The Domain-general vs. Domain-specific Role of Metacognition in Arithmetic Achievements

Naima Nigar

A thesis submitted to the University of Nottingham, UK, in partial fulfilment of the degree of PhD in Psychology.

Supervised by

Dr. Lucy Cragg
Professor
School of Psychology
University of Nottingham, UK

Dr. Claudia Danielmeier
Associate Professor
School of Psychology
University of Nottingham, UK

October 2023

Acknowledgments

First, I would like to show my earnest appreciation to the almighty, who enabled me to accomplish this research.

With immense gratitude, I acknowledge the support and help of my PhD supervisors- Dr. Lucy Cragg and Dr. Claudia Danielmeier, School of Psychology, University of Nottingham, UK, for their enthusiasm, inspiration, and constant advice. They provided intellectual suggestions, good teaching, and good company from the beginning to the end of my thesis. With their informed guidance, I could carry out this research properly. I am pleased to work with them.

I am grateful to the University of Nottingham, UK, people who created a friendly environment for international students.

I want to thank those people, especially my teachers and colleagues from the University of Dhaka, Bangladesh, for their teaching and support. I am also grateful to the authors, researchers, and publishers. I learned many things about error monitoring research from their publications and got valuable information for writing this thesis.

My special thanks go to the schools in the UK and the guardians of the child participants for their cooperation and enthusiastic participation in the study. I also want to thank my husband, who helped me with coding; it would be difficult without his support. My special thanks to my friends who motivated me to complete my thesis during the pandemic.

Finally, I want to thank my parents and elder sister. Without their support and loving environment, it would not be possible for me to come this far. It is my pleasure to dedicate this small piece of work to them.

Thesis Abstract

Error monitoring and awareness associated with meta-cognition and inhibitory processes of cognitive control also play a role in academic achievement (Young & Fry, 2008; Hirsh & Inzlicht, 2010). Despite playing an important role in academic achievement, research has yet to investigate the relationship between error monitoring and awareness and arithmetic achievement. The role of metacognition on academic performance could be domain-general vs. domain-specific (Rinne & Mazzocco, 2014), depending on the participant's age (Bellon et al., 2020). In the present study, I developed some numeric and non-numeric error monitoring and awareness tasks to test their domain-general vs. domain-specific role in arithmetic achievement. Here, error monitoring was measured by post-error slowing (PES) and post-error change in accuracy (PECA), which are the parts of behavioural adjustments after committing an error. Past research found that post-error behavioural adjustments are modulated by response stimulus intervals (RSI; Danielmeier & Ullsperger, 2011). The RSI was also manipulated in a series of short vs. long RSIs in the error monitoring tasks. Error awareness was measured as a proportion of aware errors in the numeric and non-numeric error awareness tasks.

Three studies were conducted. The first two studies (Chapters II and III) explored the relationship between error monitoring and arithmetic achievement. The third study (Chapter V) explored the relationship between error awareness and arithmetic achievement. Data were collected from children and adults to investigate the relationship between error monitoring and arithmetic achievement. However, data were only collected from adults for the error awareness study.

It is found from the present research that RSI has a significant effect on PES and PECA in adults. PES was higher under short RSI(s) than long RSI(s). The effect of RSI on PES was not viable to investigate in children due to the problem of recruiting children during the pandemic. There were also some issues with the task design, discussed in Chapter IV. However, PECA was affected by task types and RSI in adults. No improvement but decrement in accuracy was observed after error trials, and accuracy got significantly lower under short RSI. The decline in accuracy after error trials was higher for the numeric error monitoring task.

Children with better numeric error monitoring ability tended to show better arithmetic achievement. However, adults with better error monitoring showed a cautious response (Slower response) trend in solving arithmetic problems. This relationship was only found to be significant for children in Study I (Chapter II). Study 3 in Chapter V showed that adults who were more aware of their errors in the non-numeric error awareness task also showed higher arithmetic efficiency. However, this relationship was only significant when numeric error awareness was added to the model. On the contrary, adults who showed poorer arithmetic efficiency were more aware of their errors in the numeric error awareness task.

It could be concluded from this thesis that error monitoring and awareness is a complex process. More extensive research is needed in this area, especially with children. It is essential to consider the processing speed due to practice effect and response bias while designing error monitoring and awareness tasks.

Contents

List of Tables

List of Figures

1. Chapter I: Metacognition and Arithmetic Achievements

Introduction

Early mathematical ability is essential for later life achievement (Stevenson & Stigler, 1992; Young-Loveridge, Peters, & Carr, 1997; cited by Aubrey et al., 2006). Studies have shown that childhood mathematical ability is associated with academic motivation, intelligence, and obtaining qualifications (Ritchie & Bates, 2013). Numerical ability is associated with better financial decision-making (Agarwal & Mazumder, 2013), and it is also essential to maintaining health and making informed medical decisions (Reyna et al., 2009).

Due to the importance of numerical knowledge in our daily lives and later achievements, psychologists studying mathematical cognition try to comprehend the underlying mental processes that people use to understand different mathematical ideas over developmental periods. They also try to explain why there are individual differences in arithmetic performance (Gilmore, Göbel, & Inglis, 2018). One mental process that may underpin successful mathematics learning and contribute to individual differences is metacognition, defined as “thinking about thinking” (Flavell, 1979). When we try to solve an arithmetic problem, it is necessary to know how to solve the problem or which strategy to use and monitor our performance for error checks and performance improvement. That is why it is essential to understand metacognition before testing how it could be related to better mathematical skills. In the present thesis, I tried to understand how metacognitive processes of cognitive function are related to mathematics. In the following sections, I explain metacognition, its different models, and its relation with mathematics.

Chapter I

Metacognition

One of the difficulties researchers faced was defining metacognition from different perspectives. Here, I tried to comprehensively cover all the aspects of metacognition to understand its role in cognition and education.

The first person to coin the term metacognition was Flavell (1979). In simplest terms, he defined metacognition as “thinking about thinking.” The operational definition of metacognition is “Knowledge concerning one's cognitive process and products, or anything related to them, e.g., the learning-relevant properties of information or data... Metacognition refers, among other things, to the active monitoring and consequent regulation and orchestration of these processes about the cognitive objects on which they bear, usually in the service of some concrete goal or objective.” (Flavell, 1979, p.232)

After Flavell (1979), many researchers tried to define metacognition, some of which are presented in Table 1.1.

Table 1.1 Definitions of Metacognition

“The knowledge and control children have over their thinking and learning activities.”	Cross & Paris, (1988)
Self-appraisal and self-engagement: personal reflections about knowledge states and abilities; metacognitive action, or how metacognition can orchestrate cognitive aspects of problem-solving.”	Paris & Winograd (1990)
“Awareness of one’s thinking, awareness of the content of one’s conceptions, an active monitoring of one’s cognitive processes, an attempt to regulate one’s cognitive processes	Hennessey (1999)

about further learning, and an application of a set of heuristics as an effective device for helping people organize their methods of attack on problems in general.”	
‘Metacognitive’: (a) the thinking subject has some knowledge about his thinking and that of other persons; (b) the thinking subject may monitor and regulate the course of his thinking, i.e., may act as the causal agent of his thinking.”	(Papleontiou-Louca, 2003)
“Awareness and management of one’s thought”	Kuhn & Dean, (2004)
“The monitoring and control of thought”	Martinez, (2006)

It was observed from these definitions that metacognition includes both the awareness of one’s knowledge and the thinking process. Two popular metacognitive models are briefly discussed below to give a clearer idea about metacognitive knowledge and process.

Metacognitive Models

The classic model of Flavell (1979)

Flavell proposed the first classical model of metacognition in 1979, where he mentioned four phenomena of metacognition to understand the metacognitive process: (a) Metacognitive knowledge, (b) Metacognitive experiences, (c) Metacognitive goals (or tasks), and (d) Metacognitive actions (or strategies). These four components interact in an overlapping, complex manner.

In a simple form, metacognitive knowledge is a person’s knowledge about the world and self, as well as awareness about their cognitive processes to solve diverse cognitive tasks by using appropriate cognitive strategies (Flavell, 1979). On the other hand, metacognitive experience is conscious cognitive or affective experiences used to control cognitive activities to ensure goal attainment (Flavell, 1979; Livingstone, 2003). An example of metacognitive knowledge and experience would be- when a student knows that she is better at arithmetic than reading; this is called metacognitive knowledge. However, when she talked about different arithmetic problems with her friends, she realized some friends were better than her. This later awareness is called metacognitive experience. In Flavell (1979), a metacognitive goal refers to a cognitive enterprise's objectives, and action refers to the observed and unobserved (cognition) behaviour used to attain goals. Now, if someone is asked which number is larger, 8 or 3? The goal here is to identify the larger number. Now, anyone from their prior numeric knowledge could say that eight, which can be called metacognitive knowledge. What if an individual needs to solve a sum problem, i.e., 2 + 2 =? and calculate the mean of a number of series (2, 3, and 5)? For the first problem, they must apply the basic rules of addition. However, to solve the second problem, they need to be aware that only adding the number would not give the correct answer; they need to divide the summation by the total number of cases, which is three. The process of differentiating these two processes is called metacognitive experience.

According to Flavell (1979), metacognitive experience plays a role in developing and improving metacognitive knowledge by monitoring and regulating one’s cognition. It is believed that metacognitive experience could occur during task performance, which monitors task characteristics and helps to maintain task goals by taking measures on strategic components (Moritz & Lysaker, 2018). The metacognitive experiences, goals, and action components overlap with each other. I discussed these three components and how they help solve problems and make decisions through monitoring and regulation. Let us give an arithmetic example to understand how metacognitive experience, goals, and actions work together. A student is asked to solve an arithmetic problem: x= y+10 and y= 2x+5 by finding the x and y values. Here, the goal is to find the value of x and y. Metacognitive action would be to use different arithmetic strategies (addition, subtraction, multiplication, and division) in a step-by-step manner to solve the problem. Metacognitive experience would monitor and regulate our ongoing cognitive processes by detecting errors and re-evaluating metacognitive strategies to solve the problem. This self-regulatory technique helps us update our knowledge through metacognitive experience, goals, and actions.

Nelson and Narens's (1990) model of metacognition

Looking at the classical model, we would notice that so many components were used by Flavell (1979) to explain metacognition through their complex transaction. Nelson & Narens (1990) provided a model of metacognition to simplify the metacognitive process for laboratory experiments. Their proposed model is presented in Figure 1.1.

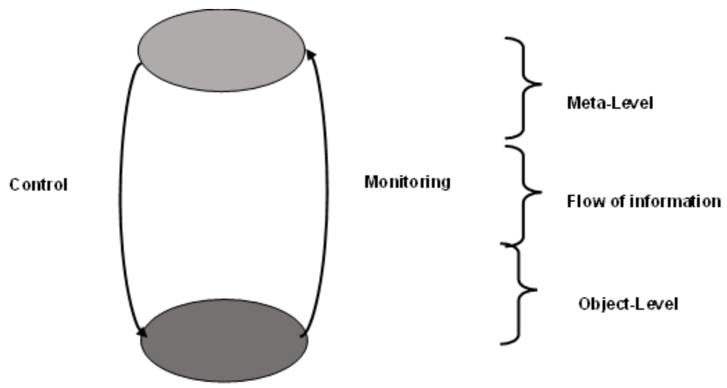


Figure 1.1 Nelson and Narens' (1990) formulation of a meta-level/object-level theoretical mechanism consisting of two structures (meta-level and object-level).

The metacognitive system has two interrelated levels, object and meta-level, with two metacognitive processes, monitoring and control. The object level contains the essential information stored in our memory through encoding, learning, and remembering. The object-level is one's current cognitive processes monitored and controlled at the meta-level (Jia, Li, & Cao, 2019). In the meta-level, the information originating from the object level is processed with top-down regulation of object-level functions (Fleur et al., 2021). Monitoring refers to the process by which the metalevel traces the accuracy of object-level performance to inform the meta-level what is occurring at the object level. In contrast, control refers to the processes by which the meta-level regulates the operation of object-level processes toward achieving different goals (Koriat, 2019). How the information between object-level and meta-level occurs is described below-

Control. It is observable from Figure 1.1 that control does not yield any information from the object level. If we look at the picture, we can see that information flow from the meta-level modifies the object level. The changes could be of two types: changes in the state of the object-level process or in the object-level process itself. Usually, some actions occur during metacognition's control process: 1) Initiate an action, 2) continue an action, and 3) terminate an action. However, this way, information processing does not work the other way around, that is, change in meta-level due to object-level.

Monitoring. On the other hand, in the monitoring aspects of metacognitive information flow, from the object level to the meta-level. There could be a change in the situation at the meta-level, or maybe not. Mostly, it is a time-dependent change that we need to make at the meta-level to make changes at the object level later when information flows from the meta-level to the object level. In the monitoring level, a person will monitor their action and performance, but initiation of action will occur in the control level of metacognition.

Now, let us think about a hypothetical situation to understand the control and monitoring aspects of information flow from meta-level to object-level and from object to meta-level. Think about a child solving simple arithmetic addition and multiplication problems with two multiple-choice answers. If the problem is an addition (2+5), the given multiple-choice answers are 7 and 10. If we think about the object level, the information about the characteristics of these numbers and mathematical strategies is stored at the object level. To solve this mathematical problem, the child must initiate some actions and choose to answer 7. After gathering information from the object level, the child might choose a right or wrong answer. If the answer is correct, then the monitoring process of metacognition would inform the meta-level that the correct answer has been chosen and to continue solving the following problem. However, instead of choosing the correct answer, if they choose the wrong answer, then the child needs to terminate their action to correct their given response. In this example, the actions of initiation, continuation, and termination are the control aspects of metacognitive information flow. However, detecting errors and evaluating answers are the monitoring aspects of metacognition.

Nelson and Narens (1990) also proposed that metacognitive monitoring in memory level could be either retrospective (i.e., confidence judgment about previous response) or proactive (i.e., judgment about future response) in nature. This judgment of performance during metacognitive monitoring could be divided into three types, discussed below.

The first one was the ease of learning (EOL). It predicts if the learning would be easy or difficult to learn and which items would be easier to learn (Underwood, 1966; cited by Nelson & Narens, 1990) or following which strategies would make the learning easier (Seamon & Virostek, 1978; cited by Nelson & Narens, 1990). This type of judgment usually occurs before learning and is mainly inferential; judgments occur before the acquisition.

The second one was Judgements of Learning (JOL). This kind of judgment usually occurs during or after knowledge acquisition, and predictions about future test performance based on presently studied items are made.

Finally, the third one was Feeling-of knowing (FOK). This kind of judgment usually occurs during or after the acquisition of present learned items, and future prediction is usually made to see whether a given currently non-malleable item is known or will be remembered on a subsequent retention test (e.g., empirical investigations of the accuracy of FOK judgments in a subsequent retention test).

Suppose we could make a comparison between these two models. It could be assumed that Nelson and Narens's (1990) object-level knowledge about the control process could relate to metacognitive knowledge components proposed by Flavell (1979). On the other hand, Nelson and Naren’s meta-level, control and monitoring process could be compared to Flavell’s (1979) metacognitive experiences, goals, and action aspects. In a simple form, the whole metacognitive process could be divided into metacognitive knowledge and metacognitive regulation components (Livingston, 2003). Metacognitive knowledge is about the knowledge of cognition and all its monitoring processes. Metacognitive regulation involves examining cognitive knowledge and its processes through controlled planning, action, and behaviour adaptation based on the outcome to meet task goals (Fleur et al.,2021).

One point is clear from the above discussion: researchers tried to define and explain metacognition in different terms. If these definitions and models are scrutinized, metacognitive components could be divided into sub-components within metacognitive knowledge and regulation components. Different researchers raised concern over the vague ideas about these metacognition sub-components, which could confuse research objectives and perplex the research findings (Brown, 1987; cited by Baker et al., 2000). According to White (1988; cited by Baker et al., 2000), the researcher must mention which area of metacognition they want to cover to make the research objectives more straightforward. That's why it is crucial to understand how different components are related to metacognition's knowledge and regulation aspects. In her 2011 review paper, Lai showed that different cognitive processes and experiences could be divided into two major metacognitive components: Knowledge of Cognition and Regulation of Cognition. Unlike Lai, here I argue that the two major components should be metacognitive knowledge and regulation since the fundamental notion of metacognition is the awareness of our cognitive knowledge and processes (Table 1.2). If we inspect the above table, we will see that different researchers introduced different terminology to reflect different aspects of metacognitive knowledge. In the classical model of metacognition, Flavell (1979) explained that metacognitive knowledge develops from knowledge about self, task, and strategy, which eventually helps people develop a global knowledge of cognitive enterprise.

Table 1.2 Types and Terminology of Metacognitive Component (Adapted from Lai, 2011)

Metacognitive Component	Constructs	Types	
Metacognitive Knowledge	Knowledge about oneself as a learner and factors affecting cognition	Person and task knowledge	Flavell, 1979
		Declarative knowledge	Cross & Paris, 1988; Schraw, & Moshman, 1995
	Awareness and management of cognition, including knowledge about strategies	Procedural knowledge	Cross & Paris, 1988; Schraw, & Moshman, 1995
		Strategy knowledge	Flavell, 1979
	Knowledge about why and when to use a given strategy	Conditional knowledge	Garner, 1990
Metacognitive Regulation	Identification and selection of appropriate strategies and allocation of resources	Self-regulatory Planning	Cross & Paris, 1988 Whitebread et al., 2009
	Attending to and being aware of comprehension and task performance	Monitoring	Cross & Paris, 1988 Schraw et al., 2006
		Metacognitive Experiences	Flavell, 1979
	Assessing the processes and products of one's learning and revisiting and revising learning goals	Evaluating	Cross & Paris, 1988 Schraw et al., 2006

It could be observed from the table that some researchers divided them into procedural, declarative, and conditioned knowledge. Declarative knowledge is self-knowledge and what influences performance while doing a task or learning (e.g., Cross & Paris, 1988; Schraw & Moshman, 1995). On the contrary, procedural knowledge means the knowledge of the execution of procedural skills or how skills are operated or applied (e.g., Cross & Paris, 1988; Schraw & Moshman, 1995). Conditional knowledge refers to knowing when and why to use declarative and procedural knowledge (Garner, 1990). Thus, the knowledge of cognition is related to all the knowledge we gain, which helps us create a global knowledge about self and the world.

Similarly, different terminologies were introduced by researchers to explain different metacognitive regulatory processes. For example- in the classical model of metacognition, Flavell (1979) talked about metacognitive experiences, which can affect metacognitive knowledge by monitoring one's knowledge about self, task, and strategy. It can add, delete, and revise the existing cognitive knowledge to create a cognitive enterprise. This monitoring and evaluation could occur before, during, or after the development of cognitive enterprise. In their model, Nelson and Naren (1990) also explained how information flow from object-level to meta-level controls the cognitive processes through initiation, action, evaluation, and termination to achieve desirable outcomes. Other researchers focused on self-regulatory aspects of metacognition, i.e., the ability to plan, monitor, and evaluate ongoing cognitive processes (e.g., Cross & Paris, 1988; Schraw et al., 2006; Whitebread et al., 2009). Metacognition is our awareness of cognitive knowledge and self-regulatory ability to monitor our cognitive processes. It is clear from the above discussion that no matter how one defines or divides metacognition into different aspects, there are two major components of metacognition- Metacognitive knowledge and regulation. The current thesis focused primarily on the regulatory aspect of metacognition. My initial aim was to investigate whether participants could monitor and evaluate their performance to change their future performance.

Chapter I

Assessment Methods in Metacognition

So far, the discussion has focused on the definition and models of metacognition. Now, I am going to discuss various measurements of metacognition. Various metacognitive measurement procedures are available, but the best measure for research depends on which metacognitive component the researcher is interested in investigating.

The most common ways to measure metacognitions are:

- Observational methods: This is more applicable to children who are eight or under eight years old (e.g., Gascoine et al., 2017)
- Verbal reports: Including questionnaires (Boekaerts & Corno, 2005)
- Survey interviews (Perry, 2002), think-aloud techniques (Ward & Traweek, 1993), and metacognitive judgment of accuracy (e.g., Rinne & Mazzocco, 2014).
- Error-detection approach: The most common approach used to assess comprehension monitoring in reading, where readers identify the errors embedded in the texts they are reading (e.g., Baker, 1979). Another method is studying error monitoring with post-error adjustment behaviours (Danielmeier & Ullsperger, 2011). Finally, another approach is measuring error awareness with a verbal report of committed error (e.g., O'Connell et al., 2007).

These are primarily measurement procedures used in the metacognitive study. However, error monitoring and awareness are the two techniques that are the focus of my thesis, which I have discussed briefly below.

Chapter I

Error Monitoring and Awareness

People tend to slow down after making an error, which is associated with our brain mechanism activated during error monitoring. For example- Overbye et al. (2019) showed stronger error-related negativity, which can be measured as an event-related brain potential reflecting the medial frontal action-monitoring processes, is associated with a slower response in the subsequent trial after an error. On the other hand, if our judgment aligns with our performance, we become confident enough to make decisions. This awareness of our mistakes helps us avoid future errors (Yeung & Summerfield, 2012). The detection and awareness of the error, then slowing down to improve our future performance, are related to metacognition's control and monitoring process.

Error monitoring can be measured by investigating post-error behavioural adjustment. There are three types of post-error behavioural adjustment: post-error slowing (PES), post-error reduction in interference (PERI), and post-error change in accuracy (PECA; Danielmeier & Ullsperger, 2011). I am interested in investigating PES and PECA as metacognitive measures in the current thesis. PES is defined as our tendency to slow down after committing an error response, i.e., the reaction time (RT) is more extended in the trial after an error response than after a correct response. On the other hand, PECA is usually defined as a change in accuracy (mostly improvement) after error trials compared to after the correct trials. Most research investigating the error-monitoring process to measure metacognition used forced choice-reaction time tasks (e.g., Flanker task; Eichele et al., 2010) where questions do not interrupt participants' performance.

The second variable I was interested in measuring in the present thesis is error awareness. It is known from the definition of error awareness that it is the ability to perceive one's mistakes consciously (Klein et al., 2013). In most metacognitive experiments, the researcher used the calibration of judgment techniques to measure metacognitive awareness of performance. The calibration of judgment can be defined as the alignment between one's confidence in judgment in a problem and the accuracy of the judgment (e.g., Rinne & Mazzocco, 2014). The experiment using this measurement procedure usually asked participants to rate their confidence in the accuracy of their responses after each trial (Bellon et al., 2020). In measuring error awareness, instead of asking participants about their confidence in their performance after each response, participants are asked to signal their error with a rating scale or button press (e.g., Endrass et al., 2005; O'Connell et al., 2007).

All the assessment approaches listed above are the most used methods in the field of metacognitive research, and they each have unique strengths and drawbacks. However, each procedure is well suited for each study based on the research questions the researcher wanted to answer. We briefly evaluated different assessment procedures in the section below to justify the current thesis's selected methods.

Evaluation of different metacognitive assessment methods

If researchers want to measure metacognitive awareness in younger children, especially for cases involving children aged eight or below, then the observational method could be the best choice. This technique is widely used among younger children who cannot report their metacognitive skills due to limited verbal abilities (Chua et al., 2006).

Most researchers rely on questionnaires and interviews to measure overall metacognitive knowledge. This assessment method had the advantage of testing reliability and validity, and it is administered to a large population at a time. However, it had some pitfalls, too. One primary concern was that participants might be reluctant to express their ideas or experiences if they were young children. Some children might need help understanding the questionnaire due to immature language development. Response bias might be challenging to eliminate for some questionnaires, such as the Motivated Strategies for Learning Questionnaire (MSLQ; Boekaerts & Corno, 2005). Survey interviews are another metacognitive assessment method where the examiner asks questions to induce the subject's metacognitive knowledge and awareness of self-regulatory strategies at various tasks. The problem with interviews is that this procedure is not structured, and the scoring procedure for the interview technique could be complicated if it were open-ended (Ward & Traweek, 1993).

However, if someone is interested in measuring the monitoring and control aspects of metacognitive, think-aloud, metacognitive judgment rating, and error-detection are the preferred approaches. In the think-aloud procedure, participants verbally reported all the experiences that came to mind while performing the task (Ward & Traweek, 1993). In metacognitive judgment, participants reported their confidence level in the accuracy of their responses after each trial (Bellon et al., 2020).

Both methods have rich verbalization data from students. Compared to the structured questionnaire, the advantage of these techniques is that the researcher can get data about students' ongoing thoughts and feelings as they occur rather than recalled after doing the task. However, these methods rely heavily on verbal reports

(Whitebread et al., 2009). Despite having advantages, these verbal report techniques have some disadvantages. Firstly, verbal reports are not an ideal measure of metacognition for younger children. Due to their limited vocabulary, it isn’t easy to describe their immediate and inner thoughts during the task. Managing a dual task requires much practice—tasks must be novel, complex, and difficult enough to measure metacognitive skills. This might also create a cognitive load, task interferences, and response biases while reporting metacognitive thoughts (Ward & Traweek, 1993; Boekaerts & Corno, 2005). Finally, it is not easy to measure implicit metacognitive processes with verbal report techniques (Lai, 2011).

Another approach for metacognitive testing is the error-detection approach (Baker et al., 2000). This method was widely used in the 1990s for listening or problem-solving mathematics tasks (e.g., Baker, 1984). In this method, the errors/problems are embedded in the test materials, and participants must detect the errors either during or at the end of the task. The problem with this approach was that there was significantly less chance that people could detect the error; only 38% of the errors were detected, and fewer than 25% were reported to have been noticed during reading (Baker & Cerro, 2000). Even though it was a widely used method in the field of metacognition in past research, detecting errors in embedded learning materials has limitations. Firstly, Baker and Cerro (2000) emphasized that when readers are informed about the text’s problems, differences in their comprehension monitoring occur. In addition, readers may use different criteria to detect errors and evaluate their understanding of the text. However, failure to detect errors in a text does not necessarily mean poor comprehension. Moreover, the error detection paradigm also needs ecological validity; regular life texts are not intentionally embedded with errors (OZTURK, 2017).

On the other hand, in the error monitoring approach, participants' behaviour adjustment, i.e., slowing down after the error and changes in performance after the error, is investigated (Jentzsch & Dudschig, 2009). This process has advantages since participants are not interrupted by their verbal reporting of metacognitive processes. Researchers can investigate the implicit metacognitive processes by measuring post-error behavioural adjustments in participants. Moreover, due to a lack of reliance on verbal reports, the measurement of post-error behavioural adjustments can also be investigated in children. However, this technique makes it difficult to measure someone's ongoing thought process. The metacognitive judgment process could easily measure ongoing thought processes. In this process, participants are asked to flag their correct or error response with a button press or a rating scale while performing a task (e.g., O'Connell et al., 2007; Bellon et al., 2019). The advantage of the error awareness task is that the researchers can measure metacognition as an ongoing process. On the other hand, assessing performance after each trial can interrupt the cognitive aspects of monitoring, as reported above for verbal reporting. Another disadvantage is that reliance on verbal reporting makes it difficult to apply to younger children.

Developmental Trajectory of Metacognition

In this thesis, child participants are recruited, and here, I will discuss how metacognition evolves over the developmental period. Since measurement of metacognition requires knowledge about the self and surrounding environment and the ability to regulate one's performance, it is hard to decide precisely at which age metacognition emerges in children. Another problem is associated with measurement methods. Different metacognitive measures are available and mostly rely on verbal abilities, so it is hard to measure metacognition in toddlers. Finally, the different aspects of metacognition could emerge at different ages of life. Still, due to their different measurement methods being age-sensitive, it is hard to decide which aspect developed at what age.

This problematic nature of measuring metacognition led researchers to varied conclusions. Many researchers pessimistically believed that young children were impoverished in reporting what they knew or did not know about learned knowledge. Based on this belief, it was assumed that metacognition is a late-developing skill and typically does not develop before children enter school (Roebers, 2014; Schneider, 2015; as cited in Roebers, 2017). Research evidence, however, showed that children's metacognitive ability could develop in their early developmental years before starting school. For example, Whitebread et al. (2009) conducted a study on children 3-5 years old. This study observed 582 videos of children with metacognitive or self-regulatory videotaped events. They concluded from their observation that children as young as 3 to 5 exhibited verbal and nonverbal metacognitive behaviours during different problem-solving stages, including articulation of cognitive knowledge, cognitive regulation, and emotion regulation.

The relationship between age and the component of metacognition is found to be complex. For example- Schneider (2008), in his review paper, discussed that declarative metacognitive knowledge shows steady improvement through childhood and adolescence because of increased knowledge of learning strategies; the results could be more similarly clear-cut for procedural metacognition. Significant cross-age differences were observed for self-control activities but not monitoring abilities.

Baker (1989), in their review paper on metacognition, comprehension, and monitoring in adult readers, concluded that older children have a better metacognitive understanding than younger children; such skills for understanding are fully developed by adulthood. However, the influence of educational level on this relationship with age and metacognitive understanding could be a possibility.

However, mixed findings on the effect of age on metacognitive monitoring and control were observed between children (5 and 7 years old) and adults (O'Leary & Sloutsky, 2019). One of their experiments found that both seven years old and adults selected the easier task to maximize performance and minimize the effort but not the five years old, even though the game difficulty level was not informed. However, the estimation of the accuracy performance was better in 7-year-olds than in five-year-olds or adults. However, this effect of age on performance monitoring was diminished when participants were reported about the task's difficulty level beforehand. From these research findings, it could be said that age's effect on metacognitive monitoring depends on the nature of the task.

Since, in the present thesis, I exclusively focused on error monitoring and awareness aspects of metacognition, I discussed a few studies associated with the developmental trajectory of error processing in the following section.

Developmental trajectory of error processing

This section discusses the studies associated with the role of developmental courses in error monitoring. One part of my thesis research focused on the relationship between error monitoring and arithmetic achievement in children and adults. Different studies showed that there is an age effect on post-error behavioural adjustment. It was found from different studies that post-error behavioural adjustments differ in different age groups. For example, Gupta et al., in their 2009 study, recruited children between the ages of 6 and 11 to perform two number-digit tasks. In this task, there was one block with single-digit numbers (1 or 3) and one with a three-digit number (111 or 333). The numbers were presented on the screen's left or right side. For the former block, they had to determine "what number?" and in the later block, they had to determine "How many numbers?". The difference between fast and slow trials was evaluated by computing RT in the 5th and 95th percentile values for different conditions. It was found from their results that the major development in error processing as measured by PES takes place between the years 6 and 10. With an initial increase in PES between 6 to 7 years of age, then a decrease in the following ages. The decrease between the ages of 7 to 8 years and 9-10 years is not uniform. With some decrease in PES between the ages of 7 and 8, there was a substantial reduction in PES between the ages of 9 and 10. They concluded that the reason for this non-linear development of PES was more likely associated with the response RT that varied between fast and slow trials for different age groups. It was evident from their research that RT in speedy trials increased for the age group 6 to 7 years old and then decreased for the age groups 7 to 8 and 9 to 10 years of age. However, RT values for slow trials decreased only from 9 to 10 years of age. PES is positively correlated with RT in fast and slow trials. Thus, the effect of age on PES is response RT dependent on fast and slow trials under different conditions.

A similar effect of age on PES and RSI was observed in the study of Smulders et al. (2016), which found that children show reduced PES as they age. In their study, they conducted two experiments with different age groups ranging from 5 to 25 years. In these experiments, participants performed a two-choice RT task with RSI manipulated among 50 ms, 150 ms, 200 ms, 250 ms, 500 ms, and 1000 ms. In Experiment-I, developmental increases in the speed of responding and developmental decreases on PES, and the effect of RSI were observed. However, this developmental trend on PES was diminished when speed differences across the age group were controlled. In addition to this, no RSI or interaction effect between RSI and age group was observed on PES. In experiment II, they ran the same experiment in four separate sessions with the same task but with larger trial numbers and two RSIs (50 ms & 500 ms) to control the response speed in different age groups. It was found from their experiment II that PES decreased with age, but in addition, PES's effect was significantly larger under shorter than longer RSI. This finding is consistent even after controlling the age effect on the response speed.

Another study (Masina et al., 2018) compared error awareness and PES in three different age groups: children (8-13 years), young adults (19-35), and older adults (61-83 years). This study found no significant differences in PES for each age group, but error awareness differed in different age groups. They modified the Go/No-go Stroop task as an Error Awareness Task (EAT). It was found from their study that younger adults were more aware of their commission errors compared to children and older adults. No significant differences were found comparing mean error awareness in children and older adults. However, PES slowing was similar for the Go-no-Go task across all the groups. They explained that the children age group they selected for their study already developed a mature error processing system like adults. In the case of error awareness, differences in different age groups are associated with an impaired attentional process in children and older adults.

In a recent online study (de Mooij et al., 2022), a group of researchers investigated the effect of PES on an online adaptive mathematics and language learning environment in children aged between 5-13 years old. The data were collected from an adaptive learning platform (Prowise Learn) with mathematics and language skills games. It was adaptive based on the accuracy and speed level of the students. There were three difficulty levels, and students could choose their level based on their motivation. It was found from their study that PES positively correlated with post-error accuracy. PES was higher when the time pressure was less in the task, participants showed a faster RT to the errors, and the game focused on mathematical rather than language skills. Individually, students who chose the most challenging level of task to practice and had higher skill ability also showed higher PES. Finally, non-linear developmental differences were found in error processing; the magnitude of PES increased between 6 and 9 and decreased between 9 and 13. They concluded from their study that the reason for such non-linear PES development over different age groups is associated with cognitive control development during childhood. Children before age 8 are most likely to rely on reactive control after an error, associated with the increased PES observed in children aged 6 to 9. There was a developmental shift from 8 when children started to shift towards more proactive control skills than reactive control, which could be reflected in the reduction of PES observed in children aged 9 to 13. (e.g., Chevalier et al., 2015; Niebaum et al., 2020; cited by de Mooij et al., 2022).

In another recent study, Dubravac et al. (2022) investigated the effect of age on PES in multi-trial (four trials after correct and error response) using three executive function tasks (Simon, Stroop, and Flanker; RSI: 250 ms). Children with three different age groups (8, 10, and 12 years old) and adults performed these tasks. PES was measured for the first, second, third, and fourth trials. Results showed that all the age groups showed significant PES. However, a decrease in PES with age was observed for the first trial, but this age effect was diminished for subsequent trials (Trials 2, 3, and 4). Their findings suggest a shift from an orienting response to more balanced cognitive control adaptations over the developmental periods.

While discussing metacognition, it is vital to discuss the domain-general vs. the domain-specific role of metacognition. The following section will briefly discuss the domain-specific vs. domain-general nature of metacognition.

Domain-specific vs. Domain-general Role of Metacognition

One of the main questions in the metacognitive research is whether metacognition is domain-general or domain-specific. Domain-general usually means when participants evaluate their performance in a particular area (e.g., memory) and are good at judging performance in other areas (e.g., perception; Schraw, Dunkle, Bendixen, & Roedel, 1995). That is the global ability of metacognition. The assumption of the domain-general nature of metacognition has been found in different behavioural and brain imaging studies (e.g., McCurdy et al., 2013; de Gardelle et al., 2016). On the other hand, according to the domain-specific notion of metacognition, our metacognitive processes are specific for each domain (Veenman. et al. 2006), i.e., metacognitive judgment of performance in a particular area (e.g., memory) is only associated with that area. However, different research findings showed that different factors determine if metacognition is domain-general or domain-specific (e.g., age or task types). Below, some research is discussed on metacognition's domain-general vs. domain-specific role.

A study by de Gardelle et al. (2016) found the domain-general role of metacognition in different domains. In their experiment, they developed a perceptual task, including visual stimulus (Gabor patch stimuli: series of black and white bars) and auditory stimulus. The task varied with the sensitivity level and had three blocks: visual, auditory, and mixed with visual and auditory stimuli. A pair of trials were presented in succession order with a fixation point coloured in red or blue as an inter-stimulus interval (500 ms). A metacognitive judgment window appeared with a pair of red and blue colour boxes associated with the colour of the fixation point. Participants had to report their confidence judgment by selecting the colour box associated with the fixation point of the trial in which they felt more confident about their judgment. It was hypothesized that the effect of confidence judgment on task sensitivity would differ across different task conditions (visual, auditory, and mixed conditions) when task-specific (domain-specific). The effect of confidence judgment on task sensitivity would not differ across different conditions when task-generic (domain-general). It was found from their study that participants with trials showing higher task sensitivity showed higher task confidence. Metacognitive ability was unchanged between the within-task (auditory and visual) and the across-task (mixed) conditions for confidence comparison. These findings indicated that the confidence judgment is task-generic (domain-general).

Another study (Scott & Berman, 2013) investigated metacognition's domain-general vs. domain-specific role in different subject areas. It was survey research with 644 college students. In their study, Scott and Berman (2013) used a questionnaire to measure metacognitive knowledge and regulation. Metacognitive judgment on accuracy was measured by comparing the predicted score of each participant after each exam with their obtained results. The courses were divided into two domains: humanities and science. Students' metacognitive judgment on accuracy was calculated in percent by comparing their predicted and actual grades from the exam. It was found from their analysis that knowledge and regulation aspects of metacognition measured in the questionnaire did not differ across different subject areas; however, metacognitive judgment on accuracy significantly varied across domains, leading to poorer judgment on science than humanities. It could be summarized that the assessment technique could influence metacognition's domain-specific vs. domain-general role on academic achievement.

However, a different scenario with mixed findings has been observed when studying domain-general vs. domain-specific aspects of metacognition in children. Geurten et al. (2018) conducted a study with three groups of children aged 8-9, 10-11, and 12-13. They used one arithmetic task with two-digit addition problems and one memory task with words. Participants were asked to use some strategies to solve arithmetic and memory problems. Then, they made metacognitive judgments after each trial to rate the ease of better strategy selection and the confidence judgment of their response. It was found that metacognition had a domain-specific role before the age of 10. As children mature, the domain-general role of metacognition is observed.

In a study conducted by Bellon et al. (2020), standardized arithmetic and spelling tests were used to measure academic achievement in these two domains. They developed a computerized arithmetic and spelling task to measure metacognitive monitoring in each domain. The first part of their study analysed the data for children 8-9 years old, and a cross-domain relationship between metacognitive monitoring and standardized tests was observed. However, the second part of the analysis was done with children between 7-8 years old. No cross-domain relationship was observed between metacognitive monitoring and the standardized tests of arithmetic and spelling. These findings indicated a domain-general role of metacognition in academic learning for children aged between 8 to 9. However, this relationship is domain-specific for children between 7 and 8 years old.

It can be summarized from the above discussion that metacognition is domain-specific in children. However, as they age, it becomes domain-general, and this relationship could vary within the metacognitive components.

So far, I have discussed the definition, models, assessment techniques, and factors associated with metacognitive development. The following section discusses the association between metacognition and arithmetic achievements.

Metacognition and Arithmetic Achievements

As discussed above, metacognition is awareness of our knowledge and making strategic action plans and future goals by monitoring our ongoing cognitive processes. Students need to know about their level of knowledge in a particular area to sit for a test. However, making an appropriate action plan to solve different academic (e.g., arithmetic) problems is more important while monitoring our performance to avoid future mistakes and perform better.

Because of the importance of metacognition in executing different functions, researchers primarily focused on the relationship between metacognition and academic achievements in adults and children. For example, Young & Fry (2008) found a positive correlation between metacognitive awareness in students' grades and GPA (Grade Point Average). Metacognition's subfactors, the knowledge of cognition and regulation of cognition, were also positively correlated with the student's GPA and end-of-course grades. Another study (Hirsh & Inzlicht, 2010)investigated the relationship between error monitoring and academic performance using neuroimaging techniques (e.g., EEG and fMRI) in mediation analysis showed that PES mediates the relationship between error-related negativity (ERN: an electrophysiological marker thought to reflect changes in dopamine when participants make errors in cognitive tasks) and academic achievements. It was found from their behavioural data that higher PES is associated with higher grades.

Studies with children also found a positive impact of metacognition on academic performance. For example- a study by Freeman et al. (2017) on metacognitive monitoring of work memory and its relationship with academic achievement in Grade 4 children. They found that children's metacognitive monitoring of their working memory was related to higher academic achievement.

Early research investigating the role of metacognition on arithmetic performance used metacognitive questionaries to explore its role in arithmetic. For example- research conducted in 2011 by Özsoy with 242 primary school students demonstrated a significant and positive relationship between metacognition and mathematics achievement, and 42% of the total variance of mathematics achievement could be explained with metacognitive knowledge and skills. In his study, he used the Turkish version of Metacognitive Knowledge and Skills Assessment (MSA-TR) to measure metacognitive knowledge and skills.

Reviews and meta-analytic papers have shown the significant contribution of metacognition to mathematics learning. For example- Muncer et al. (2022), in their systematic literature review and meta-analysis, investigated the association between metacognition and math performance in adolescence. They found a significant positive correlation between metacognition and adolescent math performance ($r = .37$, $p < .001$).

As we already know from the above discussion, different assessment techniques of metacognition informed us about different components. Moreover, suppose the researcher wants to understand how monitoring and regulation aspects of metacognition are related to mathematics. In that case, it is crucial to design studies with appropriate metacognitive tasks (e.g., error monitoring task or metacognitive judgment of confidence task). Unfortunately, only a few studies explored the role of error monitoring and metacognitive judgment of performance in arithmetic.

