FACULTY OF PSYCHOLOGY AND
EDUCATIONAL SCIENCES

PARENTING AND SPECIAL EDUCATION
RESEARCH UNIT

Metacognition counts too.

The role of metacognition in arithmetic development: Evidence from behaviour and brain.

Elien Bellon

Doctoral thesis offered to obtain the degree of

Doctor of Educational Sciences (PhD)

Supervisor: Prof. dr Bert De Smedt

Co-Supervisor: Prof. dr Wim Fias

2020

FACULTY OF PSYCHOLOGY AND
EDUCATIONAL SCIENCES

PARENTING AND SPECIAL EDUCATION
RESEARCH UNIT

Metacognition counts too.

The role of metacognition in arithmetic development: Evidence from behaviour and brain.

Elien Bellon

Doctoral thesis offered to obtain the degree of

Doctor of Educational Sciences (PhD)

Supervisor: Prof. dr Bert De Smedt

Co-Supervisor: Prof. dr Wim Fias

2020

For you, but most of all, thanks to you.

Elien Bellon - Metacognition counts too.

The role of metacognition in arithmetic development: Evidence from behaviour and brain.

Dissertation submitted to obtain the degree of Doctor in Educational Sciences, 2020.

Supervisor: Prof. dr. Bert De Smedt – Co-Supervisor: Prof. dr. Wim Fias

---

Arithmetic performance and development in children is characterized by large individual differences. Understanding what is associated with and/or drives these individual differences is of the essence, as arithmetic is hugely important in children's educational development and, more generally, in many aspects of modern life. Rightfully so, a large body of research has investigated the correlates of arithmetic performance and development. Important processes such as numerical magnitude processing and executive functions have been put forward. Yet, their unique contribution to arithmetic is not fully understood, as these processes are generally studied in isolation. The aim of the current dissertation was to further our understanding of these correlates of arithmetic by investigating different cognitive, metacognitive and affective processes in concert. The critical contribution of this dissertation was the inclusion of metacognition and thorough investigation of metacognitive monitoring in this context, which has been mostly overlooked in the existing literature.

Firstly, we ran a longitudinal panel study in primary school children that examined the unique roles of numerical magnitude processing, executive functions and metacognition in second grade, in third grade and over development between second and third grade. The results revealed that, over and above the role of numerical magnitude processing and executive functions, metacognitive monitoring was associated with performance in both early and middle primary school and predicted later arithmetic performance. Its role in predicting arithmetic development, taking into account prior arithmetic performance, needs further investigation.

In a next study, we first investigated the interrelations between metacognitive monitoring and mathematics anxiety, to then examine whether their interplay influenced the concurrent and longitudinal associations that were found between metacognitive monitoring and arithmetic. Our findings indicate that, while mathematics anxiety was related to both metacognitive monitoring and arithmetic, the association between metacognitive monitoring and arithmetic was unique and specific, without mediation or moderation of mathematics anxiety.

The fourth study addressed the currently debated extent to which metacognitive monitoring is domain-specific or reflective of a more general performance monitoring process. Investigating this in highly relevant academic domains in primary school children, we found that more domain-general metacognitive monitoring processes emerge over the ages from 7 to 9.

Finally, in a fifth study, we uncovered the neurobiological basis of metacognitive monitoring in children. We demonstrated that brain activity during metacognitive monitoring increased in de left inferior frontal gyrus and correlated with arithmetic performance. We provided the first empirical evidence in favour of the hypothesis that prefrontal cortex activity during arithmetic is related to the higher-order process of metacognitive monitoring.

Collectively, these studies demonstrated the importance of metacognitive monitoring of accuracy for arithmetic performance and development in primary school children, while confirming the role of numerical magnitude processing and mathematics anxiety, and to a lesser extent executive functions. The current dissertation thus furthered our understanding of the correlates of arithmetic. Its critical contribution was uncovering the role of metacognition on top of other crucial processes, and as such emphasizing that 'metacognition counts *too*'.

Elien Bellon - Metacognitie telt ook.
De rol van metacognitie in de rekenontwikkeling van kinderen: Evidentie vanuit gedrags- en hersenonderzoek.
Proefschrift aangeboden tot verkrijgen van de graad Doctor in de Pedagogische Wetenschappen, 2020.
Promotor: Prof. dr. Bert De Smedt – Co-Promotor: Prof. dr. Wim Fias

---

De rekenprestatie en –ontwikkeling van kinderen wordt gekarakteriseerd door grote individuele verschillen. Het goed begrijpen van wat samenhangt met deze individuele verschillen en/of wat deze individuele verschillen drijft, is essentieel gezien het uitermate groot belang van rekenen in de schoolse ontwikkeling, en in het algemeen in verscheidene aspecten van ons moderne, dagelijkse leven. Een omvangrijke onderzoeksliteratuur heeft dan ook terecht de correlaten van rekenprestaties en –ontwikkeling onderzocht. Het belang van processen zoals getalgevoel en executieve functies werd reeds benadrukt. Hun unieke bijdrage tot rekenen is echter nog niet volledig duidelijk, aangezien deze processen meestal onafhankelijk van elkaar worden bestudeerd. Het doel van dit proefschrift was om ons begrip van deze correlaten van rekenen te bevorderen, door verschillende cognitieve, metacognitieve en affectieve processen gezamenlijk te onderzoeken. De kritische bijdrage van dit proefschrift is het includeren van metacognitie in deze onderzoekslijn en het diepgaand onderzoeken van metacognitieve monitoring. Dit werd immers veelal over het hoofd gezien in de bestaande literatuur.

In een longitudinale studie bij lagere school kinderen werd de unieke rol van getalgevoel, executieve functies en metacognitie onderzocht in het tweede leerjaar, het derde leerjaar en overheen de ontwikkeling tussen tweede en derde leerjaar. Onze resultaten toonden aan dat metacognitieve monitoring, bovenop de rol van getalgevoel en executieve functies, samenhangt met prestatie in het begin en in het midden van de lagere school, en latere rekenprestatie voorspelt. Er is nood aan verder onderzoek naar de voorspellende rol van metacognitieve monitoring voor de ontwikkeling van rekenen, bovenop wat voorspeld wordt door eerdere rekenprestatie.

In een volgende studie onderzochten we eerst de onderlinge samenhang tussen metacognitieve monitoring en rekenangst, om daarna te bestuderen of deze wisselwerking een invloed heeft op de samenhang die we vonden tussen metacognitieve monitoring en rekenen. De resultaten toonden aan dat, hoewel rekenangst weldegelijk samenhangt met zowel metacognitieve monitoring als rekenen, de onderlinge samenhang tussen metacognitieve monitoring en rekenen niet gedreven of beïnvloed wordt rekenangst.

Een vierde studie richtte zich op het huidige debat over de mate waarin metacognitieve monitoring domein-specifiek is, dan wel een meer domein-algemeen monitoring proces weerspiegelt. We onderzochten dit in zeer relevante academische domeinen bij lagereschoolkinderen en vonden dat meer domein-algemene metacognitieve monitoring processen ontluiken tussen de leeftijd van 7 tot 9 jaar.

Tenslotte, in een vijfde studie achterhaalden we de neurobiologische basis van metacognitieve monitoring bij kinderen. We toonden aan dat hersenactiviteit tijdens metacognitieve monitoring toenam in de linker inferieur frontale gyrus en correleerde met rekenprestatie. Daarmee verschaften we de eerste empirische evidentie voor de hypothese dat prefrontale activiteit tijdens rekenen samenhangt met het hogere orde proces metacognitieve monitoring.

Het geheel van deze studies beklemtoont het belang van metacognitieve monitoring van accuraatheid voor rekenprestaties en –ontwikkeling en bevestigt tegelijk de rol van getalgevoel en rekenagst, en in mindere mate executieve functies. Zodoende bevordert dit proefschrift ons begrip van de correlaten van rekenen. De cruciale meerwaarde van dit doctoraat is het aantonen van de rol van metacognitie bovenop andere cruciale processen, en aldus het benadrukken dat ‘metacognitie *ook* telt.’

Table of Contents

Overview of Included Figures

Chapter 1

Chapter 2

Chapter 4

Chapter 5

Chapter 6

Overview of Included Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

CHAPTER I

General introduction & Aims

Chapter 1
General Introduction & Aims

Arithmetic is where
the answer is right and everything is nice and you can look out of the window and see the blue sky-
or the answer is wrong and you have to start over and try again and see how it comes out this time.
– C. Sandburg

From a very young age, children are surrounded by arithmetic: Starting in primary school, arithmetic represents an important part of their curriculum, and even after school, children use arithmetic day in, day out to calculate the number of sweets they have compared to a sibling or to make sure they receive the promised screen time. This omnipresence of arithmetic continues throughout adulthood as we adjust recipes to fit the number of people at dinner, calculate how delayed our train or plane is, and figure out if the end of season sale is worthwhile.

Arithmetic, i.e. the ability to mentally manipulate numbers, including addition, subtraction, multiplication and division, and its development have been major topics in research for numerous decades, studied in numerous research groups all around the world and from the perspective of numerous disciplines. Insight into arithmetic performance and development is of interest to not only educators, psychologists and educational scientists, neuroscientists, educational policy makers, or mathematicians, but in general, to all who need to learn arithmetic and to deal with numbers (Dowker, 2019c).

An extensive body of literature demonstrates that there are large individual differences in the way children acquire arithmetic abilities (e.g. Berch et al., 2016; Dowker, 2005, 2019c). In modern society, arithmetic skills are crucial abilities, related to job prospects, income and quality of life (e.g. Ancker & Kaufman, 2007; Finnie & Meng, 2001; Gerardi et al., 2013) and early individual differences in mathematics, including arithmetic, predict later adult socioeconomic status (e.g. Chiswick et al., 2003; Ritchie & Bates, 2013). Because there are such large individual differences in arithmetic abilities and because this skill has such a central role in modern society, it is not surprising that there is a long-standing tradition in developmental and educational research devoted to uncovering the cognitive correlates of arithmetic performance and development, and eventually, to developing ways to intervene in and/or stimulate arithmetic. Rightfully so, during the last decade, there has been a boost in behavioural and neuroimaging research on numerical cognition seeking to identify important predictors of mathematics, and in particular arithmetic performance (e.g. Dowker, 2019c; Gilmore et al., 2018a).

This doctoral dissertation aimed to contribute to this body of research by investigating the role of different cognitive and affective, and in particular metacognitive processes in individual differences in arithmetic performance and development. In this general introduction, first, I discuss the choice for arithmetic as mathematical domain of interest. Second, the different cognitive, metacognitive and affective processes that were investigated within this dissertation are discussed, elucidating the nature of these processes, their association with arithmetic performance and development, and how these processes are (generally) measured, while standing still at shortcomings or gaps of the existing literature. Next, I consider two important issues for research on cognitive, metacognitive and affective processes associated with arithmetic in primary school children. These concern the importance of investigating different cognitive, metacognitive and affective processes in concert, and the importance of investigating the interrelations between these developing processes longitudinally, taking into account prior performance. Subsequently, I elucidate the specific focus on metacognitive monitoring within this dissertation and discuss two important issues regarding the association between metacognitive monitoring and arithmetic that were tackled. More specifically, I will discuss why considering the role of mathematics anxiety within the association between metacognitive monitoring and arithmetic is warranted, and whether this association is specific to the arithmetical domain or whether it fits into a more general role of metacognitive monitoring in academic performance. Thereafter, I consider why including brain-imaging research techniques adds value to the existing literature on metacognitive monitoring and arithmetic in primary school children. The penultimate section focusses on several critical methodological aspects of the current dissertation. The introductory chapter ends with a disclosure of the concrete aims and the outline of this doctoral dissertation.

1 Why focus on arithmetic?

Mathematics is a complex, multidimensional skill that includes different domains (e.g. arithmetic, word problem solving, geometry, algebra), and within these domains, different skills. For instance, arithmetic comprises the use of different strategies (i.e. fact retrieval versus various procedural strategies) and includes different operations (e.g. addition, multiplication). As a result of the extensive research into mathematical performance and development, numerous processes have been identified as being associated with or predictive of mathematical performance (e.g. Berch et al., 2016; Campbell, 2005; Cappelletti & Fias, 2016; Geary, 1994). These include numerical magnitude processing (e.g. Schneider et al., 2017), executive functioning (e.g. Bull & Lee, 2014), metacognition (e.g. Schneider & Artelt, 2010), and mathematics anxiety (e.g. Mammarella et al., 2019).

Most studies in this research area have used general mathematical tests (De Smedt et al., 2013), which often yield a total score that reflects performance averaged across various mathematical domains. Furthermore, the content of these general tests may differ substantially depending on, for example, the specific test used, the age range for which it was designed, or the country in which a general test was administered. As a result, these general achievement tests differ in the mathematical skills they measure. The observed pattern of associations between these general tests and other (meta)cognitive and affective processes (such as numerical magnitude processing or executive functions) is thus determined by the set of mathematical skills included in these general mathematics tests. The use of total scores and the differences in mathematical content in these general tests make it functionally unclear how other (meta)cognitive and affective processes are involved in mathematical subdomains.

Importantly, as different mathematical subdomains are highly diverse, one could argue that not all aspects of mathematical performance are associated with these other processes or at least not to the same extent (e.g. Cragg & Gilmore, 2014, for a critical discussion). Indeed, the role of different processes in mathematical performance has been shown to change as a function of mathematical domain under investigation. For example, this has been demonstrated for the roles of numerical magnitude processing (Schneider et al., 2017) and working memory (Peng et al., 2016), but also for inhibition, as Gilmore and colleagues (2015) found different associations between inhibition and specific components of mathematics. Therefore, instead of using general mathematics achievement tests, which include diverse mathematical domains, investigating specific mathematical skills is essential to obtain a deeper, more nuanced understanding of the precise role of (meta)cognitive and affective processes associated with or predictive of mathematical performance and development.

In the current doctoral project, the focus lays on arithmetic as the mathematical domain of interest. The decision to focus my research specifically on arithmetic was driven by several characteristics of arithmetic that make it an essential subdomain for which it is critical to investigate (meta)cognitive and affective processes related to or predictive of performance and development. Firstly, arithmetic is a core element of primary school children's curriculum and thus key to their early academic success. Secondly, arithmetic represents a major building block for children's development of more complex mathematical abilities (e.g. Geary et al., 2012; Kilpatrick et al., 2001). Thirdly, difficulties in arithmetic have been considered to be the hallmark of children with dyscalculia (American Psychiatric Association, 2013). More broadly, as was discussed above, arithmetic is an important area of study because of the omnipresence of arithmetic in children's and adult's daily lives and the major impact of arithmetical ability and the ability to efficiently process numerical information later in life (e.g. Gerardi et al., 2013). Taken together, these characteristics of arithmetic performance and development make it crucial to understand the processes underlying individual differences in children's arithmetic performance. This understanding is educationally relevant because it might help develop effective instructional approaches and might contribute to designing scientifically validated remediation programs for children at risk for or with difficulties in arithmetic and mathematical performance in general.

2 Processes arithmetic counts on

A multitude of processes that are associated with or predictive of arithmetic performance and development have been identified in the vast body of existing research (e.g. Dowker, 2019c). These processes are not only cognitive in nature, but also include affective, educational, social, and cultural processes. For example, arithmetic relies on a good understanding and manipulation of numbers, e.g. the cognitive process numerical magnitude processing (Schneider et al., 2017), but is also influenced by affective processes, such as mathematics anxiety (e.g. Dowker, 2019e). Another example is the educational context in which arithmetic is taught (e.g. Campbell & Xue, 2001; Opdenakker & Van Damme, 2007), as different didactic approaches may influence arithmetic development, e.g. emphasis on rote memorization of arithmetic facts versus focus on conceptual understanding (Schoenfeld, 2004).

The current doctoral project, and by extension the existing research on arithmetic, does not exhaustively investigate all these important influencing processes in concert. Yet, it is essential to always keep in mind that performance and individual differences in arithmetic are influenced by a multitude of diverse interacting processes at multiple levels. As will be outlined in the remainder of this general introduction, the current project therefore investigates the interplay of diverse processes (i.e. cognitive, metacognitive and affective processes) and uses different methodological and analytical frameworks (i.e. behavioural methods and neuro-imaging methods).

In what follows, an overview will be given of processes that have been identified as important for arithmetic. First, a conceptual clarification of the use of the terms 'domain-general' and 'domain-specific' processes in this dissertation will be presented. Thereafter, the key processes of which their role in arithmetic was investigated within this dissertation are discussed, namely numerical magnitude processing, executive functions, metacognition, and mathematics anxiety.

2.1 Domain-general vs domain-specific processes

To successfully solve arithmetic problems, children need to be able to understand the problem, select an appropriate strategy for that problem and, eventually, perform that strategy accurately and efficiently. To do this, they not only need domain-specific mathematical knowledge and skills, i.e. processes specifically relevant for learning a particular academic skill, such as numerical magnitude processing. The involvement of other, more domain-general processes, i.e. processes relevant for learning various academic skills, such as executive functions, are also essential. Indeed, besides differentiating between categories of processes (e.g. cognitive, affective), research has also distinguished between domain-general processes and domain-specific processes (e.g. Geary & Moore, 2016; Vanbinst & De Smedt, 2016a).

One can argue, however, that using the labels 'domain-general' and 'domain-specific' might not be ideal, as they may sometimes be misleading. All processes may be general or specific to a certain extent, yet labelling them as 'domain-general' or 'domain-specific' may give the impression that a particular process is in fact either domain-general or domain-specific. While a strict distinction between the two is theoretically appealing, a nuanced use of these terms seems warranted. More specifically, the role of a domain-general process, for instance, may differ quantitatively and/or qualitatively depending on the domain it is measured in. For example, while inhibition may be an important process within both mathematics (e.g. Gilmore et al., 2015) and reading (e.g. Borella et al., 2010), it may play a stronger role in one of them, or it may be of importance in some aspects of, for instance, mathematics, and not (so much) in others (e.g. Gilmore et al., 2015). Moreover, inhibition skills in one domain may not necessarily be related to inhibition in another domain (e.g. Bellon et al., 2016) and consequently, a domain-general process may show some domain-specificity. On the other hand, a domain-specific process, while mostly important for a particular skill, may still play a role within other domains. For example, while phonological processing has been highlighted as an important domain-specific predictor for reading development (e.g. Melby-Lervåg et al., 2012), it has also been, to a much lesser extent, associated with arithmetic development (Vanbinst & De Smedt, 2016a), even after controlling for reading skills (Vanbinst et al., 2020). Hence, a domain-specific process may show some domain-generality. Against this background, I define domain-general and domain-specific processes in relative terms throughout this project. A domain-general process is thus conceptualised as a process that plays an important role in many different domains (e.g. executive functions), without clear indications that this process is a key factor in one domain, more than it is in other domains. On the other hand, a domain-specific process is defined as a process that is considered a key process for performance and development in a particular domain, more than it is in other domains. An example of such a domain-specific process for arithmetic is numerical magnitude processing.

