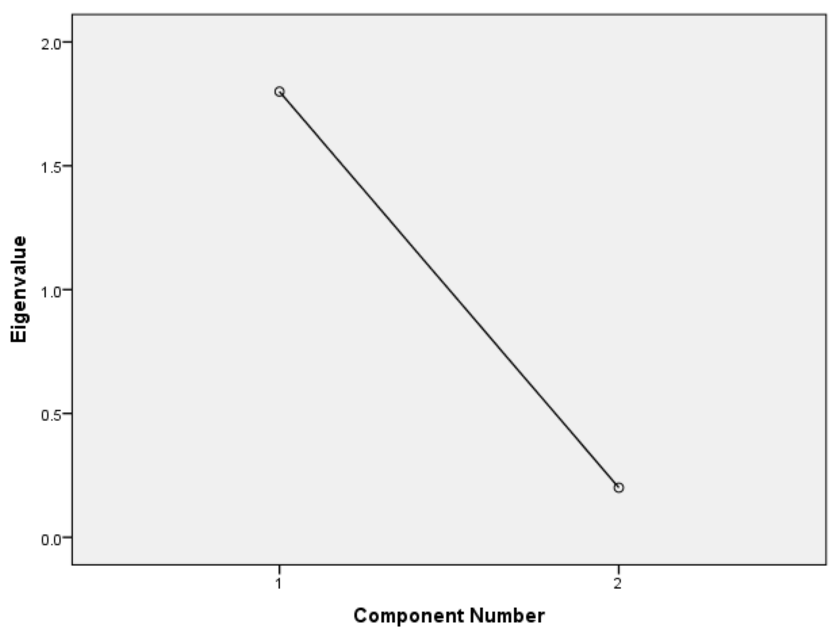

Figure A2: **Zoo AAI dimension-reduction.** Generated in SPSS.

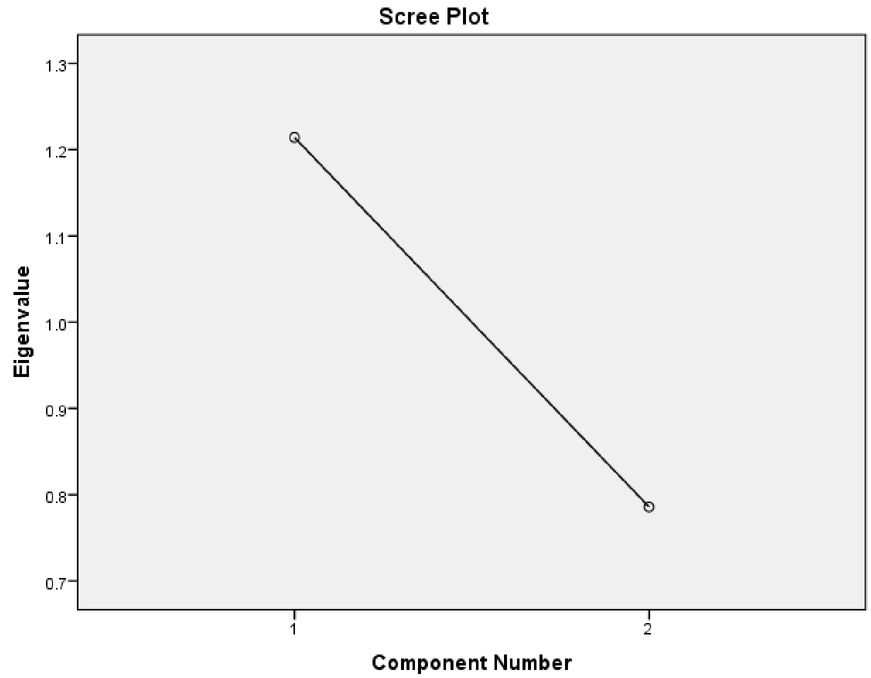

Figure A3: **Zoo BI dimension-reduction.** Generated in SPSS.

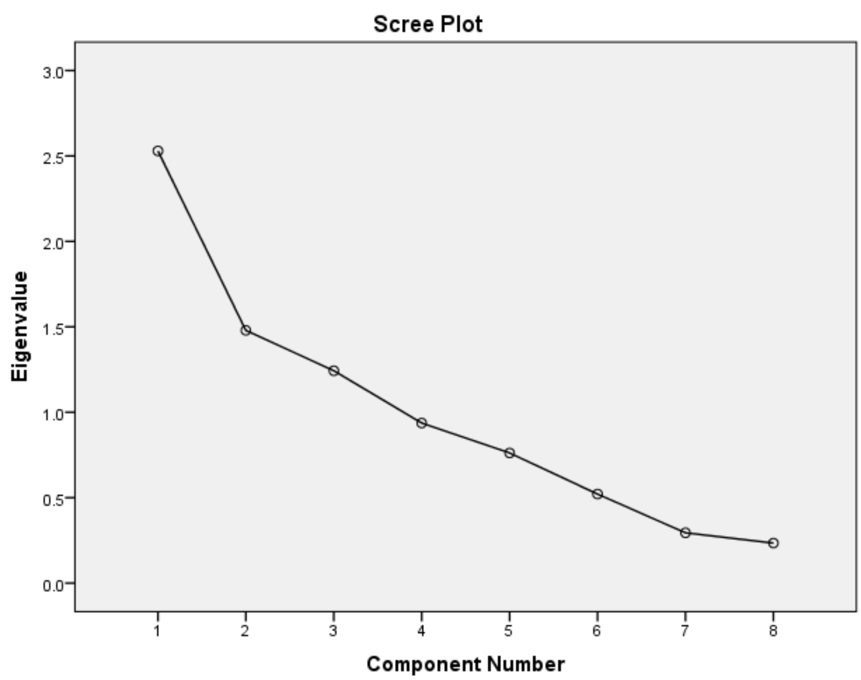

Figure A4: **Disc-task performance dimension-reduction.** Generated in SPSS.

8.1.2 Descriptive statistics

Table A1

Descriptive statistics of cognitive task variables.

	Mean (SD)	Median		Mean (SD)	Median		Mean (SD)	Median
	Zoo-task			Memory- task			Disc-task	
Efficiency	181.35 (1672.01)	366.85	$\text{Efficiency}^{\text{Set1}}$	0.37 (0.12)	0.36	Factor 1	0.53 (0.34)	0.47
MCON	1.98 (1.92)	1.50	$\text{Efficiency}^{\text{Set2}}$	0.29 (0.14)	0.32	Factor 2	-0.00 (0.08)	-0.01
MMON	17.43 (3.36)	18.00	$\text{Efficiency}^{\text{Set3}}$	0.23 (0.11)	0.24	Factor 3	0.29 (0.20)	0.26
JOK	4.37 (0.79)	4.67	JOK^{Set1}	3.73 (0.96)	3.83			
RCJ	4.35 (0.90)	4.67	RCJ^{Set1}	4.45 (0.63)	4.67			
JOK-AAI	2.71 (3.39)	0.90	RCJ^{Set2}	4.05 (1.11)	4.50			
RCJ-AAI	1.00 (1.51)	0.34	RCJ^{Set2}	4.07 (0.98)	4.40			
JOK-BI	0.03 (1.12)	0.00	$\text{JOK-AAI}^{\text{Set1}}$	2.05 (1.43)	1.96			
RCJ-BI	0.00 (1.08)	-0.04	$\text{RCJ-AAI}^{\text{Set1}}$	1.17 (1.14)	0.81			
			$\text{RCJ-AAI}^{\text{Set2}}$	2.08 (2.30)	1.40			
			$\text{RCJ-AAI}^{\text{Set3}}$	2.26 (1.99)	1.61			
			$\text{JOK-BI}^{\text{Set1}}$	-0.00 (0.56)	0.01			
			$\text{RCJ-BI}^{\text{Set1}}$	-0.01 (0.56)	0.09			
			$\text{RCJ-BI}^{\text{Set2}}$	0.02 (1.07)	0.16			
			$\text{RCJ-BI}^{\text{Set3}}$	-0.01 (0.92)	0.06			

MCON-metacognitive control, MMON- metacognitive monitoring, MEVAL-metacognitive evaluation, RCJ- retrospective confidence judgement, JOK- judgement of knowledge, BI- bias index, AAI- absolute accuracy.