One study investigated the relationship between numerical error monitoring and numerosity estimates (Duyan & Balcı, 2018). In this experiment, they developed a numerical monitoring task where participants heard sequences of beep sounds (444 Hz, 60 ms) with random interstimulus intervals (varied between 300 and 600 ms). Participants were asked to stop the sequence by pressing the space key when they thought the beep count reached the target numbers (Experiment 1: 11 or 19; Experiment 2: 7 or 11). After each response, participants were asked to rate their responses into confidence-rating pairs: Low(L), Medium(M), and High(H). The direction of the error was measured as over or under based on overestimation or underestimation of the counted numbers. To get numerosity judgments, the confidence rating was paired with the estimation of number: Under(U)-Low(L), Under(U)-Medium(M), Under(U)-High(H), Over(O)-High(H), Over(O)-Medium(M), Over(O)-Low(L). Participants' confidence judgments (low-medium and high) reflected the deviation from the target, regardless of the direction of their errors (over or under). Slopes significantly higher than zero would indicate better monitoring of the degree of deviation and direction of errors. They hypothesized that judgments on the direction and the magnitude of errors would reflect the nature of the actual estimation of errors in the numerosity domain. It was found from their study that there is a positive relationship between the estimation of error and confidence judgment, pointing to a numerical-error-monitoring ability.

Rinne & Mazzocco (2014) conducted a study with children with and without learning disabilities between Grades 5-8. on mental arithmetic judgment, investigating the relationship between the calibration of judgment (i.e., the alignment between confidence in judgments and the accuracy of those judgments) with mental arithmetic accuracy. It was found from their study that higher calibration of judgment is associated with higher mental arithmetic accuracy. Children with learning disabilities showed lower mental arithmetic scores than their typically developing peers, and their judgment calibration was also highly inaccurate.

A recent study (Denervaud et al., 2020) with children (8–12 years old) on error monitoring and the traditional vs. Montessori schooling system found that students whose learning strategy focused on mistakes and feedback (Traditional schooling) committed fewer errors in the arithmetic task than students from the Montessori schooling system. As opposed to the traditional schooling system, in the Montessori schooling system, teachers encourage children to notice their incorrect thinking or to help peers identify incorrect thinking in a pro-social manner. fMRI data revealed that traditionally schooled students showed greater functional connectivity between the anterior cingulate cortex (ACC), involved in error monitoring, and the hippocampus following correct trials. After error trials, Montessori students showed greater functional connectivity between the ACC and ventromedial prefrontal cortex. Furthermore, arithmetic efficacy was higher in Montessori students than in traditionally schooled students. The findings suggest that pedagogical experience influences the development of error monitoring and its neural correlates and arithmetic efficacy later in life. However, a solid conclusion couldn't be made based on these results since only 32 children participated in this fMRI study.

Another study (Bellon et al., 2019) investigated the role of metacognition and executive function in arithmetic in second-grade students. In that study, they measured arithmetic by a production task of addition and multiplication. A questionnaire was used to measure global metacognitive knowledge and calibration of judgment in the arithmetic task was used to measure task-based metacognitive monitoring. It was found that participants who had better general metacognition knowledge performed significantly more quickly in solving addition problems, but a non-significant relationship was observed for multiplication. Calibration of judgment contributed significantly to the accuracy of addition and multiplication problems. These findings indicated that domain-general and domain-specific role metacognition could be problem-specific.

The Current thesis

Metacognition is essential for both academic achievement and solving arithmetic problems. We use different mental strategies (e.g., counting from first for addition and separating from for subtraction; Gilmore et al., 2018) while solving different mathematical problems. It is essential to become aware of these strategies and to execute the proper response to control one's cognitive system. This awareness could only be obtained through metacognitive monitoring and regulations, which helped us solve mathematic problems using different mental strategies (e.g., Schneider & Artelt, 2010).

The primary focus of the current research is to investigate the role of metacognition in arithmetic skills. This study chooses error monitoring and awareness processes associated with metacognition's monitoring and regulation aspects to investigate metacognition. Measurement of error monitoring with post-error behavioural adjustments (PES and PECA) does not have the problems associated with traditional measurement methods, e.g., problems with cognitive process interference or requirement of mature verbal ability from children. So, it is a perfect process to measure metacognition in children. I developed some error awareness tasks to overcome the interference effect of the metacognitive judgment. Details of these task designs are discussed in Chapters II, III, and V.

Previous research in mathematical cognition has identified that metacognition is related to better arithmetic achievement. The domain-specificity of this relationship may change with age. However, this research has yet to investigate the role of error monitoring (PES and PECA) and awareness with arithmetic achievement. Two studies were conducted in the present research. The first experiment with a pilot study explored the domain-specific vs. the domain-general role of error monitoring on arithmetic achievement in children and adults. In this study, we investigated error monitoring regarding post-error slowing and post-error change in accuracy. The second study explored the domain-general vs. the domain-specific role of error awareness on arithmetic achievement in adults. This study investigated error awareness regarding the percentage of aware errors.

A general hypothesis was made for the present research based on previous research findings. However, more study-specific hypotheses are discussed in the following chapters. Participants with better error monitoring and awareness ability were hypothesized to do better in arithmetic. This relationship would be domain-specific in children and domain-general in adults (e.g., Rinne & Mazzocco in 2014; Bellon et al., 2020).

Chapter I

In Chapters II and III, the error monitoring aspect of metacognition was exclusively investigated. Chapter IV discussed some additional analysis from Chapter III. In Chapter V, the error awareness aspect of metacognition was investigated. Finally, in Chapter VI, a general discussion of an overall thesis is made.

2. Chapter II: Relationship between error monitoring and arithmetic achievements (Pilot Study)

Abstract

Error monitoring associated with meta-cognition and inhibitory processes of cognitive control also plays a role in academic achievement (Hirsh & Inzlicht, 2010). However, it has yet to be investigated if error monitoring is related explicitly to arithmetic achievements. Adults and children took part in this study. Two different kinds of error monitoring tasks (non-numeric and numeric) were developed to measure post-error slowing (PES) and post-error change in accuracy (PECA). It is known that PES is modulated by response-stimulus intervals (RSI; Danielmeier & Ullsperger, 2011); therefore, short and long RSI (200 and 750 ms) were used. Results showed that PES was larger under short RSI, irrespective of task types in adults. Adults' PECA significantly varied in short RSI but not for long RSIs in the Simon task. In the Simon task under short RSI, participants were less accurate on trials following an error than on trials following a correct response. However, for the arithmetic interference task, it decreased for both RSIs. Children's post-error accuracy decreased in both tasks under both RSIs. This decrease in accuracy after error trials is opposite to the study hypothesis. Adults showed no significant correlation between error monitoring (higher PES) and maths fluency. Children with better numeric error monitoring showed higher math fluency, but no such relationship was observed for non-numeric error monitoring. The findings indicated that the relationship between PES and arithmetic achievement is sensitive to RSI and the age of the participants. Further study with a large sample size is needed to ensure the significance of the relationship between error monitoring and arithmetic achievement.

Chapter II

Introduction

Error Monitoring

In our daily life, we always make mistakes. People usually tend to slow down after an error to avoid future mistakes, and this behaviour is associated with our brain mechanism activated during error monitoring. For example- Overbye et al. (2019) showed stronger error-related negativity, which can be measured as an event-related brain potential reflecting the medial frontal action-monitoring processes, is associated with a slower response in the subsequent trial after an error. As defined earlier, error monitoring is a metacognitive process of detecting and signalling errors once the response has been made (Yeung & Summerfield, 2012). After we commit an error, we adjust our behavioural and neural responses, known as post-error adjustments. Here, we focus on two types of post-error behavioural adjustments, post-error slowing (PES) and post-error change in accuracy (PECA).

For years, researchers have tried to explain what causes humans to slow down after an error and what behavioural changes it usually brings after the commission of an error. Most of these experiments tried to explain post-error behavioural adjustment using forced-choice RT tasks. For example, in Flanker tasks (e.g., Eichele et al., 2010), Stroop tasks (Gehring & Fencsik, 2001), Go/No-go tasks (Cohen et al., 2009), and Simon tasks (Danielmeier et al., 2011). However, none of the research could agree on explaining post-error behavioural adjustments associated with error monitoring. Some research found post-error slowing after an error of commission (e.g., Regev & Meiran, 2014). Contrary to these studies, some researchers found post-error speeding, i.e., participants respond faster after committing an error due to heightening alertness in failure to detect threatening objects (e.g., Caudek et al., 2015). Some studies showed improvement in accuracy after the aware but not unaware errors (e.g., Klein et al., 2007), and some found a decline in accuracy after the commission of error (e.g., Fiehler et al., 2005). All these contradictory findings led researchers to explain error monitoring under two broad umbrella terms: adaptive error processing and maladaptive error processing. Firstly, individual studies on adaptive and maladaptive error monitoring processes are discussed. Then, the literature that integrated both processes is discussed.

Adaptive and Maladaptive Theories of Error Processing

Studies that support the adaptive processes of error processing believe that error monitoring and its associated post-error behavioural adjustments are the product of the cognitive control process between an error response and a correct response. The first researcher who tried to explain error processing in terms of adaptive processes was Laming in 1968 (as cited in Laming, 1979; Wessel, 2018). In his information theory of choice reaction times, he argued that participants tried to inspect the task-relevant stimulus display before the reaction stimulus was presented. The preexposure field might cause a perturbation in the stimulus-triggered decision process, ultimately leading to an error response. In simple words, an error is the result of premature information sampling. That is why reaction times in error trials are faster than reaction times for correct trials. In his 1979 experiment, Laming showed that participants exhibited extra care not to make any further errors in the following five trials after committing an error as part of the cognitive process. His study showed that the number of errors is reduced following an error trial, and the reaction time following an error trial also increases. It was clear from this study that after an error, participants adjusted their decision process, resulting in a decreased error and increased reaction time in post-error trials.

In support of the adaptive process of post-error adjustment, some research studies explained error monitoring considering the conflict-related control adjustment hypothesis (Botvinick et al., 2001; Yeung et al., 2004). According to this hypothesis, when an error occurs, there is a conflict between the error and the correct response, instigating an increased response threshold in response to the conflict, creating a speed–accuracy trade-off, which is defined as *"The complex relationship between an individual's willingness to respond slowly and make fewer errors compared to their willingness to respond quickly and make relatively more errors."*- Zimmerman, (2011). Botvinick et al., in their 2001 review article, discussed that there is a chance our brain's anterior cingulate cortex (ACC) plays a role in cognitive control. It is also evident from other studies that ACC also plays a vital role in error detection and performance regulation. For example- Carter et al. (1998) found that the anterior cingulate cortex (ACC) is involved in error detection and performance monitoring. The ACC showed greater activity on incorrect trials than on correct trials but also increased activity on correct trials when there was a response conflict. These findings support the idea that the ACC plays a role in detecting and responding to errors and monitoring performance. The hypothesis that error monitoring is associated with conflict-related control processes came from the study findings of error-related negativity. It is most likely generated in the ACC. The study by Scheffers & Coles (2000) showed that ERN covaried with the perceived inaccuracy of the behaviour. Larger ERN is associated with aware errors, and a smaller ERN is associated with unaware errors. Moreover, ERN was larger for trials with higher response conflict.

Another study by Gehring & Fencsik 2001 found that error monitoring results from conflict monitoring. Their research found that ERN is higher for errors committed in incongruent trials than those committed in congruent trials. They also found that the slightest degree of response conflict and the smallest ERN was observed when the error and the correct response were dissimilar. It can be concluded from these study findings that the process of error monitoring is the product of conflict-control processing of error and response conflict. Furthermore, stronger ERN is associated with conflict arising from the trials' congruency or the response conflict between error and correct trials.

The role of the cognitive control demand on post-error adjustment was also investigated by Regev & Meiran (2014). Their experiment used two types of tasks with different cognitive loads. To ensure cognitive load, they designed two tasks. One task was a cueing task: dimension vs. mapping cue. Another one was the Stroop task with incongruent trials and neutral trials. The cue task with the mapping cue condition notified the participants which button to press (left vs. right). The dimension cue notified them which tasks would appear in the subsequent trial. Since, in the dimension cue, participants had to predict the button press, which they didn't have to predict for the mapping cue, it was assumed that cognitive load would be higher and more cognitive control is required in the dimension cue conditions than in the mapping cue conditions. Similarly, for the Stroop task, it was assumed that more cognitive control was required in the incongruent condition due to the incongruency effect than the neutral Stroop condition, which made this condition high in cognitive load. Since more cognitive control was required for the dimension cue condition and incongruent Stroop task because of their high cognitive load nature, it was hypothesized that PES would be higher under the dimension cue and incongruent

Stroop condition. It was found from Regev & Meiran's (2014) study that PES was higher in high cognitive load dimension cue tasks and incongruent Stroop tasks. However, a mixed effect of cognitive load on post-error accuracy was found. For the dimension cue, the post-error accuracy decreased significantly, but no such effect of cognitive load was found for the Stroop task. Because of the mixed findings on post-error accuracy, it is hard to conclude if it exclusively supported adaptive or maladaptive theories of error processing. However, it could be concluded that the cognitive load of the task could also affect the post-error behavioural adjustments. Higher cognitive control is needed in the highly cognitively loaded task, resulting in higher PES after an error trial. This might indicate that data from present experiment suggest the cognitive control /adaptive theory of error processing.

It could be concluded from the above discussion that error monitoring and its associated post-error behavioural adjustments are the products of the cognitive control process of monitoring errors. However, some research proposed an alternative theory of post-error behavioural adjustments, the maladaptive theories of error processing. According to maladaptive theory, not all post-error behavioural adjustments improve performance after errors (e.g., Hajcak et al., 2003). This notion is different from the cognitive control theory, which believes that we slow down after errors because our attention is oriented away from the next task due to errors. Before we could focus on the task to improve our performance, the subsequent trial appeared quickly to readjust our performance (Notebaert et al., 2009). Researchers who proposed maladaptive theories found common phenomena in describing post-error behavioural adjustments. For example- error processing occurs when the error is infrequent, and PES does not lead to improvements in accuracy after error trials.

In support of this maladaptive notion of error processing, Notebaert et al. proposed an orienting account of post-error adjustment in 2009. According to this account, people slow down after an error because of the relative infrequency between error and correct trials. Since an infrequent event cause slowing down, participants slowing down after an irrelevant signal is also the result of the orienting response. In their study, two experiments were developed: In Experiment I, they manipulated the error rates by following an adaptive program. In this perceptual task, the accuracy rates were prespecified within 35%, 55%, and 75%. They predicted post-error slowing for infrequent errors and post-correct slowing for rare correct conditions. It was found that when the accuracy rate was 75%, participants showed PES, but when the accuracy rate was 35% accuracy, participants showed post-correct slowing. They conducted another experiment with a task containing an irrelevant sound stimulus to postulate that PES results from the occurrence of an infrequent event. They predicted slowing down after infrequent sound trials. In Experiment II, the overall error rates were only 8.88%. Participants responded slower when the irrelevant stimulus was presented in 25% of trials. There was no significant improvement in performance in post-error trials. These findings suggested that PES occurs due to disruption and reorientation of the attentional processes due to the infrequent events rather than the error-related conflict response.

Another hypothesis proposed to explain why some error processing studies did not show improved performance in post-error trials is called the bottleneck hypothesis. According to this hypothesis, the reason for no post-error improvement in accuracy is related to the fact that cognitive processes might get stuck in the bottleneck stage after an error due to the depletion of limited cognitive resources (Lavro et al., 2018). Another distinctive trait of maladaptive error processing is a reduced accuracy rate after error trials, especially when RSI is short. Lavro et al., in their 2018 paper, took both behavioural and event-related potential (ERP) measures of error monitoring. It was found from the behavioural data that participants slowed down immediately after an error, and they were less accurate. ERP measure was taken for the N1 component, the early visual discrimination process (Vogel & Luck, 2000; cited by Lavro et al., 2018), and the P3 component, which represented the central processes of decision-making and memory updating (Kok, 2001; Polich, 2007; cited by Lavro et al., 2018). N1 and P3 were measured for five trials following error and correct responses. A higher negative amplitude was observed in N1 after the error than in the correct response. In P3, a higher amplitude was observed after the correct response than the error response. It was found that the amplitude of P3 decayed from the second trial, but the amplitude of N1 was persistent on 4th trial after error. The higher negative amplitude of N1 was observed until the 4th trial after an error. Since N1 is related to the attentional process (Vogel & Luck, 2016), which supports the fact from this study that participants need some time to reorient their attention to the task after an error, on the other hand, there was a difference in P3 amplitude for the post-correct and post-error responses. All these together indicate the bottleneck theory of PES, i.e., our central cognitive resources are affected after an error.

In support of the bottleneck hypothesis over the orienting hypothesis, Houtman & Notebaert (2013) conducted three experiments showing a decline in accuracy after post-error trial results from limited central cognitive resources. Participants performed a classical flanker task with arrows followed by a rapid serial visual presentation (RSVP) of numbers (1 to 9) in their Experiment. In Experiment I, in most trials, a letter was presented in three possible positions: immediately after the feedback signal of the flanker task, in the third position in the RSVP, and in or the sixth position in the RSVP. The RSVP was designed with black digits (all digits between 0 and 10 were possible), and an uppercase letter (K, L, D, or S) was presented rapidly. This was followed by the question: "Which of the four letters (KLDS) have you seen?". Possible responses could be K, L, D, S, or the space bar. This spacebar response indicated that they did not see any letter. In Experiment I, both tasks were followed by feedback. In Experiment II, the tasks were without feedback to ensure the change in accuracy after an error was not the product of the attentional blink (described as the inability to process a second target briefly after a first target has been detected; Raymond et al.,1992). They hypothesized that participants would make more errors or blank responses after the error trials. There would be reduced target detection after the error because it is reasonable to assume gaps of attention spread over several trials after committing an error. As hypothesized, it was found from both experiments that participants showed higher error rates and blank responses after error trials compared to correct trials. Experiment-I's results indicate that performance was worse after errors than after correct responses. Participants missed more targets after making errors than responding correctly. Consistent with Experiment I, the participant's performance in Experiment II was worse after the error than in the correct trials. Participants also missed more targets after errors and reported more wrong Targets because of the attentional re-orientation. To dissociate between a bottleneck and an orienting account of error monitoring, they developed Experiment III, where they introduced some infrequent trials with letters J and F with the frequency of (40% vs. 60%), and found decreased target detection after irrelevant trials, irrespective of frequency. In the first two experiments, the researcher showed that performance reduction after the error results from attentional distortion. Experiment III showed that disruption in performance after post-error trial results from bottleneck since the decline in performance in infrequent trials was not affected by the frequency of infrequent trials.

Unifying Theories of Error Processing

It is clear from the above research findings that different studies explained post-error behavioural adjustments from the perspective of either an adaptive or maladaptive error monitoring process. Recently, some researchers have proposed a unifying approach to explaining post-error behavioural adjustments. For example- Jentzsch & Dudschig, in their 2009 paper, proposed that post-error behavioural adjustments could either be the product of a capacity-limited error-monitoring process or the product of the criterion adjustment mechanism. According to the capacity-limited error-monitoring process, post-error behavioural adjustment results from disrupted central cognitive resources, which is maladaptive. On the contrary, the criterion adjustment error mechanism hypothesizes that post-error behavioural adjustments result from conflict-control processes arising from the conflicting response between the error response and the correct answer. This conflict signals decay over time. That is, the closer the subsequent trial following an error, the stronger the adjustment signal and the larger the PES. This criterion adjustment hypothesis aligned with the adaptive nature of error monitoring. Whether the error monitoring would be capacity-limited or criterion adjustment-oriented depends on the timing of the error trials and the subsequent trial after an error. To test their hypothesis, Jentzsch & Dudschig (2009) in their study manipulated the response stimulus interval (RSI) between 50 ms and 1000 ms in a categorization task with high and low perceptual contrast. They showed that under short RSI, the PES showed capacity-limited error monitoring. The closer the subsequent trial after the error trial, the more substantial the adjustment signal and, subsequently, the larger the PES. Participants must access the central processing system within this interval (50 ms) to correct the subsequent response. That is why participants had decreased accuracy rates after error trials under short RSI. Participants showed criterion adjustment error monitoring under long RSI. With time intervals that are long enough between the error and the subsequent trial, participants had enough time to show strategic influence control adjustment to error detection. This helped participants get time to correct their responses to the subsequent trials. That is why PES followed an increased accuracy after an error trial under a longer RSI.

Another study by Guan & Wessel 2022 showed motor inhibition after an error could be either nonselective (i.e., related to the task-unrelated effectors) or selective (i.e., related to the task-related effectors). In their Experiment, they used the transcranial magnetic stimulation (TMS) technique to measure the corticospinal excitability (Corticospinal pathway to relay neural signals from higher brain areas to the locomotor muscle) in participants while they were performing the Simon task. They took the corticospinal excitability measurement from task-unrelated and task-related muscle from four different time points: 150ms, 200ms, 250ms, 300ms, 375ms, and 450ms. It was found that the corticospinal excitability showed nonselective suppression at 250ms after errors compared with correct trials. This finding supports the maladaptive motor inhibition due to the attentional reorientation after an error. Next, they measured corticospinal excitability from the responding muscle but not from the task-irrelevant one. They found that at 200 ms and 250 ms points, there was a selective suppression in post-error vs. post-correct trials. This supports adaptive motor inhibition due to the strategic adjustment of cognitive control after errors. This research also demonstrated that error monitoring could be maladaptive and adaptive based on the error's nature and the subsequent trial's timing.

Another perspective on post-error behavioural adjustments was provided by Fievez et al. (2022). In their research, they explained that the nature of post-error behavioural adjustment depends on the accuracy of the task choice. Most tasks require balancing speed and accuracy by making a speed-accuracy trade-off. However, it varies with the task context, either making impulsive (a quick decision with low accuracy) or cautious (a slow decision with high accuracy) choices. With the impulsive response style, people invest less time in decision-making, and access to quality information is prohibited. This ultimately led people to make a less correct choice but a faster response. However, with a cautious response style, people invest more time in decision-making and access to quality information. This ultimately led people to make a correct choice but a slower response. Fievez et al. (2022) in their study showed that post-error behavioural adjustments could be transformed from adaptive (i.e., PES would lead to having higher accuracy after error trials) to maladaptive (i.e., PES would lead to having less accuracy after error trials). It depends on the decision made by focusing more on speed or accuracy while doing the same task in different conditions. Their Experiment used a token task with feedback under two speed-accuracy trade-off contexts. In a hasty task context, the penalty was as low as -4 cents for an incorrect response, which ultimately pushed the subject to focus more on speed over accuracy. Under the Cautious context, the incorrect response was severely penalized by being taken away 14 cents. High penalties eventually led participants to focus more on accuracy over speed. It was found from their analysis that participants only showed significant PES under a hasty context but without improvement in the post-error accuracy rate. There was no significant PES difference in the cautious context but a significant improvement in performance after error response in the cautious context. This supported their hypothesis that post-error adjustment could be adaptive and maladaptive for the same task with different speed-accuracy trade-off decision contexts.

Wessel proposed a recent adaptive orienting theory of error monitoring in 2018. He aimed to explain error processing mechanisms under both adaptive and maladaptive conditions. Previous theories of error processing were unable to explain why accuracy decreases in post-error trials when the duration between the error trial and subsequent correct trial is small (e.g., Jentzsch & Dudschig, 2009) or why the post-error adjustment is considered more adaptive when there is accuracy in post-error trials (e.g., Hajcak et al. 2003) but show greater PES when accuracy is compromised (e.g., Fievez et al. 2022). The adaptive orienting theory was proposed to address these issues and explain why the PES effect disappears when errors are no longer unexpected (e.g., Notebaert et al., 2009). This theory proposed that errors are unexpected outcomes resulting from mismatches between an action plan and the performed action. When an error occurs, it triggers an automatic cascade of two processes: an inhibitory control process that interrupts ongoing processing and shifts attention to the source of the error and an attentional orienting process that focuses on the cause of the error. The theory suggests that the frontobasal ganglia system plays a role in inhibitory control while the anterior midcingulate cortex adjusts after the error. The intensity of the orienting response that follows the error is thought to be proportional to the degree of controlled processing after the error. The degree of error awareness is believed to be related to the magnitude of the orienting response. Unlike maladaptive theories, this unifying theory suggests that the automatic cascade of processing after an error is beneficial for initiating slower, adaptive, controlled processes that aim to improve accuracy. Wessel also proposed that post-error behavioural adjustments be either strategic or orienting-related. Strategic post-error adjustment is a slow, deliberate process that improves performance on subsequent trials after an error. Orienting-related post-error adjustment occurs when there is not enough time to complete the automatic and cognitively controlled processes following an error, leading to decreased accuracy in subsequent trials.

It can be summarized from the above discussion that post-error behavioural adjustments could be maladaptive and adaptive depending on the characteristics of the task and the time duration between the error response and subsequent correct trials. I am now going to discuss the developmental trajectory of error monitoring.

Developmental Trajectory of Error Monitoring

So far, different theories of error monitoring have been discussed. In Chapter I, the developmental trajectory of error monitoring was elaborately discussed. Here, I am only going to discuss the study findings briefly. Different studies showed that post-error behavioural adjustments differ in different age groups. For example, Gupta et al., in their 2009 study, recruited children aged between 6 and 11 years old to perform two number-digit tasks with a block of single-digit numbers (1 or 3) and a block of three-digit numbers (111 or 333). It was found from their results that error processing as measured by PES has an initial increase between 6 to 7 years of age, then a decrease in the following ages. The decrease between the ages of 7 to 8 years and 9-10 years is not uniform. With some decrease in PES between the ages of 7 and 8, there was a substantial reduction in PES between the ages of 9 and 10. They concluded that the reason for this non-linear development of PES was more likely associated with the response RT that varied between fast and slow trials for different age groups. It was evident from their research that RT in speedy trials increased for the age group 6 to 7 years old and then decreased for the age groups 7 to 8 and 9 to 10 years of age. However, RT values for slow trials decreased only from 9 to 10 years of age. PES is positively correlated with RT in fast and slow trials. Thus, the effect of age on PES is response RT dependent on fast and slow trials under different conditions.

A similar effect of age on PES and RSI was observed in the study of Smulders et al. (2016), which found that children show reduced PES as they age. The PES was significantly larger under shorter than longer RSI. However, the effect of age on RSI was only observed when processing speed was not controlled.

Another study (Masina et al., 2018) compared error awareness and PES in three different age groups: children (8-13 years), young adults (19-35), and older adults (61-83 years). This study found no significant differences in PES for each age group, but younger adults were more aware of their commission errors than children and older adults. No significant differences were found comparing mean error awareness in children and older adults. They explained that the children age group they selected for their study already developed a mature error processing system like adults. In the case of error awareness, differences in different age groups are associated with an impaired attentional process in children and older adults.

In a recent study, Dubravac et al. (2022) investigated the effect of age on PES. Children with three different age groups (8, 10, and 12 years old) and adults performed these tasks. Results showed that all the age groups showed significant PES. However, a decrease in PES with age was observed for the PES measured for the first trial, but this age effect was diminished for subsequent trials. Their findings suggest a shift from an orienting response to more balanced cognitive control adaptations over the developmental periods.

Limited studies only investigated the effect of age on PES and RSI, and the findings are inconsistent. However, one thing is that before age 7, PES increased in children. Then it decreases. The effect of age on PES depends on task complexity and the response RT of the children on fast and slow trials. However, these all-confounding results are most likely to be associated with executive function development in children; that is the next point discussed below.

Executive function development and error monitoring

The age group differences in error processing are most likely because of immature executive function development in the children. Error monitoring changes over the developmental period. For example, Overbye et al. (2019) showed that ERN decreases over the ages (8-19 yrs), but error-positivity remains constant. But both post-error slowing and post-error improvement in accuracy increased with age (RSI: 1200-1800ms). Further research showed inconsistent error-related negativity before adolescence and a steady increase throughout adolescence (Davies et al. 2004). Other research has found a relationship between error monitoring and age affected by task loads. For example, a McDougle (2022) study developed an instrumental learning task shaped by low vs. high cognitive load across the blocks. The task was a simple perceptual task with blocks mixed with selecting shapes and different colour boxes. From their study, they showed that even after controlling the reinforcement learning effect, PES varied because of different working memory loads, and it was found to be decreased in high working memory load trials compared to low working memory load trials, which represents the cost of exerting cognitive control to restore successful performance after errors. This effect of task load on PES was observed. A recent study conducted only with children (4-6 years; Ger & Roebers, 2023) found that children showed PES in tasks with intertrial intervals varied between 100ms and 300ms. In their study, they used two conflict-monitoring tasks with different levels of cognitive load. The one task was Hearts and Flower (interstimulus interval: 300ms) with three blocks (congruent, incongruent, and mixed blocks).

In the congruent block with hearts, children were asked to press the button on the same side as the heart. In incongruent blocks with flower stimulus, they were asked to press the button on the opposite side of the flower. In the mixed block, congruent (heart) and incongruent (flower) trials were presented in a pseudo-randomized order, with the constraint that a flower trial always came between two heart trials or followed a heart trial as the last trial. The mixed block was designed to test rule switching. In another funny fruit task (interstimulus interval: 100ms), participants had three blocks (baseline, congruent, and incongruent). Participants had to choose a colour with a matching square in the baseline trial. In the congruent trials, participants had to choose the yellow and red colours based on the fruit on the screen (e.g., yellow banana and red strawberries). In the incongruent trial, participants had to choose yellow and red squares based on the fruit's mismatched colour (i.e., red banana and yellow strawberry). The study found that the average level of PES differed depending on the complexity of the tasks. PES was higher for the "funny fruit" task than in the "mixed block" and the "low cognitive load incongruent flower block" in the "heart flower" task. Between the "incongruent" condition of the "heart flower" task and the "mixed" condition of the "heart flower" task, PES was higher in the "incongruent" condition of the "heart flower" task., PES was higher for the complex task condition than the typical one. PES was significantly above zero across all tasks. This finding suggested that PES is task complexity difficulty sensitive in children. In a study by Hogan et al. (2005) with participants 12–22 years, a correlation between age and the ERN amplitude was found only for a more complex choice response task condition and not for more uncomplicated conditions. These findings indicate that children's PES differed depending on the executive function demands of the task.

Error Processing and Arithmetic Achievements

Research in metacognitive monitoring showed that participants who monitor their performance better showed higher arithmetic accuracy (e.g., Rinne & Mazzocco, 2014; Bellon et al., 2019). However, few studies investigated the relationship between post-error behavioural adjustments and arithmetic achievement. Most of the studies conducted in this area are mostly brain imaging studies. For example, in their study, Hirsh & Inzlicht (2010) showed that a larger ERN following errors and higher PES significantly correlated with better academic performance. Few behavioural studies have been conducted in post-error behavioural adjustments and arithmetic. However, they did not investigate the direct link between error processing and arithmetic achievement but primarily related to the developmental trajectory of error processing or the development of error monitoring tasks. For example 2012, Desmet et al. developed a mental arithmetic task to study the effects of error and conflict adaptation. It was found from their study that post-error accuracy increases after errors in mental arithmetic. Another online study by de Mooij et al. (2022) mainly focused on developmental changes of PES in mathematics and language learning in an adaptive online learning environment. This study found higher PES related to post-error accuracy for mathematics and language tasks. This finding indicates better error monitoring is associated with increased accuracy following an error in learning maths and language. It could be concluded from these studies that even though no direct relationship between error processing and arithmetic was observed, we could assume that people with better error monitoring would perform better in arithmetic from the ERN studies and error processing studies.

Domain-specific and domain-general role of error monitoring on arithmetic achievement

Chapter I already discussed the relationship between metacognition and arithmetic achievement is domain-specific, and this relationship is age-specific (e.g., Bellon et al., 2019; Bellon et al., 2020). However, the way metacognition was measured in these studies used the calibration of performance judgment, which measures the performance awareness aspects of metacognition but not the post-error behavioural adjustments. Therefore, no studies have focused on the domain-specificity role of error monitoring on arithmetic in different age groups.

The current study

Research in mathematical cognition has identified that error monitoring, as measured by post-error slowing and post-error changes in accuracy, is linked to better arithmetic achievement. The domain-specificity of this relationship may change with age. However, this research has not considered findings from the error monitoring literature demonstrating that RSI influences post-error slowing and that this influence may also depend on the participant's age. In the current study, I developed a numeric (arithmetic interference task) and non-numeric (Simon task) version of the error monitoring task to measure the role of domain-specificity in error monitoring and how it relates to maths in children and adults. Each task included trials with either a long or short RSI to provide the most accurate measure of PES and PECA for each task and age group. It was a pilot study to test the task and experimental design before the final data collection.

The previous research hypothesized that PES would be significantly higher under short than long RSI for children and adults (e.g., Jentzsch & Dudschig, 2009; Smulders et al., 2016). Increased accuracy after error trials would be visible under longer than short RSI (e.g., Jentzsch & Dudschig, 2009). The relationship between RSI and post-error behavioural adjustments would be task-specific (McDougle, 2022; Ger & Roebers, 2023). Participants with better error monitoring ability would do better in arithmetic. This relationship would be domain-specific in children (e.g., Rinne & Mazzocco in 2014; Bellon et al., 2020) and domain-general in adults.

Methodology

Participants

In the present study, I collected data from 20 adult participants (Mean age = 27.4, SD = 4.61, 10 male). All participants were postgraduate students at the University of Nottingham, UK, and completed both the arithmetic interference and the Simon task. Four out of 20 students were British, and the rest were Asian (Bangladeshi, Chinese, and Indian). None of them had any known disability. All of them had maths as a course in their GCSE.

Data from children were collected during Summer Scientist Week, a research event organized each year by the University of Nottingham, UK, to help researchers in developmental psychology collect data from child participants. In this event, each testing room was allocated for a maximum of five researchers, and each had a separate sitting place in a large room to collect data. Researchers were instructed to keep the environment quiet. With the event's help, 48 data sets were collected over Summer Scientist Week. Twenty-four data sets were collected from children for the arithmetic interference task (Mean age 9.89, SD = 1.08, 10 boys). Twenty-three data sets from children were collected for the Simon task (Mean age 8.94, SD= 1.38, 11 boys). All children's first language was English, except for three whose first language was Arabic, Tamil, and non-specified. 34 out of 47 children were white; three were Asian, one was Black, and two were mixed race and information about five children was non-specified. Forty-three children were neurotypical; one child had dyslexia, one child had learning difficulties, one had motor problems, and one had hearing loss in the right ear. Data for the Simon task and arithmetic interference task was collected separately due to the time restriction on data collection during Summer

Scientist Week. The University of Nottingham School of Psychology Ethics Committee ethically approved the study.

Design

For adult participants, the study used a 2 (short vs. long response stimulus interval) × 2 (numeric vs. non-numeric task) within-subject design for PES and 2 (short vs. long response stimulus interval) × 2 (numeric vs. non-numeric task) × 2 (trial type: post-correct accuracy vs. post-error accuracy) within-subject for PECA. The same variable interaction was used for child participants. Still, it was a mixed group design where only the task types were the between-group factor, and other factors were within-group factors.

Materials

A set of Cognitive Control tasks was used.

Simon task (Simon, 1969; Simon & Craft, 1970). A Simon task was developed for the present study, suitable for children and adults. In the original task, participants had to carry out spatially defined responses (e.g., left or right button presses) to non-spatial stimulus attributes (e.g., red or green colour) randomly appearing on the screen's left or right side. In the present task, I used two cartoon characters (minions from the movie 'Despicable 'Me'), and participants were instructed to press the left key (L) whenever they saw a single-eye minion and the right key (R) in response to a double-eyed minion, whereby stimulus location (left or right side of the screen) is entirely task-irrelevant. In the present study, I used a '4 Button Curve Right' button box manufactured by current designs. The leftmost key (Button 1) was labelled L, and the rightmost key (Button 4) was labelled R, indicating the left and right buttons.

Participants were asked to use the left-hand index figure for the left button and the right-hand index figure for the right button. RTs and accuracy were measured.

In this task, the response stimulus interval (RSI) was either 200 ms or 750 ms. In this task, 416 trials were split into four blocks (104 trials per block). Half of the trials in each block were congruent; that is, the position of the required button presses and the position of the Minions aligned in the same position.

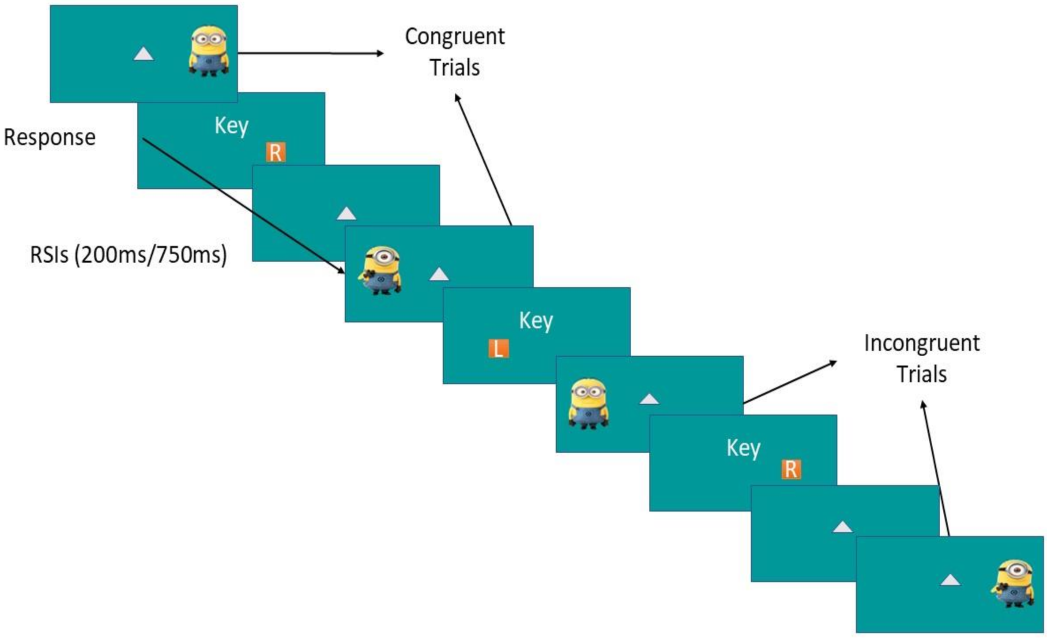


Figure 2.1 Example trials arrangement in the Simon Task

Half of the congruent trials were combined with short RSIs, and half incongruent trials were combined with long RSIs. The maximum repetition of each condition (Congruent vs. incongruent and short vs. long RSI) was eight times. The maximum repetition of the button press on the same side was three times. All the trials were randomly assigned under each block by fulfilling these criteria. The stimulus duration was 700 ms for the child version of this task and 400 ms for the adult version. The stimulus duration was decided based on a few pre-pilot tests of the Experiment on children and adults. Children found it challenging to complete the task under a stimulus duration of 400 ms.

Arithmetic Interference Task (Armitage & Cragg, 2018). This interference task was initially developed based on the work of Lemaire, Barrett, Fayol, and Abdi (1994) and LeFevre, Sadesky, and Bisartz (1996). In this Experiment, multiplication and addition problems appeared on the screen along with two possible answers, which were displayed as response options on the left and right sides of the screen, respectively. Half of the trials showed multiplication problems, and half displayed addition problems. Multiple choice answers were set in a certain way to create response conflict in participants.

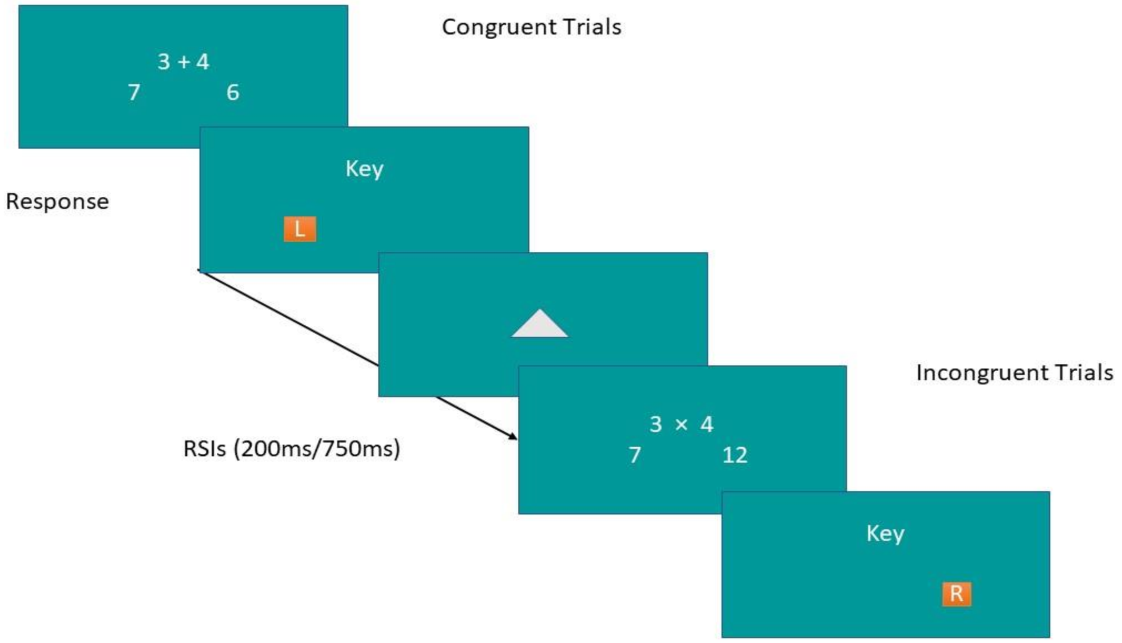


Figure 2.2 Example trials arrangement in the Arithmetic Interference Task

To create response conflict, answers for half of the trials for each problem type (multiplication or addition) had one correct answer, and another was an incorrect response from the same problem type (multiplication or addition). The other half of the trials were non-conflict, with multiple choice options set so that problem size was similar, but they did not belong to the same problem types (multiplication or addition). One multiple-choice was a correct answer, but the other was an incorrect response that did not correspond with a multiplication/addition conflicting answer. For example, in conflicting trials, 4×6= may be displayed on the computer screen, with the correct answer ("24") displayed on the right side of the screen and the incorrect answer (e.g., "12") on the left. In non-conflicting trials, 4×6= may be displayed on the computer screen, with the correct answer ("24") displayed on the right side of the screen and the incorrect answer (e.g., "15") on the left. Participants had to press the corresponding correct key, i.e., if the answer was on the right side of the screen, then the left key (L). If it was on the right side of the screen, then the right key (R). Participants were asked to use the left-hand index figure for the left button and the right-hand index figure for the right button. RTs and accuracy were measured using the same button box as the Simon task.