Domain-general and domain-specific processes have been studied in relative isolation from each other within research on arithmetic (Fias et al., 2013). As a result, little attention has been paid to the joint effects of domain-general and domain-specific processes on arithmetic, even though it is not unlikely that these processes interact with regard to their association with arithmetic. For example, numerical magnitude processing performance itself may to some extent be influenced by domain-general processes, such as inhibition (e.g. Fuhs & Mcneil, 2013; Gilmore et al., 2013). On the other hand, the domain-specific factor mathematics anxiety might explain some variance in the association between the domain-general process metacognition and arithmetic (Morsanyi et al., 2019). Therefore, the current dissertation does not uniquely focus on either domain-general or domain-specific processes, but includes processes from both categories to assess their relative importance for arithmetic and its development. As a result, a more comprehensive image can be uncovered of how arithmetic is related to different cognitive, metacognitive and affective processes, both at a more domain-specific and at a more domain-general level. This will lead to a more thorough understanding of arithmetic performance and development and individual differences herein, which in turn can contribute to devising appropriate educational approaches and interventions.

In the remainder of this section, I will first focus on both domain-general and domain-specific *cognitive* processes, followed by a discussion of *metacognitive* processes and the domain-specific *affective* process of mathematics anxiety.

2.2 Numerical magnitude processing

Numbers lie at the heart of arithmetic. Not surprisingly, a large number of studies on arithmetic thus focused on numerical magnitude processing, as the core domain-specific process that correlates with and predicts (individual differences in) arithmetic performance and development. Numerical magnitude processing can be defined as children's elementary intuitions about quantity and the ability to understand the meaning of numbers, and its crucial role in arithmetic performance and development is highlighted in an extensive body of empirical research (e.g. Bugden & Ansari, 2011; De Smedt, Verschaffel, et al., 2009; Holloway & Ansari, 2009; Kolkman et al., 2013; Lonnemann et al., 2011; Vanbinst et al., 2018). This was shown in cross-sectional (e.g. Durand et al., 2005; Holloway & Ansari, 2009; Vanbinst et al., 2012) and longitudinal research (e.g. Kolkman et al., 2013; Vanbinst et al., 2019) and to some extent in experimental studies (e.g. Booth & Siegler, 2008; but see Merkley et al., 2017 and Szűcs & Myers, 2017). Integrating results from numerous studies on numerical magnitude processing, both De Smedt et al. (2013, for a narrative review) and Schneider et al. (2017, for a meta-analysis) emphasized the important role of numerical magnitude processing in arithmetic. In line with this, in a review aimed at integrating various phenomena in numerical development into a unified framework or integrated theory, Siegler (2016) proposed the generation of increasingly precise magnitude representations for an increasingly broad range of numbers as the common core of numerical development.

In research on individual differences in mathematics performance, numerical magnitude processing is often measured using either nonsymbolic (i.e. dot patterns; e.g. Inglis et al., 2011) or symbolic (i.e. Arabic numerals; e.g. Holloway & Ansari, 2009) magnitude comparison tasks (see De Smedt et al., 2013). In these tasks, participants are asked to indicate which of two presented numerical magnitudes is the numerically larger one. There is converging evidence that fast and accurate performance on these comparison tasks coincides with higher arithmetic achievement (De Smedt et al., 2013; Schneider et al., 2017). Typically, performance on numerical magnitude comparison tasks is operationalised as overall accuracy, response time, ratio or distance effects (i.e. a subtraction of reaction times for comparisons with large versus small ratios or numerical distances; Holloway & Ansari, 2009), or the Weber fraction (i.e. the minimal amount of change in magnitude that is needed to detect a difference; Halberda et al., 2008). These measures capture different aspects of participants' performance and are not interchangeable (De Smedt et al., 2013). Importantly, particularly when operationalised as response time, performance on the symbolic comparison task has been found to be robustly and significantly correlated with concurrent and future mathematics achievement (see De Smedt et al., 2013). Accuracy, for example, can yield ceiling effects, especially from primary school onwards, and might lead to an underestimation of the relation between numerical magnitude processing and arithmetic competence

(Holloway & Ansari, 2009). Hence, in this dissertation, response time was chosen as the key performance measure of numerical magnitude processing skills.

The relations between the processing of non-symbolic and symbolic numerical magnitudes and their development constitute one of the most debated topics in the field of numerical cognition (Schneider et al., 2017), yet the existing body of evidence converges to the conclusion that symbolic abilities are the most critical for the development of mathematics (e.g. Merkley & Ansari, 2016). Indeed, comparing the role of nonsymbolic and symbolic numerical magnitude processing in mathematics, it has been consistently found in the existing literature that symbolic numerical magnitude processing has a more important role in mathematics compared to nonsymbolic numerical magnitude processing. For example, De Smedt and colleagues (2013) revealed in their review of the existent literature that results for symbolic numerical magnitude processing were consistent and robust across studies and populations, while for nonsymbolic numerical magnitude processing, many conflicting findings have been reported. In line with this, Schneider and colleagues (2017) found in their meta-analysis that especially symbolic numerical magnitude processing was important for mathematical achievement, with a smaller role played by nonsymbolic numerical magnitude processing. Hence, within this doctoral project, only symbolic numerical magnitude processing was included, operationalised with a symbolic numerical magnitude comparison task.

Different underlying mechanisms might be at play in the association between symbolic numerical magnitude processing and arithmetic. A possible driving force might be that to succeed in nearly every form of arithmetic, proficiency with Arabic symbols is required (Gilmore et al., 2018a). Furthermore, the majority of measures of arithmetic performance require the interpretation and transformation of information presented in symbolic form (i.e. Arabic numerals). As was suggested by Nosworthy et al. (2013), the unique variance in mathematical performance accounted for by symbolic processing may be related to recognizing numerals and mapping numerals to magnitudes, which are crucial skills in the mental manipulation of digits during arithmetic. In line with this, in mental arithmetic, one needs to mentally represent the magnitudes of the two operands and the magnitude of the answer (DeStefano & LeFevre, 2004). Another underlying mechanism might be that proficient numerical magnitude processing skills induce the transition to more efficient strategies (Booth & Siegler, 2008; Vanbinst et al., 2012). For example, a counting-on-larger strategy (Siegler, 1996; e.g. 2 + 5 = 5, 6, 7), which is a very common strategy in early arithmetic, requires a decision on which of the operands contains the larger number, and therefore this strategy draws on the understanding of numerical magnitudes. As such, proficient numerical magnitude processing may also be helpful for using an efficient arithmetic strategy. Another example is in the use of the so-called subtraction-by-addition strategy (e.g. Peters et al., 2014), in which one determines how much needs to be added to the subtrahend to get to the minuend (e.g. solving '72 – 64 = .' by '64 + 6 = 70 and 70 + 2 = 72, so the answer is 6 + 2 = 8'). Consequently, this strategy requires a comparison of the magnitude of the numbers in the problem (Linsen et al., 2015), and is considered to be particularly efficient on problems with a small numerical difference between minuend and subtrahend. Additionally, arithmetic facts may be stored in long-term memory in a meaningful way, i.e. according to their magnitude (e.g. Booth & Siegler, 2008; Butterworth et al., 2001; Robinson et al., 2002). More recently, evidence was found for a bidirectional association between symbolic numerical magnitude processing and arithmetic. This emphasizes that early arithmetic also predicts later symbolic numerical magnitude skills, suggesting that arithmetic development strengthens children's ability to process the numerical meaning of Arabic digits (Vanbinst et al., 2019).

Already early in childhood, numerical magnitude processing skills are present, can be reliably measured and are significantly correlated with mathematics performance (De Smedt et al., 2013; Schneider et al., 2017). Both cross-sectional (e.g. Nosworthy et al., 2013) and longitudinal studies (e.g. Matejko & Ansari, 2016; Vanbinst et al., 2018) have indicated that, in (early) primary school, numerical magnitude processing skills develop substantially (e.g. Holloway & Ansari, 2009; Sekuler & Mierkiewicz, 1977) as children become faster and more accurate at numerical magnitude processing tasks, especially on symbolic tasks. Concerning potential underlying developmental mechanisms, it has been suggested that number symbols become linked to the nonsymbolic system, the ability to rapidly and approximately estimate and compare nonsymbolic numerical quantities (i.e. the Approximate Number system, ANS). This is thought to happen through a process of mapping number symbols onto their corresponding nonsymbolic quantities (Mundy & Gilmore, 2009). Yet, more recently, a growing body of evidence questioned the direction of this underlying mechanism (e.g. Lyons et al., 2012; Matejko & Ansari, 2016). As mentioned above, there is growing evidence on bidirectional developmental dynamics between numerical magnitude processing and arithmetic, such that numerical magnitude processing skills are also strengthened in children through their arithmetic development across primary education (Vanbinst et al., 2019).

While the current body of research on numerical magnitude processing in mathematics has provided valuable insights, its narrow focus has been criticized (see Fias et al., 2013). In their meta-analysis, Schneider and colleagues (2017) demonstrated that the overall correlation between numerical magnitude processing and mathematical competence is estimated to be r = .278. For mental arithmetic in particular, the correlation is a little higher (i.e. r = .378). These correlations suggest that numerical magnitude processing skills explain a significant but only small portion of the variance in mathematical abilities. Consequently, by narrowing the scope to numerical processing, existing research has ignored other processes that might play a role in arithmetic performance and development. Moreover, it remains to be seen whether and how the association between arithmetic and numerical magnitude processing is impacted by other (meta)cognitive processes. For example, it is not fully clear whether the association between arithmetic and numerical magnitude processing remains after taking the role of these other processes into account. This issue was tackled in the current dissertation by including various processes in concert to investigate their unique associations with arithmetic in addition to each other (see below).

2.3 Executive functions

For a long time, executive functions have been identified to play an important role in mathematics performance and development (see Bull & Lee, 2014; Cragg & Gilmore, 2014; Raghubar et al., 2010, for reviews). Executive functions refer to a family of top-down mental processes required to concentrate and pay attention (Diamond, 2013). In other words, executive functions are processes that allow us to respond flexibly to our environment and to engage in deliberate, goal-directed thought and action (Cragg & Gilmore, 2014). This is particularly important in new situations and activities, which are a key characteristic of academic learning and thus essential for arithmetic in primary school children.

Executive functioning is multi-faceted and mainly consist of the processes of inhibition, shifting and updating (e.g. Baggetta & Alexander, 2016; Miyake et al., 2000). First, inhibition refers to one’s ability to control one’s attention, behaviour and thoughts, in order to override a strong internal predisposition or external lure and instead do what is more appropriate or necessary (Diamond, 2013). Two types of inhibition can be distinguished, namely response inhibition (i.e. the ability to control one’s behaviour and one’s emotions in order to control behaviour, in order to resist temptations, and to not act impulsively) and interference control or cognitive inhibition (i.e. the ability to selectively attend to or focus on what we choose and suppress attention to other stimuli). Second, shifting is defined as the disengagement of an irrelevant task set or strategy, and the subsequent initiation of a new, more appropriate set (van der Sluis et al., 2007). Third, updating, or the central executive component of working memory, involves holding information in memory and flexibly manipulating it (Baddeley & Hitch, 1994).

Importantly, there is both unity and diversity in executive functions (e.g. Baggetta & Alexander, 2016; Diamond, 2013; Miyake et al., 2000). For example, in their review, Friedman and Miyake (2017) found that, when measured with latent variables, executive functions are correlated but separable. In line with this, behavioural studies incorporating batteries of widely used executive function tasks found low or non-significant correlations between these tasks, or, using factor analysis, yield multiple factors (e.g. Baggetta & Alexander, 2016; Brocki & Bohlin, 2004; Bull & Scerif, 2001; Carlson et al., 2013; Friedman & Miyake, 2017; Huizinga et al., 2006; Lee et al., 2012; Miyake et al., 2000; Van der Ven et al., 2012). Furthermore, neuroimaging studies provide evidence in support of the multi-componential nature of executive functions: Executive functions activate both common and specific neurobiological areas (Friedman & Miyake, 2017) and different components of executive functions are seen to rely on different parts of the prefrontal cortex (e.g. Aron et al., 2004, 2014; Crone et al., 2006; Crone & Steinbeis, 2017; Narayanan et al., 2005).

Executive functions emerge during the first few years of life and from then onwards, throughout primary school and into adolescence, major advances in executive functioning occur (e.g. Best & Miller, 2010; Carlson et al., 2013; Diamond, 2013; Huizinga et al., 2006). These are partially attributed to brain maturation in the prefrontal cortex (e.g. Carlson et al., 2013; Diamond, 2002). The pronounced development of executive functioning in primary school is reflected in, for example, improvements in accuracy of performance on executive function tasks. Different components of executive functions vary somewhat in their developmental trajectories (Best & Miller, 2010). Research in young children provides evidence for dissociable, but interrelated executive functioning components, already in early childhood (e.g. Lee et al., 2013).

Executive functions are recognized as important in general academic achievement, in different domains such as reading, writing, science and mathematics (e.g. Best et al., 2011; St Clair-Thompson & Gathercole, 2006). Some studies indicated that the role of executive functions in mathematics seems to be particularly important compared to their role in other domains (Geary, 2011; Willoughby et al., 2012). Research on the associations between executive functions and mathematics has yielded important insights into mathematical development and the evidence for the importance of executive functions in mathematics performance stems from research using different designs (Gilmore et al., 2018a). One source of evidence comes from correlational or longitudinal studies, in which individual differences in executive functions are related to individual differences in concurrent or future mathematics performance. Another source of evidence stems from experimental studies. For example, in dual-task studies, participants perform tasks that tax their executive functions (e.g. loading working memory) while at the same time they perform arithmetical tasks, in order to investigate the effect of this taxation on arithmetic performance (e.g. De Rammelaere et al., 2001). In this diverse body of research, executive functions have been found to be correlated with and predictive of mathematical performance and development (e.g. Bull & Lee, 2014; Cragg & Gilmore, 2014; Friso-Van Den Bos et al., 2013), indicating that better executive functioning skills are associated with better mathematics skills.

Yet, in the existing body of research on the role of executive functions in arithmetic, different results are found depending on the component(s) of executive functions that are investigated. The majority of research on executive functions in arithmetic has focused on working memory or updating. Across these studies, this process has consistently been found to be a strong predictor of arithmetic performance (e.g. for a meta-analysis see Peng et al., 2016). Although more recently the interest is growing, historically, less attention has been devoted to the role of inhibition and shifting in arithmetic, and the evidence on their role in arithmetic is more equivocal. Some studies reported significant associations between inhibition or shifting and mathematics, while others found no relationship between the two (see Bull & Lee, 2014; Cragg & Gilmore, 2014; Friso-Van Den Bos et al., 2013, for reviews). These difference in results could be due to differences in, for example, the aspect of inhibition that was measured (e.g. response inhibition versus cognitive inhibition). Indeed, inhibition is a multi-faceted skill (Diamond, 2013; Huizinga et al., 2006), and some forms of inhibition may be more involved in arithmetic than others, consequently resulting in different findings depending on the aspect of inhibition that was measured. Furthermore, differences could be due to the executive function tasks used and the stimuli within the task (e.g. numerical versus non-numerical stimuli). For example, some evidence suggests that the relationship between inhibition and mathematics achievement is stronger when the inhibition task involves numerical rather than non-numerical stimuli (Cragg et al., 2017). Differences in the findings on the role of inhibition and shifting in mathematics could also be due to whether they were studied in relative isolation, or whether the unique contribution of these processes was considered on top of other processes such as updating or IQ. For example, Van der Ven et al. (2012) found that inhibition and shifting did not predict mathematics when updating was also considered. Another important point on which studies differ that likely influenced the results, is the mathematics task considered. Inhibition and shifting skills may have different levels of involvement in different aspects of mathematics.

In the existing literature, numerous measures of executive functions have been used (see Baggetta & Alexander, 2016, for a review). Inhibition is often measured experimentally, using tasks in which participants have to respond to certain features of a task and ignore others. Shifting is most commonly measured using sorting or classification tasks. Updating measures usually ask participants to recall information while simultaneously processing or manipulating this information. A pivotal point in the measurement of executive functions is the issue of task impurity (e.g. Bull & Lee, 2014; Miyake et al., 2000), which concerns the problem that a single task assess different components (e.g. both executive and non-executive function processes) making a pure interpretation of what the task measures troublesome. In research on executive functions, this is particularly problematic, because executive functions are higher-order processes that manifest themselves in operating on other processes. As such, tasks measuring executive functions insurmountably also tap into other (cognitive) processes that are not necessarily relevant to the targeted executive functions. To minimize interpretation difficulties concerning the importance of executive functions in arithmetic versus the importance of numerical processing, none of the executive functioning tasks in this dissertation included numerical stimuli. For inhibition, this dissertation specifically focused on measures of cognitive inhibition or interference control, because cognitive inhibition rather than response inhibition is likely to play a role in arithmetic. An example of this is when one has to inhibit competing answers while retrieving arithmetic facts (e.g. Verguts & Fias, 2005).

The association between executive functions and arithmetic is theoretically appealing. For example, cognitive inhibition (i.e. interference control) might play a role in arithmetic, as arithmetic facts may be stored in an associative network in semantic memory (e.g. Verguts & Fias, 2005). Therefore, they might be particularly prone to interference because of the number of features they share (De Visscher & Noël, 2014b). Hence, incorrect but competing answers have to be inhibited during arithmetic. This is exemplified in the types of errors typically seen when children retrieve arithmetic answers, e.g. retrieving the answer to the corresponding addition item instead of the multiplication item, such as responding ‘5’ to ‘3 × 2’. On the other hand, good executive functions might help to suppress inefficient arithmetic strategies and switch to more efficient ones (e.g. using a decomposition of operands strategy instead of finger counting), or shift between different arithmetic operations. Good updating skills might assist in keeping relevant information in mind during problem-solving, keeping intermediate solutions in mind and manipulating them while calculating, and may also be required to store and access information (e.g. arithmetic facts) in long-term memory.

Because of the multi-componential nature of executive functions (e.g. Friedman & Miyake, 2017; Lee et al., 2013; Miyake et al., 2000), because of the evidence for the importance of all three executive functioning skills for arithmetic (Gilmore et al., 2018b), and because these different components of executive functions are differently related to arithmetic (e.g. Bull & Lee, 2014; Cragg et al., 2017), it is of utmost importance to include all three aspects of executive functioning when examining their role in other processes such as arithmetic. In the current dissertation, all three well-known components, namely inhibition, shifting and updating, were therefore considered.

2.4 Metacognition

Another important domain-general process that has received less attention in research on individual differences in arithmetic performance, yet has been intensively studied in the field of mathematics education (e.g. Schneider & Artelt, 2010), is metacognition. Various conceptualisations of metacognition have been used in the existing literature. A broad definition of metacognition that has been widely used is ‘thinking about your thinking’ or “*any knowledge or cognitive activity that takes as its object, or regulates any aspect of any cognitive enterprise*” (Schneider, 2015b, p. 282).

The idea that accurate self-knowledge is meaningful and is something to strive for, has captivated thinkers since Socrates. The term *metacognition* was first introduced by Flavell (1979) and is often defined as a broader concept including, on the one hand, declarative metacognitive knowledge and, on the other hand, procedural metacognition. Metacognitive knowledge is defined as knowledge about cognition and learning (e.g. Brown, 1978) or “*knowledge or beliefs about what factors or variables act and interact in what ways to affect the course and outcome of cognitive enterprises*” (Flavell, 1979, p. 907). It includes factual knowledge about the importance of person, task and strategy variables for processing and recalling information (Flavell, 1979; Schneider, 2015b). Procedural metacognition is a collection of self-reflecting, higher-order cognitive processes, in other words, how people monitor and control their cognition on-task during ongoing cognitive processes (Flavell, 1999; Nelson & Narens, 1990). Because both aspects of metacognition have been shown to be important for mathematics, declarative metacognitive knowledge as well as procedural metacognition were included in the current dissertation.