8.1.3 within-memory task correlation analysis

Table A2

Bivariate correlations within memory-task variables

	Efficiency^{Set1}	Efficiency^{Set1}	Efficiency^{Set1}	JOK-AAI-Set1	RCJ-AAI-Set1	RCJ-AAI-Set2	RCJ-AAI-Set3	JOK-BI-Set1	RCJ-BI-Set1	RCJ-BI-Set2
Efficiency^{Set1}										
Efficiency^{Set2}	-0.04									
Efficiency^{Set3}	**-.16**$^{--}$	**.42*****								
JOK-AAI^{Set1}	**-.51*****	0.05	0.1							
RCJ-AAI^{Set1}	**-.40*****	-0.07	-0.02	**.33*****						
RCJ-AAI^{Set2}	0.1	**-.58*****	**-.34*****	-0.01	0					
RCJ-AAI^{Set3}	0.01	**-.15**$^{--}$	**-.36*****	0.09	**-.14**$^{--}$	**.58*****				
JOK-BI^{Set1}	**-.37*****	0.02	0.08	0.06	**.20***	0	-0.01			
RCJ-BI^{Set1}	**-.18***	0.06	**.20***	**.17***	-0.02	-0.11	-0.12	0.11		
RCJ-BI^{Set2}	**.14**$^{--}$	**-.45*****	**-.15**$^{--}$	-0.07	0.02	0.04	**-.19***	-0.04	**.18***	
RCJ-BI^{Set3}	**.22*****	**-.22*****	**-.52*****	**-.18***	0.02	-0.08	**-.19***	-0.09	0.16	0.53

*** *$p<0.0001$* ***$p<0.001$* **$p<0.05$* $^{--}$*$p<0.10$. RCJ- retrospective confidence judgement, JOK- judgement of knowledge, BI- bias index, AAI- absolute accuracy.*

8.1.4 within-disc task correlation analysis

Table A3
Bivariate correlations within disc-task variables.

	Disc Factor1		Disc Factor2	
Disc Factor1				
Disc Factor2	-.15			
Disc Factor3	**.81**	***	**.17**	*

***p<0.0001 **p<0.001 *p<0.05

8.2 Appendix B- Hierarchical-Multiple-Regression; zoo task predicting memory task

Table B1

Hierarchical-multiple-regression on Memory-set 3 efficiency with Zoo-task predictors.

	Zoo task variables → Memory- Efficiency^{Set3}				
	Coefficient	SE	P	Effect-size	Power
Model 1				F (3,150)=2.57; p=.06, R^2=.049	
Age	-0.06	0.08	.43	.06	.12
GCA	0.11	0.10	.31	.08	.17
Gender	**0.21**	**0.08**	**.01**	**.20**	**.73**
Model 2				**F (8,142)=3.32; p=.002, R^2=.158**	
GCA	-0.08	0.09	.33	.08	.16
Age	0.03	0.11	.81	.02	.06
Gender	0.17	0.08	.04	.16	.56
Ethnicity{A&W,B-L&AfA&B,W&W&B/L}	**-0.68**	**0.25**	**.01**	**.21**	**.78**
Zoo-Efficiency	**-0.16**	**0.08**	**.04**	**.16**	**.54**
Zoo-MCON	**-0.22**	**0.09**	**.02**	**.19**	**.68**

GCA- General Cognitive Ability, MCON-metacognitive control. Variables in bold represent statistically significant results.

Table B2

Hierarchical-multiple-regression on Memory JOK-BI set 1 with Zoo metacognition predictors.

	Zoo task variables → Memory-JOK-BI^{Set1}				
	Coefficient	SE	P	Effect-size	Power
Model 1			**F (3,150)=7.44; p=.0001, R^2=.130**		
GCA	-0.11	0.08	.15	.11	.30
Age	**0.39**	**0.10**	**.00**	**.30**	**.98**
Gender	-0.12	0.08	.12	.12	.34
Model 2			**F (7,145)=7.67; p<.0001, R^2=.270**		
GCA	-0.11	0.07	.14	.10	.31
Age	**0.25**	**0.10**	**.01**	**.18**	**.73**
Gender	-0.07	0.07	.33	.07	.17
School{R&W-T&S&H&J&F}	**-0.35**	**0.08**	**<.0001**	**.31**	**.99**
Zoo-MCON	**0.17**	**0.07**	**.02**	**.17**	**.69**

GCA- General Cognitive Ability, MCON-metacognitive control. Variables in bold represent statistically significant results.

Table B3

Hierarchical-multiple-regression on Memory RCJ-BI set 1 with Zoo metacognition predictors.

	Zoo task variables → Memory-RCJ-BI^{Set1}				
	Coefficient	SE	P	Effect-size	Power
Model 1			**F (3,150)=0.72; p=.54, R^2=.014**		
GCA	-0.11	0.08	.15	.11	.30
Age	**0.39**	**0.10**	**.00**	**.30**	**.98**
Gender	-0.12	0.08	.12	.12	.34
Model 2			**F (10,140)=2.05; p=.03, R^2=.128**		
GCA	-0.11	0.07	.14	.10	.31
Age	**0.25**	**0.10**	**.01**	**.18**	**.73**
Gender	-0.07	0.07	.33	.07	.17
School{J&W&R& F&S&T-H}	**-0.35**	**0.08**	**<.0001**	**.31**	**.99**
Zoo-BI	**0.17**	**0.07**	**.02**	**.17**	**.69**

GCA- General Cognitive Ability, BI- bias index. Variables in bold represent statistically significant results.

Table B4

Hierarchical-multiple-regression on Memory RCJ-AAI set 3 with Zoo metacognition predictors.

	Zoo task variables → Memory-RCJ-AAI^{Set3}				
	Coefficient	SE	P	Effect-size	Power
Model 1				F (3,149)=1.87; p=.14, R^2=.036	
GCA	-0.12	0.08	.13	.12	.32
Age	0.18	0.10	.09	.14	.40
Gender	-0.04	0.08	.66	.03	.07
Model 2				**F (7,143)=2.17; p=.04, R^2=.096**	
GCA	-0.12	0.08	.14	.12	.32
Age	0.14	0.11	.20	.10	.25
Gender	-0.06	0.08	.46	.06	.11
Ethnicity{B/L-&W&AfA&A&W,B&B,W}	**-1.18**	**0.55**	**.03**	**.17**	**.57**
Zoo-BI	**-0.16**	**0.08**	**.05**	**.15**	**.51**

GCA- General Cognitive Ability, BI- bias index. Variables in bold represent statistically significant results.

Table B5

Hierarchical-multiple-regression on Memory RCJ-BI set 3 with Zoo metacognition predictors.

	Zoo task variables → Memory-RCJ-BI^{Set3}				
	Coefficient	SE	P	Effect-size	Power
Model 1			F (3,150)=0.64; p=.59, R^2=.013		
GCA	-0.07	0.08	.37	.07	.15
Age	0.10	0.11	.37	.07	.14
Gender	-0.04	0.08	.66	.04	.07
Model 2			**F (10,140)=2.56; p=.007, R^2=.154**		
GCA	-0.05	0.09	.55	.05	.09
Age	-0.03	0.11	.78	.02	.06
Gender	0.02	0.08	.85	.02	.05
Ethnicity{AfA&L&B,W-B/L&W,B&W&A}	**-0.36**	**0.18**	**.05**	**.16**	**.51**
Zoo-AAI	**-0.21**	**0.09**	**.02**	**.19**	**.68**
Zoo-MMON * Zoo-AAI	**0.19**	**0.08**	**.02**	**.18**	**.62**

GCA- General Cognitive Ability, AAI- absolute accuracy, MMON- metacognitive monitoring. Variables in bold represent statistically significant results.

8.3 Appendix C- Hierarchical-Multiple-Regression; disc task predicting zoo task

Table C1

Hierarchical-multiple-regression on zoo-task MCON with disc-task predictors.

	Disc Factor → Zoo-MCON				
	Coefficient	SE	P	Effect-size	Power
Model 1			**F (3,173)=2.67; p=.05, R^2=.044**		
GCA	**0.20**	**0.07**	**.01**	**.20**	**.76**
Age	-0.09	0.10	.38	.07	.14
Gender	0.03	0.08	.65	.03	.07
Model 2			**F (5,153)=3.25; p=.01, R^2=.096**		
GCA	**0.17**	**0.08**	**.04**	**.16**	**.55**
Age	-0.14	0.10	.17	.11	.28
Gender	0.04	0.08	.64	.04	.07
Ethnicity{B,W&W,B&A&W&AfA&L-B/L}	**-1.22**	**0.51**	**.02**	**.19**	**.66**
Factor 2	**0.17**	**0.08**	**.04**	**.16**	**.52**

GCA- General Cognitive Ability, Variables in bold represent statistically significant results.

Table C2

Hierarchical-multiple-regression on zoo-task MMON with disc-task predictors.