In this task, the response stimulus interval (RSI) was either 200 ms or 750 ms. The arithmetic interference task had 352 trials across eight blocks (44 in each block). Half of the trials in each block were congruent, and the other half was incongruent. Half of the congruent trials were combined with short RSIs, and half of the incongruent trials were combined with long RSIs. The maximum repetition of each condition (addition vs. multiplication, congruent vs. incongruent, and short vs. long RSI) was eight times. The maximum repetition of the button press on the same side was three times. All the trials were randomly assigned under each block by fulfilling these criteria.

The stimulus duration was 3s, the response duration was 5s for the child version of the task, and for adults, it was 1s stimulus duration and 4s response duration. These two different stimulus durations were decided based on the data from the pre-pilot, where it was observed that children were finding it challenging to complete the task under 1s stimulus duration. Half of the trials in congruent conditions had long RSI, and the other half had short RSI. Similarly, half of the trials in incongruent conditions had long RSI, and the other half had short RSI. Similarly, the conditions of multiplication versus addition, congruence versus incongruence, and left versus right button press were equally distributed in each block.

Mathematical Ability test. The mathematical fluency subtest of Woodcock–Johnson Psycho-Educational Battery–Revised was used to assess mathematical abilities (Woodcock, McGrew, & Mather, 2001). It is a paper-pencil-based task. It was a three-minute test of simple sums. Participants were instructed to solve as many problems as accurately possible within these three minutes. The time required to complete the task was measured with a stopwatch. If participants completed this task in less than three minutes, the total duration of task completion was recorded. One point was given for each correct answer and 0 for the incorrect answer. The measures were taken as a total score of accuracy for each participant.

Variables Measured

Error monitoring was measured as PES and PECA in the present study.

Post-error slowing (PES)

The mean RT of the post-error trials was taken for the correct trials after an error response. For example- for a series of responses like this (E_1, E_2, C_1, E_3, C_2, C_3, C_4, C_5, E_4, E_5, E_6, C_6, C_7, E_7, E_8……), the post-error trials would be C_1, C_2, C_6, but not or E_5, E_8.

There are two ways correct trials can be calculated for measuring PES. One is the traditional technique (e.g., Smith & Brewer, 1995), and the other is the robust technique (Dutilh et al., 2012). In traditional PES calculations, post-correct trials are defined as all that follow another correct trial. Trials just after an error is excluded. For example, for the above series, the mean RT for post-correct trials would be the mean of trials C_3, C_4, and C_7, but not including C_1, C_2, or C_5, C_6.

On the other hand, in the robust technique, only the RT of pre-error trials preceded by a correct response is included. For example, for the above series, the mean RT for pre-error trials would be the mean of trials C_5 and C_7. In our Experiment, I used the robust technique to measure PES.

The formula for these measures is discussed below-

However, our study used different tasks with different difficulty levels and stimulus-response duration. To compare PES between the two tasks, I calculated PES in percent deviation of the individual's mean RT. Here, the mean RT for correct trials was RSI-specific:

$$PES = \frac{Mean(PostError_{RT} - PreError_{RT})}{Mean\ (CorrectRT)}$$

Post-error change in accuracy (PECA)

The formulas for post-correct accuracy and post-error accuracy are given below-

PCA= Post-Correct Accuracy

PCC= Post-Correct Correct Responses

PCE= Post-Correct Error Responses

PEA= Post Error Accuracy

PEC= Post Error Correct Responses

PEE= Post Error Error Responses

n stands for the total number of trials.

$$PCA = \frac{nPCC}{nPCC+nPCE} \times 100$$

$$PEA = \frac{nPEC}{nPEC+nPEE} \times 100$$

Procedure

Data Collected from Adults. Each adult participant was tested individually in a quiet lab room at the School of Psychology, University of Nottingham. At first, each participant was briefed about the study, and informed consent was taken. Every participant performed both the Simon task and the arithmetic interference task. The task order was counterbalanced between the participants. Between performing these two tasks, the participant completed the Woodcock–Johnson math fluency test for 3 minutes.

Before starting the main Experiment, instructions were given for each task. Participants also performed two practice trials to understand what to do in the main Experiment. Participants had to sit in front of a laptop to perform the Simon and arithmetic interference task. After completing all these tasks, information about the participant's age, gender, ethnicity, and mathematical experience was collected (e.g., up to the age they studied math and their maths scores in their GCSE and A level exams). The session lasted approximately 40 minutes.

Data Collected from Children. Data were collected during Summer Scientist Week from the child participants. The University of Nottingham arranged the Summer Scientist Week, and consent was taken from the parents of the children. Summer scientist week is like a science fair where children come to the event with their parents. Some rooms are allocated for children to participate in different games.

Children were asked to sit in front of a laptop to do the task. Instruction was read out loud to the children. They were told to ask the experimenter if they had questions about any part of the Experiment. The experimenter and parents sat outside the children's visual field while performing the task. To reduce the effect of environmental noise, experimenters were briefed before SSW started to keep the study environment as reasonable as possible. The study was conducted in a big room with a maximum of 4 to 5 children in each room. Twenty-three children participated in the Simon and math fluency tasks; another group of 24 participated in the arithmetic interference and math fluency tasks. The tasks were presented in counterbalanced order. A similar data collection procedure was followed for all the tasks for adults and children.

Results

In the present study, I analysed the data in three phases. Firstly, the effects of RSI and task type on post-error slowing (PES) were investigated. Next, the effect of RSI and task type on post-error change in accuracy (PECA) was calculated. Finally, the correlation between PES, PECA, and math achievement was calculated. In the present study, two groups of children were tested (one group for each task) due to time constraints in data collection. Therefore, data were analysed separately for children and adults. In the following section, the results are explained in detail.

Pre-processing analysis

Outliers Detection

Outliers were removed before analysing the data. The detection of outliers was done in three phases. In the first phase, trials with an RT of more than ±3 SD from the ' 'individual's mean RT were excluded from the study. Next, participants whose mean RT was more than ±3 SD from the group mean RT were excluded from further analysis. In the Arithmetic interference task, 2% of all trials were excluded from the adult participant's data, and 2.7% were excluded from the child participant's data. In the Simon task, 1.4% of all trials were excluded from the adult participant's data, and 1.1% were excluded from the child participant's data. Data from one adult participant was excluded for both Simon and the arithmetic interference task, and one child participant was removed for the arithmetic interference task who represented an outlier. One child participant was excluded from the analysis for not completing the Maths fluency test. However, there were no outliers for the Simon task.

Before doing the primary analysis of accuracy data, it was tested for potential outliers for the Mean and SD of the PCE, and the PEE accuracy rate was calculated separately. The participants with more than ±3 SD from the overall error rate were excluded from the final analysis. One adult participant's mean PCE and PEE accuracy rate was more than 3 SD, above the group mean, so data from these participants were deleted from further analysis. In children's data, no such outliers were found for both tasks.

To calculate PES, I needed to set a minimum number of post-error and pre-error trials per condition. In previous studies, no experiment was done to test the reliability of trial numbers in measuring PES. However, few event-related potential (ERP) studies tested a minimum number of trials for measuring ERN and Pe (Olvet & Hajcak, 2009; Pontifex et al., 2010). In these studies, they measured ERN and Pe for error of commission from 2 trials to 14 trials, and it was found that ERN and Pe may be accurately quantified with as few as six to eight commission error trials across the life span (Olvet & Hajcak, 2009; Pontifex, et. al. 2010). Since the current study is a behavioural study and no previous research has mentioned the minimum number of post-error and pre-correct trials per condition, it was decided to set the criteria of five trials per condition for the analysis. At first, at least five post-error and pre-error trials for each RSI (short vs. long) under each condition (congruent vs. incongruent) were decided to be retained for each task. However, some previous studies showed that task congruency does not always affect error monitoring (e.g., Desmet, Imbo, De Brauwer, Brass, Fias, & Notebaert, 2012). Even those who originally developed the arithmetic interference task found that congruency did not affect the RT (Armitage & Cragg, 2018). In the present study, we analysed the congruency effect on RT for both error-monitoring tasks was found (Appendix I; 187). Therefore, to retain large numbers of participants, it was decided to set the five trial criteria for each participant under short vs. long RSIs by collapsing across congruent and incongruent trials. It was found that three adult participants from the Simon and the Arithmetic Interference tasks had fewer than five post-error and pre-error trials under each RSI who were excluded from the final analysis. Five child participants had fewer than five trials under each RSI in the Arithmetic Interference Task and were excluded from further analysis.

Finally, 16 adult participants were retained to investigate the RSI and task effect on PES analysis, and 19 were retailed for PECA analysis. Twenty-three child participants from the Simon task and 17 from the arithmetic interference task were retained for final analysis.

RT and Accuracy in Simon and Arithmetic Interference Tasks

Before proceeding with the primary analysis, the mean RT for pre-error, error, and post-error trials for the Simon and arithmetic interference tasks were calculated separately for adults and children, as shown in Figure 2.3 and Figure 2.4. It can be observed from these figures (Figure 2.3 and Figure 2.4) that immediately after the error, the mean RT in post-error trials becomes higher compared to error and pre-error trials. The overall error rate was calculated separately for each task and age group. The overall error rate in the Simon task for adults was M= 0.16, SD= 0.37; for children, it was M= 0.29, SD= 0.45. The mean error rate for arithmetic interference tasks in adults was M= 0.12, SD= 0.33, and in children, it was M= 0.24, SD= 0.43.

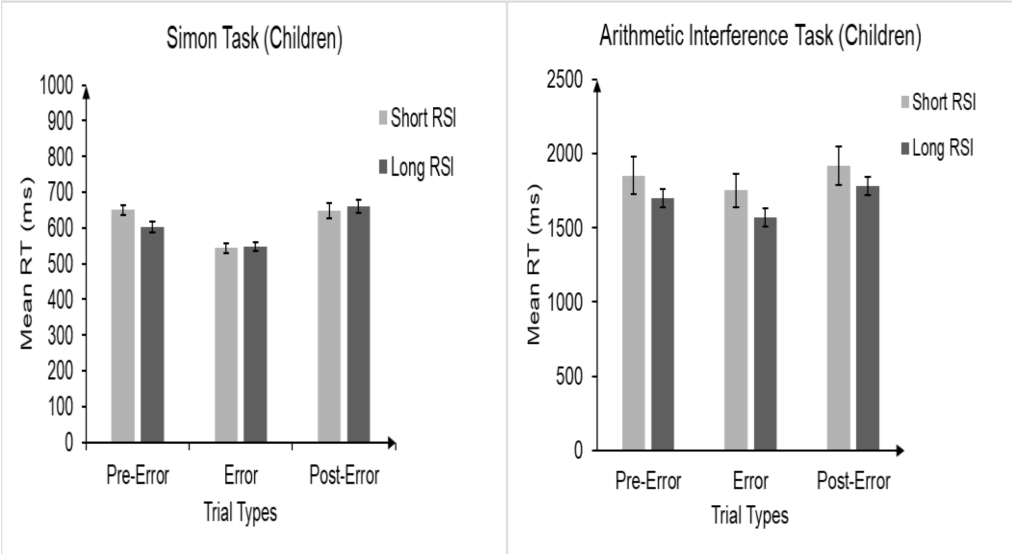


Figure 2.3 Mean RT in pre-error, error, and post-error trials in Simon Task and Arithmetic Interference Task for short and long RSIs, Error bar representing standard error (SE).

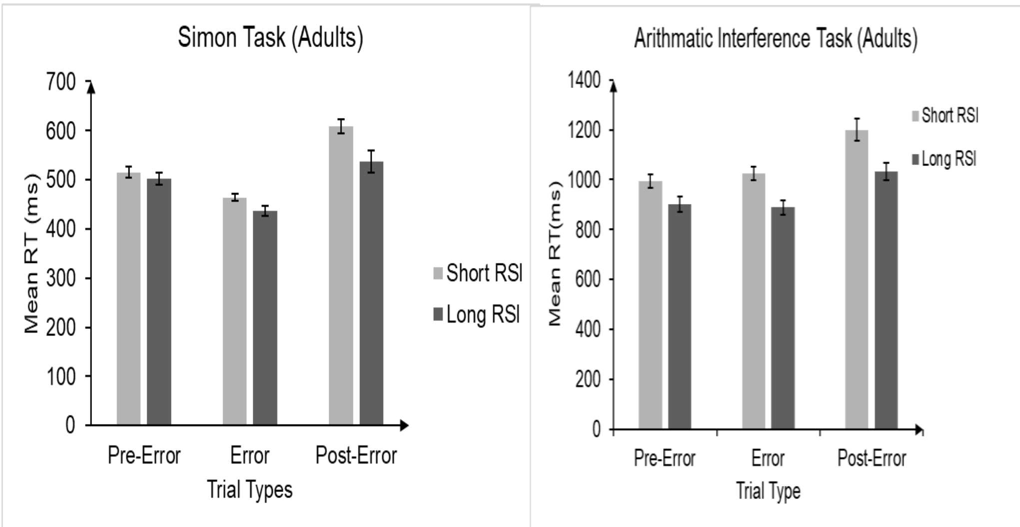


Figure 2.4 Mean RT in pre-, error, and post-error trials in Arithmetic Interference Task and Simon Task for short and long RSIs, Error bar representing standard error (SE).

Main Analysis

Association of task type, RSI, and post-error slowing (PES)

In the present study, I tested two separate groups of children due to time constraints in data collection. Therefore, data for Adult and Child participants were analysed separately to see the association of task type and RSI on PES.

Post-error slowing (PES) in adults.

PES for each task under each RSI was calculated using the formula described in the method section. The mean PES for the two RSIs for the Simon and arithmetic interference tasks are presented in Figure 2.5.

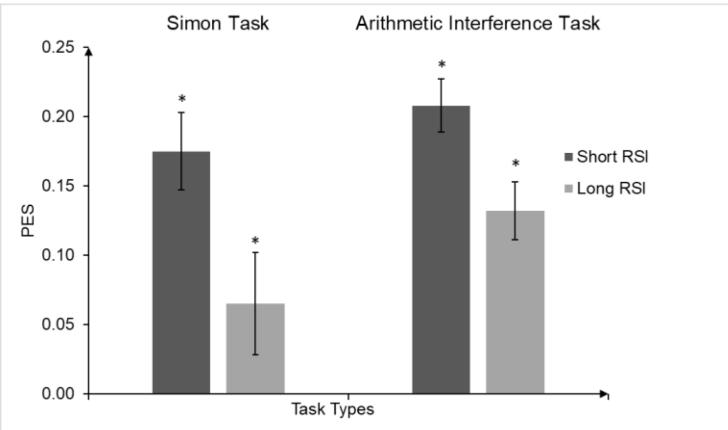


Figure 2.5 Mean PES in Arithmetic Interference Task and Simon Task for short and long RSIs, Error bar representing standard error (SE). * = *significantly different from zero.*

It can be observed from the figure that PES was higher under short RSI than long RSI for both tasks. One-sample t-tests were performed with Bonferroni-corrected p-value at a level of 0.01 to test whether PES under each RSI varied significantly from zero. Results from one-sample t-tests indicated that PES under both RSIs and for both tasks significantly varied from zero; results are presented in Table 2.1.

Table 2.1 One-Sample t-test (PES in Adults)

	RSI	*M (rates)*	*SD*	$t_{(1,15)}$	*P*
Simon	Short	0.180	0.111	6.294	<0.001
	Long	0.071	0.065	3.503	0.003
Arithmetic Interference Task	Short	0.208	0208	5.580	<0.001
	Long	0.132	0.132	6.183	<0.001

The data were analysed using two-by-two repeated measurement ANOVA to see the effect of task type and RSI on PES. Results from repeated measurement ANOVA indicated a significant main effect of RSI $F_{(1, 15)}$= 10.03, p=0.006, η_p^2= 0.40. However, the effects of task type $F_{(1, 15)}$ = 4.47, p=0.052, η_p^2= 0.23 was non-significant, but the p-value indicates a trend. Also, no interaction between task type and RSIs was observed $F_{(1, 16)}$ = 0.58, p=0.46, η_p^2= 0.04. Overall, participants showed more PES in the short RSI (*M*= 0.19, *SE*= 0.03) compared to the long RSI (*M*= 0.10, *SE*= 0.02). Based on these findings, it was decided that to investigate the correlation between PES and Maths Fluency, participants who fulfilled five trial criteria under the short RSI in each task would be retained for correlational analysis to ensure more power. For investigating the correlation between arithmetic interference task PES and maths fluency, a total number of 18 adult participants' data was retained, and to calculate the correlation between PES in the Simon Task and Maths Fluency, 17 adult participants data were retained to calculate the correlation between PES in the Arithmetic Interference Task and Maths Fluency.

Post-Correct change in accuracy (PECA) in adults

PECA in adults was measured by comparing post-correct and post-error accuracy, calculated using the formula described in the method section. To investigate if the accuracy rate between the post-error and post-correct trials varied between the tasks and RSIs, a 2(Task type) ×2(RSI) ×2(Trial Type: post-correct accuracy vs. post-error accuracy) repeated measurement ANOVA was conducted. A significant main effect of task types, $F_{(1, 18)} = 7.44$, p=0.014, $\eta_p^2 = 0.29$, and Trial types was observed, $F_{(1, 18)} = 14.21$, p=0.001, $\eta_p^2 = 0.44$. A two-way interaction between RSI and trial type was observed, $F_{(1, 18)} = 7.19$, p=0.015, $\eta_p^2 = 0.29$. A three-way interaction among task type, RSI, and Trial type was found $F_{(1, 18)} = 7.38$, p=0.014, $\eta_p^2 = 0.29$, also observed. To break down the three-way interaction, two separate 2(Trial Type) ×2(RSI) repeated measurement ANOVA analyses for each task were run to explore the effect of RSI on PES further.

PECA in the Simon Task for Adults

The mean post-correct and post-error accuracy under each RSI separately is shown in Figure 2.6. It can be observed from Figure 2.6 that for short RSI, the accuracy rate decreased after errors, while for long RSI, there is not much change.

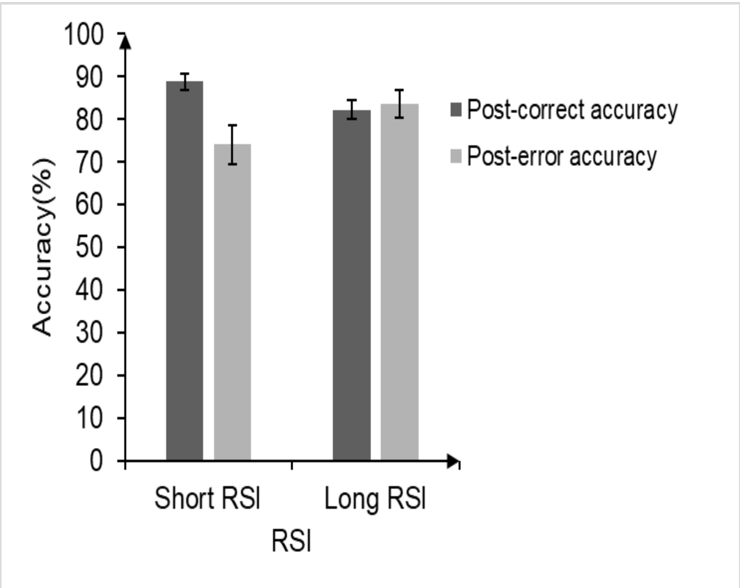


Figure 2.6 The accuracy rate in post-correct and post-error trials for the Simon Task under short and long RSIs in adults, the error bar representing standard error (SE)

A 2(Trial Type) ×2(RSI) repeated measurement ANOVA was done to test whether these mean differences were significant. There was a significant effect of Trial Type on PECA and an interaction between RSI and Trial Type on PECA for the Simon task in adults (Table 2.2). It was observed that under short RSI, the mean accuracy rate significantly declined in the post-error trials, *M*= 74.14, *SE*= 4.61, then in the post-correct trials, *M*= 88.27, *SE*= 1.20, p= .007.

Table 2.2 Analyses of Variance in RSI and Trial Type for Simon Task

Factors	$F_{(1, 18)}$	*P*	Partial η^2
RSI	0.30	0.59	0.02
Trial Type	5.72	0.03	0.24
RSI * Trial Type	10.82	<.001	0.38

However, the long RSI mean accuracy was slightly higher in the post-error trial, *M*= 83.72, *SE*= 3.28, p= .46 than in the post-correct trial *M*= 82.18, *SE*= 2.21, p= .46. Since a two-way interaction effect was observed, data were further analysed in one-way repeated measures ANOVA, where RSI was a within-subject factor, and PCA and PEA were dependent variables. It was found that accuracy after correct trials was significantly larger under short than long RSI, $F_{(1, 18)}$ = 24.94, p<0.001, η_p^2 = 0.58. However, the difference between accuracy in the post-error trial for short and long RSI was non-significant, $F_{(1, 18)}$ = 3.70, p=0.07, η_p^2 = 0.17, but it was a trend.

PECA in Arithmetic Interference Task for Adults

The mean values in Figure 2.7 indicate that participant accuracy decreased in post-error than in post-correct accuracy trials for the arithmetic interference task. This trend was observed for short and long RSI. Analysis from 2(Short vs. Long RSI)

× 2(Trial type) repeated measures ANOVA indicated that there was only a significant effect of trial type on post-error change in performance $F_{(1, 18)}$= 7.41, p< 0.014, η_p^2= 0.29). There was no significant effect of RSI $F_{(1, 18)}$= 0.51, p=.49, and no significant interaction between trial type and RSIs was observed $F_{(1, 18)}$= 0.47, p=.50. People's accuracy rate decreased after error trials (M= 86, SE=1.98) compared to correct trials (M= 90, SE=1.01).

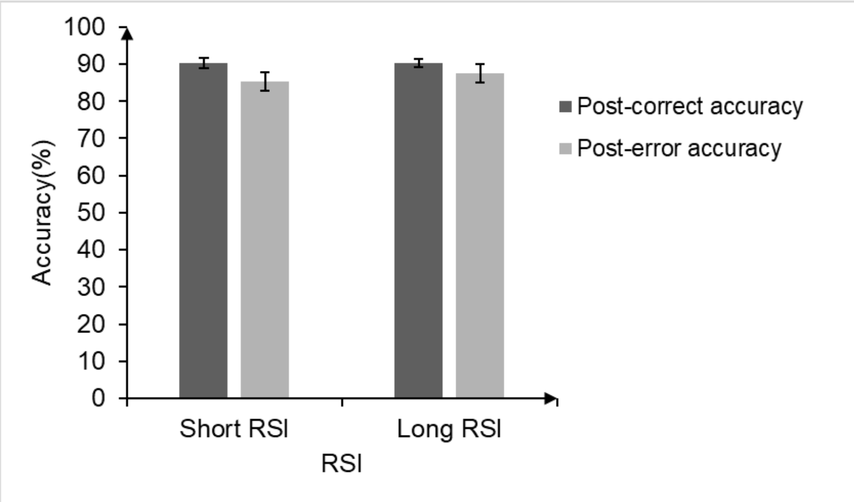


Figure 2.7 The accuracy rate in post-correct and post-error trials for the Arithmetic Interference Task under short and long RSIs in adults, the error bar representing standard error (SE)

Post-error slowing (PES) in children.

PES for each task under each RSI was calculated using the formula described in the method section. The mean PES for the two RSIs for the Simon and arithmetic interference tasks for the children are presented in Figure 2.8. It can be seen from Figure 2.8 shows that children only slowed down after an error for the long RSI in the Simon task. There was a trend of larger PES under both RSIs in the arithmetic interference task.

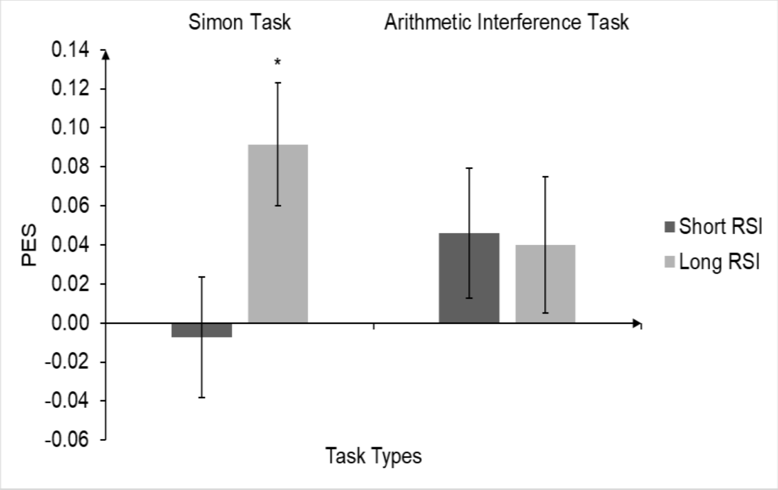


*Figure 2.8 Mean PES in Arithmetic Interference Task and Simon Task for short and long RSIs, error bar representing standard error (SE). * = significantly different from zero.*

One-sample t-tests were performed with Bonferroni-corrected p-value at a level of 0.03 to test whether PES under each RSI varied significantly from zero. The one-sample t-tests showed that PES only significantly varied above zero under long RSI in Simon Task. PES did not vary significantly from zero for all other cases. The results from the one-sample t-test are presented in Table 2.3.

Table 2.3 One-Sample Test (PES in Children)

	RSI	*M(rates)*	*SD*	*T*	*Df*	*P*
Simon	Short	-0.007	0.13	-.281	(1, 22)	0.78
	Long	0.092	0.10	4.250	(1, 22)	<0.001
Arithmetic Interference Task	Short	0.046	0.14	1.34	(1, 16)	0.20
	Long	0.040	0.19	0.88	(1, 16)	0.39

A mixed ANOVA was conducted where the task was the between-subject factor, and RSIs were the within-subject factor. No effects of task type $F_{(1, 38)}$= 0.001, p=0.97, and RSIs $F_{(1, 38)}$= 2.28, p=0.14, η_p^2= 0.06 was observed. The interaction between RSI and Task was also non-significant $F_{(1, 38)}$= 2.81, p=0.10, η_p^2=

0.07. Even though no effect of RSI was observed, the positive mean value of the PES (except for short RSI in the Simon task) indicated some individual variations. Since there are some individual variations, the maximum PES score for each individual, irrespective of the RSI condition, was calculated to see its relationship with maths fluency. Another one-sample t-test was calculated to see whether this PES varies from zero. It was found that PES for both the Simon task, $t_{(1,\ 22)}= 5.64$, $M= 0.11$, $SD= 0.09$, $p< 0.001$, and arithmetic interference task, $t_{(1,\ 16)}= 4.74$, $M= 0.14$, $SD= 0.12$, $p< 0.001$, significantly, varied from zero.

Post-error change in accuracy (PECA) in Children

Data from descriptive analysis for both Simon and arithmetic interference tasks in children are presented in Figure 2.9.

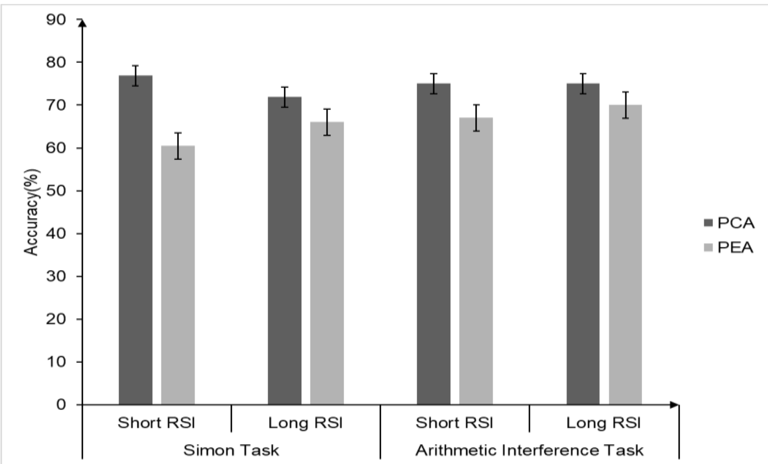


Figure 2.9 Mean accuracy rate in post-correct and post-error trials in Simon task and Arithmetic Interference task for short and long RSIs, Error bar representing standard error (SE).

To see if there was any significant post-error performance change between the Simon task and arithmetic interference task, a 2 (Task Types) × 2 (RSI: Short vs. Long) × 2 (Trial Types: post-correct accuracy vs. post-error accuracy) mixed ANOVA was performed, where RSI and Trial Types were within-subject factors and Task types were a between-subject factor. The statistical results (Table 2.4) indicated a significant effect for Trial Types, and interaction between RSI and trial types was observed.

Table 2.4 Analyses of Variance in RSI, Task Type, and Trial in Children

Factors	$F_{(1, 45)}$	P	Partial η^2
RSI	0.54	0.47	0.01
RSI * Task	0.25	0.62	0.01
Trial type	51.14	<.001	0.54
Trial type * Task	3.38	0.07	0.07
RSI * Trial type	9.00	0.004	0.17
RSI * Trial type * Task	2.82	0.10	0.06

However, no significant three-way interaction was observed among tasks, RSIs, and trial types. The Bonferroni-test of multiple comparisons indicated that under short RSI, the accuracy rate was significantly lower in post-error trials (*M*= 64, *SE*= 2.21, *p*= .03) than in post-correct trials (*M*= 76, *SE*= 1.63, *p*= .03). Similarly, in long RSI, the accuracy rate was lower in post-error trials (*M*= 68, *SE*= 2.17, *p*= .04) than in post-correct trials (*M*= 74, *SE*= 1.68, *p*= .04). These results indicated a decrease in accuracy after an error in both tasks and under both RSIs. Since a two-way interaction effect was observed between RSI and Trial types, two one-way repeated measures ANOVA was conducted separately for each trial type (PCA and PEA) where RSI was a within-subject factor. It was found that accuracy after correct trials was significantly larger under short than long RSI, $F_{(1, 45)} = 4.84$, $p<0.033$, $\eta_p^2 = 0.10$. Similarly, the difference between accuracy in the post-error trial for short and long RSI varied significantly, $F_{(1, 45)}$ = 4.61, p=0.037, η_p^2 = 0.09, and it was larger under long RSI.

It can be summarized from the above findings that PES was higher under the short RSI than the long RSI in adults. One-sample t-tests from child participants showed that PES only varied from zero under long RSI in the Simon Task. No significant effect of task types and RSI was found on PES in children. Both adults and children showed no significant improvement in accuracy after errors.

Relationship between Maths Fluency and PES

PES was most profound under the short RSI for both tasks. Therefore, it was decided to calculate the correlation between PES and math fluency for both tasks under short RSI in adults. In children, maximum PES was counted for each individual separately since PES under short RSI in the Simon task only varied from zero. No effect of RSI on PES was found for children. Before running the correlation, data were tested for potential outliers with the Mahalanobis Distance test. Mahalanobis distance is a multivariate distance measure, and it can be evaluated for each case using the chi-square distribution. The probability estimates for an outlier case when p < .001 for the produced chi-square value (Tabachnick & Fidell, 2018). The analysis also found that no adult or child participant acted as an outlier.

It was found that there was no correlation between math fluency and PES on the Simon task, $r_{(2, 15)}$ = -0.39, p = 0.12, and PES on the arithmetic interference task, $r_{(2, 16)}$ = -0.02, p = 0.93, in adults. These findings indicated that post-error slowing on the arithmetic task did not affect adult maths performance. Fisher's r-to-z tests compared the correlation values between math fluency with PES in the Simon and arithmetic interference tasks in adults. It was found that the correlations between the

Simon task and the arithmetic interference task in adults did not differ significantly, z= -1.06, p= 0.29, within the adults.

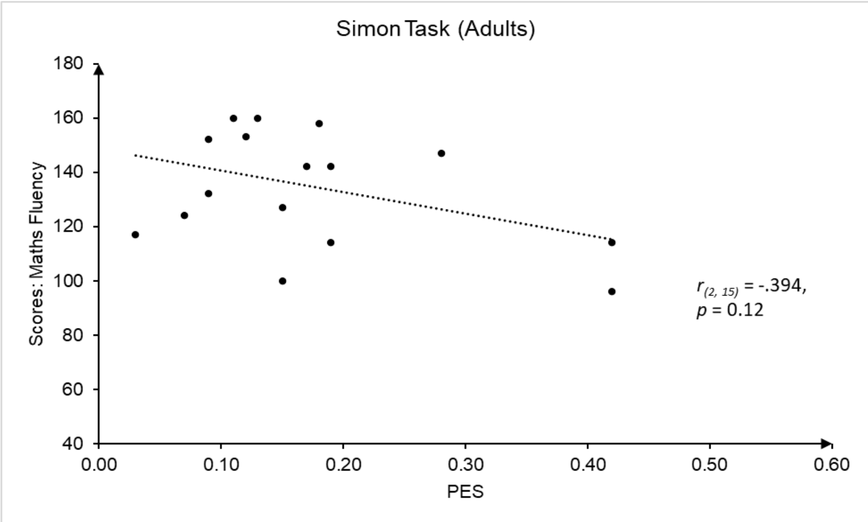


Figure 2.10 Pearson product-moment correlation between math fluency and PES under long RSI for Simon task in adults.

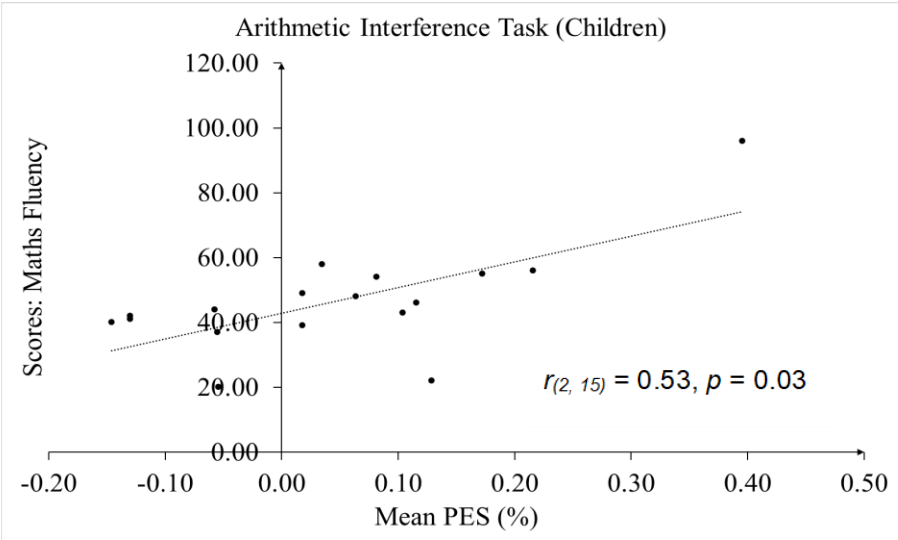


Figure 2.11 Pearson product-moment correlation between math fluency and PES under short RSI for arithmetic interference task in children.

It could be said from the correlation size of the Simon task that adult participants who showed better non-numeric error monitoring (i.e., higher PES) had a tendency to show less fluency in solving arithmetic problems.

There was a significant positive correlation between PES in the arithmetic interference task and math fluency, $r_{(2,\ 15)} = 0.53$, $p = 0.03$, but no correlation between PES in the Simon task and math fluency, $r_{(2,\ 21)} = -0.12$, $p = 0.60$, in children. Fisher's r-to-z tests were done to compare the correlation values between the maths fluency with the PES in the Simon and the arithmetic interference tasks in children, and it was found that the correlation values between the Simon and arithmetic interference tasks differ significantly, $z = 2.04$, $p = 0.04$.

Two separate Fisher's r-to-z tests were done to compare the correlation values between children and adults for maths fluency and PES in the Simon and the arithmetic interference tasks. For the Simon task, correlation values did not differ significantly, z= -0.84, $p = 0.40$, between children and adults. Similarly, it was also non-significant for the arithmetic interference task, z= 1.61, $p = 0.11$, between children and adults.

These findings indicated that children better at monitoring their errors in the numeric error monitoring task were more fluent in maths. Still, no significant relationship was observed for children's non-numeric error monitoring Simon task. No significant correlation was observed between error monitoring and maths fluency for adults. Since no improvement in performance was observed after the error in both groups, we did not calculate the correlation between PECA and math fluency.

Discussion

The study aimed to investigate the relationship between domain-general (non-numeric) and domain-specific (numeric) aspects of error monitoring and arithmetic by focusing on post-error behavioural adjustment, specifically post-error slowing (PES) and post-error change in accuracy (PECA). The response stimulus interval (RSI) was manipulated to see its effect on PES and PECA. The study also examined the relationship between error monitoring and arithmetic by comparing a numeric task (arithmetic interference task) to a non-numeric task (Simon task) and testing both adults and children.

It was found from the present study that PES was higher for adults under the short than the long RSI. One-sample t-tests from child participants showed that PES only varied from zero under long RSI in the Simon Task. No significant effect of task types and RSI was found on PES in children. Both adults and children showed no significant improvement in performance after errors.

The findings from adult participants showed that adults showed significantly larger PES under short RSI than long RSI, which is consistent with the hypothesis that PES would be larger under short than long RSI (e.g., Jentzsch & Dudschig, 2009; Smulders et al., 2016). PECA was also better under non-numeric Simon than the numeric arithmetic interference task. These findings also confirm the hypothesis that the post-error behavioural adjustments would be RSI and task-specific (Jentzsch & Dudschig, 2009; McDougle, 2022; Ger & Roebers, 2023).

The findings of RSI on PES and PECA indicate that as the time interval between the response and subsequent trial increases, the participant's error processing shifts from an orienting response pattern to a more adaptive cognitive control response pattern (e.g., Smulders et al., 2016; Dubravac et al., 2022). Even though in the present study, our focus was not to test the adaptive and maladaptive error processing, if we look at the findings of the adult participant on PES and PECA, then based on the previous findings, it could be assumed that in the Simon task under short RSI, the error processing might follow orienting error processing since PECA significantly decline under short RSI even though the PES was higher. In the long RSI, they might start to have a more cognitive control error processing because PES decreased, but no significant improvement in PECA was observed. For the arithmetic interference task, the changes in PES were the same as the Simon task, but the accuracy rate significantly declined, irrespective of the RSI. The possibility could be that the complexity of the tasks. The Simon task was more of a perceptual task. In the arithmetic interference task, participants were required to solve arithmetic problems, which might require them to invest more cognitive processes in solving arithmetic problems. Even though the task's complexity level was not tested, it was found from the adult participants' data that PECA is task-specific. This finding is consistent with Regev & Meiran (2014), who showed that post-error behavioural adjustment is affected by the complexity of the task, i.e., participants showed better post-error behavioural adjustment (Higher PES) under high than a low cognitive load of the task. Even though PES was not affected by task types, an interaction was observed between RSI and task types on PECA. It could be concluded from these findings that maybe participants needed more time to show significant improvement in the PECA for both tasks but primarily for the numeric arithmetic interference task.

These findings could also be explained by the error processing model of Wessel (2017), where he proposed that once the error is made, a conflict is created among participants due to their violation of expectancy between the expected correct response and the committed error response. This expectancy violation leads them to an automatic inhibition response action after the error response, eventually leading them to reorient their attention toward the task again. After that, the control aspect of error-specific adaptation occurs, leading them toward increased performance in post-error trials. Wessel (2017) also showed that if enough time is not given after an error, premature reorienting can occur during or before the adaptive process is started; as a result, no performance improvement would occur. Past research also indicated that post-error improvements in accuracy were most likely to be observed under a long-time duration between error response and the subsequent trials (e.g., between 900 to 2250 ms; Marco-Pallares et al., 2008). It could be summarized that RSI modulates PES, but PECA is both RSI and tasks specified in adults.

In the present study, no effect of RSI on PES was found in children, but PES varied significantly above zero for long RSI but not for long RSI in the Simon task. For the Arithmetic Interference task, it did not vary significantly above zero for both RSIs. This result contradicts previous research (Smulders et al., 2016). One of the explanations could be that children needed a longer RSI to show the effect of RSI and Task types on PES. In study II, I modified the task with a broader range of RSI to see its effect on PES.

Children showed no improvement in PECA. No effect of task types on PES and PECA was observed. One possible explanation could be that the perceived difficulty level for both tasks was similar in both child participant groups due to their still-evolving executive function skills (e.g., Hogan et al., 2005). Since two groups of child participants were recruited to perform the Simon and arithmetic interference tasks, it is hard to draw conclusions from these findings.

The primary purpose of our study was to investigate the effect of domain-specific vs. domain-general aspects of error monitoring on arithmetic. Moreover, it was hypothesized based on the previous research that participants with better error monitoring ability (Higher PES and higher PECA) would do better in arithmetic, and this relationship would be domain-specific in children (e.g., Rinne & Mazzocco in 2014; Bellon et al., 2020) and domain-general in adults. It was found from the child participants' data that children who were better at monitoring their errors in the numeric error monitoring task also showed significantly higher math fluency. However, no relationship was observed between the PES and maths fluency for the non-numeric error monitoring Simon Task. This result is consistent with the hypothesis for child participants and the previous research where the domain-specific role of metacognition was observed in children (e.g., Bellon et al., 2020). However, no significant relationship between error monitoring and maths fluency was observed in adults, contradicting the study hypothesis based on previous research.