The conceptualisation of procedural metacognition in the current dissertation is in accordance to the theoretical framework by Nelson and Narens (1990) on procedural metacognition (see Figure 1.1). This model posits that procedural metacognition encompasses two aspects, on the one hand metacognitive monitoring (top panel of Figure 1.1) and on the other hand metacognitive control (bottom panel of

Figure 1.1). Metacognitive monitoring is defined as the subjective self-assessment of how well a (cognitive) task will be/is/has been performed (Nelson & Narens, 1990). It involves judging the success and/or progress of cognitive processing (Chua et al., 2014). Metacognitive control, on the other hand, is an action-oriented component of procedural metacognition and is defined as the individual's executive activities enabling the use and adaptation of different cognitive operations with the aim to increase learning behavior or test performance (Roebers et al., 2014). Metacognitive control allows for the direction of behaviour such as strategy selection, information gathering or correction of given responses. Furthermore, the model by Nelson and Narens (1990) indicates that different kinds of monitoring and control occur during different stages of cognitive performance (e.g. acquisition, retrieval; middle panel of Figure 1.1). For example, in monitoring, judgments of learning (JOL) are predictive judgments provided during or shortly after a study phase and judge how likely someone will remember the studied information later on. Retrospective confidence judgments, on the other hand, are postdictive judgments made after performance to indicate how certain people are that their response is correct (Nelson & Narens, 1990).

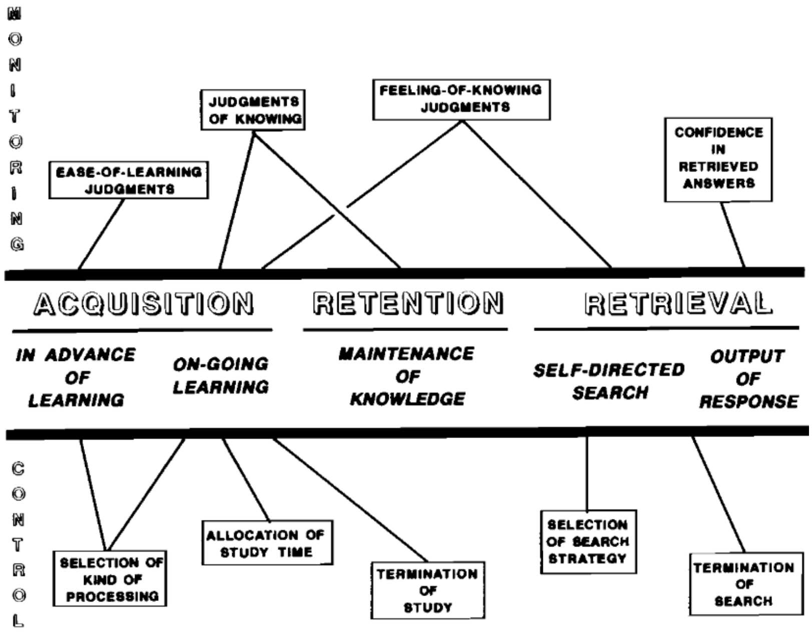

Figure 1.1. Theoretical framework of Nelson and Narens (1990, p. 129).

Because of the broadness of the concept of metacognition, including several components of metacognitive knowledge (e.g. person, task, and strategy category; Flavell, 1979) and different kinds of monitoring and control during different stages of performance (see model Nelson & Narens, 1990, Figure 1.1), the current dissertation attended to specific aspects of metacognition. Regarding metacognitive knowledge, I focused on declarative metacognitive strategy knowledge, i.e. knowledge on which strategies are likely to be effective in achieving goals in which sorts of cognitive undertakings (Flavell, 1979). The rationale behind this choice was the importance of good strategic knowledge for arithmetic performance and development (e.g. Verschaffel et al., 2007). In line with the majority of research on procedural metacognition, I focused on the metacognitive monitoring component throughout the current dissertation. Specifically, I examined retrospective metacognitive monitoring judgments on accuracy of performance, which indicates whether a child beliefs his/her answer on an item is correct or wrong. This focus on retrospective monitoring of accuracy was driven by the rationale that a good understanding of the relation between monitoring of accuracy and arithmetic can provide a foundation for intervention and/or training to enhance arithmetic accuracy. For example, it has been shown that giving feedback after performance to improve this performance is especially helpful for low confidence responses (Butler et al., 2008). In the context of arithmetic in primary school children this might provide a promising principle to utilize in interventions.

From an early age, most children are able to theorize about their own cognition (e.g. Schraw & Moshman, 1995). For example, Destan and colleagues (2014) found evidence for robust metacognitive monitoring skills in 5-year-old children. Throughout primary school, metacognitive knowledge and skills develop substantially (e.g. Lyons & Ghetti, 2010; Roebers & Spiess, 2017; Schneider, 2008, 2010), resulting in better general metacognitive knowledge (e.g. Schneider & Löffler, 2016) and improved monitoring and control skills (e.g. Garrett et al., 2006; Schneider & Lockl, 2008). For example, in primary school, metacognitive monitoring accuracy is found to increase (e.g. Ghetti, 2008; Lyons & Ghetti, 2010; Schneider, 2008, 2010; Schneider & Lockl, 2008; Schneider & Löffler, 2016). As metacognition encompasses different aspects, it is not surprising that these different aspects of metacognition follow different developmental paths (Schneider & Löffler, 2016). Compared to more clear-cut age trends for declarative metacognitive knowledge, the evidence regarding metacognitive monitoring is less univocal (Schneider, 2015b), yet, also in metacognitive monitoring critical development is observed during early to late childhood (e.g. Geurten et al., 2018; Lyons & Ghetti, 2010). Importantly, it has been suggested that a gradual shift toward domain-general metacognition occurs in children aged between 8 and 13, and that metacognition is no more bound by task content and domain knowledge after the age of 10 (Geurten et al., 2018), a possibility that will be investigated in the context of the current dissertation. Over its extended course of development during which metacognition becomes increasingly under an individual’s conscious control, metacognition becomes more explicit, powerful and effective. This metacognitive development is crucial, as age-related improvements in children’s ability to monitor and regulate their mental operations are widely recognized to be a driving force in cognitive development, underlying age-related improvements in accuracy on a wide variety of tasks (Lyons & Ghetti, 2010).

Already in 1979, in his pivotal paper, Flavell (1979) posed that metacognitive knowledge and skills have an important effect on cognitive performance. Indeed, metacognition has been regarded as a fundamental skill influencing cognitive performance and learning in diverse domains such as mathematics, memory, reading, and perception (e.g. Annevirta et al., 2007; Block & Peskowitz, 1990; Efklides & Misailidi, 2010; Freeman et al., 2017; Kuhn, 2000; Lyons & Ghetti, 2013; Özsoy, 2011; Rinne & Mazzocco, 2014; Schneider, 1998; Schneider & Artelt, 2010; Schraw et al., 2006; van Kraayenoord & Schneider, 1999; Veenman et al., 2004, 2006). The importance of metacognition in different (cognitive) domains is not surprising: Metacognitive aspects, such as knowing the limits of your own knowledge and being able to regulate that knowledge, are essential components of self-regulated and successful learning (Schraw et al., 2006), enabling learners to improve their cognitive achievements. For example, good metacognition allows learners to correctly allocate study-time, check answers when they feel unsure about the correctness of the answer or provide a learning moment when an error is detected. Moreover, there is consensus that one way in which parents and teachers can facilitate cognitive development is by the development of children's metacognition (Schneider, 2015b). Specifically investigating the role of metacognition in arithmetic thus seems a promising avenue.

Although there is a long tradition of research investigating metacognition in mathematics education (e.g. De Corte et al., 2000; Schneider & Artelt, 2010; Schoenfeld, 1992), in comparison, much less emphasis is placed on the role of metacognition in the field of mathematical cognition. This is exemplified in the fact that influential books on numerical and mathematical cognition make little mention of metacognition (e.g. Campbell, 2005; Dowker, 2019c; Geary, 1994; Gilmore et al., 2018a; Henik & Fias, 2018). Even the integrative work of Cohen Kadosh and Dowker (2015) on numerical cognition only includes very few entries on metacognition. Yet, studies on the relation between metacognition and mathematics have shown that successful mathematics performance depends not only on having adequate knowledge, but also on sufficient awareness, monitoring and control of that knowledge (e.g. Carr et al., 1994; Carr & Jessup, 1995; Freeman et al., 2017; Garofalo & Lester, 1985; Lucangeli & Cornoldi, 1997; Özsoy, 2011; Rinne & Mazzocco, 2014; Schoenfeld, 1992; Stillman & Mevarech, 2010; van der Stel et al., 2010). Moreover, very few studies specifically focused on subdomains of mathematics, such as arithmetic, to functionally specify the role of metacognition in mathematics. The current dissertation specifically placed itself within this gap in the literature by adding to the mathematical cognition literature through the specific investigation of metacognition in arithmetic performance.

There are many different ways to measure metacognition (e.g. Fleming & Lau, 2014; Lingel et al., 2019; Veenman & Van Cleef, 2019). These differences go back to, for example, the level at which metacognition is measured (e.g. global, overall judgment of performance in a whole task versus trial-by-trial measures), the source of information (e.g. observations versus self-reports), or whether metacognition is measured during performance or not (i.e. offline versus online measures). The current dissertation included different measures of metacognition, with different characteristics (e.g. both online and offline measures). In line with the current body of research (Schneider, 2015b), our measure assessing metacognitive knowledge was taken without concurrent assessment of cognitive performance (i.e. offline) using a questionnaire (Haberkorn et al., 2014), while our measure of metacognitive monitoring was collected simultaneously (i.e. online) with the measure of cognitive activity (i.e. arithmetic). To measure metacognitive monitoring I used a calibration measure (e.g. Lingel et al., 2019; Rinne & Mazzocco, 2014). This choice of measure for metacognitive monitoring is in line with the foundational model of Nelson an Narens (1990), which states that the main methodological tool for generating data about metacognitive monitoring consists of the person's subjective reports about his/her introspection. In turn, this subjective judgment is then compared to the objective performance in order to index metacognitive monitoring of accuracy. It is important to note that a large portion of research on the association between procedural metacognition and mathematics has focused on metacognitive strategy selection (e.g. Carr et al., 1994; Dowker, 2019c; Geurten & Lemaire, 2017). In this dissertation, I tackle the question on the association between metacognitive monitoring and mathematics from a novel, very specific point of view: I focused on performance outcomes of arithmetic (i.e. fluency, response time, accuracy) instead of general mathematics and (arithmetic) strategy use, and I investigated metacognitive monitoring of arithmetic accuracy. This specific focus allowed to further functionally unravel the relation between metacognition and mathematics.

The few existing studies on metacognition in arithmetic have highlighted that the successful appraisal of the accuracy of one's arithmetic answer is a powerful predictor of arithmetic performance in primary school children (Rinne & Mazzocco, 2014). Building on this work, the current dissertation tackled important gaps in the body of research on metacognition in arithmetic by investigating the unique role of different aspects of metacognition in arithmetic, in addition to other important cognitive and affective processes, in different age groups, both cross-sectionally and longitudinally, and examining it both at the behavioural and neurobiological level.

2.5 Mathematics anxiety

Development of expertise in a particular field, for example arithmetic, depends not only on cognitive skills, but also on affective processes (e.g. Batchelor et al., 2019). Such affective processes make a very important contribution to arithmetical performance (Dowker, 2019c), which is exemplified in the extensive body of research that has focused on a particular domain-specific affective process, namely mathematics anxiety. Mathematics anxiety, or "*a feeling of tension and anxiety that interferes with the manipulation of numbers and solving of mathematical problems in ordinary life and academic situations*" (Richardson & Suinn, 1972, p. 551), is systematically found to be moderately negatively related to mathematics performance in general (e.g. see meta-analyses Hembree, 1990; Ma, 1999; Namkung et al., 2019), and to arithmetic in particular (e.g. Ashcraft et al., 1998; Harari et al., 2013; Hunt et al., 2017; Sorvo et al., 2017). In line with the idea that, in general, younger children tend to have positive attitudes towards mathematics (Dowker, 2019d) and that mathematics anxiety only arises in the context of complex mathematics (e.g. algebra; Maloney & Beilock, 2012), the initial focus in research on mathematics anxiety was mainly on (young) adults (see Ashcraft et al., 2007, for a brief overview). This is also exemplified in the initial measures for mathematics anxiety, which were developed for adults (e.g. Richardson & Suinn, 1972). While much of the research on mathematics anxiety has been done in secondary school children and adults (Dowker, 2019e), more recent studies have also focused on young children (see Batchelor et al., 2019; Dowker, 2019e; Namkung et al., 2019, for discussions on this topic). This research indicates that mathematics anxiety is already reported and demonstrated in young children (e.g. Harari et al., 2013; Ma & Kishor, 1997; Ramirez et al., 2013; Vukovic et al., 2013). Rossnan (2006) suggested that mathematics anxiety can develop at any age and that its development is often linked to a child's first experiences with mathematics. The first years of formal mathematics instruction are thus of crucial importance in the development of mathematics anxiety. While some failed to find significant correlations in primary-school children (e.g. Krinzinger et al., 2009), most studies in children found a significant negative association between mathematics anxiety and arithmetic performance (e.g. Gierl & Bisanz, 1995; Harari et al., 2013; Krinzinger et al., 2009; Ma & Kishor, 1997; Petronzi et al., 2019; Ramirez et al., 2013; Vukovic et al., 2013; Wu et al., 2012).

The first clear example of an empirical study into mathematics anxiety was a research paper by Dreger and Aiken (1957) in which they referred to 'number anxiety' to label the emotional reaction to numbers and mathematics. They found it to be a separate construct that is dissociable from more general anxiety, unrelated to general intelligence, and negatively correlated with mathematics grades. These findings have been replicated many times. Indeed, although mathematics anxiety and general anxiety or test anxiety are correlated (e.g. Ashcraft & Moore, 2009; Devine et al., 2012; Park et al., 2014), mathematics anxiety overlaps only to a degree with other anxiety measures (e.g. Ashcraft et al., 2007; Ashcraft & Faust, 1994; Dew et al., 1984; Hembree, 1990) and mathematics anxiety is also found in individuals without high general anxiety or test anxiety (e.g. Passolunghi et al., 2016). Moreover, in a meta-analysis, Hembree (1990) found that mathematics anxiety is only weakly related to general cognitive skills, showing that individuals with high, average and low IQ can experience mathematics anxiety. When comparing anxiety about mathematics versus anxiety in other domains in 9-year-olds, Punaro and Reeve (2012) found anxiety for difficult problems in both mathematics and literacy. Yet, children expressed more anxiety about mathematics and this anxiety about mathematics correlated negatively with mathematics performance, whereas this was not the case for literacy anxiety and literacy performance. Hence, although children may have general anxiety and dislike of academic achievements or performance/test anxiety, mathematics is generally considered to elicit more fear than most other academic subjects (Dowker, 2019c). It is thus essential to examine the role of this domain-specific affective factor in arithmetic.

Importantly, mathematics anxiety is not a unidimensional construct: It is often conceptualised as encompassing two distinct dimensions, namely cognitive mathematics anxiety and affective mathematics anxiety (e.g. Dowker, 2019d; Ho et al., 2000; Wigfield & Meece, 1988). The cognitive component (‘worry’, ‘performance anxiety’) consists of self-deprecatory thoughts about one’s performance; the affective component (‘emotionality’) includes feelings of nervousness, fear, discomfort, tension and unpleasant physiological reactions to mathematics and/or the presence of mathematical stimuli (e.g. Dowker, 2019d; Ho et al., 2000; Wigfield & Meece, 1988). In general, the affective component is most strongly and consistently related to mathematics performance (e.g. Harari et al., 2013; Ho et al., 2000; Sorvo et al., 2017; Vukovic et al., 2013; Wigfield & Meece, 1988; Wu et al., 2012). Therefore, in this dissertation, I used a measure of mathematics anxiety that predominantly taps into this affective component of mathematics anxiety.

Mathematics anxiety is mostly measured using self-report questionnaires or rating scales in which respondents indicate how anxious they would feel (from “not anxious at all” to “very anxious”) in situations ranging from formal math settings to informal, everyday situations. The most widely used test of mathematics anxiety is probably the Mathematics Anxiety Research Scale (MARS; Richardson & Suinn, 1972) and adapted (e.g. abbreviated) versions of it. In this dissertation, I used a developmentally appropriate, adapted measure of this questionnaire.

There are several plausible reasons why mathematics anxiety is related to arithmetic performance. A first possibility is that mathematics anxiety leads to worse mathematics performance, as is explained in the Debilitating Anxiety Model (Carey et al., 2016) and the Cognitive Interference Theory (Namkung et al., 2019). For example, mathematics anxiety may lead to avoidance of mathematics. On the one hand, this could result in a local effect on the individual’s performance. This may be exemplified in rushing to finish the task to minimize the time doing arithmetic, even at the cost of accuracy (e.g. Ashcraft et al., 2007; Ashcraft & Faust, 1994; Dowker, 2019e) or in not providing an answer to an arithmetic question because math anxious people may see this as a better option than providing an entirely wrong answer (i.e. a no-attempt error; Chinn, 2012). On the other hand, there could be a more global avoidance effect, e.g. to avoid mathematics-related activities, at the cost of grasping learning opportunities (Ashcraft, 2002). Mathematics anxiety may also lead to rumination and procuring thoughts that occupy working memory resources which could otherwise be used for arithmetic (i.e. ‘Disruption Account’, Ramirez et al., 2018). This impact of mathematics anxiety on working memory is particularly problematic in arithmetic given the important role of working memory in arithmetic procedures (Ashcraft & Moore, 2009), which require the temporary storage of problem information and/or interim solutions and keeping track of counts, and thus rely on working memory. The influence of mathematics anxiety on performance is also known to increase with time pressure when completing mathematics tasks (Ashcraft & Moore, 2009).

A second possibility is that arithmetic performance deficits lead to anxiety, as is explained in the Deficit Theory (Carey et al., 2016; Namkung et al., 2019): When confronted with mathematics problems, children with low mathematics performance activate recollections of previous poor performance, which generates mathematics anxiety in the current situation (e.g. Carey et al., 2016;

Dowker et al., 2016; Hembree, 1990; Ma & Xu, 2004; Maloney et al., 2015; Sorvo et al., 2019). This may explain why levels of mathematics anxiety and their association with performance usually increase with age (Ma & Kishor, 1997). Indeed, one possible underlying mechanism therein might be that repeated experiences of failure in mathematics together with a growing awareness of this failure might lead to (more) math anxiety.

More recently, a third possibility has been coined, expressing the association between mathematics anxiety and performance as bidirectional (e.g. Dowker, 2019e; Gunderson et al., 2018), i.e. a Reciprocal Theory (Carey et al., 2016). This theory suggests that poor mathematics performance can trigger mathematics anxiety and mathematics anxiety can reduce performance, which could lead to a vicious circle (e.g. higher mathematics anxiety can lead to worse arithmetic performance, which in turn increases anxiety).

Taken together, the current, extensive body of research demonstrated the importance of mathematics anxiety for mathematics performance. Yet, in line with the concluding remarks by Mammarella and colleagues (2019) in their book on "What is known and what is still to be understood" on the topic of mathematics anxiety, there is a lack of longitudinal investigations into the directionality and development of this association. This is especially the case in young primary school children. Furthermore, there is a lack of studies that simultaneously include other processes that potentially interfere in this association (Mammarella et al., 2019), such as metacognition (Tobias, 1986). Therefore, the current dissertation examined mathematics anxiety longitudinally in young primary school children, taking into account other, potentially relevant (meta)cognitive processes, in order to further our understanding of the processes underlying arithmetic performance.