	Disc Factor → Zoo-MMON				
	Coefficient	SE	P	Effect-size	Power
Model 1			**F (3,173)=13.35; p<.0001, R^2=.188**		
GCA	**0.41**	**0.07**	**<.0001**	**.41**	**1.00**
Age	0.13	0.09	.14	.10	.31
Gender	-0.03	0.07	.63	.03	.08
Model 2			**F (9,149)=7.53; p<.0001, R^2=.313**		
GCA	**0.33**	**0.07**	**<.0001**	**.30**	**.99**
Age	0.04	0.09	.64	.03	.08
Gender	-0.02	0.07	.74	.02	.06
Ethnicity{A&AfA&W&W,B&L&B,W-B/L}	**-1.12**	**0.45**	**.01**	**.17**	**.70**
School{W&F&H-J&S&R&T}	**-0.22**	**0.07**	**.00**	**.20**	**.85**
Factor 2 * Factor 3	**0.18**	**0.08**	**.02**	**.16**	**.66**

GCA- General Cognitive Ability, Variables in bold represent statistically significant results.

Table C3

Hierarchical-multiple-regression on zoo-task AAI with disc-task predictors.

	Disc Factor → Zoo-AAI				
	Coefficient	SE	P	Effect-size	Power
Model 1			**$F\ (3,172)=3.37;\ p=.02,\ R^2=.055$**		
GCA	**0.20**	**0.07**	**.01**	**.20**	**.75**
Age	0.00	0.10	.99	.00	.05
Gender	0.12	0.08	.10	.12	.38
Model 2			**$F\ (9,148)=2.18;\ p=.03,\ R^2=.117$**		
GCA	0.08	0.08	.36	.07	.15
Age	-0.15	0.11	.15	.11	.30
Gender	0.13	0.08	.12	.12	.35
School{F&W-J&R&H&S&T}	**-0.24**	**0.10**	**.01**	**.19**	**.71**
Factor 1 * Factor 2	**0.19**	**0.09**	**.04**	**.16**	**.54**

GCA- General Cognitive Ability, Variables in bold represent statistically significant results.

Table C4
Hierarchical-multiple-regression on zoo-task BI with disc-task predictors.

	Disc Factor → Zoo-BI				
	Coefficient	SE	P	Effect-size	Power
Model 1			F (3,173)=.94; p=.42, R^2=.016		
GCA	-0.12	0.08	.11	.12	.35
Age	0.05	0.10	.59	.04	.08
Gender	-0.02	0.08	.82	.02	.06
Model 2			**F (11,147)=2.62; p=.004, R^2=.164**		
GCA	-0.05	0.08	.55	.05	.09
Age	0.10	0.11	.33	.08	.16
Gender	-0.10	0.08	.23	.09	.22
Ethnicity{B/L&A&L&AfA&W-W,B&B,W}	**-0.96**	**0.35**	**.01**	**.21**	**.79**
School{R&T&H-J&S&W&F}	**-0.30**	**0.09**	**.00**	**.26**	**.92**
Factor 2	**-0.20**	**0.10**	**.04**	**.16**	**.53**
Factor 1 * Factor 3	**-0.17**	**0.07**	**.02**	**.18**	**.63**

GCA- General Cognitive Ability, Variables in bold represent statistically significant results.

8.4 Appendix D- Primary database access request form

Mind Match Chess Database Request Form

Use this form to request access to the Mind Match Chess database collected for IES Grant R305A110932 (Exploring the Malleability of Executive Control) between July 2011 and December 2015. Completed forms should be emailed to the PI for the project, Dr Michelle Ellefson (mre33@cam.ac.uk). Decisions will be made by the PI and Co-PIs as soon as is feasible.

* *

Section A: Persons Requesting access to the Database

Please tell us who will be analyzing the data (i.e., who will have access to the requested data)?

1. Please list the name(s), position(s), and affiliation(s) of all people who are requesting access (not including PI/Co-PIs)

Name	Position	Affiliation (Department and University or equivalent)	Email
Jwalin Patel	Master of Philosophy (MPhil) Student	University of Cambridge, Faculty of Education.	jnp32@cam.ac.uk

2. If you are a student, please list the name(s), position(s), and affiliation(s) of your supervisor(s).

Name	Position	Affiliation (Department and University or equivalent)	Email
Michelle Ellefson	University Senior Lecturer	University of Cambridge, Faculty of Education.	mre33@cam.ac.uk

* *

Section B: Data Requested

1. Please identify the cohort(s), timepoint(s) and task(s) that you want to access from the database. In addition, please indicate whether you require summary or trial-by-trail data?

Cohort (A or B)	Timepoint (1,2,3, 4 or 5)	Task	Format (summary or trial-by-trial)	Specific Demographic / Dependent Variables (Age, Grade, Gender, RT, Accuracy, Efficiency etc.; overall or specific conditions)
All Phases				
B	T1	Stanford10: Literacy, Numeracy & Science Tests	Summary	Demographic: Age, Grade, Gender, School, Ethnicity Stanford: Number Correct, Standardised Score
Phase 1				
B	T1	Zoo task	Trial-by-trial and Summary	Demographic: Age, Grade, Gender, School, Ethnicity Zoo: Accuracy, RT, Efficiency, Judgements (pre and post)
Phase 2				
B	T1	Memory Challenge	Trial-by-trial and Summary	Demographic: Age, Grade, Gender, School, Ethnicity Memory: Accuracy, RT, Efficiency, Judgements (pre and post)
B	T1	Disks Challenge	Summary	Demographic: Age, Grade, Gender, School, Ethnicity Disks: Accuracy, RT, Efficiency
Phase 3 (Note: this data will only be used once the phases 1 and 2 analyses are completed)				
B	T1	Colour-Shape Challenge (Figures)	Summary	Demographic: Age, Grade, Gender, School, Ethnicity Figures: Accuracy, RT, Efficiency
B	T1	Patterns Challenge	Summary	Demographic: Age, Grade, Gender, School, Ethnicity Disks: Accuracy, RT, Efficiency for forward and backward
B	T1	Numbers Challenge	Summary	Demographic: Age, Grade, Gender, School, Ethnicity Numbers: Accuracy, RT, Efficiency and Correct inhibition
B	T1	Soccer Challenge	Summary	Demographic: Age, Grade, Gender, School, Ethnicity Soccer: Accuracy, RT, Efficiency and Stop Signal RT
B	T1	Ravens	Summary	Demographic: Age, Grade, Gender, School Numbers: Number correct, standardised score

2. In the space below, please describe the analyses that you plan to conduct with the data. This explanation should include some reference to specific variable and hypotheses that you plan to investigate.

Study Aims

This study aims to understand the relation between Metacognition (MC) and Executive Functions (EF). It hypothesizes that (a) metacognition is a domain-general skill and the metacognitive judgement measures on the memory and zoo task should correlate with each other; (b) metacognition and EF measures will have a significant overlap though will be two separate cognitive abilities; (c) Metacognitive control would be correlated with working-memory and cognitive-flexibility while metacognitive monitoring with inhibitory-control and WM.

In order to study the above hypotheses I plan to use both the current coding scheme for the zoo task and develop an alternate coding scheme for the same. The current coding scheme has judgement measures and an accuracy measure based on the number of incidences of control (strategy selection) and monitoring (not going off task). The alternate coding scheme will extend the current coding scheme and will have three outcome measures, apart from the judgement measures: (a) control, as measured by sorting, seeking / changing strategies, moving away from the order of animals listed in the question and not persisting with errors while (b) monitoring, assessed through the judgements, error detecting, not straying of any strategy they have devised and following instructions (visiting all listed animals, starting at start and ending at end, visiting zoo entrances, walking on the path and no through cages and not backtracking) and (c) accuracy measure, comparing the students' path-length with the shortest path while factoring in the number of required animals visited or missed and their relative distance from the start and end point on the map.
I will keep extensive notes on how the new coding scheme will be developed and how it could be replicated to adapt the data from the other time-points, should the new coding scheme be useful and if there is a need to replicate the same at the other time-points in the future.

Data Analysis

Establishing reliability of the Zoo task-

An initial reliability test using item response theory (Samejima's continuous response model) will be carried out.

This will be followed by a comparison of the task's accuracy measure with that of another commonly used problem-solving task (tower of Hanoi task) using linear and non-parametric regression. The analysis will compare the accuracy scores from the current and the alternative coding scheme of the zoo task to determine which can be used for further analysis

And finally both the metacognitive judgement measures (JOK and RCJ) of the two metacognitive tasks (zoo and memory-task) will be compared using MANOVA and a two level within subject factors.

Data processing for Metacognition-EF relation analysis- data for students that fit the inclusion criteria (must have completed the zoo task and atleast 3 of the 4 aforementioned EF tasks) will be used. A mean substitution will be used when the missing data are less than 10%.

Data standardization and reduction- data will be converted into standard Z scores to allow for further analysis given all the data uses different scales. CFA will be used to assess if the results on the multiple measures of the zoo task or the various EF tasks could potentially be loaded onto a single variable.