Even though no significant correlation between error monitoring and maths fluency in adults was observed for numeric and non-numeric error monitoring tasks, the correlation size was larger for the non-numeric error monitoring Simon task. It could be assumed from the direction of the correlation that adults with better error monitoring (higher PES) in the non-numeric error monitoring task tended to perform less fluency in the maths fluency test. On the contrary, children with better error monitoring in the numeric error monitoring task showed higher fluency in maths. This opposite direction could be because children were still school-going and still practicing maths, so they were already fluent because of the practice effect; as a result, they showed better maths fluency and better error monitoring (i.e., higher PES) in the numeric error monitoring task. However, very few adults practice arithmetic after O-level or GCSE. Even if they use arithmetic daily, they most likely use calculators. Despite having better error monitoring (Higher PES) ability, their numeric processing speed is still low, which might cause them to be less fluent in maths. That is why adults, even those who showed higher PES, were less fluent in maths, presumably due to their lack of arithmetic practice in daily life.

However, these all are just speculations. More extensive research with a broader range of RSI with a more challenging arithmetic achievement test is needed in larger samples to get an accurate picture of error processing and how it relates to arithmetic.

Limitations and future suggestions

This pilot study investigated the relationship between domain-specific vs. domain-general aspects of error monitoring on arithmetic achievements. Since this study was the first to explain this relationship regarding domain-specificity, a pilot study was needed to understand a few things. Since it was a pilot study, the sample size of recruited participants was insufficient. A significant number of participants are needed to be sure of the effect of RSI on post-error behavioural adjustments in children and adults. Since no significant effect of RSI was found on PES in children, further study with a wide range of RSIs is needed to identify a suitable RSI for each task separately. The test we used for measuring arithmetic achievement was a sub-test of arithmetic achievement; that is why the following study (Chapter III) planned to use a standardized arithmetic achievement test.

3. Chapter III: Relationship between Error Monitoring and Arithmetic Skills

Abstract

Like the pilot study, the relationship between error monitoring and arithmetic achievement was explored here, but with a few updates on experimental design. In this study, 95 adults and 70 children took part. Numeric and non-numeric error monitoring tasks were designed to measure post-error slowing (PES) and post-error accuracy changes (PECA). PES was correlated with participants' arithmetic skills. It was shown that PES is modulated by response-stimulus intervals (RSI; Danielmeier & Ullsperger, 2011); therefore, four RSI (200, 750, 1200, and 1750 ms) were used. Results showed that adults showed significantly more PES under the shorter RSIs (200 and 750 ms), irrespective of their task types. However, children did not fulfil the 5-trials criteria under each condition to measure PES, and data were only collected for the numeric error monitoring task. So, the effect of RSI and Task types on PES was not investigated in children. Adults and children showed no significant improvement in performance after committing errors. Adults showed a decreased accuracy rate after errors for the shortest RSI only. The correlations indicated no significant relationship between error monitoring and arithmetic skills. The findings indicate that RSI influences the PES, but both RSI and task types influence PECA in adults. Better error monitoring (larger PES) did not infer better arithmetic skills.

Introduction

The introduction to Chapter II discussed that error monitoring as post-error behavioural adjustments are affected by RSI (Danielmeier & Ullsperger, 2011), age (Gupta et al., 2009), and task types (Regev & Meiran, 2014). The relationship between error monitoring and mathematics was also discussed (e.g., Rinne & Mazzocco in 2014; de Mooij et al., 2022). Since this is a follow-up study of the pilot study in Chapter II, the literature, objectives, and hypothesis remain the same. It was found from the pilot study that adults showed more PES in the arithmetic interference task than in the Simon task. Overall, PES was higher for adults under the short than the long RSI. One-sample t-tests from child participants showed that PES only varied from zero under long RSI in the Simon Task, for the short RSI in the Simon task, and both short and long RSI in the arithmetic interference task; it was non-significant. No significant effect of task types and RSI was found on PES in children. Both adults and children showed no significant improvement in performance after errors. Only children showed a significant positive correlation between numeric error monitoring and maths fluency. In this follow-up experiment, it was decided that the RSI would be varied within a broader range. This experiment used four different RSI ranges (200ms, 750ms, 1200ms, and 1750ms). In addition, in the pilot study, a speeded arithmetic achievement test (Maths Fluency) was used, which does not measure the overall arithmetic achievement in participants. At first, for this follow-up study, the Wechsler Individual Achievement Test (WIAT) was planned to measure the arithmetic achievement test. However, due to the pandemic, I had to run the experiment online, and no online version of the WIAT was available. A Procedural skills task was later adapted for the online version. The changes to the study's design are based on the pilot study's findings discussed in the method section. The study plan was pre-registered and found in Appendix IV, see page 206.

Data from adults were collected online using the Prolific website and the University of Nottingham research participation scheme. However, to collect data from children, I contacted schools in different provinces of England. Then, the school authority forwarded our advertisement to the parents of the students with a link for them to complete the tasks online. However, when I tried to analysed child participants' data based on the pre-registration, it was found that very few children fulfilled the analysis criteria of a minimum of 5 trials under each condition. A comparison analysis was done between online and Pilot data to investigate the RT and accuracy difference and to understand what is causing deviation in child participants' data. The details of this additional analysis are presented in Chapter 4. I collected some face-to-face data from children during Summer Scientist Week. Still, only data for the arithmetic interference and procedural skill tasks were collected due to time constraints.

In this chapter, data are presented and discussed based on the pre-registration.

Methodology

Participants

The present study used GPower 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) and PANGEA (Westfall, 2016) software to determine the hierarchical multiple regression and mixed-way ANOVA sample size.

In GPower 3.1, a multiple linear regression: Fixed model, R2 increase test was selected with a [1]Cohen's d effect size of 0.10, probability of 0.05, and power of 0.90. The number of tested predictors was 4 (PES and PECA separately for the Simon and arithmetic interference tasks). Five additional predictors with age group and their interaction with the tested predictors were selected. With nine predictors, selected effect size, and probability, 160 samples (80 in each age group) were suitable for this study. A sample size of 160 also provides 0.91 power to detect a 2 x 2 x 4 interaction in a mixed-way ANOVA for PES and PECA separately with an effect size of 0.16 η^2. Pre-registration of the study is clipped to Appendix IV (page 206).

Due to the pandemic, data for adults (18-30 years) were collected from the University of Nottingham, UK participants recruitment portal SONA and another online platform, Prolific. Data were collected from 95 participants (Mean age = 19.80, SD = 2.11, 79 Female) to keep more participants from dropping out as outliers. Informed consent and demographic information were collected in Qualtrics. Seventy-three out of 95 students were white British/European, and 22 were from other races

---

[1] *Note: According to Cohen (1988), conventional values for the effect size for multiple regression are small $f^2 = 0.02$, medium $f^2 = 0.15$, and large $f^2 = 0.35$.*

(Asian, African, and Middle Eastern). Three had colour blindness. Six were neurodivergent, of which two had ASD, one with ADHD, and three were non-specified. All of them had maths as a course in their GCSE.

Data from 70 children were collected at the University of Nottingham Summer Scientist Week. It is a similar summer science fair for researchers to collect data, as was mentioned in the Chapter II method section. Data were collected from children aged 8-15 (Mean age = 10.65, SD = 1.80, and 39 were girls). Forty-three children were British, and the remaining 27 belonged to other ethnic groups. Six children had ASD with some associated disorders (ADHD, Anxiety, learning difficulties), one had ADHD, and one had learning difficulties. The rest of the 62 children were neurotypical. Sixty-five of them speak English, and five of them speak English as a second language.

The University of Nottingham School of Psychology's ethical committee approved the study ethically.

Design

The pre-registered design of the study was a 2 (numeric vs. non-numeric Task) × 2 (Age) × 4 (200ms, 750ms, 1200ms, 1750ms) mixed-group design where tasks and RSIs were within-subject factors, and age was a between-group factor. As mentioned earlier in the introduction, due to the pandemic, the data collection process was hampered. The study design was then changed to 2 task type (numeric vs. non-numeric Task) × 4 RSI (200ms, 750ms, 1200ms, 1750ms) within-group design for adults and one-way repeated measurement design for children where RSI was a within-subject factor in the arithmetic interference task.

Materials

A set of Cognitive Control tasks was used.

Simon task (Simon, 1969; Simon & Craft, 1970). The Simon task used in the present study was a modified version of the pilot experiment. The trial number, block number, and stimulus duration were identical to the pilot study. Unlike the Task designed for the pilot study here, a standard duration of 500 ms RSI was set for all trials. Previous literature observed that post-error slowing is higher when RSI is very short (e.g., ≤ 500 ms; Danielmeier & Ullsperger, 2011; Jentzsch & Dudschig, 2009). A pseudo-randomization of the RSI was done in this experiment using Psychopy software to manipulate four different RSIs. The pseudo-randomization of the RSI (200, 750, 1200, and 1750 ms) was done in this experiment using Psychopy software. A list (200, 750, 1200, and 1750 ms) with four different RSIs was created in the Psychopy. Each time the participant committed an error, one of the four RSIs was selected randomly from the list and removed to make space for other RSIs to be selected next. This random selection process of RSI was continued until all the RSIs were selected, and the list was repopulated with all 4 RSI options. The RSI only changed after the participant committed an error and for a matched future post-correct trial. For each RSI change in the post-error trial, the same RSI was set for a future post-correct trial: After committing an error, the experiment was programmed to pick the targeted post-correct trial between trial numbers 4^{th} to 8^{th} after the post-error trial. Five post-error and post-correct trials were needed to calculate the PES; programming the experiment this way helped me fulfil the analysis requirement. It also helped keep the Task shorter and more engaging for the participants.

Arithmetic Interference Task (Armitage & Cragg, 2018). The arithmetic interference task used in this experiment is like the Task used for the pilot study. All the modifications in assigning four different RSIs and selecting the post-correct trials were done in a pseudo-randomized order like the Simon task discussed above.

Arithmetic skills task (Cragg et al., 2017). This arithmetic skill task was used to measure arithmetic procedural skills. No standardized online arithmetic test was available during the pandemic, and in previous research, performance on this procedural skill task was found to correlate with standardized arithmetic tests (Cragg et al., 2017). In this procedural skills task, arithmetic problems were presented on the screen in each trial, and participants were instructed to use their preferred mental method to solve math problems. The trials given to children were composed of addition and subtraction operations, and the trials given to adults were composed of addition, subtraction, multiplication, and division operations. The problems for each age group involved a mixture of single and double-digit numbers. There were 12 problems to solve in the adult version and ten problems in the child version. Each problem appeared for a maximum of 2.5 seconds on the screen. When the problems appeared on the screen, participants were instructed to solve them and press the spacebar when they thought they were ready to type their answers on the next blank screen. They were asked to use the number keys at the top of their keyboard.

The measure for arithmetic skills was taken as mean RT for correct responses. When the participant pressed the spacebar, a blank screen appeared for the subsequent trial to type the answer. This RT is the participants' processing time of the problems but not the time they took to type their answers. Here, a lower score in RT means better performance in solving the problems.

Variables Measured

In the present experiment, error monitoring was measured using PES and PECA. The formula for these measures is discussed below-

Since the present experiment used two task types with different difficulty levels and stimulus-response duration, PES was calculated in percent deviation of the individual's mean RT, and it also controlled for the speed of responding for each participant by dividing it with standard trial (500 ms) reaction time for correct responses only.

$$PES = \frac{Mean(PostError_{RT} - PostCorrect_{RT}\ )}{Mean\ RT\ of\ Correct\ Trials\ (Trials\ with\ 500ms\ only)}$$

As Chapter II mentions (see page 63), a minimum of five post-error and pre-error trials are required per condition to calculate PES. This decision was made based on the findings from previous error monitoring studies (Olvet & Hajcak, 2009; Pontifex et al., 2010).

Post-error change in accuracy (PECA)

The formulas for post-correct accuracy, post-error accuracy, and post-error change in accuracy are given below-

PCA= Post Correct Accuracy

PCC= Post Correct Correct Responses

PCE= Post Correct Error Responses

PEA= Post Error Accuracy

PEC= Post Error Correct Responses

PEE= Post Error Error Responses

n stands for the total number of trials.

$$PCA = \frac{nPCC}{nPCC + nPCE} \times 100$$

$$PEA = \frac{nPEC}{nPEC + nPEE} \times 100$$

PECA= PEA-PCA

Here, the PCA is the percentage of post-correct correct trials relative to all post-correct trials, and PEA is the percentage of the post-error correct trials relative to all post-error trials. PECA is the difference between these two percentages. To calculate PECA, the main goal was to calculate the percentage of correct and incorrect responses after an error, and a correct trial was needed, and that's why, unlike measuring PES, the five trials rule was not required here.

Procedure

Data Collected from Adults. Participants were recruited online through the University of Nottingham's online platform (SONA system) or the Prolific website for data collection. First, they read the information sheet, completed the consent form, and filled out the demographic information in Qualtrics. Then, they moved toward the Pavlovia website to perform all three tasks: the Simon task, the arithmetic interference task, and the procedural skills task. The tasks were presented in counterbalanced order.

Participants were requested to use either a laptop or desktop to perform the tasks. Before starting the main experiment, instructions were displayed on the computer screen for each Task, and they also performed two practice blocks to get a clear idea of what to do in the main experiment. The first practice block was designed so participants could not move to the subsequent trial unless they understood the rule and responded correctly. Once they understood the experiment, they could move to the next practice block with eight trials, similar to the main experiment. The session lasted approximately 45 minutes.

Data Collected from Children. In the previous chapter, the details of SSW were already discussed. The child participants were recruited during the summer scientist week, which the University of Nottingham arranged in August 2022. The arithmetic interference task and procedural skill task were presented in counterbalanced order.

Results

This section first discussed the effect of task types and RSI on the adult participants' post-error behavioural adjustments. Since data from child participants are only available for the arithmetic interference task and procedural skills task, the effect of RSI on the child participants' post-error behavioural adjustments on the arithmetic interference task only is discussed.

Pre-processing analysis

Outliers Detection in Adults and Children

Before analysing the data, outliers were removed separately for adult and child participants' data. The detection of outliers was done in three phases. In the first phase, trials with an RT of more than ±3 SD from the individual's mean RT were excluded from the study. Next, participants whose mean RT was more than ±3 SD from the group mean RT were excluded from further analysis. Finally, participants with less than five trials under each experimental condition were excluded from the analysis.

Post-error slowing (PES). At first, adult participants' data were tested for potential outliers. Next, the child participants' data were tested.

In the Arithmetic Interference task, a mean RT of 1.5% and 0.1% of all trials had SD above and less than ±3 SD, which was later excluded from adult participants' data. In the Simon task, a mean RT of 0.9% and 0.2% of all trials had SD above and less than ±3 SD, which was later excluded from adult participants' data. One individual adult participant whose mean RT was less than -3 SD from the group's mean RT was excluded from both Tasks from further analysis.

The number of post-correct and post-error trials was calculated separately for each participant's RSI. It was written in the pre-registration that the primary analysis would run with a maximum of 80 adult participants. However, when collecting data, I missed one fact: some post-correct trials turned into error or blank trials. I only noticed this once I stopped recruiting participants with a sample size of 95. After filtering data for outliers, only 35 participants met the criteria of a minimum of five trials under each RSI for running ANOVA (See the details in Appendix II; *Table 9.4*; 199). The ANOVA was run with these 35 participants. Still, to retain a larger number of adult participants for correlational analysis, I planned to select the participants fulfilling the criteria for the RSI that would show larger PES. More information can be found on page no. 105. Finally, it was possible to have 77 participants for our final correlational analysis.

It is to be noted that for the main thesis, it was decided to follow the pre-registration and do the ANOVA on the 35 adult participants who met the five trial criteria for each RSI. However, following the suggestion of the examiners, the data were also analysed, collapsing the 200 and 750 RSI and the 1200 and 1750 RSI; these analyses can be found in Appendix II; page 193.

Finally, we further tested for potential outliers in the variables measured for PES and arithmetic skills using the Mahalanobis distance test to analyse the correlation. Mahalanobis distance is a multivariate distance measure, and it can be evaluated for each case using the chi-square distribution. The probability estimates for an outlier case when $p < .001$ for the produced chi-square value (Tabachnick & Fidell, 2018). It did not find any adult participants' data as an outlier.

The only data I had from summer scientist week in children was the arithmetic interference task. It was found that 0.99% of the trials from the arithmetic interference task had mean RT more or less than ±3 SD, and they were excluded. No child participant had an overall mean RT more or less than ±3 SD. Next, we calculated the number of trials under each RSI, and it was found that only one child fulfilled that criterion (minimum of five trials per RSI). Descriptive data for trial counts can be found in **Error! Reference source not found.**; Appendix II. Possible explanations for children not fulfilling the test criteria are discussed in Chapter IV. Since the effect of RSI couldn't be observed because of a shortage of a minimum of five trials, it was decided later to calculate the PES without considering different types of RSI. By following this method, it was possible to retain more participants for the correlational analysis. Still, 4 participants did not fulfil the criteria. Finally, we further tested for potential outliers in the variables measured for PES and arithmetic skills using the Mahalanobis distance test to analyse the regression; it did not find any child participant's data as an outlier. The data of 66 of 70 child participants was retained for correlational analysis.

It is to be noted that following the suggestion of the examiners, the child participant's data were also analysed, collapsing the 200 and 750 RSI and the 1200 and 1750 RSI; these analyses can be found in Appendix II; page 193.

Post-error change in performance (PECA). Before the primary analysis of accuracy data, it was tested for potential outliers on overall accuracy, PCA accuracy rate, and PEA accuracy rate. Eight adult participants were excluded further from the analysis because that participant's mean PCA and PEA score was less than -3 SD from the overall group mean. No participant had a mean accuracy rate of more or less than ±3 SD from all the participants' overall accuracy rates. Since the main concern was the frequency of error and correct trials under each condition, there was no requirement to fulfil a particular number of trials under each RSI for PECA; 87 out of 95 adult participants were retained for all analyses.

It was found from the child participants' data that 4 out of 70 child participants' mean score for PEA and PCP was less than -3 SD. Since there was no requirement to fulfil a particular number of trials under each RSI for PECA, 66 of 70 children's data was retained for all PECA analyses. Participants fulfilled all the criteria for analysis and completed both tasks.

RT and Error Rates in Simon and Arithmetic Interference Tasks in Adults

Before doing the primary analysis, the mean RT for error and correct trials (RSI 500ms) was calculated (Figure 3.1). It can be observed from this figure that participants responded faster in error trials than in the correct trials.

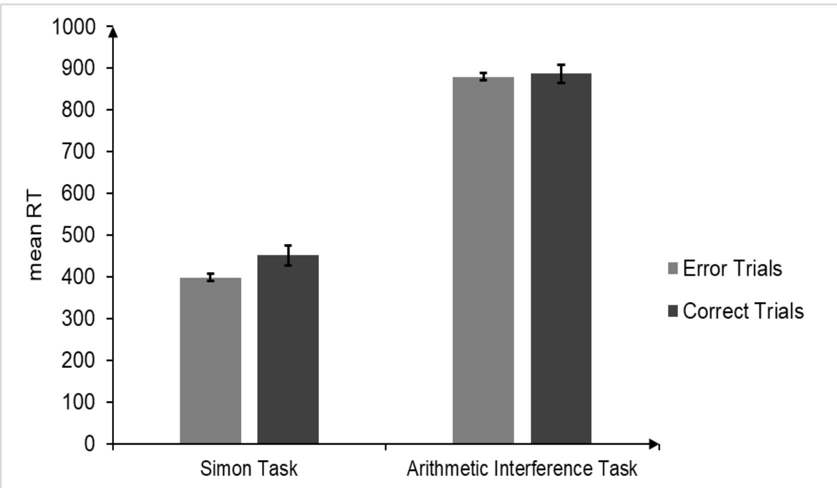


Figure 3.1 Mean RT in Error trials and Correct trials in Simon Task and Arithmetic Interference Task in Adults. Error bar representing standard error (SE).

The mean RTs for post-correct and post-error trials are presented in Figure 3.2. It can be observed from these figures that participants responded more slowly in post-error trials than in matched post-correct trials. The overall error rate in the Simon task was 22.2%, and in the arithmetic interference task, it was 16.9%.

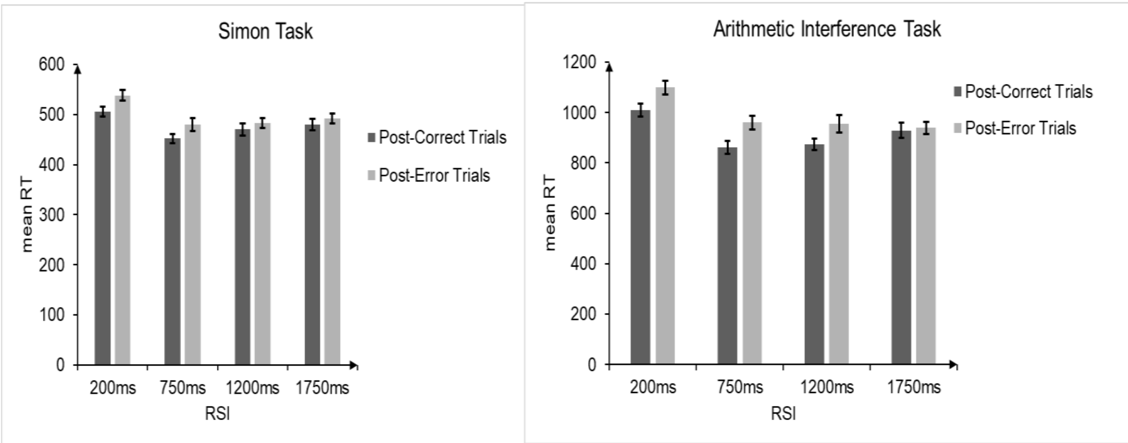


Figure 3.2 Mean RT in Post-Correct, Post-Error Trials under each RSI for Simon Task and the arithmetic interference task in adults, Error bar representing standard error (SE).

RT and Error Rate in Children for Arithmetic Interference Task

As mentioned in the methodology section, there were problems with online data collection from children due to the pandemic. It was only possible to collect data from numeric arithmetic interference tasks. Besides that, when I tried to calculate PES using the formula mentioned in the methodology section, it was found that most children did not fulfil the analysis criteria (minimum five trials per condition). Then, it was decided to collapse data across the RSI to calculate PES. Figure 3.3 presents data for mean RT for error, correct (correct trials with RSI 500 ms), post-error, and post-correct trials in the arithmetic interference task. It can be observed from Figure 3.3 that participants were faster in responding in error trials compared to correct trials. Nevertheless, once they made an error, the response time slowed down in post-error trials compared to the matched post-correct trials. The overall error rate in the arithmetic interference task was 18.68%.

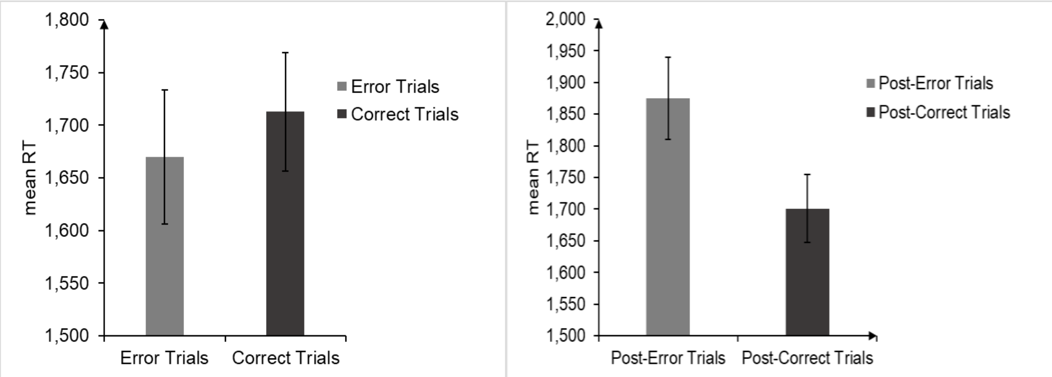


Figure 3.3 Mean RT in Error, Correct, Post-error, and post-correct trials in Arithmetic Interference Task for children, Error bar representing standard error (SE).

Main Analysis

Association of task type, RSI, and post-error slowing (PES) in adults.

Data from adult participants were analysed in 2(Numeric vs. non-numeric) × 4(RSIs: 200ms; 750ms; 1200ms; 1750ms) repeated measurement ANOVA to see the mean PES differences under four RSIs for the Simon and arithmetic interference tasks are presented in Figure 3.4.

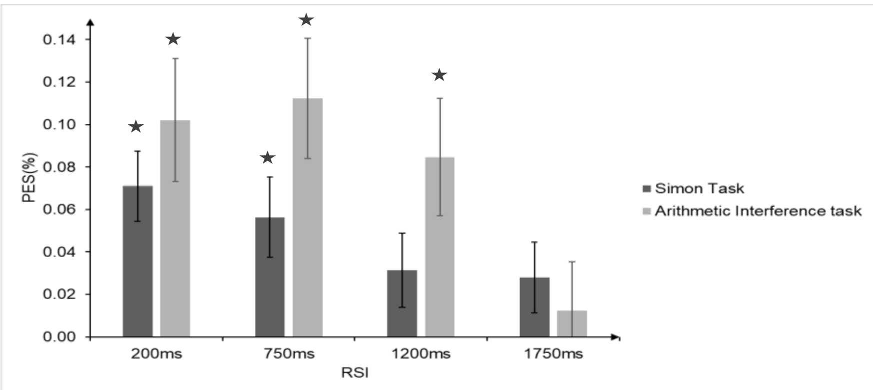


*Figure 3.4 PES in the Simon task and Arithmetic Interference task for different RSIs, the error bar representing standard error (SE). *= significantly different from zero.*

It can be observed from Figure 3.4 that PES was higher under short RSIs, but as the RSI increased, there was a trend of reduction in PES. One-sample t-tests were performed with Bonferroni-corrected p-value at a level of 0.01 to test whether PES under each RSI varied significantly from zero (Table 3.1).

Table 3.1 One-Sample Test for PES under different RSIs

Tasks	RSIs	$t_{(1,34)}$	*M*	*SD*	*P*
Simon Task	200ms	4.309	0.071	0.10	<0.001
	750ms	2.966	0.056	0.11	0.005
	1200ms	1.803	0.032	0.10	0.080
	1750ms	1.686	0.028	0.10	0.101
Arithmetic Interference Task	200ms	3.532	0.102	0.17	0.001
	750ms	3.951	0.112	0.17	<0.001
	1200ms	3.066	0.085	0.16	0.004
	1750ms	0.537	0.012	0.14	0.595

All PES significantly varied from zero apart from the PES associated with RSI 1750ms under the arithmetic interference task and PES associated with RSI 1200ms and 1750ms in the Simon task (Table 3.1). A repeated-measurement ANOVA was conducted to determine the effect of RSI and task types on PES. RSI and Task types as within-subject factors indicated significant effects of RSI, $F_{(3, 32)} = 4.65$, p=0.004, η_p^2= 0.12, on PES. However, the main effect of Task type, $F_{(1, 34)} = 2.97$, p=0.094 and the interaction effect between task type and RSI, $F_{(3, 32)}$= 0.93, p=0.43, were non-significant. Bonferroni, multiple comparison results indicated that PES with RSI 1750 ms (*M* = 0.02, *SE* = 0.14) was significantly lower than PES with RSI 200 ms (*M* = 0.09, *SE* = 0.015, *p*= 0.02) and 750 ms (*M* = 0.08, *SE* = 0.015, p=.02), but not from the RSI with 1200 ms (*M* = 0.06, *SE* = 0.014, p= 0.28). All other PES under RSIs 200ms, 750ms, and 1200ms did not vary significantly from each other.

Results from one sample t-test indicated that PES under RSI 1750 ms were non-significant for both tasks, and PES under RSI 1250 ms was non-significant in the Simon task. Moreover, Bonferroni, multiple comparison results from ANOVA indicated that the PES under RSI 1250 ms and 1750 ms did not significantly vary. In the pre-registration, I planned to select the PES for the RSI, showing the highest average group PES for the correlational analysis. However, to retain the maximum number of adult participants for correlational analysis, it was decided from the ANOVA results that those individual participants who fulfilled the minimum five trial criteria under both RSIs (200 and 750 ms) or who fulfilled this requirement either for one of this RSIs (200 ms or 750 ms). Using this method, 77 adult participants were included in the correlational analysis.

Association of task type, RSI, and post-error change in Accuracy (PECA) in adults

The mean percentage of accuracy rates after post-correct and post-error trials in the Simon and Arithmetic interference task are presented in Table 3.2.

Table 3.2
Mean percentage of accurate post-correct and post-error trials in the Simon and Arithmetic interference task

Tasks	*RSI*		*Mean* (n= 87)	*SD*
Simon Task	200ms	PCA	88.42	11.43
		PEA	77.51	14.04
	750ms	PCA	77.41	17.79
		PEA	77.00	14.11
	1200ms	PCA	81.37	14.05
		PEA	78.58	13.38
	1750ms	PCA	80.42	16.82
		PEA	80.17	13.71
Arithmetic Interference Task	200ms	PCA	84.17	15.96
		PEA	79.49	16.38
	750ms	PCA	83.72	15.54
		PEA	80.83	15.06
	1200ms	PCA	82.27	15.02
		PEA	81.71	15.43
	1750ms	PCA	84.73	15.40
		PEA	83.02	15.69

It can be observed from the following table that for all RSIs, the mean accuracy score in post-error trials is lower than in post-correct trials for both tasks. These numbers indicated that participants' performance got worse after committing errors. The mean value of accuracy difference after post-error and post-correct trials (PECA) under different RSIs is presented in Figure 3.5.

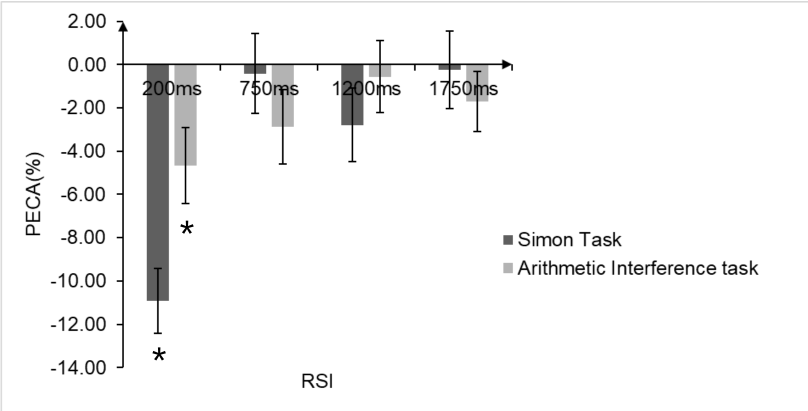


*Figure 3.5 PECA in the Simon task and Arithmetic Interference task for different RSIs, the error bar representing standard error (SE). *= significantly different from zero.*

It could be observed from Figure 3.5 that the participant's performance declined for all RSI, but it is more visible for the shortest RSI (200ms) for both tasks. A one-sample t-test was performed with a Bonferroni-corrected p-value of 0.01 for the Simon Task and arithmetic interference task to test whether PECA under each RSI significantly differed from zero. It was found that PECA varied significantly from the zero for RSI with 200ms in the Simon task only, $t_{(1, 86)} = -7.25$, $p<0.001$, and the arithmetic interference task, $t_{(1, 86)} = -2.67$, $p=0.009$.

None of the other PECAs varied significantly from zero ($p>0.01$). To see the effect of tasks and RSIs on post-error change in accuracy (PECA), a 2(Task) ×4(RSI) repeated measurement ANOVA was conducted. It was found that there was a significant two-way interaction between Task and RSI, $F_{(3, 84)} = 3.00$, $p=0.031$, $\eta_p^2=$

0.034. Since two-way interaction was observed, data were split between two tasks, and two one-way repeated measurement ANOVAs were performed separately for each task to see the RSI effect on PES. There was a significant effect of RSI on PECA for the Simon task only, $F_{(3, 84)} = 9.32$, $p < 0.001$, $\eta_p^2 = 0.10$, but not for the arithmetic interference task $F_{(3, 86)} = 1.22$, $p = 0.30$, $\eta_p^2 = 0.01$. Bonferroni corrected, multiple comparison results for the Simon task indicated that only PECA with RSI 200 ms (*M* = -10.91, *SE* = 1.50) significantly varied from PECA under RSI 750 ms (*M* = -0.41, *SE* = 1.86, *p*= 0.004), 1200 ms (*M* = -2.79, *SE* = 1.72, *p*< 0.001) and 1750 ms (*M* = -0.24, *SE* = 1.78, p<.001). No significant differences in PECA under RSIs (750 ms, 1200 ms, and 1750 ms) were observed in the Simon task.

It can be summarized from the above findings that adults showed significantly larger PES under RSIs 200 ms and 750 ms for both tasks. PES under RSI 1200 ms and 1750 ms didn't vary significantly from one another. The post-error accuracy got significantly worse under RSI with 200ms and was non-significant for other RSIs for the Simon task only. For the arithmetic interference task, the effect of RSI was non-significant. These results indicated that PES is influenced by RSI duration but not task types. The PECA is influenced by both RSI and Task.

Post-error slowing (PES) in Children.

It was impossible to measure the PES following the formula mentioned in the methodology section mentioned earlier; most children did not fulfil the criterion of a minimum of five trials under each RSI. PES in the arithmetic interference task was measured by collapsing all the data into post-error and post-correct trials. Since data were collected only for the arithmetic interference and procedural skills tasks, it was impractical to investigate the association between RSI and task types on PES in children. After measuring PES in the arithmetic interference task, a one-sample t-test was performed to investigate if the mean PES (*M*= *0.09*, *SD*= 0.20) significantly varied from zero. It was found that the mean PES significantly varied from zero. It was 9% slower $t_{(1,65)}$= 3.91, p<0.001.

Association of RSI on Post-error Change in Accuracy (PECA) in Children

The mean percentage of accurate post-error and correct trials under each RSI were calculated for the arithmetic interference task in the children (Table 3.2).

Table 3.2 Mean percentage of accurate post-correct and post-error trials in the Arithmetic interference task

RSI		*Mean* (n= 66)	*SD*
	PCA	86.68	20.84
200	PEA	82.81	19.66
	PCA	82.75	25.57
750	PEA	82.66	20.09
	PCA	83.89	26.34
1200	PEA	80.58	19.06
	PCA	83.48	21.26
1750	PEA	83.42	17.36

PECA for the arithmetic interference task in the children was calculated using the formula described in the method section, presented in Figure 3.6. Results from a one-sample t-test indicated that mean PECA under all RSI did not significantly vary from zero (all p> 0.05). A one-way repeated measures ANOVA was conducted to investigate the effect of RSI on PECA in children for the arithmetic interference task. It was found that there was no significant effect of RSI on PECA, $F_{(3,\ 65)}$= 0.42, p=0.52, η_p^2= 0.01. These findings indicated no significant accuracy improvement after error trials under each RSI for the arithmetic interference task.

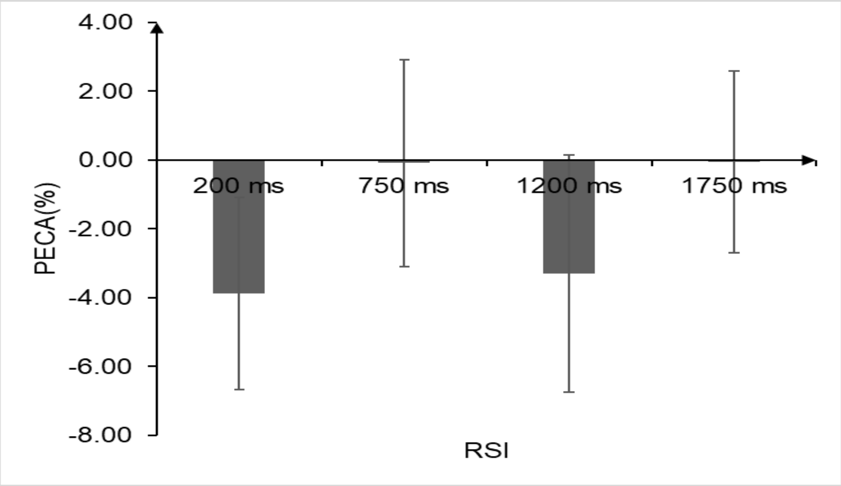


Figure 3.6 post-correct, post-error accuracy, and PECA in Arithmetic Interference task under different RSIs, Error bar representing standard error (SE).

Relationship between PES and Arithmetic skills

The correlations between arithmetic skills and PES in adults were calculated separately for each Task. The correlation between PES and mean RT on the arithmetic skills were non-significant for the Simon task, $r_{(3,\ 74)}$ = .17, *p*= .14, and for the Arithmetic Interference Task, $r_{(3,\ 74)}$= .07, *p*= 0.53. These findings indicate no significant relationship exists between error monitoring (PES) arithmetic skills in adults.

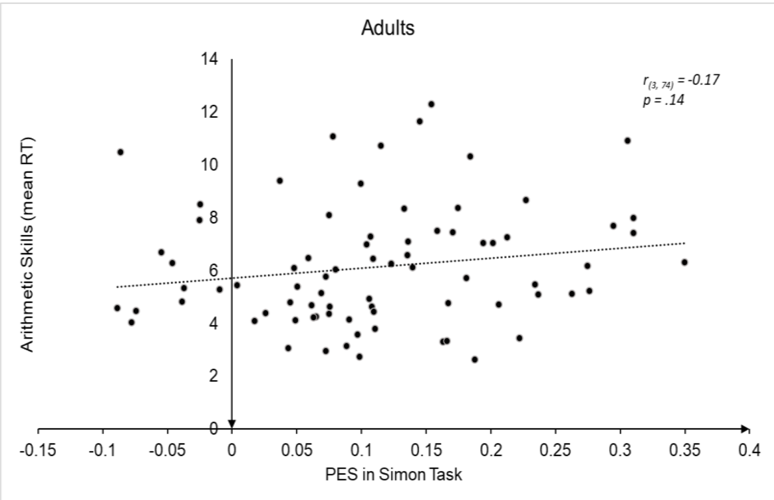


Figure 3.7 Pearson product-moment correlation between mean RT of arithmetic skills and PES in the Simon task in adults.

A Williams-Steiger test was used to investigate if the size of the correlations between the Simon task and the arithmetic interference task differed significantly in adults. It was found that the correlation scores between the tasks in adults were non-significant, $t_{(3, 74)}$= 0.60, *p*= 0.55. It was found from the child participant's data that the relationship between PES on the arithmetic interference task and arithmetic skills was non-significant, $r_{(2,64)}$= -.21, *p*= 0.09, which is a trend. The direction of the correlation indicated that children with better error monitoring tended to show better efficiency (took less time) in solving arithmetic problems. Since no improvement in accuracy was observed in post-error trials, the correlation between PECA and arithmetic skills was not calculated. A Fisher's r-to-z test showed that the correlation scores between adults and children for the arithmetic interference task differed significantly, *z*= 1.65, *p*= 0.10. These correlations indicated no significant relationship between error monitoring and arithmetic skills in adults and children.

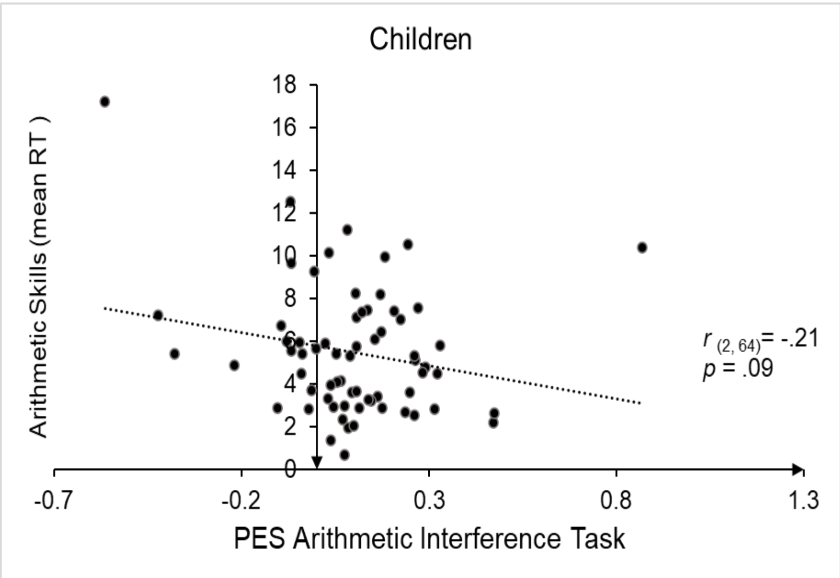


Figure 3.8 Pearson product-moment correlation between mean RT of arithmetic skills with PES of arithmetic interference task in children.

However, from the direction of the correlation, it could be assumed that adults tended to show a cautious approach (longer time) in solving arithmetic problems, despite having better non-numeric error monitoring. However, children who showed better error monitoring in the numeric error monitoring task tended to be more efficient (faster response time) in solving arithmetic problems.

Even though in the pre-registration, I did not mention the correlation between the accuracy score of the arithmetic skills task and PES. It is a procedural skills task focused on the time an individual takes to solve arithmetic problems. Since the above correlation analysis between PES and arithmetic skills, task RT in adults indicates a cautious response style, another correlation was run between PES and accuracy scores of the arithmetic skills task based on the examiner's feedback. It was found that correlations between the accuracy score in the arithmetic skills task and PES in the Simon task, $r_{(3,74)}$= -.12, p= .30, and Arithmetic interference task, $r_{(3,74)}$= -.07, p= .57, were non-significant. Even though the correlation is non-significant, the direction of the correlation in the Simon task indicates that adults who showed better non-numeric error monitoring tended to show less arithmetic accuracy. This notion also supports the cautious response style in solving arithmetic problems in adults. A Williams-Steiger test indicated that the correlation scores between the Simon task and the arithmetic interference task in adults were non-significant, $t_{(3, 74)}$= -.296, p= 0.77.