3 Investigation of processes in concert

As outlined above, a multitude of processes have an impact on arithmetic performance. Dowker firmly stated in a recent influential book on individual differences in arithmetic (Dowker, 2019c, p. 4): "*Any statement that arithmetic ability is purely the product of a single factor is oversimplified*". Yet, the vast majority of the abovementioned studies on the associations between these cognitive, metacognitive and affective processes and arithmetic have largely based their conclusions on isolated bodies of research without studying multiple processes in concert.

It is essential to simultaneously examine different processes which, when investigated in isolation, have been identified as important for arithmetic performance and development. Such simultaneous consideration of processes allows for the investigation of their unique contribution to performance in addition to each other. For example, in contrast to an extensive body of research on the role of numerical magnitude processing and executive functions in mathematics, much less is known about whether, on the one hand, executive functions continue to predict mathematics skills after taking numerical magnitude processing into account, and, on the other hand, to what extent numerical magnitude processing itself and its association with arithmetic is affected by more domain-general processes such as executive functions or metacognition.

Simultaneous investigation of different processes is especially important when it has been shown that these processes are interrelated. In the existing body of research, associations between some of the key processes that are considered in the current dissertation have already been demonstrated. For example, in their meta-analysis, Chen and Li (2014) found that the overall effect size of non-symbolic magnitude comparison and mathematical competence was significantly lower in studies controlling for general non-numerical cognitive abilities compared to studies not controlling for them. Likewise, Schneider and colleagues (2017) suggested that including other cognitive abilities (such as inhibition) in regression models might have a similar effect on the association between symbolic numerical magnitude processing and mathematics performance. Simanowski & Krajewski (2019) found that, after controlling for early numerical magnitude processing, executive functions in kindergarten were no longer predictive of mathematics skills in first and second grade.

Within this dissertation, two particularly interesting topics were identified for which simultaneous investigation seems critical, namely the interrelations between executive functions and metacognition, and the interrelations between metacognition and mathematics anxiety (see below section 5.1 ‘Affected by affect’). For example, there is large theoretical overlap between executive functions and metacognition as both are higher-order, control processes related to the regulation of behaviour. The way they are defined in research is often very similar and they follow a similar developmental trajectory (see Roebers, 2017, for an extensive review). Furthermore, studies investigating both functions simultaneously suggest that executive functions and metacognition are related (see Roebers & Feurer, 2016, for a short overview). When considered together to predict educational achievement, there is some evidence that metacognitive skills are more important than executive functions (Bryce et al., 2015). To thoroughly investigate their unique role in arithmetic, it is thus essential to investigate these processes in concert.

Despite these suggestions and observations that simultaneously investigating different processes can have an important impact on the results regarding the associations of these processes with performance, studies that include a variety of (meta)cognitive and affective processes are scarce. Hence, in the current dissertation, I have investigated different cognitive, metacognitive and affective processes in concert, which are known to play a key role in arithmetic when investigated in isolation. As such, I was able to investigate their unique role in individual differences in arithmetic performance and development, in addition to each other.

4 Longitudinal investigation

The current dissertation focused on primary school children. Throughout primary school crucial development occurs in arithmetic performance (Siegler, 1996). This arithmetic development is characterized by the transition from an effortful, slow and erroneous process to a process marked by fluency, i.e. fast and accurate processing. This is explained by a transition from initially relying on effortful strategies to solve basic arithmetic items (e.g. counting strategies), towards efficient arithmetic strategies (e.g. decomposition) or automation through arithmetic fact retrieval (e.g. Bailey et al., 2012; Siegler, 1996). Rather than this development implying an abrupt transition from *only* procedural strategies to *only* retrieval of arithmetic facts, there is a change in strategy distribution over developmental time encompassing a decrease in reliance on procedural strategies and an increase of the use of arithmetic fact retrieval (e.g. Barrouillet et al., 2008; Siegler, 1996). Importantly, large individual differences in the repertoire of strategies used by children remain present over development (e.g. Dowker, 2005). Over primary school, these dynamics result in substantial development and considerable individual differences in arithmetic skills.

As was discussed above, primary school is also a crucial period for development in various other domains and processes, including numerical magnitude processing (e.g. Matejko & Ansari, 2016), executive functions (e.g. Carlson et al., 2013; Diamond, 2013), metacognition (e.g. Schneider, 2010, 2015a; Schneider & Lockl, 2008) and mathematics anxiety (e.g. Mammarella et al., 2019). These developmental trajectories are very likely to impact the associations studied within this dissertation (e.g. Bull & Lee, 2014; Van der Ven et al., 2012). For example, growing experience with arithmetic might improve numerical magnitude processing skills, and improving numerical magnitude processing skills might facilitate the development of more mature arithmetic strategies (e.g. retrieval), both leading to a stronger association between numerical magnitude processing and arithmetic over time (e.g. Vanbinst et al., 2019). Furthermore, it is also likely that executive functions and metacognitive skills play a different role when children are in the early stages of learning arithmetic versus when they perform known arithmetic operations that are more automatized at a later developmental stage. For example, initial, effortful strategies early in arithmetic development might strongly rely on updating skills, a role which may decrease in light of the use of less effortful strategies (e.g. Raghubar et al., 2010). In the beginning stages of the use of retrieval strategies, inhibition skills might be needed more to inhibit competing, but irrelevant answers that are also activated when learning to retrieve answers to arithmetic problems. Once strong associations between arithmetic items and their respective answer are formed, inhibition skills might be less necessary (e.g. Goldfarb, 2018). The role of shifting skills in arithmetic might become more prominent over development, for example, as children learn more strategies and thus can shift between those (e.g. Yeniad et al., 2013). Improvements in metacognitive knowledge and skills and in arithmetic may also impact their relation, as, for example, improvements in arithmetic may enhance metacognitive monitoring in arithmetic, strengthening their association. Or on the other hand, gradually improving metacognitive knowledge and skills may enhance performance, for example, by providing a learning moment when an error is detected (e.g. Dunlosky et al., 2003; Schneider & Artelt, 2010). The development of mathematics anxiety and arithmetic may also affect their interrelation, which may become stronger due to increasing disruption of working memory as a result of increasing mathematics anxiety, or due to increasing experience with failure in arithmetic over development (e.g. Dowker, 2019e).

Hence, in order to obtain a thorough understanding of these interrelations in primary school, not only simultaneous investigation of these processes, but also a longitudinal research is essential. However, most existing studies on associations between (meta)cognitive and affective processes and arithmetic performance only report on concurrent relations, leaving important issues unresolved. First, it remains unknown whether the associations between arithmetic and these processes under investigation are stable over development. Second, it is unclear whether these processes (i.e. numerical magnitude processing, executive functions, metacognition and mathematics anxiety) are not only associated concurrently, but also predict later arithmetic performance. Finally, going one step further, it is yet to be examined whether these associations still hold when prior arithmetic performance is taken into account and thus whether the aforementioned (meta)cognitive and affective processes contribute specifically to the *development* of arithmetic skills.

When the associations between arithmetic and numerical magnitude processing, executive functions, metacognition and mathematics anxiety are studied in isolation from each other, longitudinal studies have confirmed the predictive power of each of these processes separately for later arithmetic performance. For example, several studies found evidence for the predictive value of numerical magnitude processing (e.g. De Smedt, Verschaffel, et al., 2009; Sasanguie et al., 2012; Schneider et al., 2017; Vanbinst, Ghesquière, et al., 2015) for later arithmetic. There is evidence as well for the predictive value of executive functions, especially updating skills, for later arithmetic performance (e.g. De Smedt, Janssen, et al., 2009; Lee & Bull, 2016; Mazzocco & Kover, 2007; Passolunghi et al., 2008; Van der Ven et al., 2012). While studies on the longitudinal associations between metacognition and arithmetic are scarce, the few available studies confirm the predictive power of metacognition for children's later arithmetic (e.g. Rinne & Mazzocco, 2014; van der Stel & Veenman, 2010). Mathematics anxiety has also been found to be predictive of later mathematical performance (Carey et al., 2016). Importantly, these longitudinal studies, again, focus mostly on one (meta)cognitive or affective process to predict later arithmetic performance. As such, they fail to identify the unique contributions of such processes when other critical (meta)cognitive or affective processes that predict arithmetic are considered. Even more critical, most of these longitudinal studies fail to include prior arithmetic performance as an important predictor in their models, and hence do not investigate the importance of these processes relative to prior arithmetic performance. This is crucial, as extensive evidence has suggested that early academic skills are the most robust indicator of later performance (e.g. Duncan et al., 2007). Moreover, including prior arithmetic performance importantly yields the possibility to investigate the predictive power of these functions for *development* in arithmetic (Duncan et al., 2007). It is important to note that, while the current doctoral project predominantly focusses on the role of different processes in arithmetic, it is nevertheless imperative to also consider the role of arithmetic in these processes and the possible bidirectionality of these interrelations. To tackle these abovementioned issues, the current dissertation includes a longitudinal panel study including arithmetic, numerical magnitude processing, executive functions, metacognition, and mathematics anxiety.

5 Metacognitive monitoring in arithmetic: A closer look

Throughout the first studies presented in this dissertation, metacognitive monitoring was found to be an important, unique process in arithmetical performance and development. As a result, the second part of this dissertation aimed to flesh out in more detail the association between metacognitive monitoring and arithmetic in primary school children. Specifically, three important gaps in the existing body of research on the role of metacognitive monitoring in arithmetic were addressed. Firstly, in line with the abovementioned arguments that it is essential to investigate different processes in concert and longitudinally, we examined whether mathematics anxiety played a role in the association between metacognitive monitoring and arithmetic performance and development. Secondly, we studied whether the importance of metacognitive monitoring was domain-specific to arithmetic, or whether it reflects a more general performance monitoring process. Lastly, we took up unanswered questions in both the field of arithmetic and the field of metacognition by investigating the neurobiological basis of metacognitive monitoring in children. As such, we also furthered our understanding of the prefrontal activation that is consistently found in the arithmetic brain network (Peters & De Smedt, 2017, for a review) and that, as metacognition has also been related to the prefrontal cortex (e.g. Vaccaro & Fleming, 2018), has been suggested to at least partially reflect metacognitive processes (e.g. Ansari et al., 2005). In the following sections, I further elaborate on these three addressed research gaps in more detail.

5.1 Affected by affect?

Research on metacognitive monitoring and mathematics anxiety has been done in isolation from each other, making their interrelation and unique contribution to the (individual differences in) performance and development of arithmetic unclear. This is particularly troublesome, as it is likely that metacognitive monitoring and mathematics anxiety are associated, because both are linked to a reflection on one’s performance. Hence, it is of importance to further our understanding of their interrelations and uncover whether their interplay has an impact on the respective associations of both processes with arithmetic. In the literature, suggestions and/or hypotheses on the importance of these interrelations can be found. For example, Ashcraft and Faust (1994), when discussing the anxiety to performance association, suggested that this relation might exist because mathematics anxious individuals feel less certain of their answers and thus, for example, engage in extensive answer changing. Furthermore, Gilmore and colleagues (2018a) included aspects of metacognitive thinking when listing possible risk factors for developing mathematics anxiety (e.g. "*student's own low perceptions of their mathematics ability*"; "*expectations about performance*"). Reviewing the literature on mathematics anxiety, Ashcraft and colleagues (2007) indicated that higher levels of mathematics anxiety are associated with lower self-confidence in mathematics, which was also found in the meta-analysis by Hembree (1990). Similarly, Dowker (2019c) indicated that mathematics anxiety is often linked to low estimates of one's own mathematics ability and expectations of failure. In a more recent account on determinants of mathematics anxiety (i.e. the interpretation account; Ashcraft, 2019; Ramirez et al., 2018), one's interpretation of previous mathematics experiences is emphasized, rather than, for example, reduced competency, and as such, this model also hints to metacognitive mechanisms. In line with this, Tobias (1986) hypothesized that being metacognitively aware of one's (poor) performance may increase feelings of pressure and mathematics anxiety. On the other hand, Morsanyi et al. (2019) suggested that mathematics anxiety may lead to a biased interpretation of one's own performance.

In spite of these frequent suggestions of a potential interrelation between metacognitive monitoring and mathematics anxiety, rigorous, empirical research on the developmental associations of these potentially interrelating processes and the impact thereof in arithmetic is lacking, especially in primary school children. The existing literature offers some suggestions, as cross-sectional associations have been observed between test anxiety and metacognitive skilfulness in secondary school students (Veenman et al., 2000), between mathematics anxiety and metacognition in Chinese 10-year-olds and Turkish 12-year-olds word problem solving (Lai et al., 2015; Özcan & Gümüs, 2019), in university students in arithmetic (Legg & Locker, 2009) and in their general mathematics achievement (Erickson & Heit, 2015). However, associations between metacognitive monitoring and mathematics, and between mathematics anxiety and mathematics are already observed in the early grades of primary school. It therefore remains unclear how metacognitive monitoring and mathematics anxiety are related to each other and to mathematical achievement in early primary school children, in whom these processes are in the midst of development. Even more critical, none of the abovementioned studies has collected longitudinal data, which renders claims on the developmental dynamics of the associations problematic. Within the current dissertation, these outstanding questions were tackled by using a longitudinal panel design to examine the associations between metacognitive monitoring, mathematics anxiety and arithmetic achievement in primary school children.

5.2 Specificity in domain-generality?

Metacognitive monitoring is generally considered to be a domain-general process that is a critical correlate and predictor of cognitive performance and learning in diverse domains (e.g. Annevirta et al., 2007; Block & Peskowitz, 1990; Efklides & Misailidi, 2010; Freeman et al., 2017; Kuhn, 2000; Lyons & Ghetti, 2013; Özsoy, 2011; Rinne & Mazzocco, 2014; Schneider, 1998; Schneider & Artelt, 2010; Schraw et al., 2006; Veenman et al., 2006, 2004). Yet, as indicated above, the extent to which a process is domain-general can differ, and can also change over time. For example, the importance of metacognitive monitoring might differ in different (academic) domains (e.g. mathematics, reading). Educationally relevant, outstanding questions are whether metacognitive monitoring in academic performance is domain-specific or whether it reflects a more general performance monitoring process, and whether this might change over primary school development. Research investigating this issue in rather distant domains (e.g. emotion versus numerical domain; Vo et al., 2014) provides evidence that metacognitive monitoring is first domain-specific, that domain-generality of metacognitive monitoring emerges over development and that a gradual shift from domain-specific towards domain-general metacognitive monitoring occurs in children aged between 8 and 13 (Geurten et al., 2018). Importantly, over development, evidence for some domain-specificity of metacognitive monitoring remains (e.g. Garcia et al., 2016; Lingel et al., 2019; Löffler et al., 2016; Schraw et al., 1995).

To thoroughly investigate the extent to which metacognitive monitoring of academic performance is a more domain-specific or domain-general process, this should be examined in related, yet distinct academic domains. Studying this in two highly related academic domains ensures a more stringent empirical test of the possible limits of domain-specificity of metacognitive monitoring, as domain-specificity is much harder to ascertain in related domains compared to distant domains for which such specificity might, on a surface level, be more easily observed (e.g. numbers vs. emotions). In this dissertation, two quintessential academic domains in primary school education were investigated, in which primary school children go through crucial developmental steps, namely arithmetic and spelling. Determining the developmental trajectory of whether and how metacognition generalizes across domains is crucial, from a theoretical perspective, as it sheds light on how metacognition develops throughout childhood and thus furthers our understanding of the functioning and cognitive architecture of metacognition. Furthermore, it is also crucial from a practical perspective, as determining when metacognition becomes domain-general, and which conditions drive such a generalization, could have important influences on, for example, how metacognition can be stimulated through educational practice. The current dissertation aims to provide a first step towards an understanding of the domain-specificity or –generality of metacognition by focusing on a narrow age range in which this development could occur, in related and highly relevant domains for children's academic development, i.e. arithmetic and spelling.

5.3 Metacognition-related brain activation during arithmetic?

An important part of the research on arithmetic has been done using brain imaging techniques (see Peters & De Smedt, 2017, for a review). While cognitive neuroscience has made considerable progress in understanding the neurobiological basis of cognitive performance in several academic domains, such as arithmetic, much less is known about how the brain generates metacognitive awareness of task performance. Within the context of the topic of the current doctoral project, i.e. the role of metacognition in arithmetic, examining the underlying neurocognitive architecture supporting metacognitive abilities in children is particularly promising. This is because, on the one hand, in the neuro-imaging literature on metacognition in adults, metacognition is considered a higher brain function that strongly depends on the prefrontal cortex or PFC (see Pannu & Kaszniak, 2005 and Shimamura, 2000, for reviews; see Vaccaro & Fleming, 2018, for a meta-analysis). On the other hand, arithmetic recruits a large set of interconnected areas, including prefrontal, posterior parietal, occipital-temporal and hippocampal areas (Peters & De Smedt, 2017, for a review), and many fMRI studies have pointed to the involvement of the prefrontal cortex control processes in arithmetic (e.g. Menon, 2015). Many suggestions on the interpretation of this prefrontal activity during arithmetic have been made in the existing literature, and one process often referred to is metacognition. While many researchers have suggested this prefrontal activation could partially reflect metacognitive (monitoring) processes (e.g. Ansari et al., 2005; Arsalidou et al., 2018; Houdé et al., 2010; Kaufmann et al., 2006, 2011; Kucian et al., 2008; Menon, 2015; Rivera et al., 2005), this suggestion has never been empirically tested. Given the central role of metacognitive processes, such as metacognitive monitoring in academic learning, and the behavioural work that has revealed that metacognitive monitoring is a unique predictor of individual differences in arithmetic in children (e.g. Rinne & Mazzocco, 2014), this suggestion is not far-fetched and empirical research into this is thus warranted.

To properly empirically study this suggested overlap, first, the neurobiological basis of metacognitive monitoring in children in a higher-order processing domain, such as arithmetic, should be examined. This is important, as the current evidence on the neurobiological basis of metacognitive monitoring is based on research in adults, and was almost exclusively done in lower-level processing domains (e.g. perception; Vaccaro & Fleming, 2018). Crucially, developing brains of children differ from those of adults. Furthermore, there is evidence to suggest that there is specificity, i.e. regional specialization within the PFC, concerning the neurobiological basis of metacognition with respect to metacognitive processes in different tasks and domains (e.g. Baird et al., 2013; McCurdy et al., 2013). Therefore, results obtained in adults in lower-level processing domains cannot be generalized to the neurobiological basis of metacognitive monitoring in children in arithmetic without thorough empirical investigation (see Ansari, 2010, for a similar discussion). Therefore, in this dissertation I empirically investigated which brain regions are involved in engaging in metacognitive monitoring within a higher-order cognitive processing domain (i.e. arithmetic), in primary school children. This allows to further our understanding of not only the neurobiological basis of metacognitive monitoring in children, but also the prefrontal activation found during the process of arithmetic, which has been suggested to at least partially reflect metacognitive processes.