Finding extraneous factors to control for- a preliminary analysis will be conducted to find the correlation of age, gender, school and general cognitive ability with the metacognition and EF to study if they need to be controlled in further analysis.

Regression analysis- Data will be tested for normality, homoscedasticity and a linear relation between the dependent and independent variables. Following which multiple regression analysis will be used to explore relation between the (a) metacognition measures- MEVAL (judgement scores from zoo and memory), MMON (zoo) and MCON (zoo) skills; and (b) metacognitive components and EF sub types while controlling for any exogenous factors.

Possible further analysis- Given the time constraints of a masters course no additional analysis might be possibly though if time permits, interactive effects between metacognitive components and between EF subtypes would be explored in the aforementioned regression analysis.

3. In the space below, please describe your plan for authorship for any dissemination of this project. In most cases, the PI/Co-PIs for the project will have the opportunity to be authors. Your explanation should include expected authorship ordering as well as names of authors. Please refer to the ***Mind Match Chess Authorship Guidelines*** for more detail regarding authorship rights.

The data will be used (1) as a part of my masters thesis for which I would be the sole author and (2) possibly a poster presentation for CogSci 2017 or for a student poster presentation at the faculty of education at University of Cambridge and for both the following authors will be acknowledged as co-authors- Jwalin Patel, Amanda Aldercotte, Teresa Parr, Zewelanji Serpell & Michelle Ellefson.

At this time there are no other dissemination plans. If, in the near future, any other opportunity for dissemination, perhaps *via* a publication arises, then I will consult the PI team

4. In the space below, please describe your plan for dissemination of the findings from the data that you are requesting. Please specify if your dissemination will include an unpublished undergraduate or graduate research projects, (e.g., theses and dissertations)

The data and its analysis will be used for the completion of my MPhil thesis in Educational Research. A part of the analysis will potentially be used for the CogSci 2017 poster presentation (Patel, Aldercotte, Parr, Serpell, Ellefson).

At this time there are no other dissemination plans. If, in the near future, any other opportunity for dissemination, perhaps *via* a publication arises, then I will consult the PI team.

* *

Section C: Requirements for Using the Mind Match Chess Database

There are specific rules that you must agree to before gaining access to this database. Some of these rules are required by the sponsor/funder of the grant awarded to collect these data and some have been devised by the PI/Co-PIs to protect integrity of the data and its dissemination as well as the authorship rights of the large team involved in data collection. You must agree to all of the terms below before your request can be evaluated.

1. Do you agree that you will not share the raw data beyond the people listed in Section A above?

Do You Agree? (delete as appropriate) YES

2. The sponsor/funder requires that all dissemination of this data include the following statement.

   “The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A110932 to the University of Cambridge. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education."

Do you agree to put this statement on all forms of dissemination, including but not limited to manuscripts, books, posters, paper presentations, power point presentations, research reports submitted for a grade/mark, unpublished theses or dissertations?

Do You Agree? (delete as appropriate) YES

3. There was a large Mind Match Chess team involved in collecting this database. A general acknowledgement of that team must be included in all forms of dissemination. Please see the ***Mind Match Chess Authorship Guidelines*** for more detail about the specific acknowledgement to use depending on the cohort data included in your analyses.

   Do you agree to put the appropriate cohort statement(s) on all forms of dissemination, including but not limited to manuscripts, books, posters, paper presentations, power point presentations, research reports submitted for a grade/mark, unpublished theses or dissertations?

Do You Agree? (delete as appropriate) YES

If Yes, please indicate which cohort specific acknowledgment you will include in any dissemination in the space below.

Cohort B Data: "The data presented here were collected by a large team of researchers from the University of Cambridge, Virginia State University, Virginia Commonwealth University and Ashley-Parr, LLC (listed alphabetically by last name): Mariah Adlawan, Annabel Amodia-Bidakowska, Cortney Anderson, Aaron Blount, Lakendra Butler, Parul Chaudhary, Laura Clarke, Jackson Collins, Aiden Cope, Amenah Darab, Asha Earle, Mary Elyiace, Sophie Farthing, Pippa Fielding Smith, Aysha Foster, Kristine Gagarin, Marleny Gaitan, Summer Gamal, Katie Gilligan, Cynthia Gino, Aditi Gupta, Jennifer Hacker, Donita Hay, Rachel Heeds, Joy Jones, Spencer Kearns, Hyunji Kim, Steven Mallis, Dr. Geoff Martin, Alexandria Merritt, Kelsey Richardson, Fran Riga, Tennisha Riley, Kristin Self, Amelia Swafford, Krystal Thomas, Quai Travis, Jorge Vargas, Tony Volley, Elexis White, Sterling Young. These researchers were directed by the PI/Co-PIs: Drs. Michelle Ellefson, Zewelanji Serpell, and Teresa Parr."

4. Specific aspects of the data collection, including the creation of the chess curricula, online website, questionnaires and cognitive tasks were specifically designed by a small number of the research team. If you are using that data, then additional acknowledgements should be made to the specific person(s) or team. Please see the ***Mind Match Chess Authorship Guidelines*** for more detail.

   Do you agree to put an additional statement acknowledging the work of a specific person/team when using the specific task acknowledgements listed on the ***Mind Match Chess Authorship Guidelines*** on all forms of dissemination, including but not limited to manuscripts, books, posters, paper presentations, power point presentations, research reports submitted for a grade/mark, unpublished theses or dissertations?

Do You Agree? (delete as appropriate) YES

If Yes, please indicate which additional acknowledgment(s) you will include in any dissemination

"Special thanks to Dr. Geoff Martin and Dr. Michell Ellefson for programming and developing the overall participant and research interfaces for the Mind Match Chess secured website, to Dr. Michelle Ellefson, Dr. Zewelanji Serpell and Aysha Foster for the development of the instructions and online presentation of the Disks, Colour-Shape, Soccer, Patterns, Numbers and Memory Challenges and to Amanda Aldercotte, Dr. Michelle Ellefson and Dr. Zewelanji Serpe for developing the Zoo Task"

5. All forms of dissemination should be shared with the PI/Co-PI, regardless of whether they are authors, with reasonable time them to review and comment on it before it is disseminated. The PI/Co-PI comments should be incorporated into final versions of publications (or polite justifications given for reasons not incorporating these comments), especially when the PI/Co-PIs are co-authors on any dissemination.

    a) Do you agree to email any conference submissions (e.g., abstracts) atleast one week before the submission deadline? Further, do you agree to incorporate PI/Co-PI comments into the final version?

Do You Agree? (delete as appropriate) Yes Note: I have already submitted an abstract for a poster presentation at CogSci 2017. A draft was approved by the PIs before submission, the submitted version of the abstract has been sent to the Co-PIs.

    b) Do you agree to email any conference presentations (e.g., posters, presentations) atleast two weeks before the conference? Further, do you agree to incorporate PI/Co-PI comments into the final version?

Do You Agree? (delete as appropriate) YES

    c) Do you agree to email any undergraduate research reports (e.g., undergraduate thesis/dissertation) to the PI and Co-PIs atleast two weeks before the submission deadline? (Note – this can be in draft form, but it should be a complete draft) Further, do you agree to incorporate PI/Co-PI comments into the final version?

Do You Agree? (delete as appropriate) Not Applicable

    d) Do you agree to share any masters theses/dissertations proposals (where proposal meetings are required by the degree) atleast two weeks before the submission deadline? (Note – this can be in draft form, but it should be a complete draft) Further, do you agree to incorporate PI/Co-PI comments into the final version?

Do You Agree? (delete as appropriate) Yes

    e) Do you agree to share any masters theses/dissertations atleast three weeks before the submission deadline? (Note – this can be in draft form, but it should be a complete draft) Further, do you agree to incorporate PI/Co-PI comments into the final version?

Do You Agree? (delete as appropriate) YES

    f) Do you agree to share any doctoral theses/dissertations proposals or upgrade vivas (where proposal meetings are required by the degree) atleast two weeks before the submission deadline? (Note – this can be in draft form, but it should be a complete draft) Further, do you agree to incorporate PI/Co-PI comments into the final version?

Do You Agree? (delete as appropriate)	Not Applicable

g) Do you agree to share any doctoral theses/dissertations atleast three weeks before the submission deadline? (Note – this can be in draft form, but it should be a complete draft) Further, do you agree to incorporate PI/Co-PI comments into the final version?

Do You Agree? (delete as appropriate)	Not Applicable

h) Do you agree to share any journal manuscript drafts atleast four weeks before submitting to peer review? Further, do you agree to incorporate PI/Co-PI comments into the final version?

Do You Agree? (delete as appropriate)	Not Applicable

6. Final versions of each of the items in Number 5 (above) should be emailed to the PI and Co-PIs within one week of the deadline or conference. The final version will be shared with the sponsor of the grant. Do you agree to this requirement?