The correlation between PES in the arithmetic interference task and accuracy scores of the arithmetic skills task also indicated a non-significant relationship, $r_{(2,64)}$= .19, p= .15, in children. However, the direction of the correlation indicated that children with better error monitoring tended to show higher arithmetic accuracy. This is similar to the above findings, where children with better error monitoring showed higher arithmetic efficiency (less time to solve arithmetic problems). A Fisher's r-to-z test showed that the correlation size between adults and children for the arithmetic interference task did not differ significantly, z= -1.53, p= 0.13.

Additional Analysis

It could be observed from the main analysis that when the effect of RSI and task types on PES was investigated by following the planned pre-registration analyses, almost half of the adult participants and nearly all child participants dropped from the analysis for not fulfilling five trial requirements under each RSI. Some additional analyses were done based on the examiners' review. Here, four RSIs collapsed into two RSIs (Short vs. Long). Trials with 200 and 750 ms RSIs were added to form the short RSI, and trials with 1200 and 1750 ms RSIs were added to form the long RSI. The analysis details are presented in Appendix II (page 193). It was found from the adult participants' data that even after splitting the RSI between short vs. long RSI, the findings remained the same for adults. Adults showed higher PES under short RSI compared to long RSI. Like the main analysis, adult participants' accuracy rate got worse under the short than the long RSI and the post-error change in accuracy only significantly varied from zero for the short RSI. It was summarized from the above findings that adults showed significantly larger PES under short RSI for both tasks. The post-error accuracy got significantly worse under short RSI. However, the difference in accuracy after error and correct trials under long RSIs did not differ. These results indicated that PES and PECA are influenced by RSI duration but not task types in adults. Consistent with the main analysis, no effect of RSI was found on PES and PECA for child participants. Finally, no significant correlation was found when the relationship between PES and meant RT for arithmetic skills was observed for both tasks in adults and children, like the main analysis.

Chapter III

Discussion

This study was the follow-up of the pilot study in Chapter II. Like the pilot study, it aimed to investigate the relationship between the domain-general vs. the domain-specific aspects' role of error monitoring on arithmetic skills. Some changes in experimental design were made based on the findings from pilot data. In the pilot study, since children did not show any RSI effect on error monitoring (PES and PECA), and the sample size for the pilot study was small, it was decided to rerun the experiment with a bigger sample size with four different RSIs. Analysis for this study was done in two phases; first, the effect of RSI and task type on post-error PES and PECA was investigated. Next, the relationship between error monitoring and arithmetic skills was tested in the second phase.

It was found from this study that in adults, RSI has a significant effect but not task types on PES, and no interaction effect of RSI and task types was also observed. Bonferroni's multiple comparisons indicated that only PES under the longest RSI was significantly smaller than PES under RSI 200 and 750 ms. However, no significant difference between RSI 1200 and 1750 ms was observed. One sample t-test indicated that PES in children significantly varied from zero. Correlations between PES for the numeric and non-numeric error monitoring task in adults showed non-significant negative correlations with arithmetic skills. Children showed a positive but non-significant correlation between PES in the numeric error monitoring task and arithmetic skills. No improvement in PECA was observed in the children and adults. One sample t-test also indicated that PECA only under RSI 200 ms significantly varied from zero for adults for both tasks. However, results from the ANOVA indicated that this difference in PECA under RSI 200 ms is only significant for the Simon task in adults. One sample t-test showed that PECA did not vary from zero in children, and the ANOVA test did not find any significant effect of RSI on PECA in children.

The findings from adult participants showed that adults showed significantly larger PES under short RSIs than longest RSI, which is consistent with the hypothesis that PES would be larger under short than long RSI (e.g., Jentzsch & Dudschig, 2009; Smulders et al., 2016). however, no improvement in PECA was observed in adults. Only in the Simon task was PECA worse under the RSI 200 ms; no effect RSI was observed for the arithmetic interference task. The findings from RSI and task type effect on PECA are consistent with the previous findings that the post-error behavioural adjustments could be RSI and task-specific (Jentzsch & Dudschig, 2009; McDougle, 2022; Ger & Roebers, 2023).

The adult findings indicate that as the time interval between the response and subsequent trial increases, the participant's error processing shifts from an orienting response pattern to a more adaptive cognitive control response pattern (e.g., Smulders et al., 2016; Dubravac et al., 2022). Even though in the present study, the focus was not to test the adaptive and maladaptive error processing, if we look at the findings of the adult participant on PES and PECA, then based on the previous findings, it could be assumed that under shortest RSI, the error processing might follow orienting error processing since PECA significantly decline under short RSI (200 ms). PES was higher under short RSIs (200 and 750 ms). As the time interval between the response and subsequent trials increased, they shifted towards a more cognitive control mode of error processing with decreased PES and less PECA. Past research also indicated that post-error improvements in accuracy were most likely to be observed under a long-time duration between error response and the subsequent trials (e.g., between 900 to 2250 ms; Marco-Pallares et al., 2008). The reason for not finding any improvement in PECA in adults and children could be that the RSIs were not long enough for the tasks to show any improvement in PECA.

These findings could also be explained by the error processing model of Wessel (2017), where he proposed that once the error is made, a conflict is created among participants due to their violation of expectancy between the expected correct response and the committed error response. This expectancy violation leads them to an automatic inhibition response action after the error response, eventually leading them to reorient their attention toward the Task again. After that, the control aspect of error-specific adaptation occurs, leading them toward increased performance in post-error trials. Wessel (2017) also showed that if enough time is not given after an error, premature reorienting can occur during or before the adaptive process is started; as a result, no performance improvement would occur.

The study's primary purpose was to investigate the relationship between domain-specific vs. domain-general aspects of error monitoring on arithmetic. The correlation between PES and arithmetic skills was investigated for both age groups. Non-significant correlations were observed between PES in numeric and non-numeric error monitoring and arithmetic skills in adults. A non-significant negative correlation was observed between numeric PES and children's arithmetic skills, which was a trend. This research finding is inconsistent with the previous research in that higher PES is related to better academic performance (Hirsh & Inzlicht, 2010). Since no significant correlation was observed, no claim could be made on metacognition's domain-specific vs. domain-general role in arithmetic.

The correlations are non-significant, but the direction of the correlation indicated that adults who tended to show better error monitoring process (Higher PES) in the non-numeric error monitoring task showed less arithmetic efficiency (slower response time). This is consistent with the pilot study, where it was found that adults who showed better error monitoring tended to have less math fluency.

Children who showed better numeric error monitoring (Higher PES) tended to show higher arithmetic efficiency (faster response time).

A possible explanation is already given in the pilot study- Children were still school-going, already practicing maths, and were faster in solving arithmetic problems. That is why a better error monitoring process indicated higher arithmetic skills. On the other hand, after the GCSE, adults rarely practice math or even use math, most likely to use calculators. As a result, they slow down when they solve arithmetic problems manually. That's why even though they showed better error monitoring processes, they showed poorer arithmetic skills due to lack of practice. As a result, adults who showed larger PES tended to be less efficient (slower response time) in solving arithmetic problems. On the other hand, children with larger PES tended to have higher efficiency (faster response time) in solving arithmetic problems.

If we compare the Pilot and current studies, we can see that all the findings are almost similar. One inconsistency is that child participants did not fulfil the minimum five trials criteria for measuring PES. The possible reason for their behaviours is discussed elaborately in Chapter IV.

Conclusion and future suggestions

In summary, post-error behavioural adjustments are complex and depend on many factors. However, more rigorous research is needed to identify a suitable RSI to measure PES with increased PECA for different age groups with different tasks to determine how PES and PECA are related to arithmetic skills. Even though the power analysis indicated that 80 adult and 80 child participants would be recruited since it was the first experiment of this kind, it was unexpected that the study condition would not be fulfilled (minimum five trials under each condition) by half of the participants. However, it is now known that RSI 200 ms and 750 ms are the best measures of PES in adults. A study could be conducted in the future with adults only by using these two RSI to measure PES in a larger sample. Still, if someone is interested in measuring PECA, a longer RSI or alternative measurement method, i.e., calculated PECA for the following few consecutive trials after an error, is needed.

Due to the time shortage, collecting data for the Simon task with the children was not possible. So, the domain-general vs. the domain-specific role of metacognition on arithmetic skills was not investigated in children. The children did not fulfill the PES measurement criteria (minimum five trials under each condition) in the arithmetic interference task. The reason could be that the overall trial number was possibly short for the child version of the arithmetic interference task compared to the adult version. It was discussed in the pilot study that with increased trial numbers in arithmetic interference tasks like adults, children found it hard to concentrate. In the future, it is necessary to repeat this study with children in person and with the arithmetic interference task with more trials or in two-interval sessions to get accurate results.

4. Chapter IV: Follow-up Study on Arithmetic Interference in Children on Different Testing Conditions

Experiment I: Comparison between Online and Pilot Data

Due to COVID-19, there was a restriction for face-to-face testing, so child participants' data were collected online by advertising the experiment in different primary and secondary schools across the country. As was mentioned in Chapters II and III, each participant had to fulfil the requirement of a minimum of 5 trials under each RSI for each task separately to see the effect of RSI on PES. When the children's online data for the arithmetic interference task were screened to see if this requirement was fulfilled, it was found that 71 out of 72 children had less than five post-correct and post-error trials under each RSI. Since no such problem was faced during the pilot study. To understand what is going wrong, I ran additional analyses and experiments to determine why children are not fulfilling the PES measurement criteria.

Some possible hypotheses were made for online child participants-

1. Children who participated in the online study were more efficient in solving arithmetic problems or were interested in mathematics, so they chose to participate.
2. Children who participated online felt more comfortable and less anxious solving arithmetic problems from their home environment, making fewer errors under each RSI.

Some comparative analyses were done to test these hypotheses.

Methodology

Participants

Here, data were compared between the Online study and the Pilot study. The participant details for the Pilot study are provided in the method section of Chapter II (Page 53). Here, only the online participant's details are given.

Online data were collected from almost 70 participants. However, only 61 participants completed all the tasks they were asked to perform. The ages of the children ranged from 8 to 15 (Mean age = 11.67, SD = 1.99, 37 boys). All participants were school-going children. Forty-seven were English, nine were mixed, and the rest were from other nationalities but now living in the UK. 28 were bilingual, where English was the first language. One had developmental disorder, and one had poor eyesight. All other participants were typical.

Design

A between-group experimental design was used where the testing environment was a between-subject factor, and age was a covariate.

Materials

Arithmetic Interference Task (Armitage & Cragg, 2018). The description of this task for the online version is already given in the method section in Chapter III. This task is the same task used for in-person study as well. However, data from the pilot version of the task was used to compare the pilot and online study. The description of the Pilot version of the arithmetic interference task is given in the Chapter II, page 54, method section.

Chapter IV

Procedure

Due to the pandemic, experiments were run online. Data were collected from different schools in the UK. The school head teachers were contacted and requested to advertise the study among parents. The instructions on filling out the consent form and participating in the experiment were available on the School of Psychology, University of Nottingham, UK website. In the online flyer allocated, the parents via schools provided the website link of the University of Nottingham, UK, where all the necessary steps to participate in the online study were given. Each participant was given a five-pound Amazon voucher for their participation in the study. The study was presented in counterbalanced order for each set of 20 children.

The University of Nottingham School of Psychology Ethics Committee ethically approved the study.

Results

The shared task between the Pilot and Online study was the arithmetic interference task. The following section investigated the mean RT and accuracy differences in these two studies. The results are explained below-

Effect of Testing Environment on Mean RT and Accuracy

The mean RT and accuracy differences between pilot data and online data for the arithmetic interference task are displayed in Figure 4.1. Here, Figure 4.1 represents data for mean RT and the accuracy of the arithmetic interference task. It can be observed from Figure 4.1 that the participants in the pilot study data took longer to respond, and their accuracy rate was less than that of participants recruited online. An independent sample t-test was conducted to see if the age of the participants varied between the pilot study (*M* = 9.46, *SD* = 1.06) and the online study (*M* = 11.70, *SD* = 1.98).

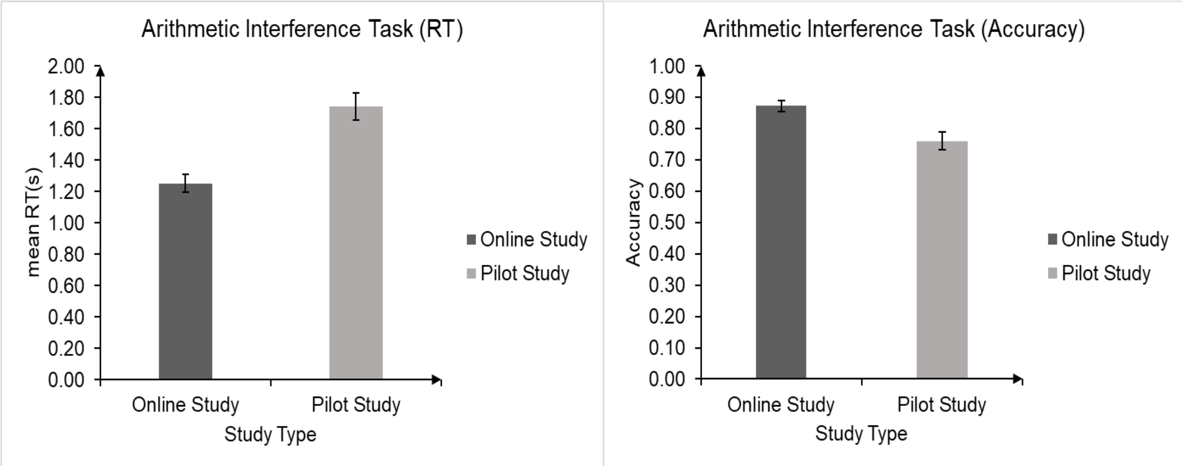


Figure 4.1 Mean RT, accuracy rate in Arithmetic Interference task for pilot and online data, and SE error bar.

Children who participated in the online study were older than those from the pilot study, and this difference was significant, $t_{(1,85)} = -5.57$, $p < 0.001$, $d_s = 1.78$. These findings showed two separate ANCOVAs between the Pilot and online data. The study type (Pilot vs. Online) was an independent factor; age was a covariate, and the mean RT and accuracy rates were dependent variables.

Results from one-way ANCOVA showed that even after controlling the effect of age, a significant differential effect of testing environment was found between children from the Pilot study and Online study, $F_{(1, 88)} = 19.198$, $p<0.001$, $\eta_p^2 = 0.18$. It indicates that children from a face-to-face condition in the Pilot study responded slower in the arithmetic interference task than those online children who participated from home. A significant effect of the testing environment was also observed in the accuracy rate of the arithmetic interference tasks $F_{(1, 88)} = 13.04$, $p=0.001$, $\eta_p^2 = 0.13$. Children who participated online showed higher arithmetic interference task accuracy compared to children who performed face-to-face in the pilot study.

It could be concluded from these findings that children who participated in the home environment did better in the task compared to children who performed in face-to-face settings. One possibility of such findings is that children who participated in this study are better at mathematics, which is why they were interested in participating. The second possibility is that children who took part in this study performed the task from their home, which is their known environment, which made them more relaxed in performing the task. As a result, they made fewer errors and did not fulfil the 5-trials condition for PES measurement.

Experiment II: Comparison between Online vs. In-person Data

The pilot study sample size was too small to draw any conclusion about the testing environment condition, so it was decided to recruit additional child participants face-to-face to understand what was going wrong. At first, I recruited five participants face-to-face for both tasks, and it was found that children fulfilled the five trial criteria for the Simon task but not for the Arithmetic interference task. Then, it was decided to stop recruiting child participants further for the Simon task since it worked fine when data were collected face-to-face. However, additional participants were recruited face-to-face during the 2022 Summer Scientist Week (SSW) to be more confident about the above findings. That is, if in-person children perform worse than online participants, then it would be concluded that children who are better in maths or who like to solve arithmetic problems took part in the online study, which is why they made fewer errors to fulfil the PES measurement criteria. But if no differences in performance for the in-person and online data were found, then it would be assumed that the task needs some modification, especially for children. If children fulfilled the five-trials criteria by making enough errors under each RSI, then it would be assumed that experimental environment is affecting the child participants by making them more nervous, leading them to commit more errors under each RSI.

However, to test these hypotheses on the effect of the testing environment on arithmetic performance, the in-person environment was manipulated into two conditions (Experimenter present vs. absent). If participants from the experimenter's present condition performed worse than those from the absent condition, then it could be claimed that in the home environment, the children were not observed by the experimenter, which made them less anxious and more comfortable performing better.

Methodology

Participants

Here, data are compared between Online and 2022 SSW participants (in-person), and details about the online participants are already given in the previous method section (Page 114). Details of the 2022 SSW were already reported in Chapter III (Page 85).

Design

A between-group experimental design was used where the testing environment was a between-subject factor, and age was a covariate.

Materials

Arithmetic Interference Task (Armitage & Cragg, 2018). The description of this task for the online version is already given in Chapter III (Page 87). This task is the same task used for the 2022 SSW.

Arithmetic skills task (Cragg et al., 2017). This is the same task used in SSW 2022. The task description can be found in Chapter III (Page 87).

Procedure

The previous methodology section gives details of the online data collection procedure. Details of the SSW 2022 in-person data collection procedure are given in Chapter III (page 90).

Results

It was found from the in-person data that children still need to fulfil the five trial criteria under each condition for measuring PES. This indicates that maybe there were some problems with the task design, which will be explored later in this section. In the following section, comparative analyses were done between online and in-person data to investigate the testing environment's effect on task performance.

Effect of Testing Environment on Arithmetic Tasks Performance

The mean RT and accuracy differences between Online and In-person data are displayed in Figure 4.2 for arithmetic interference and skills tasks. Here, the comparison between the online and SSW data on the accuracy of the arithmetic skills task was not made since the mean RT indicates performance for the arithmetic skill task. It can be observed from Figure 4.2 that those online children took less time to respond compared to in-person children.

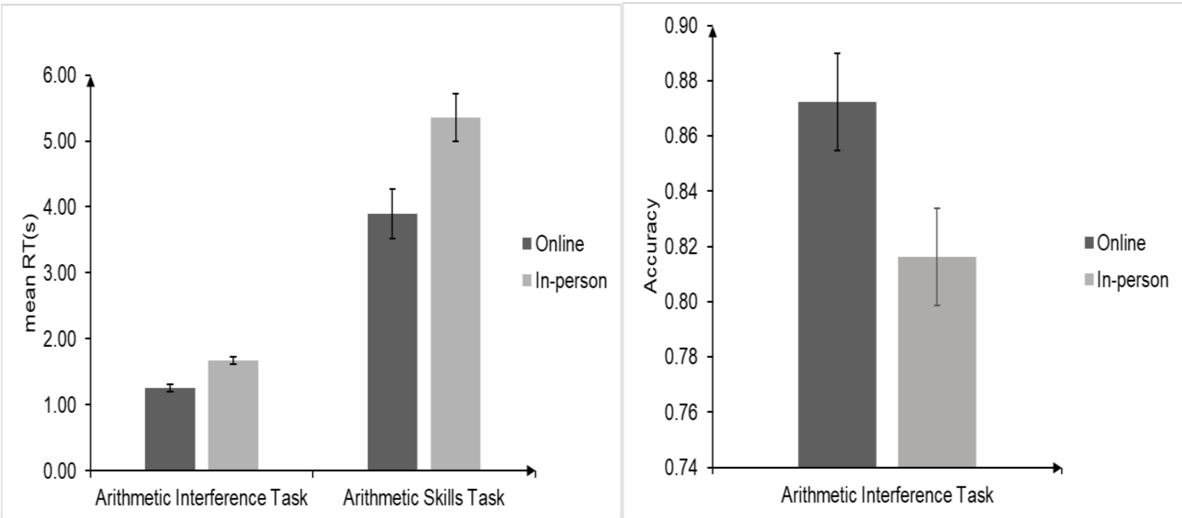


Figure 4.2 Mean RT differences in the Arithmetic Interference task and Arithmetic Skills Task and accuracy rate differences in the Arithmetic interference task between in-person data and online data, error bar representing SE.

This figure indicates that online participants' problem-solving time was quicker than in-person participants from 2022 SSW. This trend was observed for both the arithmetic interference and arithmetic skills tasks. It is observed from the accuracy rate in Figure 4.2 that child participants who took part in the online study made fewer errors than the children who participated in the 2022 SSW.

An independent sample t-test was conducted to see if the age of the participants varied between the Online study (n= 61, M = 11.70, SD = 1.98) and the In-person study (n= 70, M = 10.19, SD = 1.80). It was found that children who participated in the online study were older than those who participated in the in-person, and this age difference was significant, $t_{(1,129)} = 4.558$, $p < 0.001$, $d_s = 1.89$.

Based on these findings, three separate ANCOVAs were conducted for each dependent variable. The testing environment (Online vs. SSW) was an independent factor, age was a covariate, and dependent variables were mean RT and mean accuracy rates for the arithmetic interference task and mean RT for the arithmetic skills task. Results from one-way ANCOVA showed that even after controlling the antecedent effect of age, there was a significant testing environment effect on the mean RT of the arithmetic interference task, $F_{(1, 129)}= 16.84$, $p< 0.001$, $\eta_p^2= 0.12$, and arithmetic skills task, $F_{(1, 129)}= 8.36$, $p= 0.005$, $\eta_p^2= 0.063$. Children from the online study responded faster (M= 1.25, SD= 0.43) in the arithmetic interference task than those from the in-person setting (M= 1.67, SD= 0.47). Children from the online study had faster processing time in solving problems in the arithmetic skills task (M= 3.89, SD= 2.95) than those from the in-person setting (M= 5.36, SD= 2.98). On the other hand, the accuracy rate did not differ significantly between the online and in-person data for the arithmetic interference task $F_{(1, 129)}= 0.72$, $p= 0.39$, $\eta_p^2= 0.06$. Since only the RT but not the accuracy differs significantly for the arithmetic interference task, it is difficult to conclude what might affect the differences between the online and face-to-face study for the arithmetic interference task.

One thing could be summarized: both groups showed statistically equal accuracy in the arithmetic interference task. However, the mean RT tells a different scenario. People who performed online were more efficient (faster response time) in task performance. Data were further analysed between the experimenter's present and absent conditions to test whether the experimenter's present condition affected response time.

Effect of Experimenter Present vs. Absent Condition on Task Performance

Since significant mean RT differences were observed between online and in-person participants, to investigate further if the present experimenter condition in face-to-face settings made children more nervous, they responded slower in both arithmetic interference and skills tasks. To test the experimenter's present effect, SSW data were collected under two conditions: in one condition, the experimenter was present, and in another condition, the experimenter was absent from the child's visual area. To make the environment more interfering, the experimenter sat next to the child for half of the cases. At the same time, they performed the task, and for half of the cases, the experimenter stayed away from their visual field. The findings from the experimenter's condition on the mean RT of the arithmetic interference task and arithmetic skills task and the accuracy rate of the arithmetic interference task are discussed below.

The mean RT and accuracy differences between the experimenter's present vs. absent condition are visible in Figure 4.3 for arithmetic interference and skills tasks.

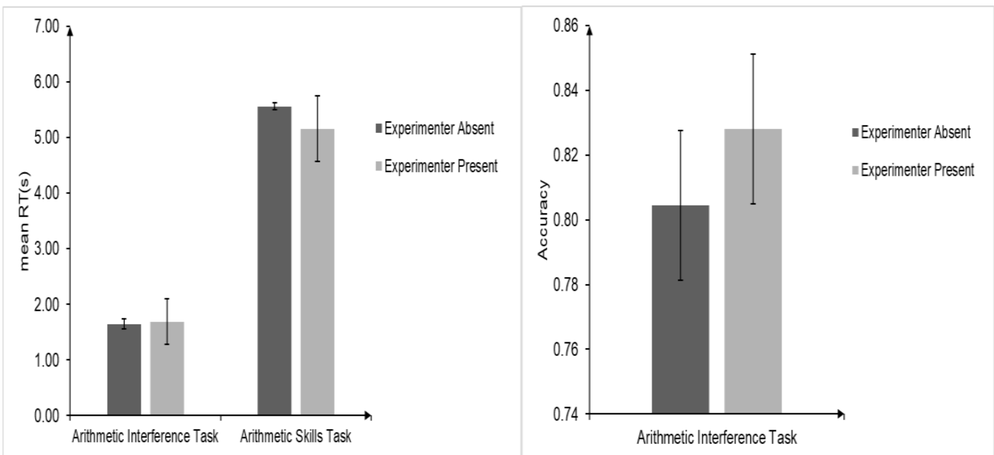


Figure 4.3 Mean RT differences in the Arithmetic Interference task and Arithmetic Skills Task and accuracy rate differences in the Arithmetic interference task between experimenter present and absent conditions, error bar representing SE.

Three independent sample t-tests were performed to test whether these mean differences were significant. The experimenter condition was an independent variable, and RT for both tasks and accuracy for the arithmetic interference task were dependent variables. It was found from the t-test that the mean RT in the arithmetic interference task, $t_{(2,68)}$ = -.396, p = 0.69, d_s = *0.47,* and the arithmetic skills task $t_{(2,68)}$ = -.563, p = 0.58, d_s = *2.99,* did not differ significantly for different conditions. Furthermore, the accuracy difference between these two conditions for the arithmetic interference task was also non-significant $t_{(2,68)}$ = -.732, p = 0.47, d_s = *0.13*. The experimenter was present, and the absent condition did not significantly affect the study's outcome. That means the environment did not impact the way children performed the task. Furthermore, it was observed from the above findings that children performed more efficiently in home conditions than in face-to-face conditions. That means children were more comfortable in their homes than in the face-to-face lab environment.

It could be summarized from the above findings that the testing environment did affect task performance but only in response time level and not in accuracy. However, in both Online and face-to-face settings, children did not fulfil the minimum 5-trials criteria. These findings indicate that there is probably a problem in designing the task. Some additional analysis was done on the arithmetic interference task, which is discussed below.

Additional Analysis

The above findings suggest that the experimental design for the online/2022 SSW arithmetic interference task needs some modification since it does not fulfil the

PES measurement requirement, i.e., a minimum of 5 trials under each of four RSIs. In the pilot study, all the trials were divided between one short RSI (200 ms) and one long RSI (750 ms) (See Chapter II). However, I designed the task with four different RSIs in this follow-up experiment (See Chapter III). The RSI in the post-error trial changed for each error, and a correct target trial was set for that corresponding post-error trial with a similar RSI to keep the tasks short and less tedious for the children (See Chapter III). that even though this whole task design kept the task short and less tedious for the children, it did not fulfil the PES measurement criteria. One possible explanation is that compared to adults, children either performed better or very poorly. If they perform better, then they will not have enough post-error or post-correct trials under each RSI. Or if they perform very poorly then they will make more consecutive errors or blank responses in post-error or post-correct trials which also lead them not to have enough post-error and post-correct trails under each RSI. In Appendix II *Table 9.6* where it was found that overall children have fewer post-error and post-correct trials. As a result, even though they made a fair amount of error in the arithmetic interference task it was not enough to fulfil the criteria of 5 trials under each RSI. Here, in Table 4.1, I presented the percentages of post-correct and post-error trials that became blank or error responses. It could be observed that post-correct trials became blank responses and children also made consecutive errors after error trials. As a result, children did not have enough post-error and post-correct trials under each RSI. After collecting online and in-person (SSW 2022) data, it could be concluded that with this minimum number of errors, many post-error and post-correct trials became blank or error responses. The percentages of consecutive errors were also visible for post-error trials.

Table 4.1 Descriptive statistics of post-error and post-correct trials turned into blanks or error responses.

Study Environment		Post-Correct Trials		Post-Error Trials	
		Blank Response	Error Response	Consecutive Errors	Blank Response
In-person (n= 71)	Mean (SD)	0.96(1.44)	6.58(6.18)	9.36(11.35)	0
	Percentages	3.78%	25.87%	5.32%	0.00%
Online (n=72)	Mean (SD)	0.65(1.20)	5.78(7.20)	6.63(12.27)	1.01(2.08)
	Percentages	2.56%	22.66%	3.80%	0.58%

As a result, it was hard to get enough post-error and post-correct trials under each RSI. Thus, it indicates that this experiment either needs modification with more extensive trials or children must be tested in two different sessions with the same task to provide more trials to fulfil the PES measurement criteria.

Discussion

In this thesis section, two comparison experiments were done to investigate the effect of the testing environment on task performance. The main reason was to discover why the modified version of the arithmetic interference task did not fulfill the PES measurement criteria (5 trials per condition). Some hypotheses were formed to explain this discrepancy between the Pilot and Online study. Firstly, it was assumed that children who were better and more interested in arithmetic participated in the online study. Secondly, it was speculated that participants were less rigid to online tasks than in-person due to the known home environment. In the first experiment of this chapter, a comparison between the Pilot and Online data was made. It was found that online participants performed better than the in-person participants from the pilot study. That means the environment influences children's performance.

Then, a second experiment was done with a larger sample to confirm our hypotheses. It was found in the second experiment that the testing environment did not influence the accuracy of the task, but it influenced the time needed to solve arithmetic problems. The response time was faster at home than in the in-person environment. A further investigation was done to see if the experimenter's presence impacted task performance, and no significant effect was found. It could be said that the testing environment influences task performance in a certain way.

Findings from the additional analysis indicated that the main problem was not the testing environment but how the task was developed for children. The primary purpose of such task design was to make it short and less tedious for children. However, it was found that children made many consecutive errors in a row, turning many post-correct and post-error trials into error trials under each RSI. Next time, the experiment needs to run in more trials or multiple sessions in the same group of children.

5. Chapter V: Relationship between Error Awareness and Arithmetic Achievement

Abstract

Metacognitive awareness is related to better academic performance (e.g., Young & Fry, 2008). One of the essential aspects of metacognition is metacognitive judgment, where people consciously monitor their performance to make future decisions (Flavell, 1979). Most studies that found that metacognitive judgment is associated with arithmetic achievement used calibration of judgment measurement technique (i.e., the alignment between one's judgment confidence in a problem and the accuracy of the judgment, e.g., Rinne & Mazzocco, 2014). In adults, this relationship appears to be domain-general rather than domain-specific (Bellon et al., 2020). However, the relationship between error awareness and arithmetic achievement was not investigated despite having an important role in performance monitoring and error processing (Kirschner et al., 2020). The present research focused exclusively on error awareness and how it relates to arithmetic achievement. Non-numeric and numeric tasks were developed to investigate the domain-general vs. the domain-specific role of error awareness on arithmetic achievement. It was found that non-numeric error awareness predicted better arithmetic efficiency, but numeric error awareness predicted less arithmetic efficacy with longer response time in arithmetic skills tasks and less accuracy in numeracy tests. It was not possible to conclude about domain-specificity since a bidirectional relation was observed between error awareness and arithmetic achievement based on task types.

Chapter V

Introduction

Metacognition monitoring and awareness are essential for academic achievement and solving math problems (e.g., Young & Fry, 2008; Özsoy, 2011). We use different mental strategies (e.g., counting from first for addition and separating from for subtraction; Gilmore et al., 2018) while solving different mathematical problems. It is essential to become aware of these strategies and to execute the proper response to control one's cognitive system. This awareness could only be obtained through metacognitive monitoring and regulations, which helped us solve mathematical problems using different mental strategies (e.g., Schneider & Artelt, 2010). Because of the importance of metacognition on arithmetic, the current thesis investigated the domain-specific vs. domain-general role of metacognition on arithmetic achievement. Here, metacognition was measured in terms of error monitoring and awareness. Chapters 2 and 3 investigated the domain-general vs. the domain-specific role of error monitoring (PES and PECA) on arithmetic in adults and children. The main reason for focusing on PES and PECA was that they do not rely on verbal reporting, which is assumed to be a reasonable measure for metacognitive monitoring in children. However, if someone wants to investigate regulatory and evaluation aspects of metacognition in adults, collecting information on ongoing performance is a reliable measure of metacognition. In this chapter, I mainly focused on investigating regulatory and evaluation aspects of metacognition regarding error awareness to see how it relates to arithmetic achievement in adults.

It is only sometimes possible to consciously perceive every error daily. For example- sometimes people break traffic rules and only become aware once they get the ticket. However, if someone became aware that they broke a traffic rule, they became more cautious at the next signal. This error awareness helps us make prudent decisions in the next event and improves our future performance.

Errors that cannot make it up to the awareness and go unnoticed are called unaware errors. An error is only aware if it filters through the highest form of consciousness, called access consciousness (Wessel, 2012).

Access consciousness can be defined as *"when a subject has a certain sort of access to the content of the state. More precisely, a state is access conscious if, by virtue of having the state, the content of the state is available for verbal report, rational inference, and the deliberate control of behaviour."* (Bayne and Chalmers, 2003; as cited by Wessel, 2012).

The present study's primary interest is the reported error that helps us monitor and regulate our performance. Access consciousness is the highest form of transitive consciousness, where the knowledge about the error is object-related. According to Wessel (2012), whether the consciousness will become transitive depends on the subjective experience of the stimulus. If the error is aware, it reaches the highest level of transitive consciousness, and when it is unaware, it can be unconscious or stuck in other levels of consciousness (Table 5.1).

Table 5.1 The typology of transitive consciousness is based on different theoretical accounts from the philosophy of mind (Wessel, 2012)

	Levels of Consciousness	
Quality of stimulus **Representation** ↑	Access Consciousness	Reportable Intentionally controlled Attentional focus
	Interoceptive Consciousness	Higher-Order Thoughts
	Phenomenality	Subjective experiences
	Unconsciousness	No representation

Brain imaging studies have also demonstrated that pre-conscious and conscious error detection mechanisms show different brain activation. For example, a study conducted by Scheffers & Coles (2000), investigated the brain's event-related potential (ERP) for accurate judgment and types of errors. They found that the largest error-related negativity (ERN) was associated with sure and unsure errors, but the smallest ERN was associated with trials rated as sure and unsure correct. However, no significant difference in error-related negativity (ERN) was observed between unsure error and correct response. The error types showed a larger amplitude for sure errors and smaller amplitudes for uncertain errors. Another experiment (O'Connell et al., 2007) investigated the ERN and error positivity (Pe) in aware error, unaware error, and correct response trials. It was found that a larger ERN was observed for error response trials than correct response trials, but it did not vary significantly between aware and unaware errors. However, the highest Pe was observed for sure error response. Pe did not differ significantly for unaware error and correct response trials.

Aware and unaware errors modulate neural activity and behavioural error processing. For example- The neural and behavioural correlates of error processing (e.g., the error-related negativity (ERN) and error positivity (Pe) and error awareness were investigated in one study (Kirschner et al., 2021). It was found from this study that error awareness and confidence had a modulating effect on both the ERN and Pe, whereby Pe was most predictive of participants' error awareness. Behavioural differences were also observed between aware and unaware errors, where it was found that only aware errors had a slowing effect on reaction times in consecutive trials. However, this slowing was not accompanied by post-error improvement in accuracy. They also found that error awareness mediates the relationship between ERN and post-error behavioural adjustments (PES).

In the present study, the focus is not on investigating the neural activity of error awareness. However, the above discussion indicated that aware error, unaware error, and correct response share common but unique underlying cognitive processes. Individuals who show a higher percentage of aware errors showed better monitoring processes. The present study chose aware error to measure error awareness as a marker of better performance monitoring.

Measurement techniques for error awareness.

Researchers who study error awareness have used different measurement techniques. For example, Scheffers & Coles (2000) used a letter version of the Flanker task to measure error awareness in their study. In this task, an array of letters was presented, and participants were instructed to use either the left (for letter H) or right (for letter S) key based on the target letter (Middle letter). For example, the stimulus was like HH**S**HH, SS**H**SS, HH**H**HH, and SS**S**SS. They used a 5-point rating scale after each trial (from left to right, "sure correct" to "sure incorrect"), which was a calibration of judgment technique. Their study to measure error awareness only took judgment ratings associated with an error response. Another technique Nieuwenhuis et al. (2001) used in this field to measure error awareness is signalling errors, where they used a visual cue task and told them to press a button (spacebar) if they thought they made an error. In another study conducted by O'Connell et al. (2007), they used a Go-no-Go task of Stroop where participants were instructed to use an error awareness button only after an error in No-go trials by abolishing the Go trial. A binary rating after each trial is another technique for measuring error awareness (Endrass et al., 2005). In this technique, participants had to respond if their response was correct or wrong after each trial. The findings from these error awareness studies were not discussed more elaborately since they all focused on brain imaging data, which is not the primary concern of the present experiment. However, some drawbacks are identified from the measurement as mentioned above techniques. Firstly, with the rating method, participants are asked to judge their performance after each trial, which is highly likely to interfere with the participant's normal cognitive process while performing the task. Secondly, with the error signalling technique, participants only reported their error response. They were highly likely to underreport errors even if they noticed them since they must respond differently after an error. As a result, if they forget to follow the instructions for the task, many noticed errors will be categorized as unnoticed. I used a measurement technique in the present study that only takes the performance rating in penultimate trials: the trial after an error trial. This method measures response awareness for error, correct responses, and post-error slowing.

Domain-general vs. Domain-specific role of metacognition

Another vital factor in metacognitive studies is metacognition's domain-general vs. domain-specific nature. For example, de Gardelle et al. (2016) found the domain-general role of metacognition in different task domains. In their experiment, they developed a perceptual task, including visual stimulus (Gabor patch stimuli: series of black and white bars) and auditory stimulus. The task varied with the sensitivity level and had three blocks: visual, auditory, and mixed with visual and auditory stimuli. A pair of trials were presented in succession order with a fixation point coloured in red or blue as an inter-stimulus interval (500 ms). A metacognitive judgment window appeared with a pair of red and blue colour boxes associated with the colour of the fixation point. Participants had to report their confidence judgment by selecting the colour box associated with the fixation point of the trial in which they felt more confident about their judgment. It was hypothesized that the effect of confidence judgment on task sensitivity would differ across different task conditions (visual, auditory, and mixed conditions) when task-specific (domain-specific). The effect of confidence judgment on task sensitivity would not differ across different conditions when task-generic (domain-general). It was found from their study that participants with trials showing higher task sensitivity showed higher task confidence. Metacognitive ability was unchanged between the within-task (auditory and visual) and the across-task (mixed) conditions for confidence comparison. These findings indicated that the confidence judgment is task-generic (domain-general).

Another study (Scott & Berman, 2013) investigated metacognition's domain-general vs. domain-specific role in different subject areas. It was survey research with 644 college students. In their study, Scott and Berman (2013) used a questionnaire to measure metacognitive knowledge and regulation. The courses were divided into two domains: humanities and science. Metacognitive judgment on accuracy was measured by comparing the predicted score of each participant after each exam with their obtained results. Students' metacognitive judgment on accuracy was calculated in percent by comparing their predicted and actual grades from the exam. It was found from their analysis that knowledge and regulation aspects of metacognition measured in the questionnaire did not differ across different subject areas; however, metacognitive judgment on accuracy significantly varied across domains, leading to poorer judgment on science than humanities.

Findings from this study indicate that the assessment technique could influence metacognition's domain-specific vs. domain-general role on academic achievement.

However, a different scenario with mixed findings has been observed when studying domain-general vs. domain-specific aspects of metacognition in children. Geurten et al. (2018) conducted a study with three groups of children aged 8-9, 10-11, and 12-13. They used one arithmetic task with two-digit addition problems and one memory task with words. Participants were asked to use some strategies to solve arithmetic and memory problems. Then, they made metacognitive judgments after each trial to rate the ease of better strategy selection and the confidence judgment of their response. It was found that metacognition had a domain-specific role before age 10. As children mature, the domain-general role of metacognition is observed.

A study conducted by Bellon et al. (2020) used a standardized arithmetic test and spelling test to measure academic achievement in these two domains. They developed a computerized arithmetic and spelling task to measure metacognitive monitoring in each domain. The first part of their study analysed the data for children 8-9 years old, and a domain-general role between metacognitive monitoring and standardized tests was observed. However, the second part of the analysis was done with children between 7 and 8 years old. A domain-specific role of metacognitive monitoring on the standardized tests of arithmetic and spelling was observed. These findings indicated that as children grow older, the relationship between metacognition and academic performance shifts from domain-general to domain-specific.

The main interest of the present study is to investigate the regulatory aspects of metacognition regarding error awareness. It could be summarized from the above discussion that the domain-specific vs. domain-general role of metacognition depends on the metacognitive measurement techniques and the studied components. It is assumed to be domain-general in adults.

Metacognitive Judgment and arithmetic achievement

As mentioned, metacognition has different components (e.g., metacognitive knowledge and regulation/experience). Past research found the significant importance of metacognition in learning (Young & Fry, 2008; Freeman et al., 2017). Still, researchers have only recently started investigating the role of metacognitive judgment on arithmetic achievement. Metacognitive judgment of performance is related to metacognitive experience, where participants must consciously monitor their performance as a step-by-step evaluative decision-making process (Flavell, 1979).

Studies found a significant relationship between arithmetic achievement and metacognitive judgment. For example- Rinne & Mazzocco (2014) conducted a study with children (Grades 5-8) with and without learning disabilities. They investigated the relationship between the calibration of judgment (i.e., the alignment between confidence in judgments and the accuracy of those judgments) and mental arithmetic accuracy. Their study found that higher judgment calibration is associated with higher mental arithmetic accuracy. Children with learning disabilities showed lower mental arithmetic scores than their typically developing peers, and their judgment calibration was also highly inaccurate.

In another study, Bellon et al. (2019) investigated the role of metacognition and executive function in arithmetic in second-grade students. In that study, they measured arithmetic by a production task of addition and multiplication. A questionnaire was used to measure global metacognitive knowledge and calibration of judgment in the arithmetic task was used to measure task-based metacognitive monitoring. It was found that participants who had better general metacognition knowledge performed significantly more quickly in solving addition problems, but a non-significant relationship was observed for multiplication. Calibration of judgment contributed significantly to the accuracy of addition and multiplication problems. These findings indicated that metacognition's domain-general and domain-specific role in arithmetic depends on the aspects of metacognition that would be measured.