Importantly, this dissertation encompassed a large dataset of longitudinal behavioural data on metacognitive monitoring in a group of primary school children. In the neuro-imaging study, the exact same paradigm as in the behavioural studies, in the same children was used. The combination of longitudinal behavioural data and the neuro-imaging data on this topic in the same children, additionally yields a unique research opportunity, as it allows to investigate the associations between brain activation during metacognitive monitoring and development in arithmetic performance over a three-year period. As such, the current dissertation aims to better understand individual differences in arithmetic and neurobiological processes that play a role when children learn school-relevant skills, in this case arithmetic. To improve the existing literature, there is a need for these kind of developmental brain imaging studies, investigating children at ages when they are acquiring a particular mathematical skill (Ansari & Lyons, 2016). Because these brain imaging studies posit themselves at the intersection of psychology, education and cognitive neuroscience, they are exceedingly promising to understand the neurobiological processes that play a role for educationally relevant knowledge and skills (De Smedt, 2018; De Smedt & Grabner, 2015), as they make it possible to investigate the brain during the learning phase of academic skills.

6 Methodology & data-analysis

Within this dissertation, different statistical and methodological frameworks were used in order to comprehensively examine our research questions. On the one hand, I did not only use the well-known and widely used frequentist statistics, but also made use of Bayesian statistics which allows one to deepen the investigation in ways that are much less possible using frequentist statistics. On the other hand, I profited from combining behavioural research methods and neuroscientific techniques. In what follows, I further explain the use of these different statistical and methodological frameworks in this dissertation. Thereafter, I comment on how I aimed to optimize reproducibility and transparency of the current dissertation. Lastly, I discuss the data that were collected for this project and on which the scientific conclusions were based.

6.1 Frequentist & Bayesian statistics

The use of frequentist analyses allowed me to explore the data by means of a well-known method to gauge statistical support for the hypotheses of interest. However, this null hypothesis significance testing relying on *p*-values has a number of statistical limitations (e.g. Andraszewicz et al., 2015). For example, *p*-values cannot quantify evidence in favour of a hypothesis, they only signal the extremeness of the data under the null hypothesis. The goal in frequentist statistics is to decide whether a particular value of a model parameter can be rejected. The *p*-value logic thus resembles a proof by contradiction: Low *p*-values indicate extreme data and usually lead researchers to reject the null hypothesis and interpret this as evidence in favour of the alternative hypothesis (Andraszewicz et al., 2015). Unlike null-hypothesis testing, Bayesian statistics allows one to test of the degree of support for a hypothesis (i.e. degree of strength of evidence in favour of or against any given hypothesis). This is expressed as the Bayes factor (BF), which is the ratio between the evidence in support of the alternative hypothesis H_1 over the null hypothesis H_0 (BF_{10}). By comparing the fit of the data under the null hypothesis to the alternative hypothesis, Bayes factors quantify the evidence in favour of these hypotheses. For example, a Bayes factor of 10 ($BF_{10} = 10$) suggests that the alternative hypothesis is 10 times more likely than the null hypothesis. Although Bayes factors provide a continuous measure of degree of evidence, there are some conventional approximate guidelines for interpretation (see Andraszewicz et al., 2015, for a classification scheme): $BF_{10} = 1$ provides no evidence either way, $BF_{10} > 1$ anecdotal, $BF_{10} > 3$ moderate, $BF_{10} > 10$ strong, $BF_{10} > 30$ very strong and $BF_{10} > 100$ decisive evidence for the alternative hypothesis; $BF_{10} < 1$ anecdotal, $BF_{10} < 0.33$ moderate, $BF_{10} < 0.10$ strong, $BF_{10} < 0.03$ very strong and $BF_{10} < 0.01$ decisive evidence for the null hypothesis. By adding these analyses, I aimed to deepen the findings from the traditional analyses, as I was able to identify which hypotheses were (not) supported, and, consequently, which hypothesis is most plausible and which predictors are the strongest. This was particularly relevant for the current dissertation, because one could, for example, compare the strength of evidence in favour of the unique role in arithmetic of the different processes under study, and compare the evidence in favour of different hypotheses (e.g. domain-specific hypothesis versus domain-general hypothesis of metacognitive monitoring in arithmetic). The use of Bayesian statistics allowed me to provide evidence for the null hypothesis, for example, to indicate a process does not uniquely explain variance in arithmetic on top of the other considered processes.

6.2 Correlational, individual differences approach

Within this dissertation, I have used both cross-sectional and longitudinal designs and different analytical techniques, aiming to provide a more comprehensive understanding of individual differences in arithmetic performance and development, numerical magnitude processing, executive functions, metacognition and mathematics anxiety.

To follow a group of children from second to third and half of that sample to fourth grade of primary school, I used a longitudinal panel design, in which the same measures of all investigated processes were administered at the different time points. Crucially, the use of this design allows to investigate development of the associations over time. This is essential given the fact that we investigated these processes in a crucial developmental time, namely primary school. As such, stability of the associations could be investigated, as well as longitudinal associations between the investigated processes. Going one step further, the use of a panel longitudinal design allows to take into account autoregressive effects (e.g. the role of prior arithmetic performance for later arithmetic performance) when investigating the associations between (meta)cognitive and affective processes and arithmetic. Hence, by means of this longitudinal panel design, I was able to provide indications of the direction of the associations and of whether the investigated processes contribute specifically to the development of the outcome measure (e.g. arithmetic).

In addition to using a longitudinal design, a cross-sectional design was used to collect data in two different age groups, namely second and third graders (i.e. 7-8 year-olds and 8-9 year-olds). The additional inclusion of these two samples, on top of the longitudinal follow up of a different cohort of the same age, enabled to investigate research questions that were not addressed within the longitudinal cohort.

To analyse these data, I used different analytical techniques, ranging from widely used (e.g. Pearson correlation coefficients, multiple regression) to some more advanced data-analysis techniques (e.g. moderation analysis, mediation analysis). Using a combination of these analytical techniques, I was able to take up my research questions from different angles to provide a more comprehensive understanding. For example, mediation analyses allows to investigate potential mechanisms by which certain effects operated, while moderation analyses allows to investigate when, i.e. at which level of another process, an effect occurred between two processes.

6.3 Neuro-imaging method

Within this dissertation, not only behavioural methods were used to thoroughly investigate my research questions, I additionally used neuro-imaging techniques, which are explained in more detail below. Using both methods provided me with different levels of analysis and measurement of the key variables of interest that cannot be accessed by either behavioural or neuro-imaging data alone. Cognitive neuroscience offers tools, methodologies and theories to investigate (meta)cognitive processes that take place during mathematical thinking and learning, which may complement and extend knowledge that has been obtained on the basis of behavioural data only (De Smedt et al., 2010). Indeed, data on brain activity might add to a complementary, detailed description of the different cognitive (sub)processes that take place during mathematical thinking and learning.

In line with this, the rationale to include brain data in the current dissertation, was twofold. Firstly, neuro-imaging techniques were used to uncover the brain basis of metacognitive monitoring processes in children. Secondly, and importantly, these brain data were also utilized to further our understanding of the prefrontal activation found during the process of arithmetic. This prefrontal activation (Peters & De Smedt, 2017, for a review) has been suggested to at least partially reflect metacognitive processes (e.g. Ansari et al., 2005), but this hypothesis has never been empirically tested. Within this dissertation, I made optimal use of the unique opportunity to build on the existing behavioural theory and the acquired behavioural data within this project, integrate this with the neuro-imaging literature on both arithmetic and metacognition, and specifically collect neurobiological data on this topic in the same sample of children on whom a large number of developmental, behavioural data was available. This was done to maximize the meaningful interpretation and the relevance of the findings.

The brain imaging data that were collected for this dissertation, were acquired by means of magnetic resonance imaging (MRI). MRI enables neuroscientists to create images of the brain anatomy based on differences in the amount of hydrogen in different tissue types, and detect task-related neural activity (functional MRI or fMRI) based on the proportion of oxygenated blood in the brain (Arthurs & Boniface, 2002). It is a non-invasive neuro-imaging method that uses powerful magnets that produce a magnetic field. This magnetic field forces hydrogen atoms in the participant's body to align with this static magnetic field, while initially hydrogen atoms are oriented randomly. The strength of the magnetic field is measured in units called tesla (T). Next, a Radio Frequency (RF) pulse is sent out, which disturbs the alignment, making the hydrogen atoms spin out of the equilibrium, and changing their spin to the opposite direction of the magnetic field. When the RF signal is stopped, the hydrogen atoms realign with the original magnetic field and release energy in the process. This release of energy is picked up and forms the basis of the MR signal. Different components of the MR signal are used to create different types of images. For example, one can distinguish between different types of tissue by using variations in the rate at which the hydrogen atoms realign with the magnetic field after the RF pulse (i.e. T1 relaxation time). These T1-weighted images are typically used for structural images of the brain, i.e. static anatomical information. A T2 constant describes how quickly the atoms emit energy when recovering to equilibrium. Taken together with the fact that deoxyhemoglobin produces distortions in this component, this forms the basis of the images in fMRI experiments (i.e. T2* image; Ward, 2015).

The fMRI method takes advantage of the fact that when neurons in the brain become active, the amount of oxygenated blood flowing through that area is increased. When neurons consume oxygen, they convert oxyhemoglobin to deoxyhemoglobin, which has strong paramagnetic properties. This change in oxygenation is measured in fMRI and is referred to as the blood oxygenation level dependent (BOLD) signal (Poldrack et al., 2011). fMRI measures temporary changes in brain physiology associated with cognitive processing, based on the Hemodynamic Response Function (HRF), i.e. the way that the BOLD signal evolves over time in response to an increase in neural activity. Compared to other brain-imaging methods, such as EEG or PET, fMRI has vastly better spatial resolution (i.e. the accuracy with which one can estimate the anatomical location of neural activation). Importantly, this hemodynamic method does not measure the activity of the neurons directly. Rather, it measures a consequence of neural activity, i.e. changes in BOLD signal. As the hemodynamic response peaks only after 4 to 5 seconds after the onset of the actual neural firing (Poldrack et al., 2011), this results in lower temporal resolution (i.e. the accuracy with which one can measure when an event is occurring).

Within this dissertation, I used cognitive subtraction, a type of experimental design in fMRI in which activity in a control task is subtracted from activity in an experimental task. Through the comparison of the activity of the brain in a task tapping into a particular (cognitive) process (e.g. metacognition) with the activity of the brain in a very similar control task that does not tap into that process, it is possible to infer which regions are specialized for this particular process (Ward, 2015). This comparison is required because the brain is always physiologically active, and regions of activity can only be meaningfully interpreted relative to a baseline or control condition. Importantly, because of the central role of the control condition in fMRI research, a good cognitive theory of the elements that comprise a task are essential. To maximally ensure this, within this dissertation, I made use of the existing behavioural theory and the data acquired in the context of the behavioural studies within this doctoral project.

6.4 Reproducibility and transparency

In line with the current, important attention that has been given to and action that has been taken for reproducible, transparent science, a comprehensive data-analysis plan of several of the studies within this dissertation was preregistered on the Open Science Framework (https://osf.io/rufxc/). The Open Science Framework (OSF) is a free and open source project management and collaboration tool, and workflow system, which is developed and maintained of the Center for Open Science (COS). The COS aims to increase openness, integrity and reproducibility of research (Foster & Deardorff, 2017). In my preregistrations on the OSF, I provided background information on the study, the existing literature on the study topic, the rationale for the analyses, and, importantly, I specified in detail which analyses would be performed to answer the research questions. Additionally, I provided the full cognitive testing battery of each preregistered study on the OSF page of each project.

6.5 Data overview

The data included in the following chapters were all specifically collected for this doctoral project by the author. The current doctoral thesis includes data of one longitudinal cohort and two age groups in the context of a cross-sectional design. In the longitudinal cohort, originally 127 primary school children were included. When these children were in the middle of second grade (7-8 year-old), a large battery of (meta)cognitive and affective data was acquired. One year later, in the middle of third grade, the same battery of tasks was administered (8-9 year-old) in 121 of the 127 children who participated in the first data collection. At the end of fourth grade, MRI data were acquired in 55 of these children (9-10 year-old), whose parents gave consent for their children to participate in the MRI study. The cross-sectional cohorts consisted of 147 third graders (8-9 year-old) and 77 second graders (7-8 year-olds). Similar to in the longitudinal cohort, in these cohorts a large battery of (meta)cognitive data was acquired. These two groups of participants were recruited to investigate research questions that were not addressed within the longitudinal cohort.

7 Research aims & outline of the doctoral dissertation

Prior research has uncovered large individual differences in children's arithmetic performance and development (e.g. Dowker, 2005, 2019c). While firstly, I aimed to increase our understanding of the complex interacting roles of key processes that have been identified as important for arithmetic performance and development, the ultimate purpose of the current collection of studies was to provide a deeper, more comprehensive understanding of the role of metacognition in arithmetic. After metacognitive monitoring was identified as a very promising predictor of arithmetic on top of other key processes, I aimed to more thoroughly investigate its role in arithmetic performance and development in detail. Through different studies, I examined this role of metacognitive monitoring in arithmetic in addition to other important cognitive and affective processes in primary school children. I further studied the extent to which this role of metacognitive monitoring was domain-specific or domain-general in different age groups in primary school. Finally, the role of metacognitive monitoring was studied both at the behavioural and neurobiological level to further our understanding of the underlying neurocognitive architecture supporting metacognitive abilities in children as well as the arithmetic brain network in children. The five specific research aims addressed in the current doctoral project are discussed in detail below.

*7.1 Research Aim 1: Simultaneous investigation of the role of numerical magnitude processing, executive functions and metacognition in arithmetic in primary school children (**Chapter 2**).*

The associations between different (meta)cognitive processes (e.g. numerical magnitude processing, executive functions, metacognition) and arithmetic have been extensively investigated in isolation. In contrast to this extensive body of research, little attention has been paid to examining the joint effects of different (meta)cognitive, domain-specific and domain-general processes in arithmetic. Especially in primary school children this research is limited, yet these children are in the middle of a crucial developmental period for both arithmetic and (meta)cognitive processes. The first research aim of the current doctoral project was to tackle this gap in the literature by simultaneously studying numerical magnitude processing skills, executive functions, metacognitive knowledge and skills and arithmetic performance of 7-8 year-old children (second grade of primary school) and as such unravel the unique contributions of these processes to individual differences in arithmetic performance. This first research aim provides the groundwork for the following research aims included in this dissertation.

*7.2 Research Aim 2: Assessing the longitudinal interplay between numerical magnitude processing, executive functions, metacognition and arithmetic in primary school children (**Chapter 3**).*

When the role of (meta)cognitive processes in arithmetic is investigated, studies mostly use concurrent or sometimes cross-sectional research designs. Yet, using a longitudinal design has major advantages, as this design, especially a longitudinal panel design, allows researchers to thoroughly investigate stability of results over (developmental) time, examine predictive associations between variables and take into account prior performance. Consequently, the second research aim of the current doctoral project was to additionally uncover developmental dynamics within these associations, using a longitudinal design. This was done by building on the study of the first research aim. To fully grasp crucial developmental periods of the key processes investigated, primary school children were followed up from second to third grade (i.e. 7-9 year-old). Using a longitudinal panel design, the current aim was to unravel the stability of the associations found in early primary school over time, to uncover the longitudinal associations of symbolic numerical magnitude processing, executive functions and metacognition with arithmetic, and to investigate these predictive associations when taking prior arithmetic performance into account.

*7.3 Research Aim 3: Examining the longitudinal interplay between metacognitive monitoring, mathematics anxiety and arithmetic in primary school children (**Chapter 4**).*

Although there is a large body of research on the role of affective processes such as mathematics anxiety (studied in isolation), there is a lack of inclusion of such affective processes in studies investigating (meta)cognitive processes in arithmetic. This results in an incomplete understanding of the role of both processes in arithmetic. Especially concerning the interplay between metacognitive monitoring and mathematics anxiety in young primary school children, the lack of simultaneous investigating of these processes is worrisome, because these processes are both related to thinking about your performance, and, in primary school, crucial development occurs for both metacognitive monitoring and mathematics anxiety. Within the current doctoral project, the third research aim was therefore to simultaneously consider metacognitive monitoring, mathematics anxiety and arithmetic achievement in young primary school children using a longitudinal panel design. Going beyond traditional correlational analyses and regression models, I used mediation and moderation models to further our understanding of the developmental dynamics of these processes in young primary school children, which is critical to develop effective educational interventions.

*7.4 Research Aim 4: Investigating the domain-specificity of the role of metacognitive monitoring across domains in primary school children (**Chapter 5**).*

As was amply shown in the literature, metacognitive monitoring is important in different academic domains. In adults, associations were found between monitoring in these different domains, pointing toward domain-generality, while in children, such associations seem to be absent, pointing to domain-specificity. Furthermore, research has suggested a shift from more domain-specificity towards domain-generality of metacognitive monitoring throughout primary school. Importantly, research on this topic is mostly based on empirical investigation in rather distant domains. I deliberately included a different, yet correlated academic domain (i.e. spelling) in addition to arithmetic, to thoroughly investigate the extent to which metacognition might be domain-specific or domain-general and whether a gradual shift towards more domain-generality occurs in primary school. Additionally, the domains were chosen because of their high relevance for primary school children. Importantly, domain-generality and domain-specificity of processes are not all-or-nothing concepts. As such, in line with the general aim of this dissertation to investigate the role of metacognitive monitoring in arithmetic, the fourth research aim was to examine to what extent this association was domain-general and/or domain-specific.

*7.5 Research Aim 5: Uncovering the neurobiological basis of metacognitive monitoring during arithmetic in the developing brain (**Chapter 6**).*

Metacognition is considered a higher brain function that strongly depends on the prefrontal cortex. Yet, this knowledge is based on studies in the adult population and on monitoring in lower-level cognitive processes, such as perceptual decisions. Therefore, the results cannot be generalized to the neurobiological basis of metacognition in children in academic domains without thorough empirical investigation. By combining neuro-imaging data with the insights from the behavioural work in this dissertation, the fifth research aim of this dissertation was to empirically investigate which brain regions are involved in engaging in metacognitive monitoring within a higher-order cognitive processing domain of arithmetic, and to do so in primary school children. Additionally, this sheds light on the frequently suggested, but never empirically tested, hypothesis that metacognitive monitoring processes could partially explain the increases in prefrontal activation that are often observed when doing arithmetic.

To conclude, in ***Chapter 7***, the findings of these doctoral dissertation are discussed, standing still at potentially fruitful avenues for future research. The discussion of the scientific contribution of the abovementioned studies is followed by a discussion of both theoretical and methodological as well as educational considerations.

CHAPTER 2

More than number sense

The additional role of executive functions and metacognition in arithmetic.

The content of this chapter has been published as:

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.2019.01.012

Chapter 2

More than number sense:

The additional role of executive functions and metacognition in arithmetic.

Abstract

Arithmetic is a major building block for children's development of more complex mathematical abilities. Knowing which cognitive factors underlie individual differences in arithmetic is key to gaining further insight into children's mathematical development. The current study investigated the role of executive functions and metacognition (domain-general cognitive factors) as well as symbolic numerical magnitude processing (domain-specific cognitive factor) in arithmetic, enabling the investigation of their unique contribution in addition to each other. Participants were 127 typically developing second graders (7- and 8-year-olds). Our within-participant design took into account different components of executive functions (i.e. inhibition, shifting, and updating), different aspects of metacognitive skills (i.e. task-specific and general metacognition), and different levels of experience in arithmetic, namely addition (where second graders had extensive experience) and multiplication (where second graders were still in the learning phase). This study reveals that both updating and metacognitive monitoring are important unique predictors of arithmetic in addition to each other and to symbolic numerical magnitude processing. Our results point to a strong and unique role of task-specific metacognitive monitoring skills. These individual differences in noticing one's own errors might help one to learn from his or her mistakes.