Do You Agree? (delete as appropriate)	YES

7. If you are using the database for a project that is not large enough for a journal publication (as might be the case for undergraduate, masters or doctoral research projects, theses, and dissertations), then the authorship plans apply only to the data described here and not if your work is incorporated into a manuscript that contains more data. If your work does become part of a larger project, then you will be contacted about authorship rights and responsibilities related to that separate dissemination. Any authorship agreements for that larger project will follow the ***Mind Match Chess Authorship Guidelines*** and will be agreed early in the process.

a) Do you agree that authorship for projects that go beyond the data requested here will not have the same authorship plan and that the revised authorship plan will be agreed before proceeding to write the manuscript?

Do You Agree? (delete as appropriate)	YES

b) Do you agree to make sure that the PI/Co-PIs have up-to-date contact information for you so that you might be contacted for any subsequent dissemination that relates to the analyses proposed here?

Do You Agree? (delete as appropriate)	YES

8. During the course of any research project, plans might change as the project moves forward. Do you agree to notify the PI/Co-PIs if there are any changes to the details outlined in this request?

Do You Agree? (delete as appropriate)	YES

9. It is the responsibility of the PI/Co-PIs to monitor the dissemination from this database. If you do not adhere to the agreements above, including submitting drafts and final versions of dissemination according to the timeframes above, then the PI/Co-PIs have the right to take sufficient time to review that material before granting approvals for submission or dissemination – potentially resulting in a delayed or missing the submission deadline. Do you agree to adhere to the deadlines above and that not adhering to the deadlines could negatively affect submission timelines?

Do You Agree? (delete as appropriate)	YES

* *

Section D: Signatures

All people named in Section A must sign and date this form.

Name (printed)	Signature	Date
Jwalin Patel		28th February, 2017
Michelle Ellefson		28th February, 2017

* *

Section E: PI/Co-PI Approvals

Once a fully completed form is submitted to the PI (Dr Michelle Ellefson), then the PI and CO-PIs (Drs Zewelanji Serpell and Teresa Parr) will make a decision about access as soon as is feasible. Additional information will be requested, as required. Once approved, each PI/Co-PI will sign below and return a fully completed copy back to the person(s) requesting access to the database.

PI: Dr Michelle Ellefson

Date:

Signature:

Co-PI: Dr Zewelanji Serpell

Date:

Signature:

Co-PI: Dr Teresa Parr

Date:

Signature:

* *

8.5 Appendix E- Zoo task code book

1. MCON and MMON Coding

For each trial, children were given a rating of 0 - 2 for the following displays:

- Clear route: Basically, if you can trace the route with your finger and not get lost, then it's clear, if you get a little lost a couple of times, it's a 1, and if you don't even know what's going on, then a 0.
  **If there's no route – just connecting two animals here and there = 0
  - 0 = Lines overlapping, hard to see where they went and when; has multiple routes present; did one route but has multiple branches of other routes off of it (not just a little branch to get to the dot); if there's no route, just connecting the dots of all the animals.
  - 1 = A few instances of overlapping lines, lose their path when tracing it; incomplete route that is confusing or does too little (e.g., stops after going to one or two animals);
  - 2 = May have overlapping but it's easy to see the route they to; incomplete routes that are clear (e.g., starts and goes to all animals but missing the finish).
- All animals seen:
  - 0 = Missing 2 animals on Trial 1 OR missing >2 animals on Trials 2 or 3; Alternatively, code 0 if they go to a list of animals of their choosing;
  - 1 = Missing 1 animal on Trial 1 OR 1-2 animals on Trials 2 or 3;
  - 2 = Missing 0 animals.
- Start and finish at correct places:
  - 0 = Neither start or finish are correct;
  - 1 = Incorrect start OR finish; if they go to and from either repeatedly;
  - 2 = Correct start and finish; counted as 2 points if they did multiple routes from 'Start=>Animal=>Finish' (unless they left one a start or finish for a couple of routes).
- Use of dots to feed animals:
  - 0 = No use of dots, goes through cages;
  - 1 = Some use of dots, mostly on path;
  - 2 = Consistent use of dots, stays on path.
- List order:
  - 0 = Saw animals in order presented in list; went 'Start=> Animal=>Finish' for each animal on the list; If they did separate lines for 'sets' of animals and these sets are in the order of the list;
  - 1 = Couple of animals seen out of order;
  - 2 = One or no animal(s) seen out of order.
- Backtracking: Defined as going backwards over to an area of the zoo they'd already been to, but do keep in mind the efficient route to all of the animals (so it's not backtracking if they went over to the llamas but then had to go back to the right to get to the finish for example).
  - 0 = Cannot see route, lines overlapping back and forth between animals;
  - 1 = Less than 2 displays of backtracking;
  - 2 = No displays of backtracking.

- Use of strategies:
    - 0 = No strategy present;
    - 1 = One strategy apparent;

  2 = Multiple strategies apparent.

2. Accuracy Score Coding

The accuracy score is based on the path-length chosen, the number of required animals visited and their relative distance from the start point. There are two manual inputs that are required to calculate the accuracy measure. They are based on the routes children draw on paper.

2.1 Tracing the child's path

For every trial you are required to trace the path that children take along the map. Start with either of the two ends of the route and trace a rough mental path that the child may have drawn. Once you are confident of the path chosen, retrace the path and enter the names of the animal cage entrances that the path crosses in the accompanying excel sheet. A few coding decisions to take care of include:

- The animal cage entrances passed should be entered in adjoining cells, in the relevant section of the sheet (divided into 3 trials), and the whole path shouldn't be entered in a single cell.
- Children may not necessarily start at the start and the end point, in which case just try picking out a start/end to the path and use the same.
- Children at times will draw a path like a tree diagram where the branches stem out from the main path and are represented by a 'one-way' path and there is no return route drawn. Consider the branched paths as a two route and the child would have effectively backtracked on them.
- Many a times students will draw disjoint paths: either (a) starting with start and ending at certain animals/ the end point or (b) between sets of animals. In both the cases while coding leave a blank cell between the codes for the two paths.
- If there is an unclear path in trial 3 it can at times help to visit the order of animals listed in the questions and checking if the children follow the exact path; please be aware of the bias in this coding process and only use it as a last resort.
- For an extremely convoluted map that you cannot trace at all (usually <2% of the maps), please use the placeholder "c" and the maps will be later revisited using an alternative coding system that will be developed in the coming fortnight. The scoring maybe based on the strong correlation between trials 2 and 3 or the map would be addressed as missing data.
- For maps where the path cannot be traced at all due to **most of the** pencil marks fading away over time please mark the trial with an "m", the placeholder for missing data. Do note that any missing data could potentially mean that the child might have to be excluded from a large amount of analysis, drastically reducing statistical power of the analysis.

2.2 Animals Visited

The excel sheet *et al*so requires you to list the cage entrances that the child visits for every trial. This will be later used to determine if the child has just passed by a required animal or "has entered the cage to feed it".

A coding decision that you might need to be mindful of includes that at times children don't stick to the spaces between cages to map out the path but instead "walk through animal cages". In such cases the benefit of the doubt should be given to the child and decisions should be made on a case by case basis for example:

- Children that normally walk through cages but specifically make sure to walk through the zoo cage entrances, marked by black dots, and should be marked as if they visit the given cage.
- There are very few children (<2%) that don't use the zoo cage entrances and walk through animal cages but the animals visited can be decided when looking at the path holistically (start and end points) and any deviations from the main path.

Don't worry about down marking children that don't stick to the given instructions, about not walking through animal cages, as it has already been factored into another part of the coding scheme for metacognitive monitoring.

An animal is only considered to be visited if the child connects 3 of the required points in the path; i.e. a path that connects start, a required animal and finish is considered while a path that connects start, a required animal and n number of non-required animals isn't considered (the path-length is calculated however the animal isn't considered to be visited).
Additionally animals are only considered to be visited if the path connects three of the required paths

3. Abbreviations

Both the inputs require coding short hand abbreviations for longer animal cage names. Please use the following list of abbreviations to aid with the same.

Start:	St
Birds:	B
Cheetahs:	Ch
Cows:	Co
Elephants:	E
Fish:	F
Giraffes:	Gi
Goats:	Go
Hippos:	Hi
Horses:	Ho
Jellyfish:	J
Lions:	Li
Llamas:	Ll
Monkeys:	M
Pandas:	Pa
Penguins:	Pe

Pigs:	Pi
Polar bears:	Po
Seals:	Se
Snakes:	Sn
Walrus:	W
Zebra:	Z
Finish:	Fin

8.6 Appendix F- Zoo task data entry and management file

1. Raw data entry

The accuracy score is based on the path-length chosen, the number of required animals visited and their relative distance from the start point. There are two manual inputs that are required to calculate the accuracy measure. *They are* based on the routes children draw on paper.