The current study

It is also clear from the above discussion that metacognitive judgment of performance is an essential marker for arithmetic achievement (Rinne & Mazzocco, 2014). Past studies found that the relationship between metacognition and arithmetic could be domain-specific or domain-general, depending on the assessment techniques and the component of metacognition being investigated (Scott & Berman, 2013; Bellon et al., 2019). However, this research did not investigate the domain-general vs. the domain-specific role of error awareness on arithmetic achievement. Even though different brain and behavioural studies also indicated that error awareness is an essential marker for performance monitoring (e.g., Kirschner et al., 2021).

A pair of numeric error awareness (arithmetic interference) and non-numeric (modified Stroop task) error awareness tasks were developed to fill this research gap in the current study. It was the first time I investigated the relationship between error awareness aspects of metacognition and arithmetic achievement.

It was hypothesized from the previous research that better error awareness would predict better arithmetic achievement (Rinne & Mazzocco, 2014), and this relationship would be more domain-general than domain-specific in adults (de Gardelle et al., 2016; Bellon et al., 2020). If relationship would be domain-general, then both numeric and non-numeric error awareness would predict better arithmetic achievement.

Chapter V

Methodology

Participants

In the present study, to determine the sample size, GPower 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) software was used to do a priori power analysis for 1 – 2 step hierarchical linear regression. In GPower, it was estimated that analysis for a small effect size of 0.02666667 for regression in R^2 change (based on the data from similar studies, e.g., Bellon & De Smedt, 2019) with a sample of 297 adult participants was sufficient to achieve a power of 0.80. However, 300 adult participants were recruited for this study. It was found that seven participants had colour blindness, and they were excluded from further analysis. Finally, 293 participants were retained for the final analysis (Mean age = 25, SD = 8.43). Among these participants, 142 were female, 145 were male, one was non-binary, and five were non-specified. Two hundred fifty-six participants were monolingual, and the rest 37 were multilingual. The primary ethnicity of two hundred nine participants was British, 20 were Asian, 27 were European, 23 were mixed, and 14 were other ethnicities. The pre-registration of this study is attached to Appendix III.

Design

This experiment used a multi-level regression model design where two error awareness tasks (numeric error awareness task and non-numeric error awareness task) were predictors. Scores from two different arithmetic tests were dependent variables.

Materials

In the present study, two error awareness tasks were developed. One was a non-numeric error awareness task, and another was a numeric error awareness task. Since the most popular arithmetic achievement tests don't have any online but booklet versions, in the present study, it was decided to use two online mathematical measures that measure arithmetic skills and achievement separately. The first test is called the arithmetic skills task for measuring arithmetic procedural skills, and the numeracy test is for measuring arithmetic achievement. Arithmetic skill task was already tested in a previous study and correlated well with standardized arithmetic achievement tests (Cragg et al., 2017). The numeracy test (Chinn, 2020) is a newly developed test that covers a wide range of arithmetic problems and is designed to measure overall performance in arithmetic.

Error Awareness Task

Non-numeric Error Awareness Task. This task is a modified Hester et al. (2005) Stroop task. Like in the Stroop task, participants were shown the words "Blue," "Red," "Green," and "Yellow" presented in different ink colours, and an additional criterion was the use of three possible combinations of upper-case and lower-case letters (e.g., GREEN, green, Green). There was a total of 8 blocks with 32 trials in each block. Among the 32 trials, 16 were congruent, and 16 were incongruent. In congruent trials, the colour of the words and the name of the colour in the words were matched. In incongruent trials, the colour of the words and the name of the colour in words were mismatched. The response stimulus interval (RSI) was 500 ms. Participants were asked to press certain buttons [(m), (,), (.), (/)] for the colours Blue, Green, Red, and Yellow in alphabetic order. Since it was a modified version of the

Stroop task, we created eight target trials among these 32 trials. Here, targets were specific words written in a particular colour and with a specified form of capitalization. Two specific targets were associated with each block (fixed targets per block), with one target staying the same from one block to the next and the other changing (so participants only needed to update one target per block). For example, GREEN (in green colour) and blue (in yellow colour) words could be target words for a particular block. However, green, BLUE, Green, and Blue were not target words for that block. Participants were instructed to press the spacebar each time they saw target words instead of any other predefined keys [(m), (,), (.), (/)]. In the target trials, participants had to remember some extra rules apart from traditional Stroop task rules, which were assumed to require more working memory load and created a dual-task paradigm. For example- In block one, if the target words are GREEN (in green colour) and blue (in yellow colour), then participants were asked to press the spacebar, which is different from predefined keys [(m), (,), (.), (/)] to perform the Stroop task. Even though the working memory load was not calculated for these target trials, it was assumed to create an extra working memory load. Due to this extra memory load, participants are likelier to make unaware errors in target trials due to failing to remember the extra rules other than Stroop rules. Of these eight target trials, half were congruent trials, and half were incongruent trials written in letters either with upper-case or lower-case or a combination of upper- and lower-case letters. The trials were wholly randomized within a given block and differed across participants. The experiment was coded in Psychopy-2022-2.4 (Peirce, Hirst, & MacAskill, 2022) so that each time participants committed an error, a judgment question appeared on the E(error trial)+2 trial after an error response

(Figure 5.1). Participants were asked ", "Was your penultimate response incorrect or correct?" They were given four multiple-choice answers- 'Definitely Correct,' 'Probably Correct,' 'Probably Incorrect,' and 'Definitely Incorrect' from which they had to choose their response. This question only appeared after a correct response followed by an error response in the penultimate trial. For each post-error judgment, this same question appeared for matched post-correct trials, selected randomly among the 4th to 8th trials after an error trial.

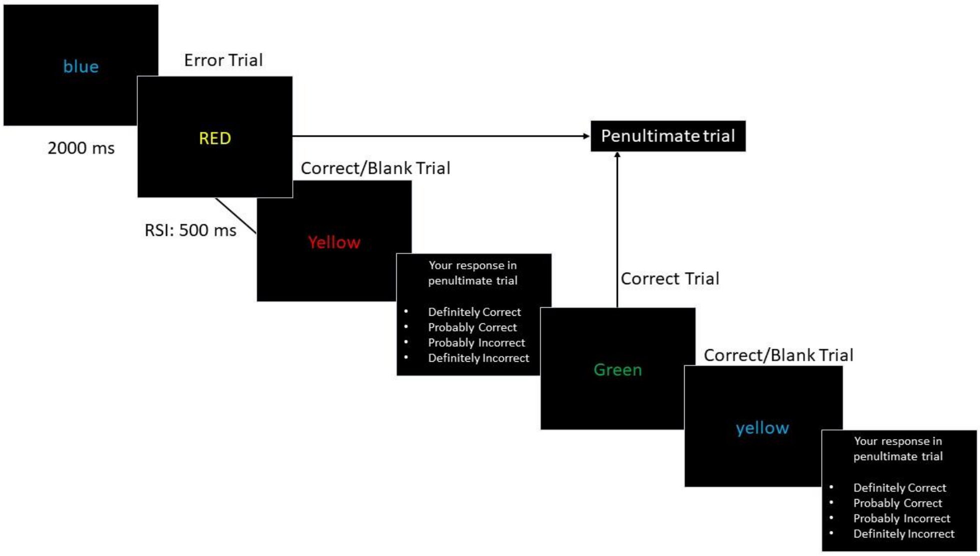


Figure 5.1 Example trials for non-numeric error awareness Stroop task

To match the numbers between the post-error and post-correct judgment questions and to avoid the chance appearance of two/three post-correct trials simultaneously, the post-correct trials were not selected randomly. The reason for selecting one trial after a post-error trial was to keep the option open for measuring PES in the future.

In our analysis, only the percentage of aware error was calculated. If the response to the participant's error awareness question matched with their performance in error trials, then participants were given scores of 1 for their

'Probably Incorrect' response and 'Definitely Incorrect' response, and a score of 0 was given for all 'Probably correct' response and 'Definitely correct' response. Here, variable measures were taken as a percentage of aware and unaware errors. For further clearance, the instructions for the task are provided in Appendix III (page 203).

Numeric Error Awareness Task. This was an ordinary arithmetic task where participants had to solve simple addition and multiplication problems comprised of all numbers 2 to 9 combinations. Tie problems (e.g., 2+2 or 5x5) were excluded. There was a total of 8 blocks with 32 problems in each block. Among these 32 trials, 16 were addition problems, and 16 were multiplication problems. An example of a trial is given below in Figure 5.2.

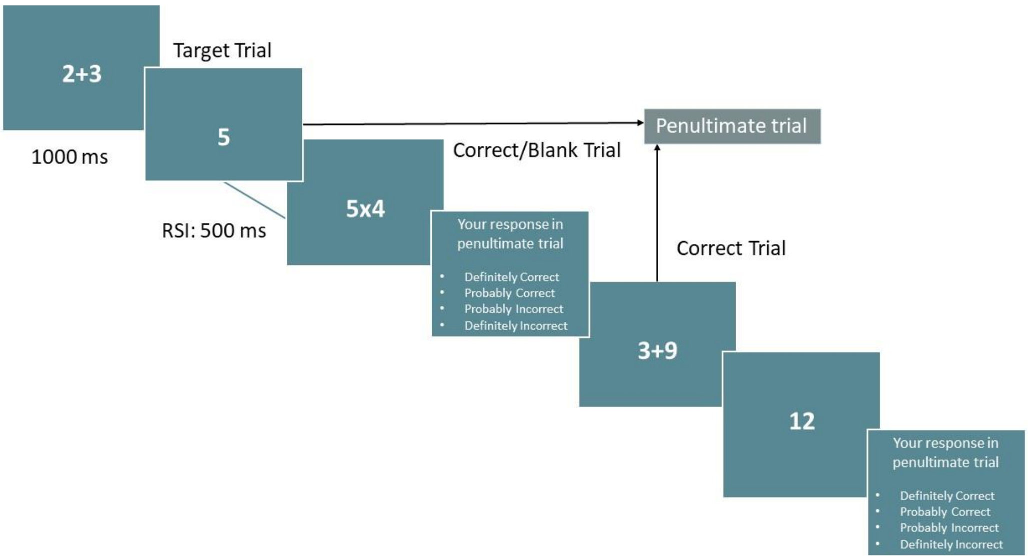


Figure 5.2 Example trials for non-numeric error awareness Stroop task

Participants saw the arithmetic problems (addition and multiplication) in the center of the screen for one second. The RSI was set to 500 ms, and participants were given 3 seconds to type in their responses. We also created eight target trials for each block to increase the task difficulty by introducing some rules to make it like the non-numeric task. Participants see one target number (any number between 2 and 9) in each block. They were asked to do the alternative maths operation of what they were asked for the target trials, i.e., if the target number is 9 in block-1 and the problem on the screen appeared was 9+3, they had to perform multiplication instead of addition (i.e., calculate 9 x 3). So, the correct answer would be 27, not 12. The number used as a target was not used for any other sums in the same block. Each block's selection of target numbers was randomized, but they were presented in the same order for all participants. The introduction of Target trials also partly led me to believe that this dual-task paradigm would let participants create more unaware errors. The trials in each block were randomly presented to the participants. The post-error trials, post-correct trials, and error judgment questions were coded similarly to the non-numeric error awareness task in Psychopy-2022-2.4 (Peirce, Hirst, & MacAskill, 2022). The measurement of error awareness was also the same as the non-numeric error awareness task.

Mathematics Tasks

The arithmetic achievement of the participants was measured with two different types of tasks. One is the arithmetic skills task that measures arithmetic procedural skills, and the other is the numeracy test that measures arithmetic achievement. Both are discussed below-

Arithmetic skills task. This is an arithmetic procedural skills task developed initially by Cragg et al. (2017). An adapted online version of the task was created. This is the same task we used previously; the descriptions are available in the

Chapter III method section. The variable measured was mean RT for correct responses.

Numeracy Test. This test is a modified version of the 15-minute norm-referenced mathematics test based on a population of over 2,500 students and adults across the UK (Chinn, 2020). The primary purpose of this test was to compare the performance level of the subjects with their peers. The 15-Minute Mathematics Test covers a range of arithmetic and algebra tasks (44 items), but it is primarily a test of arithmetic. The content is more procedures-based rather than facts-based. The level of test difficulty rises with the test's length.

Morsanyi, Trickett, and Chinn (2020; unpublished) developed an online version of 42 arithmetic and algebra tasks. In the present study, I programmed their task in Psychopy using the same stimulus. In this online version, participants had 15 minutes to complete the task and 3 minutes for each item. If participants made five consecutive errors, the test would automatically terminate. Like the offline version, participants used paper and pencil to solve maths problems but not a calculator. The measure for this test was taken as accuracy.

Procedure

This study recruited participants from the University of Nottingham, UK, the online SONA system, and the online participant recruitment platform Prolific. Participants were provided an Information sheet, consent form, and demographic questionnaire in Qualtrics. After filling out the form in Qualtrics, they were redirected to the Pavlovia website to participate in the experiments. Participants who participated in the SONA system were credited 1.25 points for their participation, and participants recruited by Prolific received 10 pounds each.

The University of Nottingham School of Psychology's ethical committee approved the study ethics.

Results

Some preliminary analyses were done before presenting the primary findings, which are discussed below.

Outliers Tested

As mentioned in pre-registration, data were tested for potential multivariate outliers with the Mahalanobis distance test. Sixteen participants whose probability estimation for the X^2 value was less than 0.05 were excluded from the final analysis. Finally, 274 participants were retained for the final analysis.

Pre-processing Analysis

The mean error rate was calculated for numeric and non-numeric error awareness tasks (Figure 5.3).

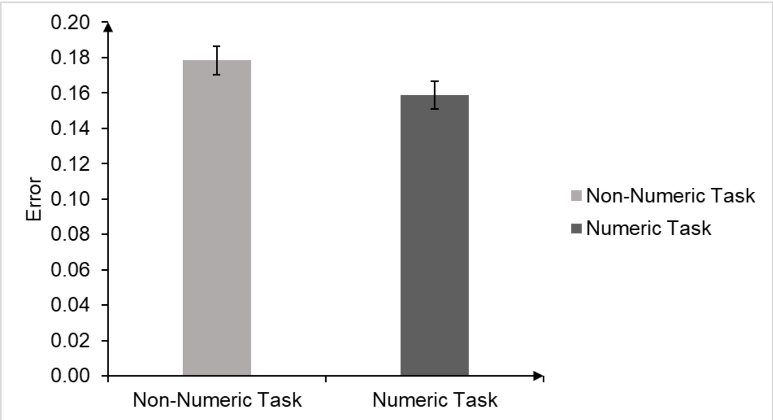


Figure 5.3 Mean accuracy rates in numeric error awareness task and non-numeric error awareness task, error bar representing SE.

It can be observed from Figure 5.3 that the rate of overall errors in the numeric (M= 0.16, SD= 0.13) error awareness task is higher than the non-numeric (M= 0.18, SD= 0.13) error awareness tasks. The paired-sample t-test also indicated this difference was significant, $t_{(1,\ 273)}$= -2.38, p< 0.02, d_s= .14. To test the differences between aware and unaware detection of errors a paired-sample t-test was done (Table 5.2). It can be observed from Table 5.2 that participants were significantly more aware than unaware of their errors in both tasks. It could be assumed from these findings that participants were attentive in doing the tasks. In addition to this, another pair-sample t-test indicated that the mean percentage of aware errors was significantly higher in non-numeric (*M*= 0.59, *SD*= 0.271) than in numeric error awareness tasks (*M*= 0.54, *SD*= 0.273), $t_{(1,\ 273)}$= -2.83, p= 0.005.

Table 5.2 The paired-sample t-test between different error awareness in both tasks

	Error Awareness	Mean	SD	$t_{(1,\ 273)}$	*P*
Non-numeric Error Awareness Task	Unaware Error	0.41	0.272	-2.556	0.011
	Aware Error	0.59	0.272		
Numeric Error Awareness Task	Unaware Error	0.46	0.273	-5.740	<.001
	Aware Error	0.54	0.273		

In the pre-registration, it was said that the accuracy of the arithmetic skills task would not be included as a measure because it was designed to measure how much time participants needed to process each problem and solve it. It was believed that this 12-item task with 2.5 seconds of solving time for each problem was easy enough to acquire higher accuracy for all participants. To test this, I calculated the percentages of participants who scored nine and above in the arithmetic skills task, and it was found that 78.5% of the participants scored above 9 out of 12. This percentage indicates a probable ceiling effect for accuracy measurement. On the other hand, for the Numeracy test score, no such ceiling effect was observed. Only 24.45% of participants obtained scores between 31 and 33 out of 44. Based on these findings, I did not include an accuracy score for the arithmetic skills task. The descriptive statistics from the arithmetic skills task and numeracy test are presented in the Table 5.3.

Table 5.3 Descriptive statistics on Numeracy test and Arithmetic Skills Task

	Numeracy Test Score (Accuracy)	Arithmetic Skills Task (Mean RT)	Arithmetic Skills Task (Accuracy)
Mean	25.50	5.188	9.10
SD	7.209	2.242	1.200
The percentage of participants who obtained higher scores	24.45% scored between 31 and 33	78.5% scored 9	

Main Analysis

Correlations

The primary purpose of this study was to investigate the relationship between the domain-general and domain-specific aspects of error awareness in arithmetic. Firstly, we calculated the Pearson product-moment correlation among aware errors in numeric vs. non-numeric error awareness tasks, mean RT for arithmetic skills tasks, and accuracy on the numeracy test. Correlations are presented in Table 5.4. It can be observed from Table 5.4 that the relationship between error awareness in numeric and non-numeric error awareness tasks is positive, i.e., participants who were aware of their errors in the numeric error awareness task were also aware of their errors in the non-numeric error awareness task.

Table 5.4 Pearson-product moment correlation among error awareness in numeric vs. non-numeric error awareness task with Arithmetic skills and Numeracy test score

	Non-numeric Error Awareness Task	Numeric Error Awareness Task	Numeracy Test Score	Arithmetic Skills
Non-numeric Error Awareness	1			
Numeric Error Awareness	.361**	1		
Numeracy Test Score	.100	-.066	1	
Arithmetic Skills	-.095	.180**	-.240**	1

*Note: ** Correlation is significant at the 0.01 level (2-tailed).*

A significant negative correlation was observed between the arithmetic skills task and the numeracy test performance. It indicates that participants who did well in the numeracy test also performed efficiently (took less time correctly solving problems) in the arithmetic skills task.

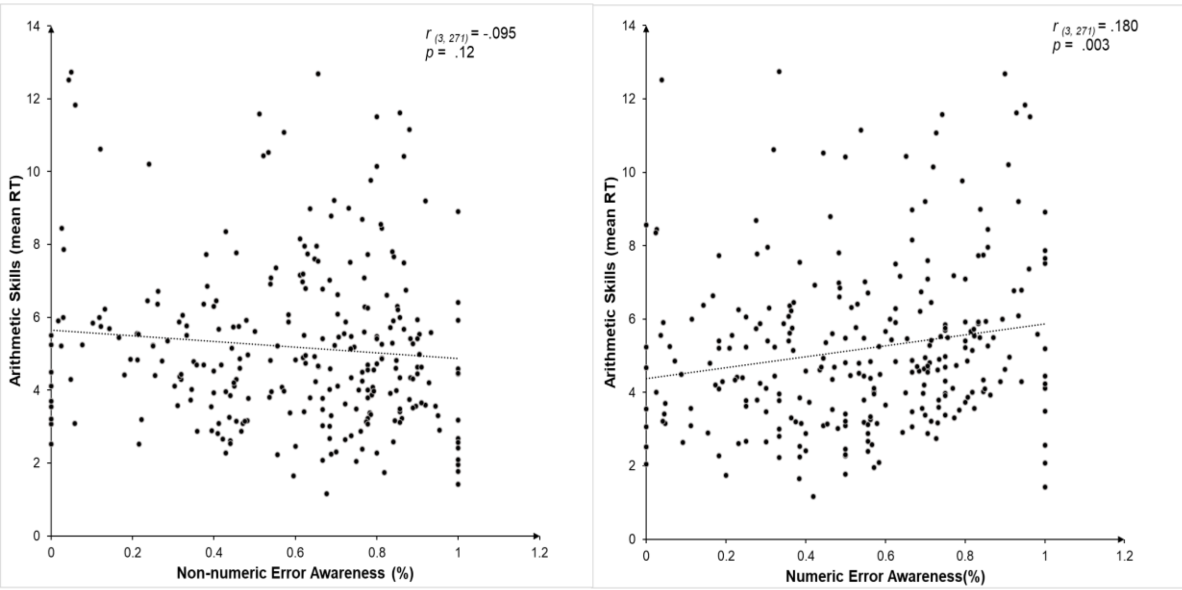


Figure 5.4 Pearson product-moment correlation between mean RT of arithmetic skills with aware error in the non-numeric and numeric error awareness tasks.

However, a significant positive correlation was observed between numeric error awareness and arithmetic skills tasks. Participants who were more aware of their errors in the numeric error awareness task took more time to correctly solve the problems in the arithmetic skills task. One reason could be that participants who are good at detecting their numeric errors are careful when solving math problems, so they take an extended time to solve arithmetic problems. However, the correlation between aware errors in the non-numeric error awareness tasks and arithmetic skills was non-significant. No significant relationship was observed between numeracy and error awareness.

Relationship between Non-numeric and Numeric Error Awareness of Arithmetic Achievement

Hierarchical regressions were carried out to assess the role of error awareness in arithmetic skills only. No regression analysis was run for numeracy test scores since there was no significant correlation between error awareness and Numeracy (Table 5.5). In the first hierarchical regression analysis to investigate the role of error awareness in arithmetic, non-numeric error awareness was entered into the first model.

Table 5.5 Hierarchical linear regression predicting arithmetic achievement by non-numeric and numeric error awareness

Dependent Variable	Arithmetic Skills (mean RT)	
Predictors	Model I ß	Model II ß
Non-numeric Aware Error	-.095	-.184**
Numeric Aware Error		.247**
R^2	.009	.062
F for R^2 Change	2.49	15.33**

*Note: ** p< 0.01*

As shown in Table 5.5, the non-numeric error awareness measures alone explained 0.9% of the variance in arithmetic achievement, and the model was non-significant, $F_{(1,272)} = 2.49$, $p = 0.116$. In the second model, numeric error awareness was added to the analysis to investigate the error awareness. Model II of hierarchical linear regression analysis revealed that non-numeric and numeric error awareness significantly explains 6.2% of the variation in arithmetic skills. However, the unique variance of numeric error awareness was 5.3%, and this model II was significant, $F_{2,271} = 8.98$, $p < 0.001$. The regression coefficient associated with non-numeric error awareness, $\beta = -.184$, 95% CI: -(.50 - 2.53), $p = 0.004$, suggests that the arithmetic skill decreased by approximately 0.184 units with each additional non-numeric error awareness unit. The regression coefficient, $\beta = .247$, 95% CI: 1.01-3.06, $p < 0.001$, associated with numeric error awareness suggests that with each additional unit of numeric error awareness, the arithmetic skills increased by approximately .247 units.

It could be summarized from the above findings that participants who were more aware of their errors in the non-numeric error awareness task took less time to solve arithmetic problems. Still, this relationship is only significant when error awareness on the numeric task was added to the model. On the contrary, those more aware of their errors in the numeric error awareness task were cautious in solving arithmetic problems, so they took an extended time to solve them. Since a bidirectional relationship between error awareness and arithmetic skills was observed based on the task types (non-numeric vs. numeric), it was hard to conclude about domain-specificity.

Chapter V

Exploratory Analysis

The primary purpose of this study was to investigate the relationship between the domain-general and domain-specific aspects of error awareness in arithmetic. In our pre-registration, it was decided that error awareness would be calculated for all the trials, irrespective of trial type. However, the above regression analysis findings were unexpected compared to our hypothesis. Especially for the Numeracy test, a significant correlation was expected between error awareness and arithmetic achievement. The type of errors investigated in the error awareness tasks could have something to do with the outcome of the results. In the present error awareness tasks, some "Target Trials" were introduced to ensure the tasks were challenging enough to produce more unaware errors by creating a dual-task paradigm. It was later realized that 'Standard Trials' errors in the numeric error awareness task were associated with arithmetic procedural skills errors and attention errors. Errors in the 'Standard Trials' of the non-numeric error awareness task were associated with the Stroop effect and attention-related inhibition errors. On the other hand, "Target Trials" errors were mostly assumed to be associated with rule-based and attention-related errors in both tasks since participants were asked to respond entirely differently than the standard trials in these trials. Because of this, there is a chance that participants were mostly unaware of their errors in the "Target Trials" because they failed to follow task instructions. Since "Target Trials" only covered a small number of trials (25%), they were expected to produce more unaware errors, which was not the primary purpose of the study to be investigated. The proof for this claim can be found in Table 5.6, where it is shown that both overall error rates and awareness are significantly higher in standard than target trials.

Table 5.6 Results from the Paired-sample t-test

Variables	Task Types	Trial Type	*M(Proportion)*	*SD*	$t_{(2, 273)}$	*P*
Error	Non-numeric	Standard	.123	.120	8.28	<.001
		Target	.065	.047		
	Numeric	Standard	.107	.101	9.32	<.001
		Target	.069	.056		
Error Awareness	Non-numeric	Standard	.478	.232	24.06	<.001
		Target	.116	.120		
	Numeric	Standard	.450	.239	24.94	<.001
		Target	.092	.091		

These findings were interesting from a performance monitoring perspective. Monitoring processes failed more often in target trials than in standard trials, as indicated by the higher percentage of unaware errors. From these findings, the following correlation and regression analysis was decided to run by including 'Standard Trials' only; the main reason for this changed analysis is to measure numeric related error awareness as accurately as possible.

Relationship between error awareness and arithmetic achievement

Like the main analysis, some correlation analyses were done. I did not include the accuracy score for the arithmetic skills task as a measure of arithmetic achievement in the pre-registration. However, based on the examiner's review, some Pearson product-moment correlations are calculated among aware errors in numeric vs. non-numeric error awareness tasks, mean RT for arithmetic skills tasks, and accuracy of the arithmetic skills tasks and the numeracy test. Correlations are presented in Table 5.7.

Table 5.7 Pearson-product moment correlation among errors awareness in numeric vs. non-numeric error awareness task with Arithmetic skills and Numeracy test score for standard trials only.

	Numeric Error Awareness	Non-numeric Error Awareness	Arithmetic Skills (mRT)	Arithmetic Skills (Accuracy)	Numeracy test score
Numeric Error Awareness (standard trials)	1				
Non-numeric Error Awareness (standard trials)	.303**	1			
Arithmetic Skills (mRT)	.227**	-.067	1		
Arithmetic Skills (Accuracy)	-.081	.114	-.212**	1	
Numeracy	-.119*	.131*	-.240**	.197**	1

*Note: *p≤.05; **p≤.001; here, only the "Standard Trials" error awareness is included in the analysis; as a result, correlation values differ from Table 5.4.*

Two hierarchical regression analyses were conducted where error awareness from both tasks was a predictor. The mean RT of the arithmetic skills task and the accuracy of the Numeracy test were dependent variables. However, regression was not run for the accuracy in the arithmetic skills task since no significant correlation was observed with error awareness. In both hierarchical regression analyses, a non-numeric error awareness task was entered in the first block to investigate the role of non-numeric error awareness on arithmetic achievement. Numeric error awareness was entered in the second block to investigate the numeric and non-numeric error awareness as predictors. As shown in Table 5.8, the non-numeric error awareness measures alone explained 1.7% of the variance in Numeracy, and the model was significant, $F_{(1,272)} = 4.75$, $p = .03$.

Table 5.8 Hierarchical linear regression predicting arithmetic achievement by non-numeric and numeric error awareness for standard trials only

Dependent Variable	Numeracy		Arithmetic Skills	
Predictors	Model I ß	Model II ß	Model I ß	Model II ß
Non-numeric Error Awareness (standard trials)	.131*	.184**	-.067	-.149*
Numeric Error Awareness (standard trials)		-.175**		.273**
R^2	.017	.045	.004	.072
F for R^2 Change	4.75*	7.85**	1.21	19.70**

*Note: ** p< .05, ** p< .01*
Numeracy was measured in accuracy; arithmetic skills were measured in mean RT.

Model II of hierarchical linear regression analysis revealed that all the predictors significantly explain 4.5% of Numeracy. However, the unique variance of numeric error awareness on standard trials was 2.8%, and this change was significant, $F_{(2,271)} = 6.36$, $p = .002$. There was a significant association between non-numeric error awareness on standard trials and Numeracy, $\beta = .184$, 95% CI: 1.90 - 9.53, $p = 0.003$. This suggests that with each additional unit of non-numeric error awareness on standard trials, the numeracy score increased by approximately 0.184 units. Similarly, there is a significant association between numeric error awareness and Numeracy, $\beta = -0.175$, 95% CI: -(8.98 - 1.57), $p = 0.005$. This suggests that with each additional unit of numeric error awareness, the Numeracy decreases by approximately 0.175 units.

The second hierarchical regression analysis investigated the role of error awareness in arithmetic skills. The arithmetic skill was a dependent variable, and non-numeric error awareness was entered as a predictor on the first block. As shown in Table 5.8, the model was non-significant, $F_{1,272} = 1.21$, $p = .272$. In the second block, numeric error awareness was added to the analysis to investigate the role of numeric error awareness on arithmetic skills. Model II of hierarchical linear regression analysis revealed that non-numeric and numeric error awareness on standard trials significantly explains 7.2% of the variation in arithmetic skills, and this model was significant, $F_{2,271} = 10.50$, $p < 0.001$. However, the unique variance of numeric error awareness was 6.8%, and this change was significant, $F_{2,271} = 19.70$, $p < .001$.

The regression coefficient associated with non-numeric error awareness and arithmetic skills, $\beta = -0.149$, 95% CI: -(.27 - .26), $p = 0.02$, suggests that the mean RT in the arithmetic skill decreased approximately by .149 units with each addition of non-numeric error awareness on standard trials. The regression coefficient, $\beta = .273$, 95% CI: 3.69 - 1.43, $p < 0.001$, associated with numeric error awareness suggests that with each additional unit of numeric error awareness on standard trials, the mean RT of arithmetic skills decreases by approximately .273 units.

It could be summarized from the above findings that adults with better non-numeric error awareness on standard trials showed higher arithmetic accuracy in the Numeracy test and higher arithmetic efficiency in the arithmetic skills task. On the other hand, participants with better numeric error awareness on standard trials showed less arithmetic accuracy in the numeracy test. Moreover, solving arithmetic problems in the arithmetic skills task took longer. However, the relationship between arithmetic skills and non-numeric error awareness was only significant when numeric error awareness was added to the model. It could be concluded from the above findings that individuals who were more aware of their errors chose a cautious response style by taking longer to solve arithmetic problems to ensure better accuracy. The possibility of speed-accuracy trade-off could be assumed from the negative correlation between arithmetic skills problem-solving RT and accuracy of the arithmetic skills task (Table 5.8). However, it is to be noted that the accuracy score in the arithmetic skills task was very high, and so this claim of speed-accuracy trade-off is hard to make due to the ceiling effect.

Additional Analysis

It was assumed from the previous studies that, like metacognitive judgment, the relationship between error awareness and arithmetic achievement would be domain-general (e.g., Bellon et al., 2020), and better metacognitive judgment would predict better arithmetic achievements (Rinne & Mazzocco, 2014). However, the above findings from our analysis tell an inconclusive story. The direction of the relationship between error awareness and arithmetic achievement was the opposite for numeric vs. non-numeric error awareness tasks. Better non-numeric error awareness predicted better arithmetic performance, but better non-numeric error awareness predicted cautious response style, i.e., speed-accuracy trade-off to ensure better arithmetic performance by choosing a longer response time. However, another possible explanation could be response bias besides having this speed-accuracy trade-off as a cause of a negative relationship between numeric error awareness and arithmetic achievement. Participants with lower error awareness could have a response bias about their performance judgment. As a result, people who were better at detecting correct responses also showed an awareness bias for error responses. This means they underestimated their error response because of their judgment bias for always being right; as a result, a participant who overestimated their correct response awareness might underreport their error responses. Maybe because of that, people who are more aware of their errors showed poor arithmetic skills and numeracy because they had more response bias. However, since this is the first experiment investigating the relationship between error awareness and arithmetic achievement, response biases were not considered an issue. That is why no confidence judgment was added to the measure. To test the effect of response bias, another set of correlations was calculated between response certainty and arithmetic achievement to ensure that the relationship between error awareness was affected by response biases. The response certainty is the participant's awareness of errors and correct responses. The logic behind this additional analysis was to reduce bias by combining correct and error response awareness by taking a single measure of response certainty. The awareness question presented in the current study was four multiple choice question answers. However, to measure response bias, I modified the Likert scoring system to have a continuous spectrum in judgment response. A Likert scoring system is developed where the accuracy of response judgment varies based on their choice of responses: *"2 and 1 scores were assigned for participants' "definitely" and "probably" correct judgment, and -2 and -1 scores were assigned for their "definitely" and "probably" incorrect judgment."* In this scoring system, participants over and under-estimation of their performance was reduced by negative scoring of their wrong judgment. This scoring was done for both correct and error response trials. After scoring the data, a correlation analysis was run. Before running the correlation, analysis data were tested for potential outliers using the Mahalanobis distance test. After excluding all the outliers, 274 data were retained for analysis. It is to be noted that since the variable is response certainty now, not error awareness, the data set after excluding outliers won't be the same as previous data on error awareness. In this analysis, we also included 'Standard Trials' only, as suggested by the examiners. The correlation results are presented in Table 5.9.

Table 5.9 Pearson-product moment correlation among response certainty in numeric vs. non-numeric error awareness task with Arithmetic skills and Numeracy test scores

	Response certainty Numeric Task	Response certainty Non-numeric Task	Numeracy	Arithmetic Skills (mRT)	Arithmetic Skills (Accuracy)
Response certainty Numeric Task	1				
Response certainty Non-numeric Task	.464**	1			
Numeracy	.144*	.243**	1		
Arithmetic Skills (mRT)	-.028	-.127*	-.236**	1	
Arithmetic Skills (Accuracy)	.078	.190**	.057	-.050	1

*Note: * p < .05; ** p < .001; here, only the response certainty is a different variable than error awareness and is only calculated for the "Standard Trials." As a result, correlation values are different from Tables 5.4 and 5.7.*

It can be observed from this table that participants with higher response certainty in non-numeric error awareness tasks showed significantly higher arithmetic accuracy in numeracy and arithmetic skills tasks. They also took significantly less time to solve arithmetic problems in arithmetic skills tasks. For the numeric error awareness task only, a significant relationship was observed between the accuracy of the numeracy test and response certainty. This means participants who showed less response bias in the numeric task also showed higher accuracy in the numeracy test. It is to be noted that in the planned analysis, the accuracy score of arithmetic skills was not included because of the ceiling effect. However, as the examiner asked to add the accuracy score to the analysis, it was included in this analysis.

Since some significant relationship was observed between response certainty and arithmetic achievement, data were further analysed for regression analysis. Three hierarchical regression analyses were conducted where response certainty in both tasks was a predictor. The mean RT of the arithmetic skills task, accuracy of the arithmetic skills task, and numeracy test were dependent variables. In the hierarchical regression analyses, non-numeric response certainty was entered in the first block to investigate the role of response certainty, and numeric response certainty was entered in the second block to investigate the role of response certainty as a predictor. As shown in Table 5.10, the non-numeric response certainty measures alone explained 5.9% of the variance in Numeracy, and the model was significant, $F_{1,272} = 17.19$, $p < 0.001$.

Table 5.10 Hierarchical linear regression predicting arithmetic achievement by non-numeric and numeric response certainty

Dependent Variables	Numeracy		Arithmetic Skills (mRT)		Arithmetic Skills (Accuracy)	
Predictors	Model I ß	Model II ß	Model I ß	Model II ß	Model I ß	Model II ß
Non-numeric Aware Error	.243**	.225**	-.127*	-.145*	.190**	.196**
Numeric Aware Error		.040		.039		-.012
R^2	.059	.060	.016	.017	.036	.036
F for R^2 Change	17.19**	.353	4.45*	.34	10.18**	.034

*Note: ** p< 0.05, ** p< 0.01; Numeracy was measured in accuracy; arithmetic skills were measured in mean RT and accuracy.*

There was a significant association between non-numeric response certainty and Numeracy, $\beta = 0.243$, 95% CI: 3.86 - 1.37, $p < 0.001$. This suggests that with each additional unit of non-numeric response certainty on standard trials, the numeracy score increased by approximately 0.243 units. Model II of this hierarchical linear regression analysis revealed that the unique variance of numeric response certainty was non-significant, $F_{2,271} = 0.353$, $p = 0.55$. There was a significant association between non-numeric response certainty and Numeracy, $\beta = 0.243$, 95% CI: 3.86 - 1.37, $p < 0.001$. This suggests that with each additional unit of non-numeric response certainty on standard trials, the numeracy score increased by approximately 0.243 units. Model II of this hierarchical linear regression analysis revealed that the unique variance of numeric response certainty was non-significant, $F_{2,271} = 0.353$, $p = 0.55$.

The second hierarchical regression analysis investigated the domain-general vs. domain-specific role of response certainty with mean RT of the arithmetic skills. As shown in Table 5.10, the non-numeric response certainty measures alone explained 1.6% of the variance in arithmetic skills. The model was significant, $F_{1,272} = 4.45$, $p = .04$. The regression coefficient associated with non-numeric response certainty and arithmetic skills, $\beta = -.127$, 95% CI: -(.03 - .77), $p = .04$, suggests that the mean RT in the arithmetic skill decreased approximately by .127 units with each addition of non-numeric response certainty on standard trials. Model II of hierarchical linear regression analysis revealed that the relationship between non-numeric response certainty was non-significant, $F_{2,271} = .34$, $p = .56$.

The third hierarchical regression analysis investigated the domain-general vs. the domain-specific role of response certainty with the accuracy of the arithmetic skills. As shown in Table 5.10, the non-numeric response certainty measures alone explained 3.6% of the variance in arithmetic skills, and the model was significant, $F_{1,272} = 10.18$, $p = .002$. The regression coefficient associated with non-numeric response certainty on standard trials and arithmetic skills, $\beta = .190$, 95% CI: .49 - .12, $p = .002$, suggests that the mean RT in the arithmetic skill decreased approximately by .127 units with each addition of non-numeric response certainty on standard trials. Model II of hierarchical linear regression analysis revealed that the relationship between non-numeric response certainty was non-significant, $F_{2,271} = .034$, $p = .85$.

It could be summarized from the above findings that adults with better response certainty (low response bias) in the non-numeric error awareness task showed higher arithmetic efficiency, i.e., higher accuracy and quicker solving of arithmetic problems.

Chapter V

Discussion

This study investigated the domain-general and domain-specific role of error awareness in arithmetic. Based on the previous literature, it was predicted that better error awareness would predict better arithmetic achievement (Rinne & Mazzocco, 2014). This relationship would be more domain-general in adults rather than domain-specific (de Gardelle et al., 2016; Bellon et al., 2020). Both numeric and non-numeric error awareness would predict better arithmetic achievement.

The analysis was done in two phases. The first phase calculated the relationship between error awareness on both tasks and arithmetic achievement. In the second phase, the relationship between error awareness and arithmetic achievement was calculated, but this time, trials exclusively associated with numeric error and Stroop-based error were considered. In the first phase, no significant relationship was observed when tested for a solo relationship between non-numeric error awareness and arithmetic achievement. However, the addition of numeric error awareness made a significant contribution for both non-numeric and numeric error awareness in explaining arithmetic skills in participants. Both variables made opposite predictions on arithmetic skills; non-numeric error awareness predicted more efficient arithmetic skills (quicker time for correct response), and higher numeric error awareness predicted less efficiency in solving arithmetic problems (slower in responding accurately).

In the second analysis, we only focused on the numeric-related errors in the numeric error awareness task and inhibition-related errors in the non-numeric error awareness task. A similar pattern and direction of the relationship between error awareness and arithmetic skills were observed in the first analysis. In addition, a significant relationship between error awareness and accuracy in numeracy tests was found. Non-numeric error awareness alone could predict Numeracy. However, adding numeric error awareness in the model strengthened this relationship. However, the pattern and direction of the relationship are still like the arithmetic skills. Higher non-numeric error awareness predicted higher accuracy in the numeracy test, but higher numeric error awareness predicted lower accuracy in the numeracy test. Because of this opposite direction of the relationship, it was hard to conclude about the domain-specificity nature of error awareness. These findings contradict previous findings where it was found that better metacognition is related to better arithmetic achievement, and this relationship is domain-general in adults (Rinne & Mazzocco, 2014; Bellon et al., 2020).

One possible explanation for finding a negative relationship between numeric error awareness and arithmetic achievement could be that people who are more aware of their numeric errors choose to be careful in solving arithmetic problems, which is why they took longer time to solve arithmetic problems in the arithmetic skills task and showed less accuracy in the speeded Numeracy test. The proof of this observation could be found in the negative correlation score between the accuracy of the arithmetic skills task and error awareness in the numeric error awareness task (*Table 5.7*). On the contrary, errors committed in the non-numeric Stroop task are more related to Stroop-related errors (Lack of inhibition, attention, and flexibility; Scarpina & Tagini, 2017), and the test materials are unrelated to arithmetic. As a result, adults did better in detecting non-numeric errors (in this study, Stroop-related) and did better in arithmetic. Past studies also found that task types could affect the relationship between metacognition and academic achievement (Bellon et al., 2019).