Introduction

Whereas waiters know within seconds how much the different drinks you ordered cost together, others shudder when only thinking about math class. Experience with mathematics could be a part of the explanation for these huge differences in math skills in adults, yet these differences already appear early in development (e.g. Duncan et al., 2007). An extensive body of literature demonstrates that there are large individual differences in the way children acquire mathematical abilities (e.g. Berch et al., 2016). When looking for cognitive explanations for these differences, research has investigated both domain-general cognitive factors (i.e. factors relevant for learning various academic skills) and domain-specific cognitive factors (i.e. factors specifically relevant for mathematics learning) (e.g. Geary & Moore, 2016; Vanbinst & De Smedt, 2016b), yet these types of factors have been investigated in relative isolation from each other. Over the past decade, research has mainly focused on one domain-specific cognitive skill, numerical magnitude processing (i.e. children's elementary intuitions about quantity and the ability to understand the meaning of numbers), as a core factor of individual differences in mathematical abilities (e.g. see De Smedt et al., 2013, for a review; see Schneider et al., 2017, for a meta-analysis). By narrowing its scope in this way, this research ignores other critical, particularly domain-general, cognitive factors that might play a role in (a)typical mathematical development (see Fias et al., 2013). Schneider and colleagues (2017) revealed in their meta-analysis that the overall correlation between numerical magnitude processing and mathematical competence was only r = .278, suggesting that numerical magnitude skills explain a significant but only small part of the variance in mathematical abilities. Consequently, other important factors that drive these individual differences need to be considered.

Some potentially important factors have been identified such as working memory (e.g. Peng et al., 2016) and other executive functions (e.g. Bull & Lee, 2014). However, the precise roles of these factors in mathematics in addition to numerical magnitude processing remain unclear. Executive functions refer to a family of top-down mental processes needed when one needs to concentrate and pay attention (Diamond, 2013), skills required to monitor and control thought and action, processes that allow one to respond flexibly to his or her environment and engage in deliberate, goal-directed thought and action (Cragg & Gilmore, 2014). Executive functions mainly consist of the processes of inhibition, shifting, and updating (e.g. Miyake et al., 2000). Inhibition refers to one's ability to control his or her attention, behaviour, and thoughts to override a strong internal predisposition or external lure and instead do what is more appropriate or needed (Diamond, 2013). Two types of inhibition can be distinguished, namely response inhibition (i.e. control over one's behaviour and emotions in the service of controlling one's behaviour to resist temptations and not act impulsively) and interference control (i.e. the ability to selectively attend, focusing on what one chooses and suppressing attention to other stimuli). Shifting is defined as the disengagement of an irrelevant task set or strategy and the subsequent initiation of a new, more appropriate set (van der Sluis et al., 2007). Updating involves holding information in working memory and flexibly manipulating it (Baddeley & Hitch, 1994).

Although the existing literature has provided some understanding about the association between executive functions and mathematics (e.g. Cragg & Gilmore, 2014), there is insufficient insight into the specificity of these associations. Some studies found a significant association between executive functions and mathematics, whereas others did not (e.g. Bull & Lee, 2014; Cragg & Gilmore, 2014; Friso-Van Den Bos et al., 2013).

This inconclusiveness might be explained by four limitations in the existing literature. First, mathematical ability consists of various different abilities (e.g. arithmetic, problem solving, geometry), all of which could be differently related to certain cognitive factors (e.g. see Cragg & Gilmore, 2014, for a critical discussion). Consequently, focusing on different math subcomponents or tasks could be the reason for these discrepancies in existing research. For example, Gilmore et al. (2015) found different associations between inhibition and specific components of mathematics. Therefore, investigating specific mathematical skills is essential when investigating which cognitive factors predict the development of mathematical skills. The current study focused on arithmetic, a major building block for children's development in more complex mathematical abilities (Kilpatrick et al., 2001). The association between executive functions and arithmetic is theoretically appealing. For example, cognitive inhibition (i.e. interference control) might play a role in arithmetic given that arithmetic facts are stored in an associative network in semantic memory (e.g. Verguts & Fias, 2005) and, thus, are particularly prone to interference because of the number of features they share (De Visscher & Noël, 2014b). Hence, during arithmetic, incorrect but competing answers need to be inhibited. On the other hand, good executive functioning might help to suppress inefficient arithmetic strategies in favour of more efficient ones (e.g. using a decomposition of operands strategy instead of finger counting) or might help to keep intermediate solutions in mind and manipulating them while calculating (i.e. updating).

Second, the discrepancy between results on the relation between executive functions and mathematics might be due to the multi-componential nature of executive functions. Executive functioning includes different dimensions, yet many studies take only one subcomponent into account. Different components of executive functions are differently related to mathematics (Bull & Lee, 2014; Cragg et al., 2017). Because there is both unity and diversity in the executive functioning components (Friedman & Miyake, 2017; Lee et al., 2013), it is important to investigate them simultaneously. In this study, we included all three well-known components of executive functions, namely inhibition, shifting, and updating.

Third, the inconclusiveness might be due to the fact that relevant moderating variables have been overlooked. One such variable that has often been neglected in theoretical models of executive functions (see Desender et al., 2014, for a critical discussion), but has received a lot of attention in more general educational research (e.g. Schneider & Artelt, 2010), is metacognition. Metacognition was first introduced by Flavell (1979) as a broader concept that encompasses monitoring and regulation of cognitive performance. It involves the ability to assess one's own cognitive knowledge and ability (Vo et al., 2014) and how people monitor and control their cognition on-task (Bryce et al., 2015). An important distinction is made between metacognitive monitoring and metacognitive control (Flavell, 1979; Nelson & Narens, 1990). The former is the meta level that is informed by the object level, and the latter is the meta level that modifies the object level, with the object level being the cognitive process or task at hand. Metacognition can be measured by means of general metacognitive knowledge/skills measures (e.g. Haberkorn et al., 2014) or by asking participants to indicate their confidence on a task-specific level (e.g. Rinne & Mazzocco, 2014).

The association between metacognition and general academic performance has been extensively studied. Vo et al. (2014) found that children's metacognitive ability in the numerical domain predicted their general school-based mathematics knowledge and suggested that children's metacognition is a domain-dependent cognitive ability in children. Freeman et al. (2017) found that calibration of confidence in a working memory task was related to academic achievement. Some studies have specifically focused on the association between metacognition and arithmetic. At a theoretical level, Shrager and Siegler (1998) identified a metacognitive system in their model of children's strategy choices and discoveries in arithmetic (i.e. SCADS [strategy choice and discovery simulation] model), where a confidence criterion (i.e. a randomly varying threshold for stating a retrieved answer) was one of the most important probabilistic components. Rinne and Mazzocco (2014) more recently showed that, in children in late elementary school, on-task metacognitive judgments were strongly related to concurrent mental arithmetic and that early metacognitive monitoring ability predicted subsequent growth in mental arithmetic performance. However, it remains to be determined whether these associations remain when the different aspects of executive functions as well as numerical magnitude processing are included and whether the same associations can be observed in younger children.

Importantly, metacognition and executive functions are closely linked. Both are higher-order factors related to the regulation of behaviour, they share theoretical features (e.g. controlled processing), they undergo similar developmental trajectories, and they are associated with comparable brain regions (Roebers & Feurer, 2016). Metacognition plays a critical role in executive functioning processes because subjective experiences (i.e. metacognition) allow for top-down control of behaviour (Desender et al., 2014). Indeed, effective executive functioning requires an accurate determination of when and which type of control is needed. On the other hand, Bryce et al. (2015) suggested that executive functions might be necessary but not sufficient antecedents to metacognitive skills. Lyons an Zelazo (2011) characterized metacognitive monitoring as a reflective process. This reflective process is explicitly conceptualised as an integral part of metacognition (Nelson & Narens, 1990) but is only implicitly assumed to take place in executive functions, for which post-error slowing can be an indicator. Because executive functions and metacognition skills are closely related (see Roebers, 2017, for a review), the lack of inclusion of metacognition in existing studies on the association between mathematics and executive functions may have led to inconclusive results.

Lastly, the inconclusiveness in the current literature on the association between executive functions and mathematics might be due to the age of the participants. Because arithmetic (e.g. Vanbinst, Ceulemans, et al., 2015), executive functions (e.g. Carlson et al., 2013), and metacognition (e.g. Schneider, 2010) are still in the midst of development during primary school, different associations might be present at different ages. Executive functioning develops substantially during primary school (e.g. Diamond, 2013), with increases in inhibition, shifting, and updating skills allowing for better executive functioning along with increasing automaticity and efficiency (Carlson et al., 2013). Executive functioning depends on neurobiological networks involving the prefrontal cortex that continue to improve into early adulthood. Metacognition also relies on the prefrontal cortex (Fleming & Dolan, 2014) and continues to develop through primary school (e.g. Schneider, 2015a; Schneider & Lockl, 2008).

Arithmetic skills also show large development in primary school, and a gradual change from procedural strategies to retrieval of arithmetic facts is observed (Siegler, 1996). This might lead to different associations with executive functions and metacognition over time. For example, the need for inhibition skills might be particularly high when children have already learned (some) arithmetic facts because similarity between these facts provokes interference (De Visscher & Noël, 2014a). Consequently, the association between executive functions and arithmetic could change across development. One prediction might be that inhibition skills have a time-limited role in arithmetic; at the early stages of the development from procedural strategies to full automatization of arithmetic fact retrieval, more interference control skills will be needed, whereas at the end of development the need for control would be lower as a semantic network is automatically activated (see Moors & De Houwer, 2006 and Siegler, 1996, for a theoretical discussion).

Because these three cognitive factors (i.e. arithmetic, executive functions, and metacognition) develop dramatically throughout primary school, the association of executive functions and metacognition with arithmetic might be variable across development. Hence, the level of experience in arithmetic is important to take into account when examining the association among executive functions, metacognition, and arithmetic. Accordingly, in this study, for second graders we included an arithmetic operation with which they had extensive experience (i.e. addition) and for which retrieval of single-digit arithmetic items is very high (De Smedt, 2016) as well as an operation for which participants were in the learning phase (i.e. multiplication) and there is no full automatization of fact retrieval present yet.

1 The current study

Little attention has been paid to investigating the joint effects of domain-specific and domain-general correlates of arithmetic. In addition, it is not clear how these factors are associated in young children (i.e. early elementary school) and, consequently, how these factors play a role in function of lesser versus greater experience in arithmetic. To tackle this important gap in the literature, the current study in typically developing second graders investigated executive functions (by using a more fine-grained operationalisation including inhibition, shifting, and updating through the use of tasks that tap into those different aspects) and metacognition (both general metacognitive knowledge and on-task metacognitive monitoring) as domain-general cognitive factors and symbolic numerical magnitude processing as a domain-specific cognitive factor, examining their unique contributions to arithmetic (i.e. addition and multiplication) using a within-participant design. We specifically recruited second graders to examine these associations in arithmetic operations with which they had extensive experience (addition) and for which they were still in the learning phase (multiplication).

We predicted that executive functioning would be associated with arithmetic (i.e. better executive functioning is associated with better, faster, and more accurate arithmetic performance). Specifically, we expected the strongest association to be between inhibition skills and the arithmetic task with which participants had the most experience (i.e. addition). Second, we predicted that metacognition would be an important predictor of arithmetic (i.e. better metacognitive skills are associated with better arithmetic performance). Based on the suggestion of domain-dependent metacognitive abilities, we predicted that this association would be the strongest for the on-task metacognitive monitoring. Third, we predicted that executive functions, metacognition, and symbolic numerical magnitude processing each would explain unique variance in individual differences in arithmetic.

Method

1 Participants

Participants were 127 Flemish second graders (64 girls and 63 boys; M_{age} = 7 years 11 months, SD = 4 months, range = 7 years 4 months to 8 years 5 months). They all were typically developing children who had no diagnosis of a developmental disorder, and had a dominantly middle-to-high socioeconomic background. For every participant, written informed parental consent was obtained. The study was approved by the social and societal ethics committee of KU Leuven.

2 Materials

Materials consisted of standardized tests, paper-and-pencil tasks, and computer tasks designed with E-Prime 2.0 (Schneider et al., 2002) and Affect 4.0 (Spruyt et al., 2009).

2.1 Arithmetic

Arithmetic was assessed with two single-digit computerized production tasks, namely addition and multiplication. Additions and multiplications were presented in separate tasks. Arithmetic problems with 0 and 1 as one of the operands were excluded. Both the addition and multiplication items comprised all combinations of the numbers 2–9 as operands, and all commutative pairs (e.g. 3 + 9 and 9 + 3) were presented, yielding 64 additions and 64 multiplications. After fixation, stimuli occurred for 2000 ms in white on a black background (Arial font, 72-point size). Afterwards, a black screen appeared until response. Children were asked to answer verbally and as quickly and accurately as possible. This specific protocol was used to foster retrieval and increase the likelihood of making errors (see below). The experimenter registered the response time (RT) and answers via the computer. For each task, stimuli were pseudo-randomly divided into two blocks (i.e. one of each commutative pair in each block). During the second block of each arithmetic task, a specific metacognitive monitoring measure was added to the task (see below). Each task began with 6 practice items to familiarize the children with the task requirements. Performance measures were the average RT and accuracy of the answers, which were calculated for each operation separately.

2.2 Executive functions

Executive functioning was measured with inhibition, shifting, and updating tasks. None of the executive functioning tasks included numerical stimuli, which allowed us to investigate executive functions without numerical processing confounds.

Inhibition. We used a classic arrow Flanker task (Eriksen & Eriksen, 1974; Huizinga et al., 2006), a speeded choice reaction time task where participants needed to respond to target stimuli (i.e. a left or right pointing arrow presented at the centre of the screen) flanked by distractors (i.e. two arrows on each side of the target). These four distractor arrows either pointed in the same direction as the target, and thus indicated the same response (i.e. →→→→→or ←←←←←; congruent condition), or in the opposite direction, and thus indicated the response that was incompatible with the target (i.e. →→←→→ or ←←→←←; incongruent condition). Children needed to indicate the direction of the middle arrow by pressing the corresponding key (i.e. left/right key) while suppressing the direction of the distractors. In total, 40 test items were administered, with 50% of the items being incongruent and the direction of the middle arrow being balanced. After fixation, stimuli appeared for 2500 ms, followed by a black screen that remained visible for 1000 ms. Before the task started, 12 trials were presented—used as a baseline condition for the analyses of this task—on which only one arrow was presented in the centre of the screen (neutral condition). Children needed to indicate the direction of the arrow. RT and accuracy of the answers were registered by the computer.

We also used an animal Stroop task (based on Szűcs et al., 2013). The task consisted of comparing two simultaneously presented images of coloured animals arranged on either side of the centre of the screen. One animal was selected from a set of large animals (e.g. giraffe, moose), and the other was selected from a set of small animals (e.g. rabbit, frog). Participants needed to select the animal that was larger in real life by pressing the corresponding key (i.e. left/right key). They were required to ignore the size of the images on the screen and to respond based on the animal size in real life only. To ensure that all children had the real-world knowledge needed for this task, participants were shown the animal images one by one in a standard size prior to the task and were asked whether the animal was large or small in real life. After this, children were presented with 20 trials—used as a baseline condition for the analyses of this task—in which the animals occupied the same area on the screen (neutral condition). On each test trial, one animal image was presented with an area on the screen four times larger than the other image. This yielded two conditions (Figure 2.1): a congruent condition (i.e. larger animal in real life = larger image on the screen) and an incongruent condition (i.e. larger animal in real life = smaller image on the screen). The position of the larger animal (both in real life and the size of the image) was balanced. In total, 40 test trials were administered, with 50% of the items being incongruent and items being presented in a random order. After fixation, stimuli appeared for 2000 ms, followed by a black screen that remained visible for 1000 ms. RT and accuracy of the answers were registered by the computer.

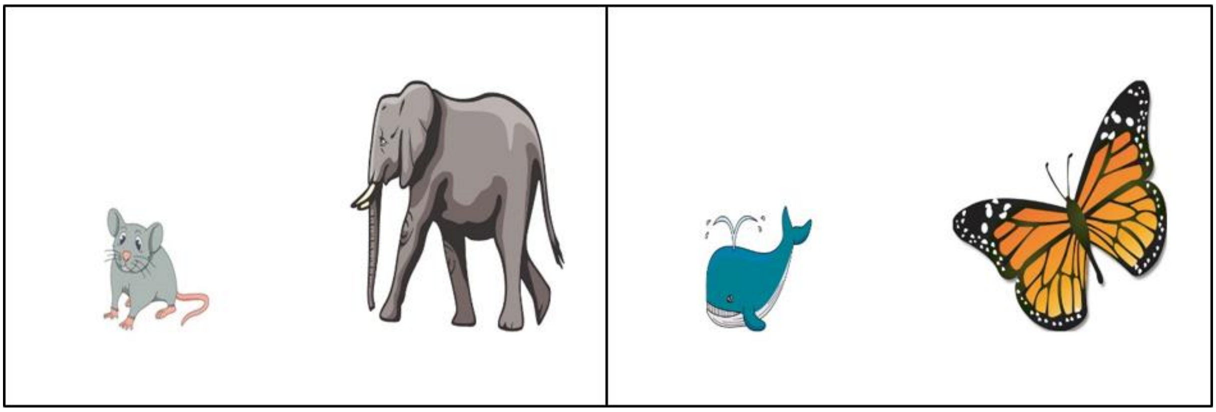

Figure 2.1. Examples of stimuli for the animal Stroop task. Left: congruent condition; Right: incongruent condition.

Because both accuracy and RT constitute essential parts of inhibition, we calculated inverse efficiency (IE) scores (i.e. RT divided by accuracy) for both inhibition tasks. These scores were used as a performance measure for both tasks (see also online supplementary material). The use of two inhibition tasks enabled us to differentiate between the information that needs to be inhibited in the two tasks, i.e. visually, stimulus driven information in the Flanker task vs. linked to acquired knowledge in the animal Stroop task.

Shifting. To measure shifting skills, we used the Wisconsin Card Sorting Task (WCST; Grant & Berg, 1948). In this sorting task, participants needed to determine how to sort cards on the basis of unspecified categories (i.e. shape and colour). We used only two sorting rules from the original WCST to minimize working memory demands (Watson et al., 2006). On each trial, three cards were simultaneously presented on a computer screen, one at the top and two at the bottom (i.e. left and right). Each card consisted of a figure with a specific shape and colour. Children needed to indicate which of the two bottom cards matched the top card. The bottom cards were either identical to the top card in shape but different in colour or identical to the top car in colour but different in shape. On 2 successive items, the cards had different colours and different shapes. Children were given no explicit instructions about the sorting rules; these needed to be inferred based on the feedback that was given after every item. Without notice, the sorting rule changed after a variable number (seven, eight, or nine) of consecutive correct responses. After this switch item, children needed to disengage from the previous sorting rule and discover and/or apply the other sorting rule. The items were randomly divided into five runs, each consisting of five blocks. In each block, a particular sorting rule needed to be applied. A new block began after the sorting rule changed. In total, the sorting rule changed four times in each run; thus, it changed 20 times during the whole task. When a child failed to reach seven, eight, or nine consecutive correct responses within 50 trials, the task was discontinued. To familiarize the children with the task, one practice block was presented in which cards needed to be sorted by shape; after eight consecutive correct responses, this block was completed. Children needed to answer by pressing the corresponding response key (i.e. left/right key). The item remained visible until response. The performance measure was the average number of items a child needed to switch between rules (i.e. total number of items needed to switch divided by number of blocks completed).