The raw data needs to be entered in the following format into three different sheets each sheet representing each trial. Table 1 represents a typical data entry sheet where raw data needs to be entered in columns colored blue, columns in white are involved in intermediate processing and are hidden from view and columns in orange represent the final output columns that will be involved in further calculations.

Partici-pant ID	School ID #	Badge Color	Object	10 columns listing the required Animals Visited	16 columns for weight-age calcu-lations	Weight-age	50 columns listing Entrances that the Path Passes by	Notes	Exc-lude	8 columns with look up tables	50 columns with pair-wise path lengths	Sum of pair wise path lengths
1												
2												
3												

Figure 1

2. Data coding process

Raw data are processed into absolute path-length and net weightage multiplier which will be further processed to create an absolute accuracy score. In order to calculate absolute path-length and net weightage multiplier of the chosen path from the raw data the zoo map has been converted into a block based map.

2.1 Entrances' absolute positions

Automated calculations will involve calculating the horizontal distance and vertical distance between the zoo entrances that the path crosses for path-length and horizontal distances from the start point for weightage. The following two alternatives block based maps were generated. The outline shown in figure 3 is used rather than that shown in Figure 2 albeit less detailed is quicker and simpler to use. Additionally it doesn't suffer from the challenge of computational backtracking as faced in figure 2 for example while calculating distances between snake and jellyfish

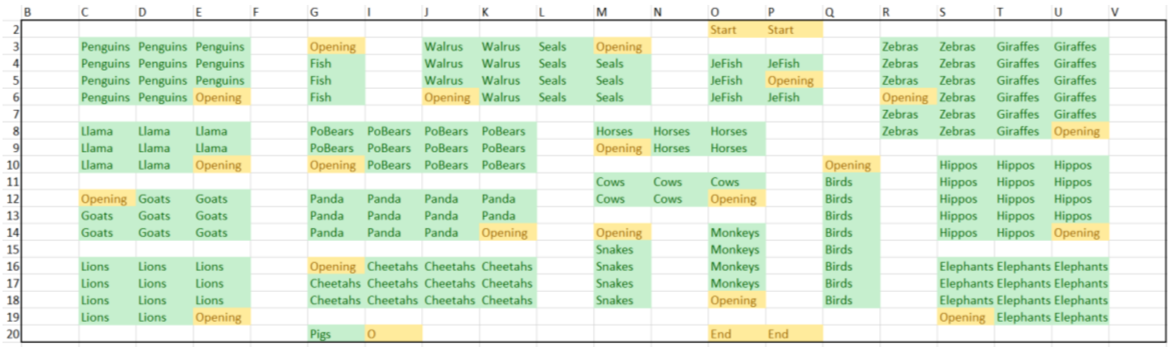

Figure 2

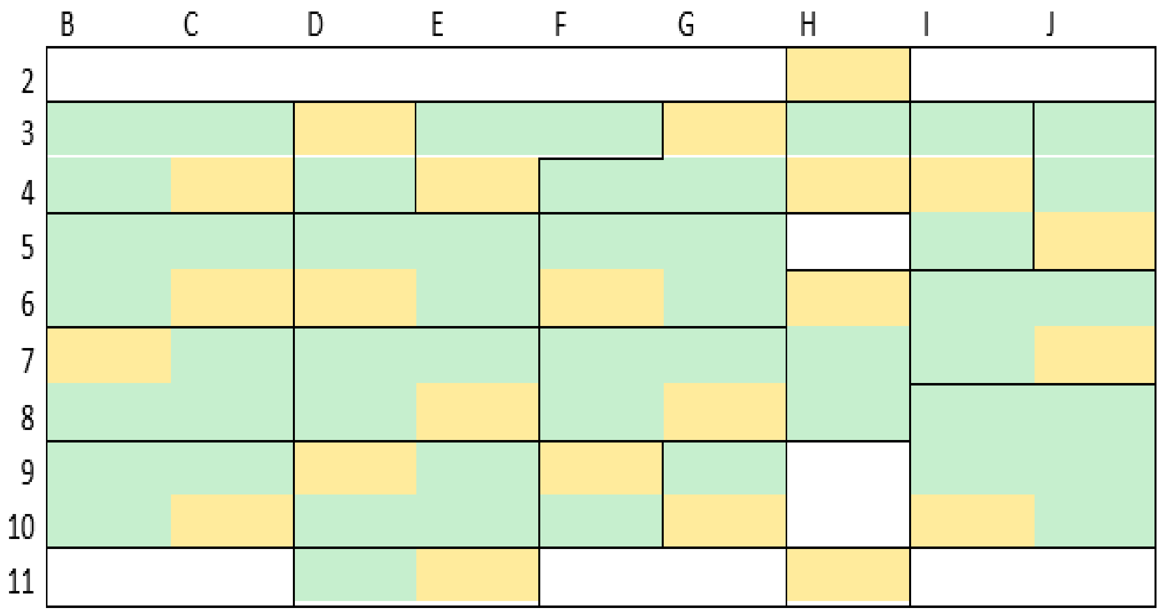

Figure 3

Using the figure y the absolute positions are provided as below.

Zoo entrance name	Position in figure 3	Vertical Position	Horizontal Position
St	H2	7	2
B	H6	7	6
Ch	D9	3	9
Co	G8	6	8
E	I10	8	10
F	D3	3	3
Gi	J5	9	5
Go	B7	1	7
Hi	J7	9	7
Ho	F6	5	6
J	H4	7	4

Li	C10	2	10
Ll	C6	2	6
M	G10	6	10
Pa	E8	4	8
Pe	C4	2	4
Pi	E11	4	11
Po	D6	3	6
Se	G3	6	3
Sn	F9	5	9
W	E4	4	4
Z	I4	8	4
Fin	H11	7	11

Table 1

The aforementioned table has been used as a reference for the vlookup function while performing further calculations of both path-length and weightage multiplier of the chosen path.

2.2 Intermediate variables

Weightage multiplier

The task involves visiting multiple animal cages and given that they are situated all across the map a few are harder to spot and would involve longer path lengths. As a result of which the final accuracy score would need to factor in the weightage given to visiting a specific animal cage and the absolute path length. All animal cages have been assigned a weightage multiplier (the sum of the weightages in a given trial = 1) based on the horizontal distance (figure 3) from start / finish as both start and finish lie on the same vertical line.

The following table represent weightages for the animal cages that required to be visited in the three trials. Start and Finish have both been given an average weightage for any given trial and is equal to weightage of 1/(total number of points to visit).

Trial 1			
Req Visit	St Fin weight	Horizontal D	% weightage
Start	0.166667	3.75	0.166667
Snakes		3	0.133333
Birds		1	0.044444
Penguins		6	0.266667
Polar Bears		5	0.222222
Finish	0.166667	3.75	0.166667
Trial 2			
Req Visit	St Fin weight	Horizontal D	% weightage
Start	0.125	3.333333	0.125
Monkeys		2	0.075
Seals		2	0.075
Lions		6	0.225

Horses		3	0.1125
Cheetahs		5	0.1875
Zebras		2	0.075
Finish	0.125	3.333333	0.125
Trial 3			
Req Visit	St Fin weight	Horizontal D	% weightage
Start	0.1	3.875	0.1
Goats		7	0.180645
Fish		5	0.129032
Elephants		2	0.051613
Walrus		4	0.103226
Pandas		4	0.103226
Cows		2	0.051613
Jellyfish		1	0.025806
Llamas		6	0.154839
Finish	0.1	3.875	0.1

Table 2

Calculation steps. In order to calculate the net weightage of the chosen path the following calculation steps are followed.

- The weightage of each animal visited is referenced from a look up table (table 4 in this document. Each trial uses a different part of the table 4). The formula used for referencing the weightage is as follows
  =IF(ISBLANK(F3),0,VLOOKUP(F3,$Q$4:$T$14,4,FALSE))
  Equation 1
- A sum of the lookup values is used in order to calculate the net weightage using the following formula
  =SUM(U3:AD3)
  Equation 2

If the net weightage formula results in a N/A then the entered raw data would need to be checked for typing errors.

Formulae explained. The Sum of if blank and Vlookups formula explained

The raw data for path lengths is entered in columns F-O (10 columns), look up tables (table 3 in this document) with spaces are entered in Q-T columns, weightage for the animal visited in columns U-AD and sum of weightages of all the animals in column AE.

Equation 3 returns a “0” when the data entry cell is left blank in instances where the child visits fewer then the required animals (represented by commands “IF” and “ISBLANK”) and when data are entered in it looks up the animal abbreviation in the look up table to finally return the weightage of visiting the given animal.