It was also assumed that people who are poor in maths often overestimate their arithmetic accuracy; as a result, they underreport their error response. That's why participants who showed higher error awareness due to response biases were the poor arithmetic performers. Even though confidence judgment was not the focus of our analysis at the beginning of the study, it was found from the additional analysis that response bias indeed influences performance awareness. It was found that better non-numeric response certainty is related to better arithmetic achievement, but this relationship is non-significant for numeric response certainty. It means that the effect of numeric response certainty diminished once the confidence measurement was added to the measures. Considering the above discussion, it could be said that both non-numeric error awareness and response certainty predict better arithmetic achievement irrespective of response bias. However, the response bias affects numeric response certainty, so no significant relationship was observed between numeric response certainty and arithmetic achievement. It could be assumed that due to response biases, numeric and non-numeric error awareness may have distinct cognitive and neural processes for predicting arithmetic achievement. However, the present study did not investigate the underlying cognitive and neural processes of error awareness and response biases. The previous brain and behavioural studies (e.g., Scheffers & Coles, 2000; Kirschner et al., 2021) also indicated that aware error, unaware error, and correct response share common but unique underlying cognitive processes. However, a proper brain and behavioural study must be conducted before making such a claim.

In summary, it could be said that non-numeric error awareness predicts better arithmetic achievement, but numeric-related error awareness predicts cautious response patterns in solving arithmetic problems. Since the bidirectional relationship between error awareness and arithmetic achievement for the numeric and non-numeric error awareness task was observed, a conclusion about the domain-specificity couldn't be made. Only the relationship between numeric response certainty and arithmetic achievement is affected by the problem of over-confidence in evaluating performance. In the future, if someone wants to explore the role of numeric error awareness on arithmetic achievement, they must consider taking confidence as an additional measure.

Limitations and future suggestions

One limitation that needs to be overcome in the future is to recruit the exact number of participants, as we mentioned in the pre-registration—due to time and money shortages, recruiting some extra participants after the outliers' exclusion was interrupted. In the present study, the results differ from what we hypothesized based on previous research. In the future, a more thorough investigation into this area is needed, especially considering the error awareness and confidence judgment together. Developing error awareness tasks based on different subject areas (e.g., reading and arithmetic) could help us understand the domain-general vs. the domain-specific role of error awareness and response certainty on arithmetic achievement.

6. Chapter VI: General Discussion

The primary purpose of the thesis was to investigate the domain-general vs. the domain-specific role of metacognition on arithmetic. Here, metacognition was measured in terms of error monitoring and awareness. Three studies were conducted for this purpose. Chapters II and III investigated the domain-general vs. the domain-specific role of error monitoring on arithmetic achievement. Chapter II was the pilot study; based on the pilot study's findings, the experiments were modified for the main study (Chapter III). In addition, some follow-up experiment was done to answer some questions from child data in Chapter IV. Then, another experiment was designed to investigate the domain-general vs. the domain-specific role of error awareness on arithmetic achievement, discussed in Chapter V.

It can be summarized from Study I (Chapter II) that adults showed higher PES under shorter RSI, irrespective of task types. In children, PES was observed for all the RSI in both tasks, apart from short RSI in the Simon task, where post-error speeding was observed. However, PES varied significantly from zero under the long RSI in the Simon task only. Children showed no effect of RSI and task types on PES. Adults and children showed no improvement in accuracy after error trials. In adults, the decline in performance was comparatively higher under the numeric error monitoring task than the non-numeric error monitoring task. No such task effect on PECA was observed in children. The relationship between error monitoring and math fluency indicated that children who were better in numeric error monitoring (higher PES) also showed significantly higher math fluency. On the other hand, adults who had a trend in showing better non-numeric error monitoring showed less maths fluency, which could be assumed to be a cautious response style to acquire better performance within a limited time. However, this relationship was non-significant in adults.

A follow-up study (Chapter III) was conducted later based on the findings from the Pilot study (Chapter II). Since children were not showing the effect of RSI and Task types on PES like adults, it was assumed that children needed a longer RSI to show its effect. That's why a follow-up study with a larger sample and a broader range of RSI was decided to be conducted. Like the pilot study, it was found from this study that PES was significantly larger under shorter RSIs in adults, irrespective of task types. Similar to the pilot study, adults showed a significant interaction effect between RSI and task types on PECA in the present study. Separate ANOVA analysis indicated that after error trials, the accuracy rate declined in adults under the shortest RSI in the non-numeric error monitoring task (Simon task). No RSI effect on PECA was observed for the numeric error monitoring task (arithmetic interference). However, in adults, the decline in performance was comparatively higher under the numeric error monitoring task than the non-numeric error monitoring task, which is also consistent with the pilot study. Due to data collection and task design difficulties, data was only collected for the arithmetic interference and skills tasks. It was found that children must fulfil the measurement criteria for the arithmetic interference task. The RSI effect was not investigated in children. No significant effect of RSI on PECA was observed for children, even though mean values showed a decline in accuracy after error trials. Like the pilot study, adults showed a non-significant relationship between error monitoring (PES) and arithmetic skills. However, the correlation size indicated that there is a trend of negative relationship between non-numeric error monitoring (PES) and arithmetic skills.

Adults with better error monitoring had a trend to choose a cautious response style by taking more time to solve arithmetic problems. Unlike the pilot study, children showed a non-significant relationship between error monitoring and arithmetic achievement. However, the p-value indicated a trend in children with better numeric error monitoring to solve arithmetic problems more efficiently.

It could be concluded by comparing the findings from these two studies that in adults, PES is affected by RSI, which is greater under short RSI(s). These findings are consistent with the previous research findings (e.g., Jentzsch & Dudschig, 2009). However, unlike previous research, no improvement in post-error performance was observed under different RSI, and the accuracy rate declined mostly under short RSI. The reason could be that a longer duration of RSI is needed to improve accuracy after error trials. Past research also indicated that the error monitoring process is time-sensitive and depends on the task's complexity level (McDougle, 2022; Ger & Roebers, 2023). This could also be explained by the unifying model of Wessel (2017), where he proposed that error monitoring is a complex cognitive process, and participants need time to process the expectancy violation between the error response and the correct answer once the error is made. This process leads to an automatic inhibition response action after the error response, eventually leading them to reorient their attention toward the task again. Previous research also indicated that as the time interval between the response and subsequent trial increases, the participant’s error processing shifts from an orienting response to a more adaptive cognitive control response pattern (Smulder et al., 2016; Dubravac et al., 2022).

Children in both studies did not perform in the arithmetic interference task as expected. In the pilot study, no significant effect of RSI on PES was observed in children, which is inconsistent with the findings of Smulder et al. (2016). On the other hand, in the follow-up experiment, they did not fulfil the measurement criteria of PES. In Chapter IV, when additional analyses were done to discover the problem, it was found that children made more errors in a row, turning many post-error and post-correct trials into error trials. Some conclusions are made from the data of child participants. Since children's error monitoring functions are still developing (e.g., Davies et al. 2004; Overbye et al. 2019), running the task with larger trial numbers or into two separate sessions to see how children behave after an error is better.

It was found from both error monitoring studies that in adults, there was a trend of better non-numeric error monitoring (Higher PES) to be related to a cautious response style by showing less arithmetic fluency and a slower response time. However, in children, there was a trend of better numeric error monitoring (higher PES) related to efficient arithmetic problem-solving skills by showing better maths fluency and faster arithmetic problem-solving time. Possible explanations for these different trends could be the fact that children were still in school; they practiced maths more frequently than adults. The practice effect had a beneficial effect on arithmetic achievement in children compared to adults. Conversely, less maths practice leads to slower processing speed in solving arithmetic problems, even though adults have better error monitoring processes (i.e., higher PES). That's why adults showed a cautious response style in solving arithmetic problems.

In the end, it could be said that the current experiments on error monitoring were the first to investigate both numeric and non-numeric post-error behavioural adjustment in children and adults. However, error monitoring is complex, depends on multiple factors, and is highly age-sensitive. In the future, extensive research with larger trial numbers should be conducted. In particular, different age groups in children should consider investigating the relationship between error monitoring and arithmetic achievement.

Next, another study (Chapter V) investigated the domain-general vs. domain-specific role of error awareness on arithmetic achievement in adults. It was found from this study that people with better non-numeric error awareness did better in maths. Still, participants with better numeric error awareness choose a cautious response style in solving arithmetic problems. The current research findings are inconsistent with the previous findings (Rinne & Mazzocco, 2014; Bellon et al., 2019; Bellon et al., 2020), where a domain-general relationship between metacognition and arithmetic achievement was found, and adults showed better arithmetic performance irrespective of types of error awareness tasks (numeric vs. non-numeric). However, when overconfidence judgment bias was considered, the relationship between non-numeric response certainty and arithmetic achievement was non-significant. In contrast, the relationship between non-numeric response certainty and arithmetic achievement was significant.

The possible explanation for finding a bidirectional relationship between numeric and non-numeric error awareness with arithmetic achievement could be that numeric and non-numeric errors function differently at cognitive level, where numeric awareness is affected by overconfidence response bias, but non-numeric awareness doesn't. The relationship between numeric error awareness and arithmetic achievement is more affected by response biases. In the current study, no brain imaging studies were conducted, and our focus was not to differentiate between error awareness and response bias. Still, the present study opens a new window to investigate how our brain processes different errors (e.g., numeric vs. non-numeric). A possible assumption could be made from past behavioural and brain research where awareness, unaware error, and correct response share common but unique underlying cognitive processes (e.g., Kirschner et al., 2021). Behavioural explanation for two opposite directions of the relationship between arithmetic achievement and error awareness measured in numeric and non-numeric error awareness tasks could be that participants who overestimate their performance are poor in detecting errors; as a result, they also performed poorly in the arithmetic test since they are unable to notice their errors due to response biases. On the contrary, the non-numeric task was an ordinary executive function task, and the stimulus was not composed of educational materials like a numeric error awareness task; as a result, error awareness in this task was not affected by response bias and overall participants who are good in detecting non-numeric errors also did well in the arithmetic achievement test.

Conclusion

It could be summarized from this thesis that error processing is affected by RSI in adults and is larger under short RSI. Furthermore, PECA is affected by task types and RSI in adults, with a significantly larger decline in accuracy rates after error trials under short RSI. It could be said from this that as the time duration between RSI increases, the error processing slowly shifts from orienting to cognitive control processing. Adults with error processing and awareness tended to have less arithmetic efficiency, i.e., they chose a cautious response style by solving arithmetic problems slowly with less fluency. On the other hand, children who were better at error processing tended to have higher arithmetic efficiency. The opposite pattern of relationships in adults could result from slower processing time due to a lack of practicing maths in adulthood. Task types and response bias affect the relationship between error awareness and arithmetic achievement. Adults better in non-numeric error awareness showed better arithmetic efficiency, but due to overestimated arithmetic performance, adults with better numeric error awareness showed less arithmetic efficiency. In the future, combined research on error monitoring and awareness and response certainty could adequately explain how this process works in adults and children.

Future Implications

This thesis is the first to investigate the numeric and non-numeric aspects of error processing and its relationship with arithmetic in different age groups. It tried to answer some questions for future research. For example- does error monitoring relate to arithmetic differently in adults and children? Is this relationship domain-specific? Is it enough to predict arithmetic efficiency in children and adults, or is a combined approach with performance monitoring, working memory load, and processing speed needed? Does the underlying cognitive mechanism for error processing and response biases affect arithmetic efficiency differently? My research would help future researchers to look at the error-monitoring process from different angles. It would also help teachers and educational psychologists to understand how different types of awareness about performance could affect learning outcomes. Especially in the classroom, teachers often find it challenging to teach children deficient in maths. If they help children practice better error monitoring and awareness by overcoming their response biases will help them self-regulate their performance to do maths more efficiently. In the case of adults, they need to do small daily life calculations manually to get the benefits from the practice effect.

7. References

Agarwal, S., & Bhashkar, M. (2013). Cognitive Abilities and Household Financial Decision Making. *American Economic Journal: Applied Economics, 5*(1), 193-207.

Armitage, R. L., & Cragg, L. (2018). Does the competition between stored addition and multiplication facts cause interference? Master thesis dissertation submitted to University of Nottingham, UK. (Unpublished).

Aubrey, C., Godfrey, R. & Dahl, S. (2006). Early Mathematics Development and Later Achievement: Further Evidence. *Mathematics Education Research Journal*, *18,* 27–46.

Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. *Journal of Reading Behavior, 11*(4), 365–374.

Baker, L. (1984). Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology, 38*(2), 289–311.

Baker, L. (1989). Metacognition, Comprehension Monitoring, and the Adult Reader. *Educational Psychology Review*, *1*, 3–38.

Baker, Linda & Cerro, Lorraine C. (2000). *"3. Assessing Metacognition in Children and Adults" (2000). Issues in the Measurement of Metacognition. 4.*

Bellon, E., Fias, W., & de Smedt, B. (2019). More than number sense: The additional role of executive functions and metacognition in arithmetic. *Journal of Experimental Child Psychology*, *182*, 38–60.

Appendix I

Bellon, E., Fias, W., & de Smedt, B. (2020). Metacognition across domains: Is the association between arithmetic and metacognitive monitoring domain-specific? *PLoS ONE*, *15*(3).

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, *54*(2), 199–231.

Botvinick, M. M., Carter, C. S., Braver, T. S., Barch, D. M., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*(5364), 747–749.

Caudek, C., Ceccarini, F., & Sica, C. (2015). Posterror speeding after threat-detection failure. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 324–341.

Chinn, S. (2020). *More Trouble with Maths: A Complete Manual to Identifying and Diagnosing Mathematical Difficulties (2nd ed.).* Routledge.

Chua, E. F., Schacter, D. L., Rand-giovannetti, E., & Sperling, R. A. (2006). *Understanding metamemory : Neural correlates of the cognitive process and subjective level of confidence in recognition memory*. *29*, 1150–1160.

Cohen, M. X., Van Gaal, S., Ridderinkhof, K. R., & Lamme, V. A. (2009). Unconscious errors enhance prefrontal-occipital oscillatory synchrony. *Front. Hum. Neurosci., 3,* 54.

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, *162*, 12–26.

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.

Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology*, *2*(SEP), 1–10.

Davies, P. L., Segalowitz, S. J., & Gavin, W. J. (2004). Development of error-monitoring event-related potentials in adolescents. *Ann N Y Acad Sci., 1021*, 324-8.

De Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a common currency between vision and audition. *PLoS ONE*, *11*(1).

de Mooij, S. M. M., Dumontheil, I., Kirkham, N. Z., Raijmakers, M. E. J., & van der Maas, H. L. J. (2022). Post-error slowing: Large scale study in an online learning environment for practising mathematics and language. *Developmental Science*, *25*(2), 1–15.

Denervaud, S., Fornari, E., Yang, X. F., Hagmann, P., Immordino-Yang, M. H., & Sander, D. (2020). An fMRI study of error monitoring in Montessori and traditionally-schooled children. *Npj Science of Learning*, *5*(1).

Desmet, C., Imbo, I., De Brauwer, J., Brass, M., Fias, W., & Notebaert, W. (2012). Error adaptation in mental arithmetic. *Quarterly journal of experimental psychology (2006)*, *65*(6), 1059–1067.

Dubravac, M., Roebers, C. M., & Meier, B. (2022). Age-related qualitative differences in post-error cognitive control adjustments. *The British journal of developmental psychology*, *40*(2), 287–305. https://doi.org/10.1111/bjdp.12403

Duyan, Y. A., & Balcı, F. (2018). Numerical error monitoring. *Psychonomic Bulletin and Review*, *25*(4), 1549–1555.

Appendix I

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E. J. (2012). Testing theories of post-error slowing. *Attention, Perception, and Psychophysics, 74*(2), 454–465.

Eichele, H., Juvodden, H. T., Ullsperger, M., & Eichele, T. (2010). Maladaptation of event-related EEG responses preceding performance errors. *Frontiers in Human Neuroscience*, *4*.

Endrass, T., Franke, C., & Kathmann, N. (2005). Error awareness in a saccade countermanding task. *Journal of Psychophysiology*, *19*(4), 275–280.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behaviour Research Methods, 41*(4), 1149-1160.

Fiehler, K., Ullsperger, M., & von Cramon, D. Y. (2005). Electrophysiological correlates of error correction. *Psychophysiology, 42*(1), 72-82.

Fievez, F., Derosiere, G., Verbruggen, F., & Duque, J. (2022). Post-error Slowing Reflects the Joint Impact of Adaptive and Maladaptive Processes During Decision Making. *Frontiers in Human Neuroscience*, *16*.

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

Fleur, D. S., Bredeweg, B., & van den Bos, W. (2021). Metacognition: ideas and insights from neuro- and educational sciences. *NPJ science of learning, 6*(1), 13.

Freeman, E. E., Karayanidis, F., & Chalmers, K. A. (2017). Metacognitive monitoring of working memory performance and its relationship to academic achievement in Grade 4 children. *Learning and Individual Differences*, *57*, 58–64.

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Research, 60,* 517–529.

Gascoine, L., Higgins, S., & Wall, K. (2017). *The assessment of metacognition in children aged 4 – 16 years: a systematic review*. *5*(1), 3–57.

Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. Journal of Neuroscience, *21*(23), 9430-9437.

Ger, E., & Roebers, C. (2023). Hearts, flowers, and fruits: All children need to reveal their post-error slowing. *Journal of Experimental Child Psychology, 226,* 105552.

Geurten, M., Meulemans, T., & Lemaire, P. (2018). From domain-specific to domain-general? The developmental path of metacognition for strategy selection. *Cognitive Development*, *48*, 62–81.

Gilmore, C., Göbel, S.M., & Inglis, M. (2018). *An Introduction to Mathematical Cognition (1st ed.)*. Routledge.

Guan, Y., & Wessel, J. R. (2022). Two types of motor inhibition after action errors in humans. *Journal of Neuroscience, 42*(38), 7267–7275.

Gupta, R., Kar, B.R. & Srinivasan, N. (2009). Development of task switching and post-error-slowing in children. *Behavioral and Brain Functions, 5,* 38.

Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*, *40*(6), 895–903.

Hennessey, M. G. (1999, April). *Probing the dimensions of metacognition: Implications for conceptual change teaching-learning.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.

Hester, R., Foxe, J. J., Molholm, S., Shpaner, M., & Garavan, H. (2005). Neural mechanisms involved in error processing: a comparison of errors made with and without awareness. *NeuroImage*, *27*(3), 602–608.

Hirsh, J. B., & Inzlicht, M. (2010). Error-related negativity predicts academic performance. *Psychophysiology*, *47*(1), 192–196.

Hogan, A. M., Vargha-Khadem, F., Kirkham, F.J., Baldeweg, T. (2005). Maturation of action monitoring from adolescence to adulthood: an ERP study. *Developmental Science, 8,* 525–534.

Houtman, F., & Notebaert, W. (2013). Blinded by an error. *Cognition, 128*(2), 228-36.

Jentzsch, I., & Dudschig, C. (2009). Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *Quarterly Journal of Experimental Psychology*, *62*(2), 209–218.

Jia, X., Li, W., & Cao, L. (2019). The role of metacognitive components in creative thinking. *Frontiers in Psychology, 10,* Article 2404.

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *Neuroimage, 34*(4), 1774-81.

Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Assessing error awareness without relying on introspective judgment? In *Frontiers in Neuroscience* (Issue 7 JUL). Frontiers Research Foundation.

Koriat, A. (2019). Confidence judgments: The monitoring of object-level and same-level performance. *Metacognition and Learning*, *14*(3), 463–478.

Kirschner, H., Humann, J., Derrfuss, J., Danielmeier, C., & Ullsperger, M. (2021). Neural and behavioral traces of error awareness. *Cognitive, affective & behavioral neuroscience*, *21*(3), 573–591.

Appendix I

Kuhn, D. & Dean, D., Jr. (2004) Metacognition: A Bridge Between Cognitive Psychology and Educational Practice, *Theory Into Practice, 43*(4), 268-273.

Lai, E. (2011). *Metacognition: A Literature Review Research Report Metacognition: A Literature Review. Pearson Assessments Research Reports.*

Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica, 43*(3), 199–224.

Lavro, D., Ben-Shachar, M. S., Saville, C. W. N., Klein, C., & Berger, A. (2018). Testing the bottleneck account for post-error slowing beyond the post-error response. *Biological Psychology*, *138*, 81–90.

Lemaire, P., Barrett, S. E., Fayol, M., & Abdi, H. (1994). Automatic activation of addition and multiplication facts in elementary school children. *Journal of Experimental Child Psychology, 57,* 224-258.

LeFevre, J. A. Sadesky, G. S. & Bisartz, J. (1996). Selection of procedures in mental addition: reassessing the problem-size effect in adults. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*(1), 216-230.

Livingston, J. A. (2003). Metacognition: An Overview. *Psychology, 13,* 259-266.

McCurdy, L.Y., Maniscalco, B., Metcalfe, J., Liu, K.Y., de Lange, F.P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *Journal of Neuroscience, 33,* 1897-1906.

Marco-Pallares, J., Camara, E., Munte, T. F., & Rodriguez-Fornells, A. (2008). Neural mechanisms underlying adaptive actions after slips. *Journal of Cognitive Neuroscience, 20,* 1595–1610.

Martinez, M. E. (2006). What is metacognition? *Phi Delta Kappan*, *87*(9), 696-699.

Masina, F., Di Rosa, E., & Mapelli, D. (2018). Intra-individual variability of error awareness and post-error slowing in three different age-groups. *Frontiers in Psychology*, *9*(JUN), 1–10.

McDougle, S. D. (2022). Post-error Slowing During Instrumental Learning is Shaped by Working Memory-based Choice Strategies. *Neuroscience*, *486*, 37–45.

Moritz, S., & Lysaker, P. H. (2018). Metacognition – What did James H. Flavell really say and the implications for the conceptualization and design of metacognitive interventions. *Schizophrenia Research*, *201*, 20–26.

Muncer, G., Higham, P. A., Gosling, C. J., Cortese, S., Wood-Downie, H., & Hadwin, J. A. (2022). A Meta-Analysis Investigating the Association Between Metacognition and Math Performance in Adolescence. *Educational Psychology Review*, *34*(1), 301–334.

Nelson, T. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation, 26,* 125–173.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology, 38,* 752–760.

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: an orienting account. Cognition, *111*(2), 275-279.

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., Robertson, I. H., & Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: A high-density electrical mapping study. *European Journal of Neuroscience*, *25*(8), 2571–2579.

O'Leary, A. P., & Sloutsky, V. M. (2019). Components of metacognition can function independently across development. *Developmental Psychology*, *55*(2), 315–328.

Overbye K, Walhovd, K. B., Paus, T., Fjell, A. M., Huster, R. J., Tamnes, C. K. (2019). Error processing in the adolescent brain: Age-related differences in electrophysiology, behavioral adaptation, and brain morphology. *Developmental Cognitive Neuroscience, 38,* 100665.

Özsoy, G. (2011). An investigation of the relationship between metacognition and mathematics achievement. *Asia Pacific Education Review*, *12*(2), 227–235.

OZTURK, N. (2017). Assessing Metacognition: Theory and Practices. *International Journal of Assessment Tools in Education*, 134–134.

Olvet, D. M., & Hajcak, G. (2009). The stability of error-related brain activity with increasing trials. *Psychophysiology*, *46*(5), 957–961.

Papleontiou-Louca, E. (2003). The concept and instruction of metacognition. *Teacher Development*, *7*(1), 9–30.

Paris, S. G., & Winograd, P. (1990). Promoting metacognition and motivation of exceptional children. *RASE: Remedial & Special Education, 11*(6), 7–15.

Peirce, J. W., Hirst, R. J. & MacAskill, M. R. (2022). *Building Experiments in PsychoPy.* 2nd Edn, London: Sage.

Perry, N. E. (2002). Introduction: Using qualitative methods to enrich understandings of self-regulated learning. *Educational Psychologist*, *37*(1), 1–3.

Peirce, J. W., Hirst, R. J. & MacAskill, M. R. (2022). Building Experiments in PsychoPy. 2nd Edn London: Sage.

Pontifex, M. B., Scudder, M. R., Brown, M. L., O'Leary, K. C., Wu, C. T., Themanson, J. R., & Hillman, C. H. (2010). On the number of trials necessary for stabilization of error-related brain activity across the lifespan. *Psychophysiology*, *47*(4), 767–773.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849–860.

Regev, S., & Meiran, N. (2014). Post-error slowing is influenced by cognitive control demand. *Acta Psychologica,152,*10-8.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How Numeracy Influences Risk Comprehension and Medical Decision Making. *Psychological Bulletin*, *135*(6), 943–973.

Rinne, L. F., & Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS ONE*, *9*(7).

Ritchie, S. J., & Bates, T. C. (2013). Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status. *Psychological Science*, *24*(7), 1301–1308.

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45,* 31–51.

Scarpina, F., & Tagini, S. (2017). The Stroop Color and Word Test. *Frontiers in psychology*, *8*, 557.

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 141–151.

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. In *Mind, Brain, and Education* (Vol. 2, Issue 3, pp. 114–121). Blackwell Publishing.

Schneider, W., & Artelt, C. (2010). Metacognition and mathematics education. *ZDM - International Journal on Mathematics Education*, *42*(2), 149–161.

Schraw, G., Crippen, K.J. & Hartley, K. (2006). Promoting Self-Regulation in Science Education: Metacognition as Part of a Broader Perspective on Learning. *Res Sci Educ, 36*, 111–139.

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

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351–371.

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.

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.

Simon, J. R. (1969). The Effects of an Irrelevant Directional CUE on Human Information Processing. *Advances in Psychology, 65,* 31–86.

Simon, J. R., and Craft, J. L. (1970). Effect of an irrelevant auditory stimulus in visual choice reaction time. *Journal of Experimental Psychology, 86,* 272-274.

Smith, G. A., & Brewer, N. (1995). Slowness and age: Speed-accuracy mechanisms. *Psychology and Aging, 10*(2), 238–247.

Smulders, S. F. A., Soetens, E., & van der Molen, M. W. (2016). What happens when children encounter an error? *Brain and Cognition*, *104*, 34–47.

Tabachnick, B. G., & Fidell, L. S. (2018). *Using multivariate statistics* (6th ed.). Pearson.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1,* 3-14.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203.

Ward, L., & Traweek, D. (1993). *Application of a Metacognitive Strategy to Assessment, Intervention, and Consultation: A Think-Aloud Technique*. *31*, 469–485.

Wessel, J. R. (2012). Error awareness and the error-related negativity: Evaluating the first decade of evidence. *Frontiers in Human Neuroscience*, *6,* 88.

Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology, 55* (3).

Westfall, J. (2016). PANGEA: Power Analysis for General ANOVA designs. Unpublished Manuscript. DOI: https://jakewestfall.org/publications/pangea.pdf

Whitebread, D., Coltman, P., & Demetriou, D. (2009). *The development of two observational tools for assessing metacognition and self-regulated learning in young children*. 63–85.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Examiner's Manual. *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959.

Appendix I

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321.

Young, A., & Fry, J. D. (2008). *Metacognitive awareness and academic achievement in college students*. *8*(2), 1–10.

Zimmerman, M.E. (2011). Speed–Accuracy Tradeoff. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds) *Encyclopedia of Clinical Neuropsychology.* Springer, New York, NY.

8. Appendix I: Additional Analysis for Chapter II

Effect of Congruency on RT and Accuracy

Data were analysed in paired sample t-tests to see the effect of congruency on RT and accuracy in each age group separately. Here, conditions (congruent vs. incongruent) were an independent factor, and the dependent variables were RT (for correct trials) and accuracy. The mean value indicated that the accuracy rate was higher in congruent than incongruent trials. It can be observed from the adult participant's data that there was a significant effect of congruency on accuracy for both tasks, but no significant effect was observed for RT (Table 8.1).

Table 8.1 Association Between Conditions of the Tasks with RT and Accuracy for Adult Participants

Arithmetic Interference Task	Conditions	*Mean RT(SD)*	*T*	*df*	*P*
	Incongruent Congruent	1004 (290) 998 (296)	-0.35	6592	0.73
		Mean Accuracy (SD)	*T*	*df*	*P*
	Incongruent Congruent	0.86 (0.34) 0.89 (0.31)	3.80	6520	<.001
Simon Task	Conditions	*Mean RT(SD)*	*T*	*df*	*P*
	Incongruent Congruent	530 (113) 524 (112)	1.58	7842	0.11
		Mean Accuracy (SD)	*T*	*df*	*P*
	Incongruent Congruent	0.80 (0.40) 0.88 (0.33)	-9.09	7559	<.001

Similarly, two paired sample t-tests of the child data revealed a significant congruency effect on accuracy for both tasks (Table 8.2). The mean accuracy rate indicated children performed better in congruent trials than incongruent trials, like adult participants. However, no significant congruency effect was observed for RT in both tasks.

Appendix I

Table 8.2 Association Between Conditions of the Tasks with RT and Accuracy for Child Participants

Task	Conditions	*Mean RT (SD)*	*t*	*Df*	*P*
Arithmetic Interference Task	Incongruent	1710 (754)	1.43	3841	0.15
	Congruent	1745 (761)			
		Mean Accuracy (SD)	*t*	*Df*	*P*
	Incongruent	0.73 (0.45)	4.58	3824	<.001
	Congruent	0.79 (0.41)			
Simon Task	Conditions	*Mean RT (SD)*	*t*	*Df*	*P*
	Incongruent	615 (184)	-0.72	8250	0.47
	Congruent	617 (178)			
		Mean Accuracy (SD)	*t*	*Df*	*P*
	Incongruent	0.63 (0.48)	-17.21	7993	<.001
	Congruent	0.8 (0.40)			

Effect of Switching on PES for Arithmetic Interference Task in Adults and Children

A set of addition and multiplication problems were used in this task. One-way repeated measurement ANOVA was performed to see if the switching between addition and multiplication problems affected PES for the arithmetic interference Task. As previously set, the criteria of a minimum number of 5 trials were required per switching condition to conduct this analysis. Based on that, 2 out of 20 adult participants and 4 out of 24 child participants were excluded from the ANOVA analysis. Mean PES in switch and no-switch trials for child and adult participants are presented in Figure 8.1. It can be observed from Figure 8.1 that PES was higher in no-switch trials than in the switching trials for both age groups. Data from Repeated measurement ANOVA indicated only a significant mean difference in PES for switch vs. no-switch trials $F_{(1, 15)}$= 7.93, p=0.012, η_p^2= 0.32 in adult participants but not in child participants $F_{(1, 15)}$= 2.97, p=0.101, η_p^2= 0.14.

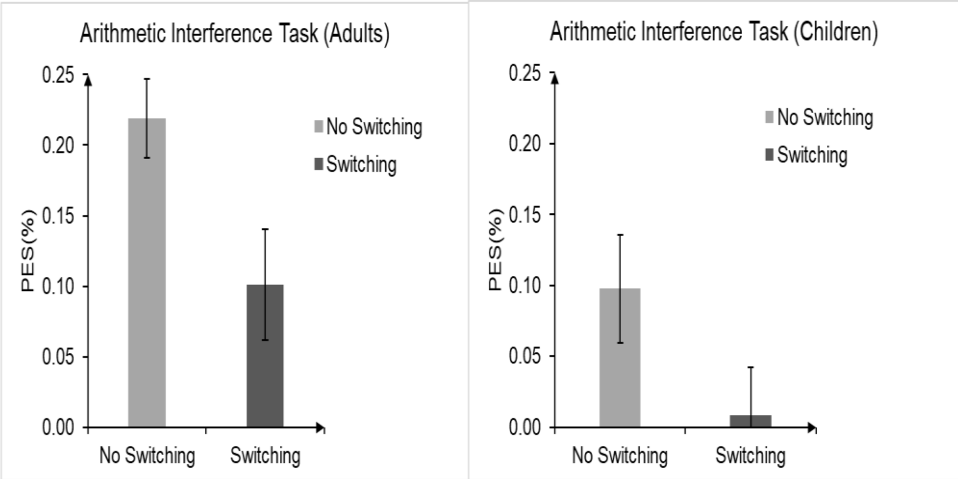


Figure 8.1 Mean PES in conflict vs. no-conflict trials for switching and no-switching trials; Error bar represents a standard error (SE)

Effect of switching on PECA

Data were analysed separately for adults and children in 2(Switching)×2(Trial Type: post-correct accuracy vs. post-error accuracy) repeated measurement ANOVA to see the effect of switching on PECA. Data from the descriptive analysis are given in Figure 8.2.

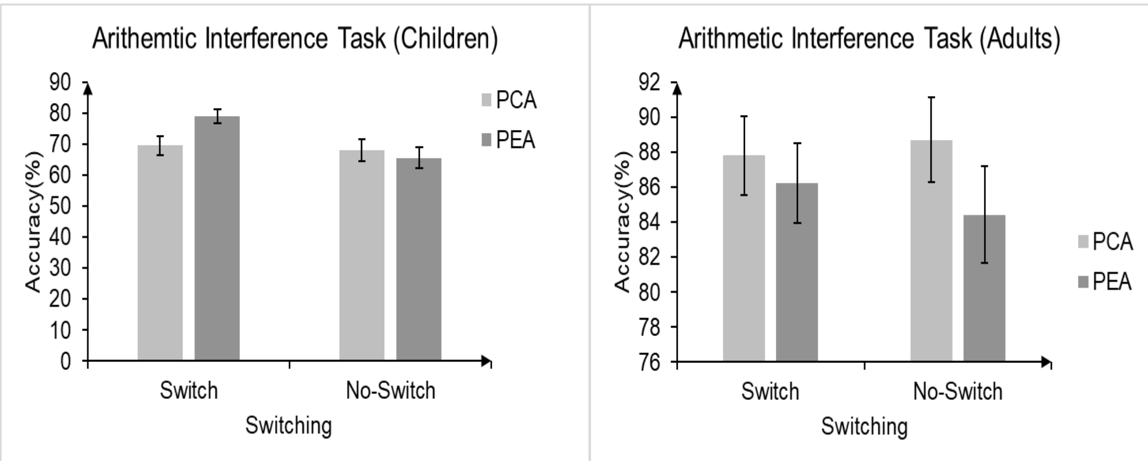


Figure 8.2 The bar diagram represents Mean PCA in shift and no-shift trials, and the error bar represents a standard error (SE).

Appendix I

First, data from adult participants were analysed in repeated measurement ANOVA, and no significant effect of Switching was found $F_{(1, 19)}$= 0.13, p= 0.74. Still, a significant difference was observed for the Trial type, $F_{(1, 19)}$= 5.33, p=0.032, η^2= 0.22. No significant interaction between switching and trial type was found $F_{(1, 19)}$= 0.904, p=0.35, η_p^2= 0.045. Mean values indicated that participants' accuracy rate was lower in post-error trials (Mean= 85, SE 2.24) than post-correct trials (Mean= 88, SE= 2.20). However, data from children indicated a significant effect of Trial type and the interaction effect between switching and trial type. However, the no-significant effect of switching was observed (Table 8.3).

Table 8.3 Analysis of Variance between Switching and Trial Type

Factors		F	df	P	η_p^2
Switching	Sphericity Assumed	3.31	1.00	0.08	0.14
Trial Type	Sphericity Assumed	12.42	1.00	0.00	0.37
Switching * Trial Type	Sphericity Assumed	6.21	1.00	0.02	0.23

Since two-way interaction was observed between RSI and trial type, data were further analysed using a Bonferroni-corrected test of simple main effects. It was observed that the mean accuracy rate of post-correct trials and post-error trials differed significantly for no switching condition $F_{(1, 21)}$= 32.28, p< 0.001, η_p^2= 0.61 but not for switching condition $F_{(1, 21)}$= 0.15, p= 0.70, η_p^2= 0.01. In the no-shift condition, the accuracy rate was higher in post-correct trials (M= 79, SE= 2.18) than in post-error trials (M= 66, SE= 3.43).

Appendix I

Descriptive Analysis of Trial Counts

Following the examiners' suggestions, descriptive analyses were done on the trial counts under each condition for adults and children separately (Table 8.4).

Table 8.4 Trial statistics in post-error and post-correct trials under each RSI for each task separately in adults.

Tasks	5-trials Criteria Fulfilled	Statistics	Post-Error Trials	Post-Correct Trials	Post-Error Trials	Post-Correct Trials
			Long RSI		Short RSI	
Arithmetic Interference Task	No	Mean(n=2)	4.00	4.00	6.50	6.00
		SD	0.00	0.00	4.95	2.83
		Minimum	4	4	3	4
		Maximum	4	4	10	8
	Yes	Mean(n=18)	17.17	13.06	16.33	13.67
		SD	9.076	4.345	9.146	5.434
		Minimum	7	7	7	5
		Maximum	47	20	43	24
Simon Task	No	Mean(n=2)	10.00	5.50	4.00	7.50
		SD	0.00	0.71	0.00	0.71
		Minimum	10	5	4	7
		Maximum	10	6	4	8
	Yes	Mean(n=18)	24.50	19.56	19.72	21.56
		SD	11.78	7.16	9.84	7.78
		Minimum	8	6	5	9
		Maximum	46	33	40	35

It can be observed from Table 8.4 that those adults who dropped out from the analysis committed few errors; as a result, they did not have many post-error trials. Similarly, I did some descriptive analysis with child participants. It is to be noted that no child participants were dropped out from the Simon task since they fulfilled the five trial criteria. The data from child participants are presented in Table 8.5. It could be observed from Table 8.5 that children who dropped out from the arithmetic interference task committed fewer errors. As a result, they had fewer post-error and post-correct trials.

Appendix I

Table 8.5 Trial statistics in post-error and post-correct trials under each RSI for each task separately in adults.

	5-trials Criteria Fulfilled	Statistics	Post-Error Trials (Long RSI)	Post-Correct Trials (Long RSI)	Post-Error Trials (Short RSI)	Post-Correct Trials (Short RSI)
Arithmetic Interference Task	No	Mean (n=7)	7.29	5.71	7.71	5.86
		SD	5.02	3.95	4.50	3.72
		Minimum	2	2	2	2
		Maximum	15	14	15	12
	Yes	Mean (n=17)	15.76	9.65	14.35	10.82
		SD	4.15	1.77	4.64	2.79
		Minimum	9	7	6	6
		Maximum	23	13	22	15
Simon Task	Yes	Mean (n=23)	55.96	36.48	50.09	39.13
		SD	19.108	9.477	18.002	11.745
		Minimum	26	19	19	21
		Maximum	99	51	83	61

9. Appendix II: Additional Analyses from Chapter III

In Chapter III, the main analysis of adult participants was done based on the pre-registered study design. As mentioned earlier in Chapter III (see page- 92), only 35 adult participants out of 90 fulfilled the five trial criteria under each condition and no child participants fulfilled the five. However, based on the examiner's feedback, some additional analyses were done by collapsing the post-error and post-correct trials under the RSI 200 and 750 ms as short RSI and under the RSI 1200 and 1750 ms as long RSI.[2] Examiners suggested collapsing the four RSI into two (Short vs. Long) so that more participants could be retained for the factorial analysis to see the effect of RSI and task types on PES. Details of descriptive and factorial analyses on the effect of task types and RSI (Short vs. Long) on PES and PECA are discussed below.

Post-error slowing (PES) in adults.

PES for each task under each RSI was calculated using the formula described in the method section. The mean PES for the two RSIs for the Simon and arithmetic interference tasks are presented in Figure 9.1. It can be observed from the figure that PES was higher under short RSI than long RSI for both tasks. One-sample t-tests were performed with Bonferroni-corrected p-value at a level of 0.01 to test whether PES under each RSI varied significantly from zero. Results from one-sample t-tests indicated that PES under both RSIs and for both tasks significantly varied from zero; results are presented in Table 9.1.

---

[2] *Note: A detailed description of mean trial counts under each condition can be found in Tables 9.7 and 9.8*

Appendix II

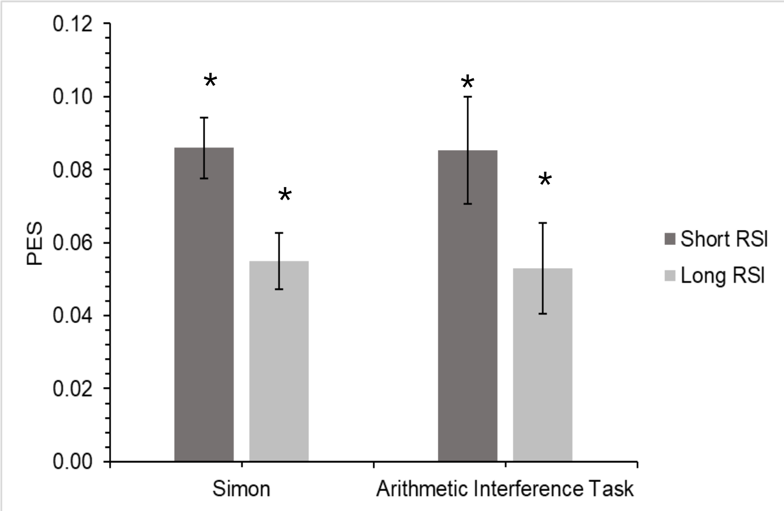


Figure 9.1 The PES for the Simon and Arithmetic Interference Task under short and long RSIs in adults, the error bar representing standard error (SE)

Table 9.1 One-Sample t-test (PES in Adults)

	RSI	*M*	*SD*	$t_{(1,88)}$	*P*
Simon	Short	0.086	0.079	5.813	<0.001
	Long	0.055	0.073	4.239	<0.001
Arithmetic Interference Task	Short	0.085	0.139	7.120	<0.001
	Long	0.053	0.118	10.328	<0.001

The data were analysed using 2 (Task Types) x 2(RSI: Short vs. Long) repeated measurement ANOVA to see the effect of task type and RSI on PES. Results from repeated measurement ANOVA indicated a significant main effect of RSI $F_{(1, 88)}$= 13.52, p<0.001, η_p^2= 0.13. Participants showed more PES in the short RSI (*M*= 0.09, *SE*= 0.009) compared to the long RSI (*M*= 0.05, *SE*= 0.008).However, no significant effects of task type; $F_{(1, 88)}$ = 0.012, p=0.915, η_p^2< 0.01 and interaction between task type and RSIs was observed $F_{(1, 88)}$ = 0.005, p=0.94, η_p^2< 0.01. This result is consistent with the main analysis where only a significant effect of RSI on PES was found, and it was larger under short RSIs.