Updating skills. Updating was assessed by means of a standard 2-back task (Pelegrina et al., 2015). In this continuous recognition task, a sequence of items is shown. For each item, participants needed to indicate whether the presented stimulus was identical to the stimulus presented 2 trials back by pressing a green or red key, thereby answering yes or no, respectively. Hence, children needed to store two elements in their memory and update these elements given that each time they needed to eliminate the previous item, add the new one, and maintain the presentation order of the items. Items were coloured images (e.g. rocket, sweater) presented one by one in the centre of a white screen. In total, 40 items, divided into two blocks consisting of 20 trials each, were presented. In each block, 30% of the trials were target trials (i.e. correct answer = yes). The first 3 trials of each block were always nontarget items. After fixation, the stimulus appeared for 3000 ms, followed by a black screen that remained visible for 1000 ms. To familiarize the children with the task, one practice block (20 items) was presented. The performance measure was the accuracy of the answer.

2.3 Metacognition

Two aspects of metacognition, namely general metacognitive knowledge and task-specific metacognitive judgments incorporated into the arithmetic task, were included in this study to explore both domain-general and domain-specific aspects of metacognition. Both dimensions on which metacognitive judgments can vary, time of judgments (i.e. prediction vs. postdiction) and granularity of judgments (i.e. local judgments specific to one item vs. global holistic judgments across multiple items), were included (Pieschl, 2009). On the one hand, one local metacognitive question measuring calibration of confidence was asked after every arithmetic item in the second block of the arithmetic task. On the other hand, two global metacognitive questions were asked, one before the start of the arithmetic task and one after the entire task was finished. By selecting confidence judgments as our task-specific metacognitive measure and not asking participants to exhibit control actions (e.g. make modifications to or correct their answers), we focused on the monitoring aspect of metacognition rather than on its control mechanism (Nelson & Narens, 1990).

General metacognitive knowledge. To measure metacognitive abilities independent of arithmetic, we used a general metacognitive questionnaire (Haberkorn et al., 2014). In this questionnaire, 15 situations involving mental performance (e.g. ‘‘Which strategy do you think is better to make sure you won’t forget to take your skates to school the next day?”) were described and three possible answers (e.g. ‘‘Write a note on a piece of paper”, ‘‘Think strongly about the skates”, ‘‘Both proposed strategies are equally good/bad”) were presented. The researcher read the situations and the corresponding options aloud one by one. Children were given a response form with pictures of the three possible answers (see

Figure 2.2) so that they could follow each item and indicate their answer. The performance measure was the number of correct answers.

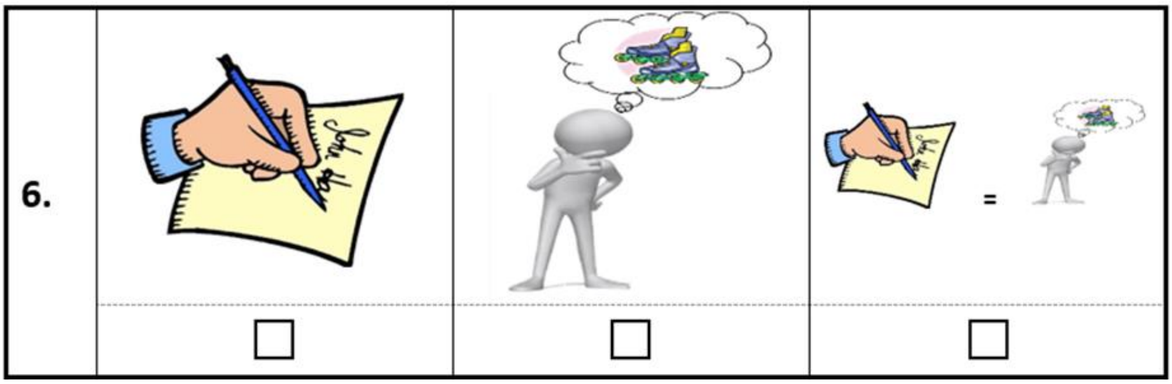

Figure 2.2. Example of the response form of the general metacognitive knowledge questionnaire.

Local on-task metacognitive monitoring. Calibration of confidence (i.e. the alignment between one's confidence in solving a problem and the accuracy of the answer) was measured as in Rinne and Mazzocco (2014) by asking children on a trial-by-trial basis to report their confidence in the accuracy of their answer during the arithmetic tasks. These data were collected for every arithmetic item in the second block of both arithmetic tasks (n = 32 in each task). A specific protocol was used in the arithmetic tasks to foster retrieval and increase the likelihood of making errors and, thus, to maximize the variability in this metacognitive monitoring measure. After giving their answer to the arithmetic problem, children needed to indicate how confident they were that their answer was correct (i.e. ''Correct,'' ''I am not sure,'' or ''Incorrect''). Calibration scores were the alignment between children's confidence rating and the accuracy of their answer (i.e. correct arithmetic answers yielded a score of 2 if children said they were correct, 1 if they said they were not sure, or 0 if they said they were incorrect; this scale was reversed when the arithmetic answer was incorrect). The calibration score per child was the mean of all calibration scores (i.e. calibration score per item; n = 32) and was calculated for each arithmetic task separately. The higher the calibration scores, the better the on-task metacognitive monitoring skills. To familiarize the children with the task, 6 practice items were presented. We used emoticons in combination with the options (e.g. ''☺'' and ''Correct'') to make the task more attractive and feasible for children.

Global domain-specific metacognitive monitoring. Two global metacognitive questions were asked in both arithmetic tasks. Before each task started, children needed to report how they thought they would perform on the task (i.e. prospective question). After the task was finished, they needed to report how they thought they had performed on the task (i.e. retrospective question). During data analyses, we observed that the scoring of these items was problematic, and so we decided to remove these two questions (see supplementary material).

2.4 Numerical magnitude processing

A numerical magnitude comparison task was used to assess children's numerical magnitude processing skills. In view of the meta-analysis by Schneider et al. (2017), which showed that symbolic measures were stronger predictors of mathematical performance than nonsymbolic measures, we administered a symbolic version of the task using Arabic digits as stimuli. This task consisted of comparing two simultaneously presented numerical magnitudes (i.e. Arabic digits) on either side of the centre of the screen. Participants needed to select the numerically larger numerical magnitude by pressing the corresponding response key (i.e. left/right key). Stimuli comprised all combinations of 1–9, yielding 72 trials. The trials were randomly divided into two blocks. After fixation, stimuli appeared until response. The position of the largest digit was balanced. To familiarize the children with the task requirements, 3 practice trials were presented. The performance measure was average RT of correct responses.

2.5 Control variables

Intellectual ability was assessed by means of the Raven's Standard Progressive Matrices (Raven et al., 1992). Children were given 60 multiple-choice items in which they needed to complete a pattern. The performance measure was the number of correctly solved patterns.

A motor speed task was included as a control for children's response speed on the keyboard (De Smedt & Boets, 2010). Two shapes were simultaneously presented on either side of the screen, and children needed to indicate which of the two shapes was filled by pressing the corresponding key (i.e. left/right key). All shapes were similar in size, and each shape occurred four times filled and four times nonfilled, yielding 20 trials. The position of the filled shape was balanced. After fixation, stimuli appeared until response. To familiarize the children with the task, 3 practice trials were included. The performance measure was average RT of correct responses.

3 Procedure

All participants were tested at their own school during regular school hours. They all completed three sessions: an individual session including the arithmetic tasks; a session in small groups of 5 children including the motor speed task, the symbolic numerical magnitude processing task, and the executive functioning tasks; and a group-administered session including the metacognitive questionnaire and intellectual ability task. The sessions took about 40, 45, and 60 min, respectively. All children completed the tasks in the above-mentioned order.

4 Analyses

We ran frequentist analyses using both univariate and multivariate techniques as well as Bayesian analyses. Frequentist analyses allowed us to explore our data by means of a well-known method to gauge statistical support for the hypotheses of interest. However, this *p*-value null hypothesis significance testing has a number of statistical limitations (Andraszewicz et al., 2015). For example, *p*-values cannot quantify evidence in favour of a null hypothesis, they only signal the extremeness of the data under the null hypothesis, and *p*-value logic resembles a proof by contradiction (i.e. low *p*-values indicate extreme data and usually lead researchers to reject the null hypothesis and interpret this as evidence in favour of the alternative hypothesis). Unlike null hypothesis testing, Bayesian statistics allow one to test the degree of support for a hypothesis (i.e. degree of strength of evidence in favour of or against any given hypothesis). This is expressed as the Bayes factor (BF), which is the ratio between the evidence in support of the alternative hypothesis over the null hypothesis (BF_{10}). By comparing the fit of the data under the null hypothesis with the alternative hypothesis, Bayes factors quantify the evidence in favour of these hypotheses (see Andraszewicz et al., 2015). For example, a Bayes factor of 10 ($BF_{10} = 10$) suggests that the alternative hypothesis is 10 times more likely than the null hypothesis. Although Bayes factors provide a continuous measure of degree of evidence, there are some conventional approximate guidelines for interpretation (see Andraszewicz et al., 2015, for a classification scheme); $BF_{10} = 1$ provides no evidence either way, whereas $BF_{10} > 1$ provides anecdotal evidence, $BF_{10} > 3$ provides moderate evidence, $BF_{10} > 10$ provides strong evidence, $BF_{10} > 30$ provides very strong evidence, and $BF_{10} > 100$ provides decisive evidence for the alternative hypothesis. By adding these Bayesian analyses, we deepened our findings from the traditional analyses because we were able to identify which hypothesis is most plausible (i.e. alternative hypothesis vs. null hypothesis) and which predictors are the strongest.

Results

1 Descriptive statistics

The descriptive statistics of the primary measures are presented in Table 2.1. A table with descriptive statistics of all administered measures—including control variables and baseline conditions—can be found in Appendix A.

2

Table 2.1
Descriptive statistics of the key variables

	n	*M*	*SD*	Range	Reliability
Arithmetic					
Addition					
Accuracy	127	.94	.07	[.66-1]	.95
Response time (ms)	127	3654.22	1445.67	[1716.01-8948.59]	.80
Multiplication					
Accuracy	127	.84	.12	[.36-1]	.94
Response time (ms)	127	7141.84	4134.53	[2484.21-31130.19]	.87
Executive functions					
Inhibition					
Flanker task – incongruent condition					
Accuracy	126	.91	.13	[.15-1]	.79
Response time (ms)	126	956.12	236.39	[560.60-1729.25]	.90
Inverse efficiency (RT/accuracy) [a]	126	1115.01	550.80	[604.50-5047.67]	.
Animal Stroop task – incongruent condition					
Accuracy	126	.91	.09	[.60-1]	.58
Response time (ms)	126	1000.10	167.46	[619.25-1393.40]	.80
Inverse efficiency (RT/accuracy)[a]	126	1106.09	196.96	[709.26-1732.23]	.
Shifting					
WCST (average # items needed to switch) [b]	123	7.03	5.01	[2-33]	.
Updating					
2-back task (accuracy)	127	.72	.09	[.15-.85]	.62
Metacognition					
General metacognitive knowledge					
General metacognitive questionnaire (# correct) [c]	127	9.20	2.47	[3-15]	.49
Calibration of confidence [d]					
In addition task	127	1.85	.15	[1.31-2]	.78
In multiplication task	127	1.74	.18	[0.81-2]	.74
Numerical magnitude processing					
Symbolic numerical magnitude comparison task					
Response time (ms)	127	856.51	223.15	[465.73-2037.15]	.92

Note. [a] Inverse efficiency scores were calculated by dividing the response time by the accuracy; the higher the score, the worse the performance; [b] Score = total number of items needed to switch divided by number of blocks completed; [c] Number of correct answers; [d] Alignment between children's confidence rating and the accuracy of their arithmetic answer, i.e. correct arithmetic answers yielded a score of 2 if children said they were *Correct*, 1 if they said *I am not sure*, and 0 if they said they were *Incorrect*; this scale was reversed when the arithmetic answer was incorrect. The higher the score, the better the calibration of confidence.

2 Preliminary analyses

Before conducting the subsequent correlation and regression analyses, we conducted a series of preliminary analyses (e.g. strategy use, performance measures in the inhibition tasks) to rule out potential alternative explanations for the current results. These analyses are included in the supplementary material. None of these alternative explanations accounted for the findings that are reported below.

3 Correlational analyses

Pearson correlation coefficients were calculated to examine the associations between the different variables under study and addition performance (i.e. RT and accuracy [Table 2.2a]), and multiplication performance (i.e. RT and accuracy [Table 2.2b]). A full matrix of all intercorrelations is provided in Appendix B. For the Bayesian analyses, we used a default prior provided by the statistical program JASP (JASP, 2019). The default prior width was set to 1 for Pearson correlations.

Table 2.2a
Correlational analyses of the response time (RT) and accuracy of addition

Variable	Addition RT			Addition accuracy		
	r	*p*	BF_{10}	*r*	*p*	BF_{10}
Executive functions						
Inhibition – Flanker (IE) [a,b]	-.156	.083	0.501	-.040	.659	0.123
Inhibition – Stroop (IE) [a, c]	.120	.183	0.270	-.188	.035	1.002
Shifting [a] (average # items needed to switch)	-.050	.585	0.131	.002	.987	0.113
Updating (accuracy) [d]	-.101	.259	0.208	.341	<.001	>100
Metacognition						
General metacognitive knowledge (# correct) [d]	-.243	.006	4.701	.114	.201	0.249
Calibration of confidence – Addition [d]	-.289	.001	23.385	.502	<.001	>100
Symbolic numerical magnitude processing (RT)[a, e]	.361	<.001	>100	.028	.757	0.116

Note. [a] The higher the scores, the worse the performance; [b] Controlled for the Flanker task baseline condition (IE); [c] Controlled for the animal Stroop task baseline condition (IE); [d] The higher the score, the better the performance; [e] Controlled for performance on the motor speed task (RT).

2

Table 2.2b

Correlational analyses of the response time (RT) and accuracy of multiplication

Variable	Multiplication RT			Multiplication accuracy		
	r	*p*	BF_{10}	*r*	*p*	BF_{10}
Executive functions						
Inhibition – Flanker (IE) [a,b]	-.142	.114	0.386	.085	.349	0.173
Inhibition – Stroop (IE) [a, c]	.001	.981	0.111	-.081	.371	0.167
Shifting [a] (average # items needed to switch)	-.090	.324	0.182	-.060	.508	0.140
Updating (accuracy) [d]	.053	.556	0.132	.278	.002	15.478
Metacognition						
General metacognitive knowledge (# correct) [d]	-.147	.099	0.426	.077	.391	0.160
Calibration of confidence – Addition [d]	-.238	.007	4.058	.791	<.001	>100
Symbolic numerical magnitude processing (RT)[a, e]	.251	.005	6.086	.008	.929	0.111

Note. [a] The higher the scores, the worse the performance; [b] Controlled for the Flanker task baseline condition (IE); [c] Controlled for the animal Stroop task baseline condition (IE); [d] The higher the score, the better the performance; [e] Controlled for performance on the motor speed task (RT).

3.1 Executive functions

Performance (IE) on the animal Stroop task was significantly correlated with addition accuracy, indicating that children with better inhibition skills performed more accurately on the addition task. The Bayes factor indicated only anecdotal evidence for this association. The n-back task (accuracy) correlated with both addition and multiplication accuracy, with Bayes factors indicating strong to decisive evidence. This correlation was driven by omission errors (and not commission errors) in the n-back task. There were no other significant correlations with the remaining executive functioning variables. Moreover, the Bayes factors indicated evidence in favour of the null hypotheses for these remaining associations.

3.2 Metacognition

General metacognitive knowledge was significantly associated with addition RT, indicating that children with better global metacognitive knowledge performed faster on the addition task. The Bayes factor indicated moderate evidence in favour of this association. There was no evidence for the null hypothesis or for the alternative hypothesis for the association between multiplication RT and general metacognitive knowledge. The null hypotheses of no association was supported for the associations between general metacognitive knowledge and addition and multiplication accuracy. Calibration of confidence was significantly correlated with addition and multiplication performance, indicating that children with better on-task metacognitive monitoring skills performed better (i.e. faster and more accurately) on the arithmetic tasks. Bayes factors indicated moderate to decisive evidence for these associations.

3.3 Symbolic numerical magnitude processing

Addition and multiplication RT were significantly related to symbolic numerical magnitude processing; children with better symbolic numerical magnitude processing skills performed faster when doing arithmetic. Bayes factors indicated moderate to decisive evidence. The null hypotheses were supported for the associations between addition and multiplication accuracy and symbolic numerical magnitude processing RT.

4 Regression analyses

Regression analyses were performed to assess the unique contribution of our different cognitive variables to arithmetic performance. Therefore, all variables that were significantly related to addition and/or multiplication were entered simultaneously into the regression model. To quantify the evidence in favour of our hypotheses, we calculated Bayes factors for each predictor and identified the best model to predict each dependent variable in JASP (JASP, 2019). Specifically, a $BF_{inclusion}$ was calculated for every predictor in the model, which represents the change from prior to posterior odds (i.e. BF_{10}), where the odds concern all the models with a predictor of interest to all models without that predictor (i.e. a Bayes factor for including a predictor averaged across the models under consideration). We used a default prior width provided by JASP of .354 (prior for *r* scale covariates) for the linear regression analyses. Table 2.3 presents the results of our regression analyses.

Table 2.3a
Regression analysis of addition RT (R^2 = .225)

Variable	β	t	p	ΔR^2	$BF_{inclusion}$
Control variables					
Motor speed task (RT) [a]	-.114	-1.321	.189	.024	0.474
Intelligence (# correct) [b]	.020	0.247	.806	0	0.254
Primary variables					
General metacognitive knowledge (# correct) [b]	-.178	-2.143	.034	.028	1.493
Calibration of confidence – Addition [b]	-.291	-3.651	<.001	.082	76.675
Symbolic numerical magnitude processing (RT) [a]	.356	4.080	<.001	.103	>100

Note. [a] The higher the score, the worse the performance; [b] The higher the score, the better the performance.

Table 2.3b

Regression analysis of addition accuracy (R^2 = .338)

Variable	β	t	p	ΔR^2	$BF_{inclusion}$
Control variables					
Stroop task Baseline (IE) [a]	.023	0.250	.803	0	0.224
Intelligence (# correct) [b]	.176	2.237	.027	.028	3.885
Primary variables					
Inhibition – Stroop task (IE) [a]	-.071	-0.731	.466	.003	0.300
Updating (accuracy) [b]	.202	2.576	.011	.037	4.673
Calibration of confidence – Addition [b]	.414	5.238	<.001	.151	>100

Note. [a] The higher the scores, the worse the performance; [b] The higher the score, the better the performance.

Table 2.3c

Regression analysis of multiplication RT (R^2 = .186)

Variable	β	t	p	ΔR^2	$BF_{inclusion}$
Control variables					
Motor speed task (RT) [a]	-.079	-0.860	.392	.005	0.36
Intelligence (# correct) [b]	.223	2.635	.010	.047	4.84
Primary variables					
General metacognitive knowledge (# correct) [b]	-.118	-1.358	.177	.012	0.63
Calibration of confidence – Multiplication [b]	-.271	-3.202	.002	.069	16.74
Symbolic numerical magnitude processing (RT) [a]	.232	2.540	.012	.044	6.69

Note. [a] The higher the score, the worse the performance; [b] The higher the score, the better the performance.