The equation 4 provides the net weightage by summing up weightages of individual animals that are visited provided by (equation 3).

Absolute path-length

The animal entrance position table along with the zoo entrances crossed allows for the absolute path-length determination.

Calculation steps. The following steps are followed for the same:

- The path lengths between pairs of various entrances on the chosen path is calculated using table 2.
  Absolute path-length between two points
  =IF(OR(ISBLANK(AH3),(ISBLANK(AI3))),"",ABS(VLOOKUP(AH3,$CI$2:$CL$48,3,FALSE)-
  VLOOKUP(AI3,$CI$2:$CL$48,3,FALSE))+ABS(VLOOKUP(AH3,$CI$2:$CL$48,4,FALSE)-
  VLOOKUP(AI3,$CI$2:$CL$48,4,FALSE)))

  Equation 3
- The absolute path-length is found using the sum function on distance between the various pairs of entrances on the path.
  The sum of all path lengths between pairs =SUM(CO3:EL3)

  Equation 4

If the absolute path-length formula denotes a N/A then the entered raw data would need to be checked for typing errors. To find the typing error during data entry you could trace back the calculation error to an error in the calculation of distance between the pairs which could later be traced to the exact data cell.

Formulae explained. The raw data for path lengths is entered in columns AH-CE (50 columns), look up tables (table 3 in this document) is entered in CI-CL, pairwise path lengths in columns CO-EL and sum of pairwise path-lengths in EC.

Equation 1 returns pair wise path lengths. It returns a blank cell when the either of the original pairs is blank (represented by commands “IF” and “ISBLANK”) and a numerical sum of the vertical and horizontal distances between two animal entrances when neither is blank (represented by sum of “ABS” formula).

The “ABS” formula returns the absolute value of the difference between the vertical and horizontal positions. The difference in positions (distances) are calculated using a “VLOOKUP” command looks up the abbreviation entered during data entry and compares it with the lookup table (table 3 in this document) and returns the value in column 3 (for vertical positions) or 4 (for horizontal positions). The “False” in the formula asks the software to look for an exact match rather than the starting letter.

The equation 2 provides a sum of the pair wise distances that are calculated using equation 1.

2.3 Accuracy Score

The accuracy score use the different between the shortest path-length and the absolute path-length and factors in the number of require animals visited and their weightage.

The shortest path-lengths for trial one, trial 2 and trial 3 are 23, 23 and 25, respectively.

Calculation steps

The following set of equations have been converted into excel formulae.

$$If\ Path\ chosen\ \leq 2 \times Shortest\ Path$$

$$Accuracy = \left[1 - \left(\frac{|Shortest\ Path - Path\ chosen|}{Shortest\ path}\right)\right] \times 100 \times \left[\sum_{i}^{n}(Distance\ of\ cage\ from\ start \times x_i)\right]$$

Equation 5

$$If\ Path\ chosen\ \geq 2 \times Shortest\ Path$$

$$Accuracy = \left[1 - \left(\frac{|Shortest\ Path - Path\ chosen|}{Shortest\ path}\right)\right] \times 100 \times \left[\frac{1}{\sum_{i}^{n}(Distance\ of\ cage\ from\ start \times x_i)}\right]$$

Equation 6

Where i-n represents the required animals the child visits and x_i represents the relative weightage of animal i and can be found from Equation 3.

$$\sum_{i}^{n}(Distance\ of\ cage\ from\ start \times x_i) = 1$$

Equation 7

The following excel formulae are used for the three different levels

- For level 1: =IF(G2>46,(1-(ABS(EM3-23)/23))*1/AE3,1-(ABS(EM3-23)/23)* AE3)
  Equation 8
- For level 2: =IF(G2>46,(1-(ABS(EM3-23)/23))*1/AE3,1-(ABS(EM3-23)/23)* AE3)
  Equation 9
- For level 3: =IF(O2>50,(1-(ABS(EM3-25)/25))*1/AE3,1-(ABS(EM3-25)/25)* AE3)
  Equation 10

If the absolute path-length formula denotes a N/A then the entered raw data would need to be checked for typing errors. To find the typing error during data entry you could trace back the calculation error to an error in the calculation of distance between the pairs which could later be traced to the exact data cell.

Formula Explained

The absolute path lengths have been calculated in column EM3 and weightage calculated in AE3. The 'If formula' checks if the absolute path-length is more than twice the shortest pathlength and if so equation 6 is used else equation 5 is used.

3. Preparing data for import

Up to this point excel has been used for data entry and initial data processing to create the accuracy measure. For consistency with the whole database this data would need to be imported into JMP, used to create efficiency measures and compare with results from other tasks.

3.1 Summary sheet for import

A summary sheet should be created in excel and the following columns from the various trials should be copied into it

- Participant ID
- School ID
- Badge Color
- Animal/Object
- Weightage (from sheet T1 and should be renamed to T1 Weightage)
- Pathscore (from sheet T1 and should be renamed to T1 Pathscore)
- Accuracy (from sheet T1 and should be renamed to T1 Accuracy)
- Exclude (from sheet T1 and should be renamed to T1 Exclusion)
- Weightage (from sheet T2 and should be renamed to T2 Weightage)
- Pathscore (from sheet T2 and should be renamed to T2 Pathscore)
- Accuracy (from sheet T2 and should be renamed to T2 Accuracy)
- Exclude (from sheet T2 and should be renamed to T2 Exclusion)
- Weightage (from sheet T3 and should be renamed to T3 Weightage)
- Pathscore (from sheet T3 and should be renamed to T3 Pathscore)
- Accuracy (from sheet T3 and should be renamed to T3 Accuracy)
- Exclude (from sheet T3 and should be renamed to T3 Exclusion)

Save the summary sheet as a .txt (tab delimited file).

3.2 jmp data management legend

In the text below, the formatting of the text is used to identify:

- Double quotes are used for indicating column and file names – note when you type in these names you should only type in what is in the double quotations and not the quotations themselves.
- Single quotes are used when you need to look for a specific bit of information in the JMP file -- again – the information in the file should not include the quotations.
- **Bold** is used to indicate JMP menus and submenus.
- ***Bold with italics*** is used to indicate text or options to be selected within a JMP dialog box.

3.3 Zoo_Alt_T1b

1) Open jmp and a new database (using control + N).
2) Import the data
    a. Go to: **File** → **Open** → Navigate to the tab delimited file and select it → **Open**.
3) Check the file if it has been imported properly
    a. The number of rows imported
    b. Check the column names and that they are have appropriate data types for later analyses
        i. Select all columns → **Right click** → **Column Info…**
            1. For “Participant ID”, “School ID”, “Badge Color”, “Animal”, “T1 Exclusion”, “T2 Exclusion” and “T3 Exclusion”
                a. ***Data Type*** should be set at ***Character***

b. ***Modelling Type*** should be set at ***Nominal***
c. Data/modelling type can be changed by clicking on the boxes with down arrows.

2. For "T1 Weightage", "T1 Pathscore", "T1 accuracy", "T2 Weightage", "T2 Pathscore", "T2 accuracy", "T3 Weightage", "T3 Pathscore" and "T3 accuracy"
   a. ***Data Type*** should be set at ***Numeric***
   b. ***Modelling Type*** should be set at ***Continuous***
   c. Data/modelling type can be changed by clicking on the boxes with down arrows.

c. Check for data that will later be excluded
   i. Go to **Table → Summary**
      1. ***Group*** by "T1 Exclusion", "T2 Exclusion" and "T3 Exclusion"
      2. Click ***OK***

4) Add data to "Zoo_T1bRaw.jmp" to create "Zoo_RawT1bAlt.jmp" using the JMP Update feature
   a. Open the appropriate "Zoo_T1bRaw.jmp" and "Zoo_Alt_T1b.jmp"
   b. Go to the "Zoo_T1bRaw.jmp" file
   c. Select **Table → Update**
      i. In the box on the upper left-hand side you should see the .jmp demographic file
      ii. Click on the tick box in front of ***Match Columns***
      iii. Select "subject_id" and "Participant_id" in the respective files, click on ***Match*** (now you'll see 'subject_id=participant_id' in the white box)
      iv. Under ***Add Columns from Update table***, click on the button in front of ***Selected***
      v. Select the following columns
         1. "T1 Weightage"
         2. "T1 Pathscore"
         3. "T1 Accuracy"
         4. "T1 Exclude"
         5. "T2 Weightage"
         6. "T2 Pathscore"
         7. "T2 Accuracy"
         8. "T2 Exclude"
         9. "T3 Weightage"
         10. "T3 Pathscore"
         11. "T3 Accuracy"
         12. "T3 Exclude"
      vi. Click ***Create***
   d. ***Save*** file as "Zoo_RawT1bAlt.jmp".