Appendix II

Association of task type, RSI, and PECA in adults

Post-correct and post-error accuracy were calculated using the formulae described in the method section. Changes in accuracy rate at post-error and post-correct trials are depicted in Figure 9.2. It could be observed from Figure 9.2 that there is a higher decline in performance under the short RSI for both. A one-sample t-test was performed with Bonferroni-corrected p-value at a level of 0.01 for the Simon Task and arithmetic interference task to test whether PECA under each RSI was significantly different from zero. It was found that PECA differed significantly from zero for the short RSI only for the Simon task $t_{(1, 91)} = -4.65$, $p<0.001$, and the arithmetic interference task, $t_{(1, 91)} = -3.42$, $p<0.001$. PECAs did not vary significantly from zero ($p>0.01$) for the long RSI.

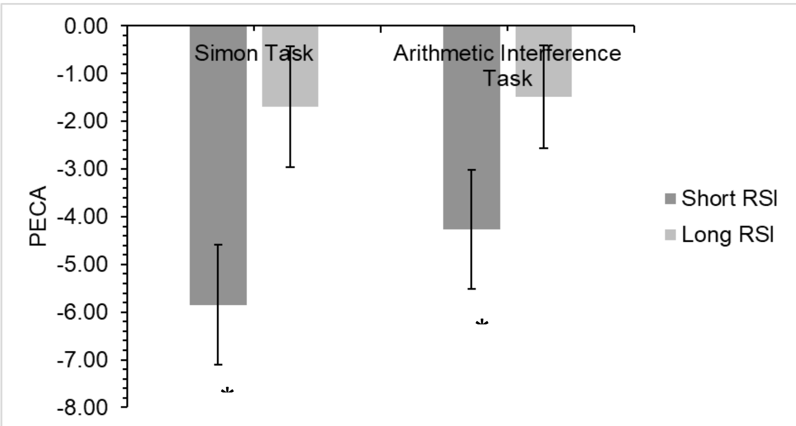


*Figure 9.2 PECA in the Simon and Arithmetic Interference tasks for different RSIs, the error bar representing standard error (SE). *= significantly different from zero.*

To see the effect of tasks and RSIs on post-error change in accuracy (PECA), a 2(Task) ×2(RSI) repeated measurement ANOVA was conducted. It was found that there was only a significant effect of RSI on PECA for $F_{(1, 91)} = 10.66$, $p= 0.002$, $\eta_p^2= 0.11$. No tasks effect, $F_{(1, 91)} = 0.45$, $p= 0.50$, and interaction effect was observed, $F_{(1, 91)} = 0.32$, $p= 0.57$. PECA was higher under short ($M = -5.06$, $SE = 0.81$) than long RSI (*M* = -1.59, *SE* = 0.82). Unlike main analysis, no significant interaction between task types and RSI was observed. However, similar to the main analysis in the chapter, PECA also declined largely under the short RSI.

It can be summarized from the above findings that adults showed significantly larger PES under short RSI for both tasks. The post-error accuracy got significantly worse under short RSI. However, the difference in accuracy after error and correct trials under long RSIs did not differ. These results indicated that PES and PECA are influenced by RSI duration but not task types. These findings are similar to the main analysis, except PECA is unaffected by task types.

Post-error slowing (PES) in children.

The PES for the two RSIs for the arithmetic interference task was calculated on 36 children who fulfilled the five trial criteria. A one-sample t-test was performed with Bonferroni-corrected p-value at a level of 0.03 to test whether PES varied significantly from zero under each RSI. Results from one-sample t-tests (Table 9.2) indicated that PES under both RSIs did not vary significantly from zero.

Table 9.2 One-Sample t-test for PES in children

	RSI	*M (n= 36)*	*SD*	$t_{(1, 35)}$	*P*
Arithmetic Interference Task	Short	0.045	0.275	0.984	0.166
	Long	0.066	0.256	1.543	0.066

The data were analysed using one repeated measurement ANOVA where RSI was the factor and PES was the dependent variable to see the effect of RSI on PES. Repeated measurement ANOVA results indicated no significant effect of RSI $F_{(1, 35)}$= 0.12, *p*=0.73, on PES.

Appendix II

Association of RSI and PECA in Children

Mean post-error change in accuracy was calculated for the arithmetic interference task, and one-sample t-tests (Table 9.3) were performed with Bonferroni-corrected p-value at 0.03 to test whether PES varied significantly from zero under each RSI. Results from one-sample t-tests indicated that PES under only long RSIs varies significantly from zero. The data were analysed using one repeated measurement ANOVA where RSI was the factor, and PECA was the dependent variable to see the effect of RSI on PECA.

Table 9.3 One-Sample t-test for PECA in children

	RSI	*M*	*SD*	$t_{(1, 69)}$	*P*
Arithmetic Interference Task	Short	-3.18	16.58	-1.60	0.11
	Long	-4.67	16.53	-2.37	0.02

Repeated measurement ANOVA results indicated no significant effect of RSI $F_{(1, 69)}$= 0.27, p=0.60, on PECA. These findings are similar to the main analysis where no effect of RSI on PECA was found. It can be summarized from the above findings that children showed no significant effect of RSI on PES and PECA, which is similar to the main analysis.

Relationship between PES and arithmetic skills

PES was most profound under the short RSI for both tasks. Therefore, it was decided to calculate the correlation between PES and arithmetic skills for both tasks under short RSIs in adults. It was found that there was no significant relationship between the arithmetic skills and PES on the Simon task, $r_{(3, 85)}$ = .09, p = 0.42, and the arithmetic interference task, $r_{(3, 85)}$ = .14, p = 0.20, in adults. A Williams-Steiger test was used to investigate if the size of the correlations between the Simon task and the arithmetic interference task differed significantly in adults. It was found that the correlation scores between the tasks in adults were non-significant, $t_{(3,\ 85)}$= 0.33, *p*= 0.55. These findings indicated that post-error slowing did not affect adult maths performance. These findings are also like the main findings from adult data analysis.

Since no effect of RSI was observed in children, maximum PES, irrespective of RSI, was calculated for individual participants. This maximum PES of the arithmetic interference task was correlated with the arithmetic skills in children. A non-significant correlation, $r_{(1,\ 35)}$ = -.09, *p* = 0.60, was observed. This finding is similar to the main analysis. A Fisher's r-to-z test showed that the correlation size for the arithmetic interference task between adults and children differs significantly, *z*= .25, *p*= 0.80. Since no improvement in accuracy was observed in post-error trials, the correlation between PECA and arithmetic skills was not calculated.

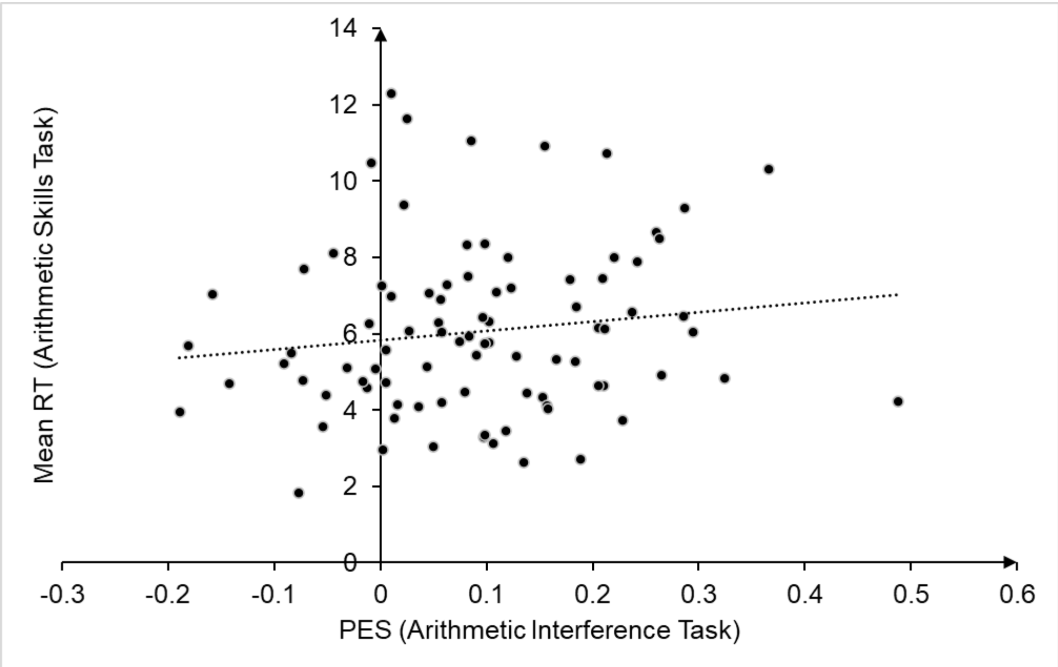


Figure 9.3 Pearson product-moment correlation between mean RT of arithmetic skills and PES in the Arithmetic Interference task in adults.

Appendix II

Descriptive Analysis of Trial Counts

Following the examiners' suggestions, descriptive analyses were done on the trial counts under each condition for adults (Table 9.4).

Table 9.4 Mean trial count in post-error and post-correct trials under each RSI for each task separately in adults (n= 35).

Tasks	Trials	RSI (ms)	Mean Trial Count	SD
Simon Task	Post-Error	200	13.71	4.78
	Post-Correct	200	9.37	2.66
	Post-Error	750	14.54	4.66
	Post-Correct	750	8.03	1.96
	Post-Error	1200	15.03	5.51
	Post-Correct	1200	8.71	2.01
	Post-Error	1750	15.17	4.91
	Post-Correct	1750	9.06	2.14
Arithmetic Interference Task	Post-Error	200	10.43	3.14
	Post-Correct	200	7.06	1.33
	Post-Error	750	10.71	3.20
	Post-Correct	750	6.91	1.40
	Post-Error	1200	11.40	4.14
	Post-Correct	1200	7.74	2.11
	Post-Error	1750	11.03	3.30
	Post-Correct	1750	7.09	1.34

It can be seen from Table 9.4 that trial numbers in the post-error trial are higher than trial numbers in post-correct trials. One of the reasons is that many post-correct trials turned into error trials or blank responses because of how participants responded. Another reason could be assumed that since overall mean post-error and post-correct trials are not that much higher due to individual variations, a more significant number of adults did not have a reasonable distribution of post-error and post-correct trials under each RSI. To test these assumptions, additional descriptive analysis was done with adult participants who dropped out from the ANOVA analysis for not fulfilling the five post-error and post-correct trial criteria under each RSIs.

Appendix II

Here, data show how many post-correct trial responses turned into blank and error responses. At the same time, to calculate how many consecutive errors and blank responses participants made in a row after an error response are presented in the table (Table 9.5).

Table 9.5 Mean trial count in post-error and post-correct trials that turned into blank, and error responses or consecutive errors between the participants who fulfilled the 5-trial criteria and who did not

Tasks	5-trials Criteria Fulfilled	Post-Correct Trials		Post-Error Trials	
		Blank Response	Error Response	Consecutive Errors	Blank Response
		Mean (SD)		Mean (SD)	
Arithmetic Interference Task	Yes	0.089 (0.29)	7.36(6.09)	11.71(10.65)	0.24(0.96)
	No	0.68 (1.62)	12.28(13.24)	24.58(29.22)	1.94(4.56)
Simon Task	Yes	1.39(1.37)	12.25(8.83)	21.65(16.69)	5.05(5.70)
	No	2.17(2.85)	14.63(14.26)	26.86(30.11)	6.46(6.20)

It could be observed from the above table that participants who dropped out from the analysis made more blank and error responses than those who fulfilled the 5-trial criteria. Similarly, dropped-out participants made more consecutive errors or blank responses in a row than the selected participants. This pattern of response observed more for the arithmetic interference task than the Simon task. These findings are indicating that adults who were poor performer could not fulfil the five trials criteria under each condition.

Following the examiners' suggestions, descriptive analyses were done on the trial counts under each condition for children (Table 9.6). It could be observed from Table 4.1 that mean accuracy rates under each RSI for post-error and post-correct trials were already low. That means overall children made a fair number of errors, but it is not enough to fulfil the five trials criteria under each RSI.

Table 9.6 Mean trial count in post-error and post-correct trials under each RSI for arithmetic interference task in children (n= 66).

Trial Types	RSI (ms)	Mean	SD
Post-Error	200	4.86	2.739
Post-Correct	200	2.70	1.252
Post-Error	750	5.05	2.787
Post-Correct	750	2.92	1.216
Post-Error	1200	4.94	2.705
Post-Correct	1200	2.56	1.083
Post-Error	1750	5.02	2.709
Post-Correct	1750	2.74	1.004

In addition to this descriptive analysis based on main analysis some additional descriptive analysis was done based on examiners review where the RSI was collapsed between short and long RSI. They are displayed in the following tables.

Table 9.7 Mean trial count in post-error and post-correct trials under short vs. long RSI for each task separately in adults.

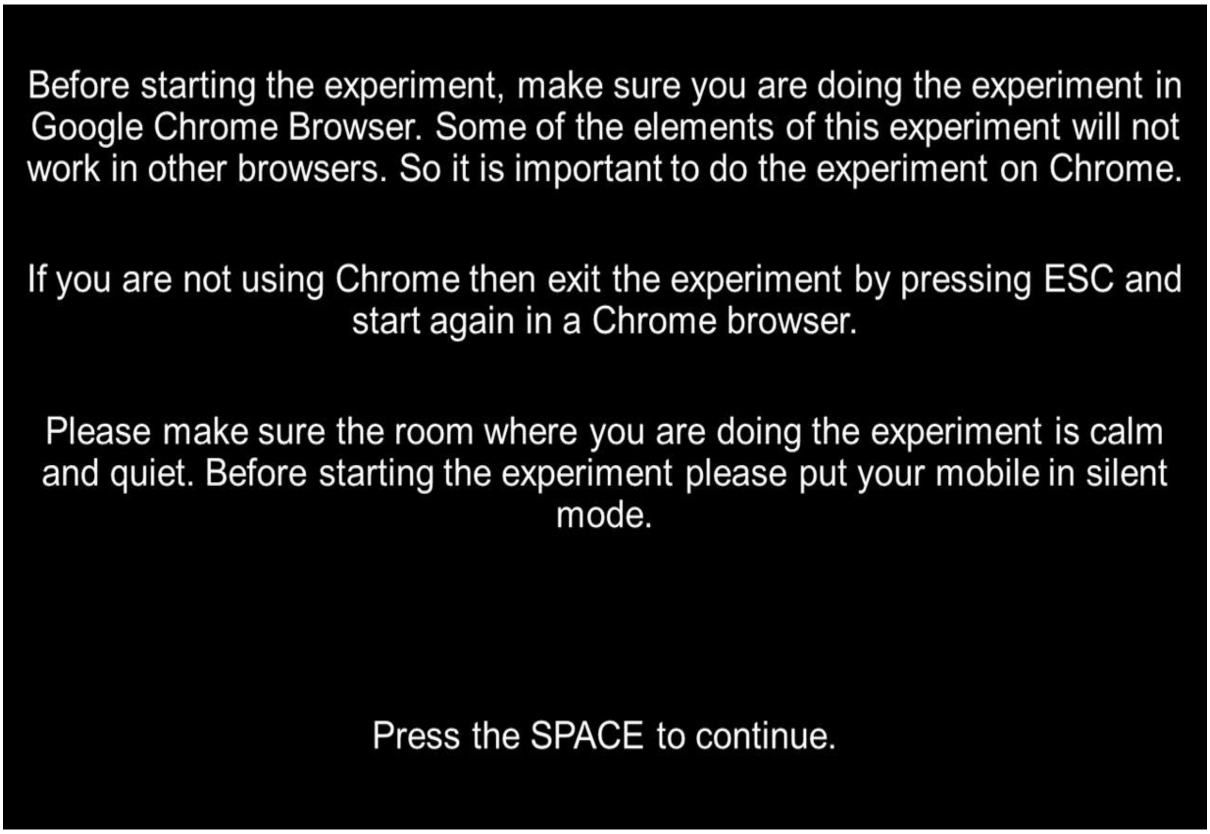
Tasks	RSI	Trials	Mean Trial Count	SD
Simon Task	Short	Post-Error	26.87	10.918
		Post-Correct	19.81	6.103
Arithmetic Interference Task	Long	Post-Error	28.81	11.124
		Post-Correct	20.74	6.182

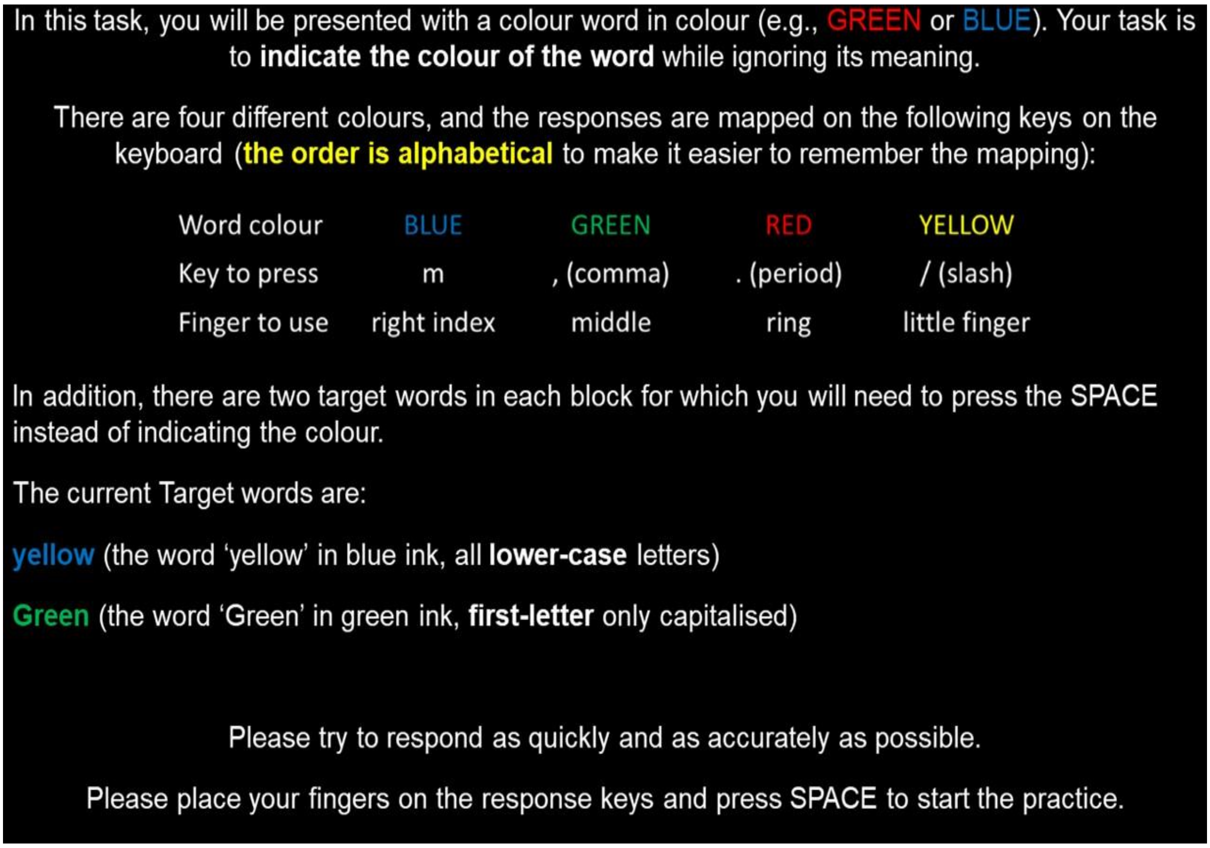
Appendix II

Table 9.8 Mean trial count in post-error and post-correct trials under short vs. long RSI for the arithmetic interference task in children.

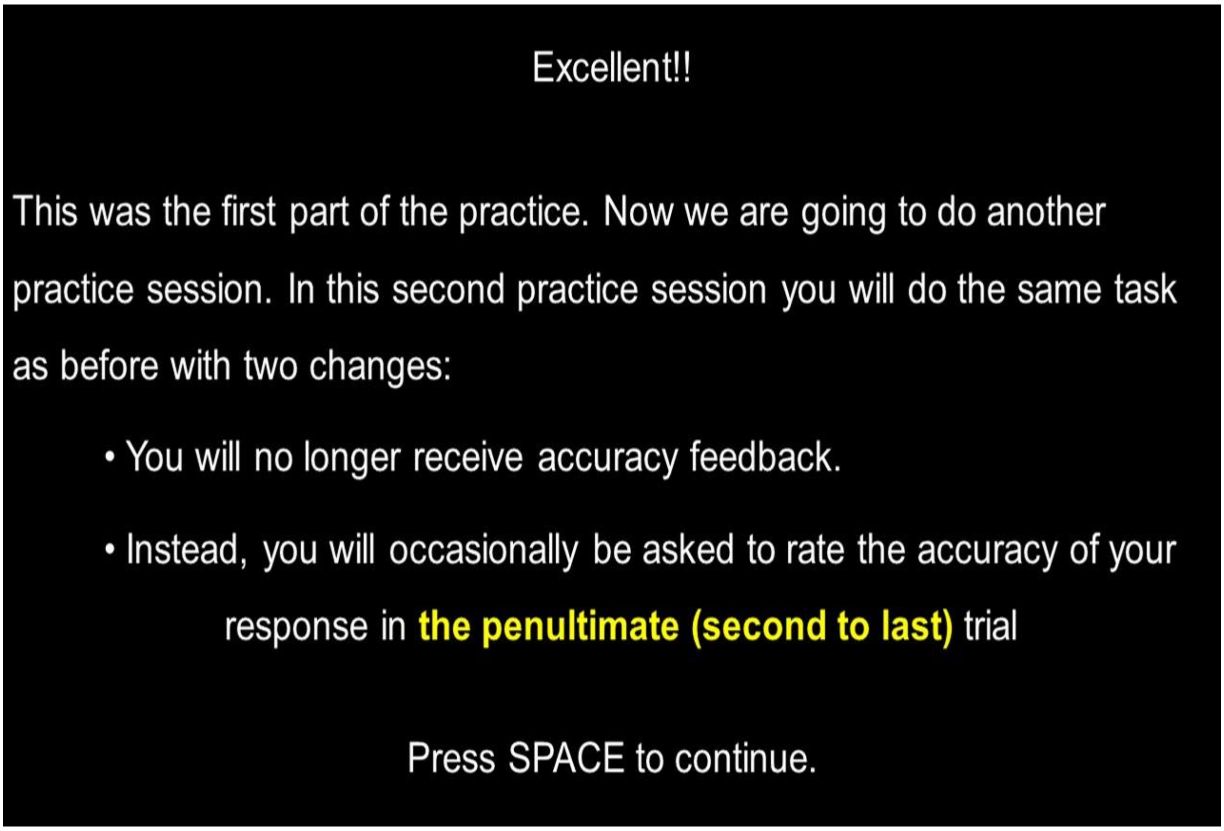
RSI	Trial Types	Mean Trial Count	SD
Short RSI	Post-error trials	12.78	4.48
	Post-correct trials	6.92	1.83
Long RSI	Post-error trials	12.58	4.45
	Post-correct trials	6.14	1.25

10. Appendix III: Instructions for Study III (Chapter V)

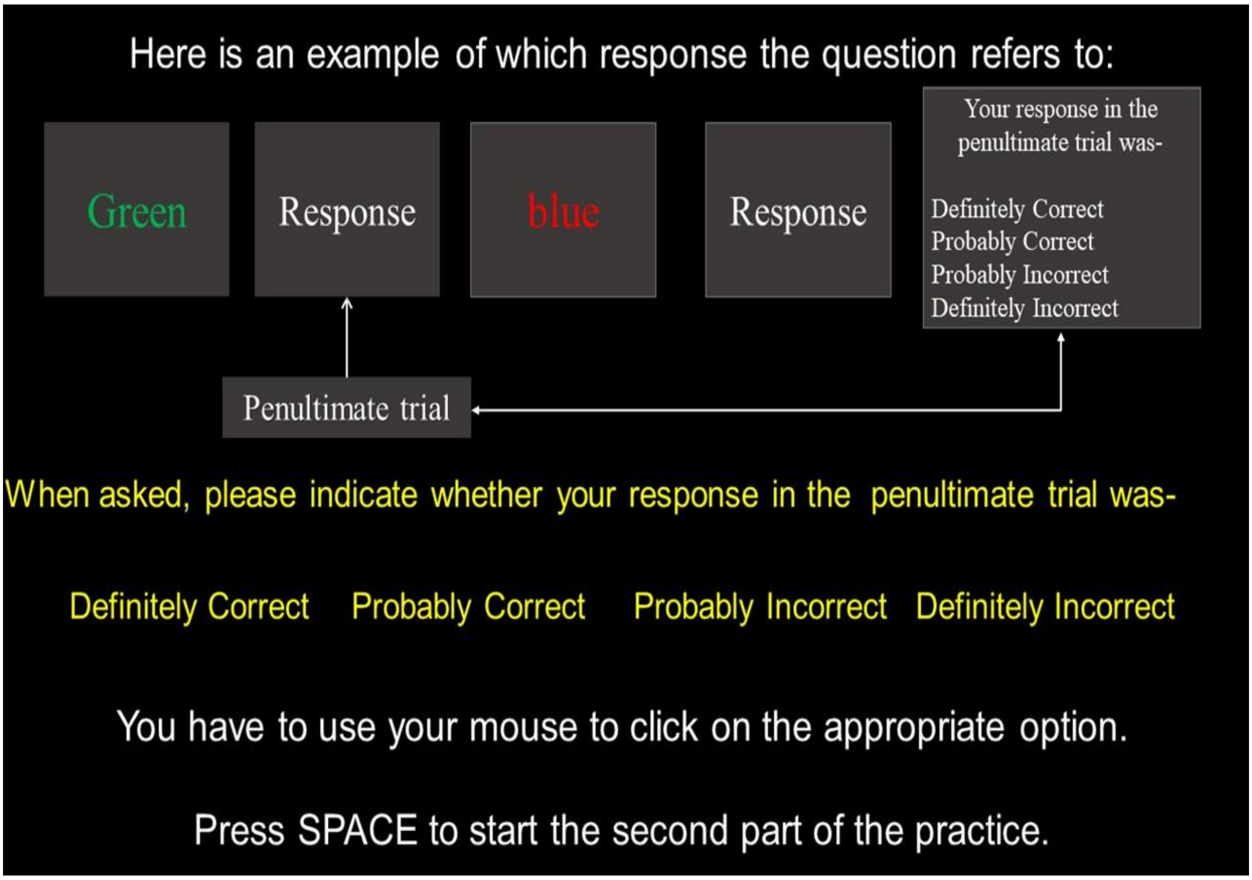


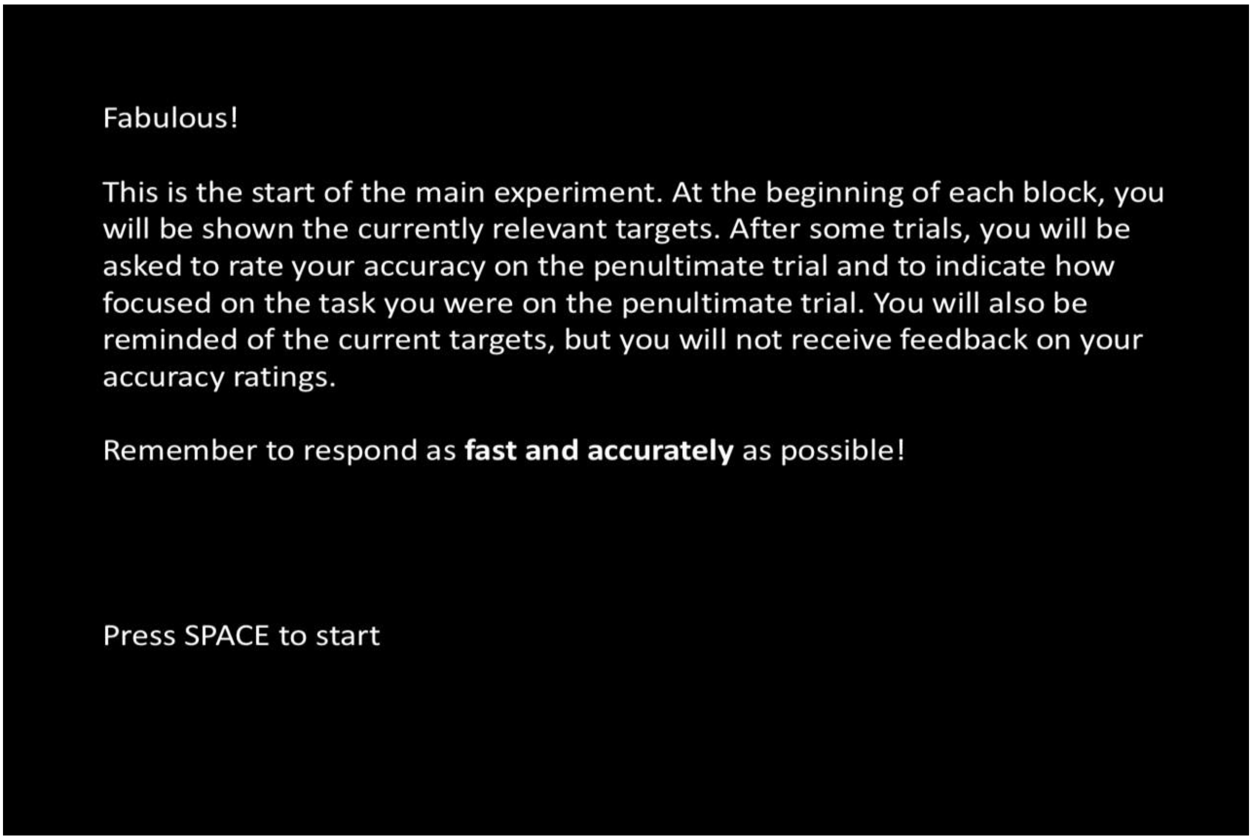


Appendix III



Appendix III





11. Appendix IV (Pre-registration Study II)

'*The Role of Metacognition in Mathematics Achievement*'
(AsPredicted #49581)

Created: 10/13/2020 04:04 PM (PT)

Author(s)
Naima Nigar (University of Nottingham, UK) - Naima.Nigar1@nottingham.ac.uk
Claudia Danielmeier (University of Nottingham, UK) - claudia.danielmeier@nottingham.ac.uk
Lucy Cragg (University of Nottingham, UK) - lucy.cragg@nottingham.ac.uk

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
In the present study, we will investigate if metacognition is related to maths achievement. Metacognition will be measured in terms of post-error slowing (PES) and Post-error improvement in accuracy (PIA). Two error monitoring tasks will be used: a non-numeric task (Simon task) and a numeric task (number fact interference
task). Since it has been shown that PES and PIA are influenced by the response stimulus interval (RSI, e.g. Danielmeier & Ullsperger, 2011), four RSIs will be used (200, 750, 1200, and 1750 ms). Two groups of participants (8-15-year-old children and adults) will be recruited. We will investigate which RSI is most suitable for measuring PES and PIA in each age group for each task. The relationship between error monitoring (PES and PIA) with overall mathematics achievement will also be determined and whether this relationship is influenced by age group and tasks.
Hypothesis:
1. Participants with better metacognitive abilities (higher PES and PIA) will perform better in maths.
2. Metacognitive abilities which will measure in numerical context will be a stronger predictor for maths achievement compared to metacognitive abilities which will measure in non-numerical contexts.
3. Children will show more post-error adjustments in longer RSIs, whereas adults will show more PES in shorter RSIs and more PIA in longer RSIs.
Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. Frontiers in Psychology, 2, 233. https://doi.org/10.3389/fpsyg.2011.00233

3) Describe the key dependent variable(s) specifying how they will be measured.
The dependent variables for the Simon task and the number fact interference task will be post-error slowing (PES) and post-error improvement in accuracy (PIA) scores.

PES will be calculated as follows:

PES=Mean(〚PostError〛 _RT- 〚PostCorrect〛 _RT)/(Mean (PostCorrectRT))

Here for PES the post-error trials are defined as correct trials that follow an error and post-correct trials are matched with corresponding post-error trials with their assigned RSI using a psychopy program. Wherever, a participant committed an error and immediately after that error a target future post-correct trial (within 4th to 8th trials next to that error) with

Appendix IV

corresponding post-error RSI was selected.

PIA will be calculated as follows:

PCA= Post-Correct Accuracy
PCC= Post-Correct Correct Responses
PCE= Post-Correct Error Responses
PEA= Post-Error Accuracy
PEC= Post-Error Correct Responses
PEE= Post-Error Error Responses
n = total number of trials.

PCA=(nPCC/n(PCC+PCE)) x 100
PEA=(nPEC/n(PEC+PEE)) x 100.

PIA=PEA-PCA

The dependent variable for the maths achievement task will be RT for correct responses.

4) How many and which conditions will participants be assigned to?
Analysis for PES: 16 conditions in a mixed design lab study. Each group (Adults and Children: Between-subject condition) of participants will take part in two different tasks (2 within-subject conditions) with four different RSIs (4 within-subject conditions).
Analysis for PIA: 16 conditions in a mixed design lab study. Each group (Adults and Children: Between-subject condition) of participants will take part in two different tasks (2 within-subject conditions) with four different RSIs (4 within-subject conditions).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
At first, we will see which RSI is suitable to measure PES and PIA for each task (Simon Task and Math Task) and for each age group. For that, we are going to do a 4 (RSIs) by 2 (Tasks) by 2 (Group) mixed ANOVA test, where RSIs and Tasks will be within-subject factors, and Group will be a between-subject factor. Separate analyses will be run with PES and PIA as the dependent variable.
Following these analyses, we will select the RSI that gives the best measure of PES separately for each task and age group to use in subsequent analyses. This will be repeated for PIA.

Finally, we will do a Hierarchical multiple regression. In the first step the predictors will be Age group, PES(Simon), and PIA(Simon), PES (maths), and PIA (maths). In the second step, the interactions between age group and the task scores will be added. RT on the maths achievement task (Procedural math task) will be the dependent variable.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
The detection of outliers will be done in three phases for measuring PES. In the first phase, trials with a mean RT of more than ±3 SD from the individual's mean RT will be excluded from the study. Next, we will exclude any participants who will have fewer than 5 trials under any of the four RSIs that will be excluded further for ANOVA analysis. Finally, participants whose mean RT will be more than ±3 SD from the group mean RT will be excluded from ANOVA analysis.
The detection of outliers in PIA will be done based on their overall accuracy rate, that is if any participant's mean accuracy rate will be more than ±3 SD then they are going to be excluded from ANOVA analysis.
To detect univariate and multivariate outliers in the Hierarchical regression analysis box plot technique and Mahalanobis distance technique will be used.

7) How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.
After doing power analysis with a small to medium effect size (0.10) for regression (small f2 = 0.02 & medium f2 = 0.15; Cohen, 1988), it was found that for this experiment 160 participants (80 children and 80 adults) will be enough with a power of 0.90.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Appendix IV

8) Anything else you would like to pre-register?

(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We already have started collecting a few data. Due to COVID, our university has arranged a summer scientist online program for a one-month data collection. So we had to program everything and launch the experiment online in a very short time. And the data collection has just begun.

This pre-registration currently has an anonymized pdf. Once the paper is accepted, you should make the pre-registration public. This will add the authors' name to the .pdf, drop the 'CONFIDENTIAL' warning from the top of the page, and ensure a copy of the pre-registration is sent to the Web Archive.

If you click below you will see more information about this process, and can still change your mind.

MAKE PUBLIC (request_public.php?a=cVJYSmdSZFJnYXZyL3FzNkM1MUxzZz09)

12. Appendix V (Pre-registration Study III)

'*The Relationship between Error Awareness and Arithmetic*'
(AsPredicted #81071)

Created: 11/25/2021 08:55 AM (PT)

Author(s)
Naima Nigar (University of Nottingham, UK) - naima.nigar1@nottingham.ac.uk
Claudia danielmeier (University of Nottingham, UK) - claudia.danielmeier@nottingham.ac.uk
Lucy Cragg (University of Nottingham, UK) - lucy.cragg@nottingham.ac.uk

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
In the present research, we are going to investigate the role of error awareness in mathematics achievement and if this relationship is domain-general or domain-specific. Data will be collected from adult participants using two error awareness tasks one non-numeric (modified version of the Stroop task) and one numeric (modified number fact retrieval task) and two mathematics tests. We are hypothesizing that error awareness will have a positive relationship with mathematics achievement and this relationship will be more domain-general rather than domain-specific.

3) Describe the key dependent variable(s) specifying how they will be measured.
Error awareness tasks
To measure error awareness the percentage of aware errors will be calculated for both numeric and non-numeric tasks. Aware and unaware errors will be determined based on a participant's response to a question following the trial after an error: "Was your penultimate response correct or incorrect?". Participants will be asked to indicate if their response had been 'Definitely Correct', 'Probably Correct', 'Probably Incorrect' or 'Definitely Incorrect'. If a participant's error response matches with their subjective response judgment (i.e., 'Definitely Incorrect' or 'Probably Incorrect') the error will be categorized as an 'aware error'. If there is a mismatch between the actual response and the subjective response evaluation, the error will be categorized as an 'unaware error'. For each participant, the percentage of aware errors (in relation to all errors) will be calculated for each task.

Mathematics tasks
As the coronavirus pandemic has affected face-to-face testing opportunities, we chose mathematics measures that can be completed online. The first measure is an online version of the procedural skills task (Cragg et al., 2017). This is composed of 12 addition, subtraction, multiplication, and division operations. The participants will be instructed to type their answers using their computer keyboard. The dependent variable used will be RT (in the previous research it was found that RT is the best measure for this test). The second measure is an online adaptation of Chinn's Mathematics test (Chinn, 2020). This is a 15-Minute test that covers a range of arithmetic and algebra questions. Participants are presented with 4 possible answers and have to choose the correct one. They complete as many questions as they can within the 15-minute time limit. The dependent variable used for this test will be accuracy.

4) How many and which conditions will participants be assigned to?
It will be a repeated measurement design study and each participant will take part in two different error awareness tasks

(numeric error awareness task and non-numeric error awareness task) as well as the two different mathematics tests.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

At first, Pearson product-moment correlation coefficients will be calculated between the four variables. Secondly, we will perform a hierarchical multiple regression model separately for each of the mathematics measures if there is a significant positive correlation between error awareness (either numeric vs non-numeric) and that mathematics measure. In this model, the predictors will be the percentage of aware errors in the non-numeric error awareness task and the percentage of aware errors in the numeric error awareness task.

At first in model 1 we will include the non-numeric task as a predictor to see how much variance will be explained by the model and then, in model 2, a numeric error awareness predictor will be included to investigate its additional contribution to the model besides the non-numeric error awareness.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

To detect univariate and multivariate outliers in the hierarchical regression analysis, the box plot technique and Mahalanobis distance will be used to determine how far each case is from the centroid of all cases for the predictor variables.

Participants who are classed as multivariate outliers (Mahalanobis distance with $p < 0.05$) will be excluded from all analyses, without replacement. Mahalanobis distance will be calculated using four variables; the proportion of aware errors on the numeric task, proportion of aware errors on the non-numeric task, RT on the procedural mathematics task, and accuracy on the Chinn mathematics test.

7) How many observations will be collected or what will determine sample size?

No need to justify decision, but be precise about exactly how the number will be determined.

After conducting a power analysis with a small effect size (0.02666667) for regression (based on the data from similar studies, e.g., Bellon & De Smedt, 2019), it was found that for this experiment 297 adult participants will be sufficient to achieve a power of 0.80. However, in a total of 310 participants will be collected to allow for the exclusion of outliers.

8) Anything else you would like to pre-register?

(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

As part of our secondary analysis, we have decided to run two additional analyses.

In addition to measuring error awareness, we have designed the experiment in such a way that two additional variables could be measured on the error-awareness tasks: Post-error slowing (PES) and mind wandering.

PES will be calculated as follows:

PES=Mean(〚PostError〛 _RT- 〚PostCorrect〛 _RT)/(Mean (PostCorrectRT))

Mind Wandering will be measured with 14 points Likert Scale in response to the question 'How focused on the task were you on the penultimate trial?' asked after the trial following an error trial or matched correct trial.

Here we will investigate if there is any relationship between error awareness and PES and mind-wandering. To analyze this, we will do two separate 2 (aware vs unaware error) χ 2(numeric vs non-numeric task) ANOVAs where PES and mind wandering will be dependent variables, respectively.

Correlations between PES and mind wandering and the two mathematics tests will also be calculated.

References

Bellon, E., Fias, W., & De Smedt, B. (2019). More than number sense: The additional role of executive functions and metacognition in arithmetic. Journal of Experimental Child Psychology, 182, 38–60. https://doi.org/10.1016/j.jecp.2019.01.012Chinn, S. (2020). The 15-Minute norm-referenced Mathematics Test. More Trouble with Maths: A Complete Manual to Identifying and Diagnosing Mathematical Difficulties (3rd ed.; pp). Routledge. https://doi.org/10.4324/9781003017721

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. Cognition, 162, 12–26. https://doi.org/10.1016/j.cognition.2017.01.014

Appendix V

If you click below you will see more information about this process, and can still change your mind.

MAKE PUBLIC (request_public.php?a=dDhmTmNtSEd6NVFMMktsTzdMTDVodz09)

Wharton CREDIBILITY LAB UNIVERSITY of PENNSYLVANIA (http://credlab.wharton.upenn.edu)

Report a bug (bug.php) | Make a suggestion (suggestion.php) | Terms of Use (https://aspredicted.org/messages/terms.php) | Support (https://aspredicted.org/messages/faq.php) | About (https://aspredicted.org/messages/about.php)