3.4 Zoo_SummaryAlt_T1b

This file will recompute the previously used coding to create MMON and MCON scores and use the new alternate accuracy coding to create efficiency scores.

1. Open "Zoo_RawT1bAlt.jmp" and ***save the file as*** "Zoo_SummaryAlt_T1b.jmp"

2. Deleting columns that are no longer going to be used ACC
    a. Select following columns from the list of columns
        i. "Trial1_ClearRoute"
        ii. "Trial1_TotalPoints"
        iii. "Trial2_ClearRoute"
        iv. "Trial2_TotalPoints"
        v. "Trial3_ClearRoute"
        vi. "Trial3_TotalPoints"
        vii. "ACC"
    b. Right click and click on ***Delete Columns***
3. Adding the recomputed "Trial#_ChangeStrategy" variable
    a. Open "ZooEntry_T1b_21Mar2016.xlsx"
    b. Copy paste the change strategy data from the "nominal scores for strategy change" into the jmp file.
    c. ***Rename*** the columns as T2_StrategyChange and T3_Strategy change and ensure that the ***data type*** is ***numeric*** and the ***modelling type*** is ***ordinal***.
4. Create new MCON and MMON columns
    a. Create 2 new columns after "Trial1_strategyType", "Trial2_strategyType" and "Trial3_strategyType" by righting clicking on the adjacent column header and clicking on ***insert column***.
    b. Double click on the columns, ensure the data type is numeric and modelling type ordinal. Rename the columns to
        i. "Trial1_MCON"
        ii. "Trial1_MMON"
        iii. "Trial2_MCON"
        iv. "Trial2_MMON"
        v. "Trial3_MCON"
        vi. "Trial3_MMON"
    c. Recomputing previously stored variables
        i. Right click on the MCON columns and click on ***formula…***
        ii. Click on "Trial#_StrategyPresent", ***plus symbol***, "Trial#_ListOrder", ***plus symbol***, "Trial#_StrategyChange" and ***Apply***.
        iii. Right click on the MMON columns and click on ***formula…***
        iv. Click on "Trial#_AllAnimalsSeen", ***plus symbol***, "Trial#_StartAndFinish", ***plus symbol***, "Trial#_GoingToDots" ***plus symbol***, "Trial#_AnyBacktracking" and ***Apply***.
5. Create Efficiency variable ("Trial#_EFF")

    The efficiency variable is generated by taking a ratio of "Trial#_Acc" and "RT"
    a. Double click on the "Trial#_EFF" and "Trial#_RT" columns and ensure that the data type is numeric and modelling type continuous.
    b. Right click on the EFF columns and click on ***formula…***
    c. Click on "Trial#_ACC", ***division symbol*** and ***Apply***
6. Standardizing variables for further calculations
    a. Select the following columns "Trial#_ACC", "Trial#_EFF", "Trial#_PreMetaQ", and "Trial#_PostMetaQ".
    b. ***Right click*** on one of the headers and go to ***New Formula Column***, ***Distributional*** and click on ***Standardize***.

   c. ***Rename*** the columns to "Std T# ACC", "Std T# EFF", Std T# JOK" and "Std T# RCJ",

7. Creating Judgement outcome measures
   a. Creating "JOK-BI" and "RCJ-BI"
      i. JOK-BI
         1. Create three temporary new columns named "T# JOK-ACC"
         2. Right click on the column headers and click on ***formula…***
         3. Click on "Std T#_JOK", ***minus symbol*** and "Std T# ACC"
         4. Create "JOK-BI"
         5. Right click on the column headers and click on ***formula…***
         6. Type ***Mean***, select the three "T# JOK-ACC" columns and apply.
      ii. RCJ-BI
         1. Create three temporary new columns named "T# RCJ-ACC"
         2. Right click on the column headers and click on ***formula…***
         3. Click on "Std T#_RCJ", ***minus symbol*** and "Std T# ACC"
         4. Create "RCJ-BI"
         5. Right click on the column headers and click on ***formula…***
         6. Type ***Mean***, select the three "T# RCJ-ACC" columns and apply.
   b. Creating "JOK AbsAcc" and "RCJ AbsAcc"
      i. JOK AbsAcc
         1. Create three more temporary new columns named "T# JAbsAcc"
         2. Right click on the column headers and click on ***formula…***
         3. Click on "T#_JOK-Acc", ***multiplication symbol*** and "T#_JOK-Acc"
         4. Create "JOK AbsAcc"
         5. Right click on the column headers and click on ***formula…***
         6. Type ***Mean***, select the three "T# JAbsAcc" columns and apply.
         7. Delete the three temporary columns developed in step 7ai. and the three developed in 7bi.
      ii. RCJ AbsAcc
         1. Create three more temporary new columns named "T# RAbsAcc"
         2. Right click on the column headers and click on ***formula…***
         3. Click on "T#_RCJ-Acc", ***multiplication symbol*** and "T#_RCJ-Acc"
         4. Create "RCJ AbsAcc"
         5. Right click on the column headers and click on ***formula…***
         6. Type ***Mean***, select the three "T# RAbsAcc" columns and apply.
         7. Delete the three temporary columns developed in step 7aii. and the three developed in 7bii.
8. Reorder the various columns
   a. Select columns that you want to move, click on ***columns***, click on ***move selected columns***
   b. Match the following order of columns

1	Cohort	2	Timepoint	3	subject_id	4	Colour
5	Object	6	DoB	7	T1 DoT	8	T1 Age (years)
9	Gender	10	Ethnicity	11	Lunch	12	School
13	Grade	14	Mindset Condition	15	Consent	16	SchoolID (ZooEntry)
17	Color (ZooEntry)	18	Animal (ZooEntry)	19	PracticeTrial_1	20	PracticeTrial_2
21	Trial1_PreMetaQ	22	Trial1_PostMetaQ	23	Trial1_RT (mm:ss.s)	24	Trial1_AllAnimalsSeen

25	Trial1_StartAndFinish	26	Trial1_GoingToDots	27	Trial1_ListOrder	28	Trial1_AnyBacktracking
29	Trial1_StrategyPresent	30	Trial1_StrategyType	31	Trial1_MCON	32	Trial1_MMON
33	T1 Weightage	34	T1 Path Score	35	T1 Accuracy	36	T1 Exclusion
37	Trial2_PreMetaQ	38	Trial2_PostMetaQ	39	Trial2_RT (mm:ss.s)	40	Trial2_AllAnimalsSeen
41	Trial2_StartAndFinish	42	Trial_GoingToDots	43	Trial2_ListOrder	44	Trial2_AnyBacktracking
45	Trial2_StrategyPresent	46	Trial2_StrategyType	47	T2_StrategyChange	48	Trial2_MCON
49	Trial2_MMON	50	T2 Weightage	51	T2 Path Score	52	T2 Accuracy
53	T2 Exclusion	54	Trial3_PreMetaQ	55	Trial3_PostMetaQ	56	Trial3_RT (mm:ss.s)
57	Trial3_AllAnimalsSeen	58	Trial3_StartAndFinish	59	Trial3_GoingToDots	60	Trial3_ListOrder
61	Trial3_AnyBacktracking	62	Trial3_StrategyPresent	63	Trial3_StrategyType	64	T3_StrategyChange
65	Trial3_MCON	66	Trial3_MMON	67	T3 Weightage	68	T3 Path Score
69	T3 Accuracy	70	T3 Exclusion	71	Notes	72	T1_EFF
73	T2_EFF	74	T3_EFF	75	Std T1 Acc	76	Std T2 Acc
77	Std T3 Acc	78	Std T1 EFF	79	Std T2 EFF	80	Std T3 EFF
81	Std T1 JOK	82	Std T2 JOK	83	Std T3 JOK	84	Std T1 RCJ
85	Std T2 RCJ	86	Std T3 RCJ	87	JOK-BI	88	RCJ-BI
89	JOK AbsAcc	90	RCJ AbsAcc				

9. Remove cases to be excluded
    a. Select the down arrow next to **Rows** (in the lower left-hand corner of the JMP window) → **Row Selection** → **Select Where …**
    b. In the dialog box, select "Trial#_Exclusion" from the left-hand box (which lists the column names)
    c. Go to the box to the right of that, it should say ***equals***, if it does not, then click on the down arrow and select ***equals***.
    d. In the text box to the right of that, type in 'EXCLUDE'
    e. Click ***OK***
        i. The appropriate rows should now be highlighted on the data file. Scroll down to do a quick spot check to confirm.
        ii. Select the down arrow next to **Rows** (in the lower left-hand corner of the JMP window) → **Delete Rows**

    In this dataset 4 cases have been excluded.

    f. Summarize the table to make sure that all participants have provided consent by going to ***tables, summary***, clicking on "consent", ***group*** and ***okay.***