Table 2.3d

Regression analysis of multiplication accuracy (R^2 = .661)

Variable	β	t	p	ΔR^2	$BF_{inclusion}$
Control variables					
Stroop task Baseline (IE) [a]	.006	0.086	.932	0	0.279
Intelligence (# correct) [b]	.119	2.115	.036	.013	1.539
Primary variables					
Inhibition – Stroop task (IE) [a]	.024	0.353	.725	.001	0.260
Updating (accuracy) [b]	.131	2.376	.019	.016	1.572
Calibration of confidence – Multiplication [b]	.759	13.906	<.001	.547	>100

Note. [a] The higher the scores, the worse the performance; [b] The higher the score, the better the performance.

4.1 Addition

Addition RT was significantly predicted by general metacognitive knowledge, calibration of confidence, and symbolic numerical magnitude processing (RT). Bayes factors indicated very strong to decisive evidence for the latter two variables, whereas the Bayes factor for general metacognitive knowledge indicated only anecdotal support for the alternative hypothesis. The strongest model to predict addition RT was the model including general metacognitive knowledge, calibration of confidence, and symbolic numerical magnitude processing (BF_{10} = 85387.75). The R^2 of the overall model was medium.

Addition accuracy was significantly predicted by the 2-back task (accuracy), calibration of confidence, and intelligence, and the evidence for the contribution of these predictors was moderate to decisive. There was moderate evidence for the null hypothesis concerning the influence of the animal Stroop task (IE). The strongest model to predict addition accuracy was the model including intelligence, updating, and calibration of confidence (BF_{10} = 1.270e+8). The R^2 of the overall model was large.

4.2 Multiplication

Multiplication RT was significantly predicted by calibration of confidence, symbolic numerical magnitude processing (RT), and intelligence. Bayes factors indicated moderate to strong evidence for these predictors. General metacognitive knowledge did not predict multiplication RT. The strongest model to predict multiplication RT was the model including intelligence, calibration of confidence, and symbolic numerical magnitude processing (BF_{10} = 464.81). The R^2 of the overall model was medium.

Multiplication accuracy was significantly predicted by the 2-back task (accuracy), calibration of confidence, and intelligence. Only for calibration of confidence did the Bayes factors indicate decisive evidence. The evidence for the contribution of the 2-back task accuracy and intelligence was only anecdotal. The null hypothesis was supported for no influence of the animal Stroop task (IE) in multiplication accuracy. The strongest model to predict multiplication accuracy was the model including updating and calibration of confidence (BF_{10} = 2.102e+25). The R^2 of the overall model was large.

Discussion

Why are some children so good in math and do others struggle their entire lives with basic math competencies? A large body of literature has already identified different cognitive factors that play an important role in mathematical development (e.g. Berch et al., 2016; Dowker, 2005; Duncan et al., 2007). These factors can be subdivided into two groups, namely domain-general cognitive factors and domain-specific cognitive factors (e.g. Geary & Moore, 2016). Existing research has largely focused on one domain-specific cognitive factor, numerical magnitude processing (Schneider et al., 2017), neglecting other critical domain-general cognitive factors that might play a role in (a)typical mathematical development. The current study included two domain-general factors, executive functions and metacognition, to identify their unique contribution in arithmetic in addition to each other as well as in addition to symbolic numerical magnitude processing as an important domain-specific factor. This study closes an important gap in the literature by focusing on executive functioning in a comprehensive way (i.e. including inhibition, shifting, and updating) while also investigating the role of metacognition. By combining these two factors, this study sheds new and important light on how these different variables cooperate in arithmetic. In addition, to thoroughly investigate the role of executive functioning, the current study took into account different degrees of experience with arithmetic, high (addition) versus low (multiplication), because the level of experience might moderate the association between the cognitive factors with arithmetic (e.g. Laski & Dulaney, 2015). On the other hand, we included both general metacognitive knowledge and on-task metacognitive monitoring to investigate the extent to which domain-general and domain-specific measures of metacognition were related to arithmetic.

Metacognition, and specifically on-task metacognitive judgments, was found to be a strong, stable, and unique predictor of both addition and multiplication. In line with Rinne and Mazzocco (2014), we found that local metacognitive judgments were strongly related with concurrent mental arithmetic. By uncovering its unique role in arithmetic above executive functions and symbolic numerical magnitude processing, and by testing this association in early primary school children, we extended Rinne and Mazzocco's (2014) findings in an important way. The current cross-sectional data do not allow us to derive strong conclusions about the direction of the association between metacognition and arithmetic. It might be that better performance on arithmetic has a positive influence on metacognitive skills and that more experience with correct arithmetic answers (i.e. due to better arithmetic skills) makes detecting incorrect answers easier, which leads to better later metacognitive performance. This hypothesis is in line with the results of Roebers and Spiess (2017), who found that earlier domain-specific skills in a first-order task (i.e. spelling) predicted later metacognitive monitoring abilities. On the other hand, it might be that better detection of errors (i.e. due to better metacognitive skills) creates more learning moments in which incorrect arithmetic answers are detected and new, potentially correct arithmetic answers are learned, which in turn leads to better later arithmetic performance. This hypothesis is in line with the results of Rinne and Mazzocco (2014), who found that better calibration predicted later gains in mental arithmetic. Lastly, the relationship between metacognition and arithmetic performance might also be reciprocal, whereby arithmetic performance and metacognition both influence each other. Future longitudinal studies are needed to investigate this.

In contrast to the associations between on-task metacognitive skills and arithmetic which were unique and strong, general metacognitive knowledge was significantly correlated with addition RT, but once symbolic numerical magnitude processing RT was considered, there was no clear evidence for an effect of this factor. This might suggest that children’s metacognition is more domain specific, which is in line with observations by Vo et al. (2014) and Neuenhaus et al. (2011). Importantly, the general metacognitive questionnaire differed in two considerable ways from the local, on-task metacognitive monitoring measure. Compared with the general metacognitive questionnaire, on-task metacognitive monitoring was measured online and in more detail (i.e. on a trial-by-trial basis). This could limit the interpretation of specificity of metacognition based on our results. The obtained results might be specific to the age of our participants given that it is theoretically assumed that the development of metacognitive knowledge and skills begins highly domain and situation specific and becomes more flexible and domain general with practice and experience (Borkowski et al., 2000).

This study also investigated different components of executive functioning, namely inhibition, shifting, and updating. Based on the lack of significant correlations between our executive functioning measures (see Appendix B), our results confirm that inhibition, shifting, and updating are separate aspects of executive functioning. This result is in line with existing research (e.g. Bull & Scerif, 2001; Cragg & Gilmore, 2014; Lee et al., 2013; Miyake et al., 2000; van der Sluis et al., 2007) and indicates that in the operationalisation of executive functions, it is essential to differentiate between different components.

In line with previous research (e.g. Bull & Scerif, 2001; Cragg & Gilmore, 2014; Peng et al., 2016), we found that updating accuracy was a stable and unique predictor of both addition and multiplication accuracy. Updating skills may be required to recall arithmetic facts from long-term memory, in line with evidence suggesting that individuals with low updating skills are less likely to retrieve answers to simple arithmetic problems (e.g. Barrouillet & Lépine, 2005; Geary et al., 2012). On the other hand, arithmetic accuracy may rely on updating skills in order to represent the arithmetic item and to store interim solutions while using procedural strategies in multistep problems.

Whereas our frequentist analyses pointed to a significant—yet weak—association between addition accuracy and inhibition, the Bayes factor indicated that the hypothesis of an association can be neither confirmed nor rejected. Consequently, based on the results of the current study, no firm conclusions can be drawn on the association between inhibition and addition accuracy. On the other hand, the Bayes factors of the association between inhibition and multiplication indicate that there is evidence in favour of no association. Based on the current results, there is not much support for a time-limited role of inhibition in arithmetic performance, as we originally predicted. Future research is needed to further investigate the association between inhibition and arithmetic.

We did not observe an association between shifting and arithmetic. Moreover, Bayes factors indicated moderate evidence in favour of no association between these variables. This result is in line with previous research (e.g. Bull & Lee, 2014; van der Sluis et al., 2007; Van der Ven et al., 2012; Yeniad et al., 2013) but differs from other existing literature (e.g. Bull et al., 1999). This difference could be due to the difference in the operationalisation of mathematics given that different mathematical abilities all could be differently related to certain cognitive factors. On the other hand, the need for shifting (e.g. between strategies) was limited in our single-digit arithmetic task because of the curricular focus on retrieval of single-digit arithmetic (De Smedt, 2016) and the homogeneity of the arithmetic tasks (i.e. addition and multiplication items were presented in separate tasks, requiring no shifts between operations in a task). The impact of shifting skills on arithmetic might become more clear when different math components are considered together. The absence of a significant correlation of shifting skills with arithmetic in this study does not indicate their lack of influence in arithmetic, yet it suggests that individual differences in shifting skills are not a strong explanation for individual differences in single-digit arithmetic.

Whereas both executive functions and on-task metacognitive monitoring contributed to arithmetic performance, on-task metacognitive monitoring proved to be more important. This could be due to the fact that the sensitivity of the executive function tasks to capture individual differences was lower because these tasks have low between-participant variability and, hence, their usefulness in studying individual differences might be limited (Hedge et al., 2017).

In line with a large body of research (see Schneider et al., 2017, for a meta-analysis), the current findings support the unique role of symbolic numerical magnitude processing in arithmetic. The current data clearly indicate that such associations do not merely arise as a result of a common reliance on executive functions (e.g. Gilmore et al., 2013).

The results of this study revealed that there is overlap in the cognitive processes that are associated with addition and multiplication performance (e.g. updating skills accuracy, on-task metacognitive monitoring, symbolic numerical magnitude processing RT) but that the strength of the association differed across the two operations. This indicates that subtle differences (e.g. level of experience) within one mathematical skill (i.e. arithmetic) could account for different associations with other cognitive variables. It is important to note that the difference between addition and multiplication is not purely one of experience and that there could also be differences in task demands. On the other hand, the kind of instruction the children received in the Belgian school system, namely the curriculum’s strong emphasis on automatization of arithmetic facts for both operations (*Onderwijsdoelen Vlaanderen*, 2018), makes these task demands rather similar.

Future research should examine these associations investigated in the current study longitudinally to examine the directions of the associations and how such associations evolve over time. Supplementary reaction time-based analyses (see supplementary material) indicated that different strategies were used to solve the arithmetic items and that indices reflective of retrieval and procedure use both were correlated with on-task metacognitive judgments and symbolic magnitude processing. However, these metrics might mask important individual differences in strategy use, and future research should investigate strategy use at a trial-by-trial level, for example, by collecting verbal report data.

Conclusion

To conclude, the results of this study show that updating, metacognition, and numerical magnitude processing all are significant predictors of arithmetic performance even in addition to each other. Importantly, these results emphasize the importance of including metacognition in cognitive research. Metacognition has received a lot of attention in more general educational research yet is mostly ignored in developmental research in the field of mathematics. Importantly, this study reveals that noticing one's own errors (i.e. task-specific metacognitive monitoring) is an important unique predictor of arithmetic performance. Knowing about one's own tendency to commit easy errors may lead to increased self-regulatory activities (Schneider, 2010) and to improvements in (arithmetic) performances over time (Rinne & Mazzocco, 2014). Hence, these individual differences in noticing one's own errors might help one to learn from his or her mistakes.

Appendixes

1 Appendix A – Descriptive statistics of all the administered measures

Table 2.A1
Descriptive statistics of all the administered measures

	n	*M*	*SD*	Range
Arithmetic				
Addition				
Accuracy	127	.94	.07	[.66-1]
Response time (ms)	127	3654.22	1445.67	[1716.01-8948.59]
Multiplication				
Accuracy	127	.84	.12	[.36-1]
Response time (ms)	127	7141.84	4134.53	[2484.21-31130.19]
Executive functions				
Inhibition				
Flanker – incongruent condition				
Accuracy	126	.13	.91	[.15-1]
Response time (ms)	126	956.12	236.39	[560.60-1729.25]
Inverse efficiency [a]	126	1115.01	550.80	[604.50-5047.67]
Flanker – baseline condition				
Accuracy	126	.94	.08	[.60-1]
Response time (ms)	126	647.82	87.39	[453.00-887.78]
Inverse efficiency [a]	126	689.36	101.26	[479.80-986.42]
Animal Stroop – incongruent condition				
Accuracy	126	.91	.09	[.60-1]
Response time (ms)	126	1000.10	167.46	[619.25-1393.40]
Inverse efficiency [a]	126	1106.09	196.96	[709.26-1732.23]
Animal Stroop – baseline condition				
Accuracy	126	.97	.05	[.70-1]
Response time (ms)	126	867.47	155.13	[508.05-1297.38]
Inverse efficiency [a]	126	895.11	182.83	[508.05-1651.02]
Shifting				
WCST (average # items needed to switch) [b]	123	7.03	5.01	[2-33]
Updating				
2-back task (accuracy)	127	.72	.09	[.15-.85]
Metacognition				
General metacognitive knowledge				
General metacognitive questionnaire (# correct) [c]	127	9.20	2.47	[3-15]
Calibration of confidence [d]				
In addition task	127	1.85	.15	[1.31-2]
In multiplication task	127	1.74	.18	[0.81-2]
Numerical magnitude processing				
Symbolic numerical magnitude comparison task				
Response time (ms)	127	856.51	223.15	[465.73-2037.15]
Control variables				
Intelligence (# correct) [c]	127	34.50	8.35	[10-50]
Motor speed task (RT)	126	593.86	153.92	[339.75-1151.56]

Note. [a] Calculated by dividing the RT by the accuracy; the higher the scores, the worse the performance; [b] Score = total number of items needed to switch divided by number of blocks completed; [c] Number of correct answers; [d] Alignment between children's confidence rating and the accuracy of their arithmetic answer, i.e. correct arithmetic answers yielded a score of 2 if children said they were *Correct*, 1 if they said *I am not sure*, and 0 if they said they were *Incorrect*; this scale was reversed when the arithmetic answer was incorrect. The higher the scores, the better the calibration of confidence.

2 Appendix B – Correlational analyses between the administered measures

Table 2.B1

Correlational analyses between the administered measures

			1a	1b	2a	2b	3	4	5	6	7	8	9	10	11
Arithmetic	Addition	1a. Addition RT													
		1b. Addition ACC													
		r	-.247												
		p	.005												
		BF_{10}	5.317												
	Multiplication	2a. Multiplication RT													
		r	.558	.153											
		p	<.001	.086											
		BF_{10}	>100	0.476											
		2b. Multiplication ACC													
		r	-.525	.413	-.168										
		p	<.001	<.001	.058										
		BF_{10}	>100	>100	0.653										
Executive functions	Inhibition	3. Flanker task IE													
		r	-.156	-.040	-.142	.085									
		p	.083	.659	.114	.349									
		BF_{10}	0.501	0.123	0.386	0.173									
		4. Stroop task IE													
		r	.120	-.188	.002	-.081	.220								
		p	.183	.035	.981	.371	.015								
		BF_{10}	0.270	1.002	0.111	0.167	2.254								
	Shifting	5. WCST													
		r	-.050	.002	-.090	-.060	.121	-.081							
		p	.585	.987	.324	.508	.186	.377							
		BF_{10}	0.131	0.113	0.182	0.140	0.274	0.167							
	Updating	6. 2-back task ACC													
		r	-.101	.341	.053	.278	-.118	.014	-.091						
		p	.259	<.001	.556	.002	.192	.877	.319						
		BF_{10}	0.208	>100	0.132	15.478	0.262	0.113	0.184						

			1a	1b	2a	2b	3	4	5	6	7	8	9	10	11
Metacognition	General metacognitive knowledge	7. Metacognitive questionnaire													
		r	-.243	.114	-.147	.077	-.145	-.116	.022	.084					
		p	.006	.201	.099	.391	.107	.199	.809	.345					
		BF_{10}	4.701	0.249	0.426	0.160	0.408	0.255	0.116	0.172					
	Task-specific metacognition	8. Calibration of confidence - Addition													
		r	-.289	.502	-.030	.396	-.104	-.205	.082	.262	.065				
		p	.001	<.001	.735	<.001	.250	.022	.369	.003	.468				
		BF_{10}	23.385	>100	0.117	>100	0.216	1.531	0.168	8.843	0.144				
		9. Calibration of confidence - Multiplication													
		r	-.451	.286	-.238	.791	.039	-.092	-.082	.170	.055	.507			
		p	<.001	.001	.007	<.001	.666	.310	.367	.056	.536	<.001			
		BF_{10}	>100	20.796	4.058	>100	0.122	0.187	0.169	0.675	0.134	>100			
Domain-specific factor	Numerical magnitude processing	10. Symbolic numerical magnitude processing RT													
		r	.361	.028	.251	.008	-.081	-.010	-.004	.021	-.158	-.013	.052		
		p	<.001	.757	.005	.929	.369	.909	.963	.816	.079	.887	.564		
		BF_{10}	>100	0.116	6.086	0.111	0.167	0.112	0.113	0.114	0.516	0.112	0.131		
Control Variables	Intelligence	11. Raven													
		r	.034	.275	.221	.221	-.310	-.275	-.084	.158	.079	.118	.115	.146	
		p	.705	.002	.013	.012	<.001	.002	.357	.076	.379	.186	.199	.105	
		BF_{10}	0.119	14.099	2.392	2.449	53.106	13.534	0.171	0.525	0.163	0.263	0.251	0.415	
	Motor speed	12. Motor reaction time task RT													
		r	.087	.036	.098	-.207	-.058	.117	-.136	-.186	-.242	-.083	-.184	.375	.055
		p	.331	.690	.273	.020	.521	.195	.134	.037	.006	.358	.039	<.001	.544

Note. ACC = accuracy; All correlations with the Flanker task incongruent condition are controlled for the performance (IE) on the Flanker baseline condition; All correlations with the animal Stroop task incongruent condition are controlled for the performance (IE) on the animal Stroop baseline task; All correlations with the numerical magnitude processing task are controlled for performance (RT) on the motor speed task, except for the correlations with the inhibition tasks, which are controlled for their respective baseline (IE).

Supplementary materials

1 Global metacognitive questions

Two global metacognitive questions were asked in both arithmetic tasks. Before the task started, children had to report how they thought they would perform on the task (i.e. prospective question). After the task was finished, they had to report how they thought they had performed on the task (i.e. retrospective question). We used emoticons corresponding with the possible answers (e.g. ☺ corresponded with "Good") to make the task more attractive and feasible for children.

Our rationale for the scoring of this task was the same as for the local, on-task metacognitive monitoring, namely giving the best score (i.e. 2) to correct extreme judgements (i.e. "Good" or "Not good"), the lowest score (i.e. 0) to incorrect extreme judgements, and an average score (i.e. 1) when the child did not make an extreme judgement. However, to define the correctness of the judgement, an indicator of actual performance has to be taken into account. In preparatory analyses, we tried to find such an indicator (1) by comparing children's performance to the performance of all participants (i.e. their percentile score), and (2) by comparing their performance to certain standards (i.e. accuracy of 1-.71 = Good; .70-.51 = Average; .50-0 = Not good). The children, however, were not aware of these indicators, and therefore they were not able to effectively rate their performance. It could be that they were thinking about "What does my teacher thinks good performance is?" or "What do I think the experimenter thinks about my performance?", or that they compared to some internal reference point (e.g. their typical performance on similar tasks). On the other hand, the use of one single item to measure a variable is not reliable.

Against this background, we decided to remove these two questions (i.e. prediction and postdiction) from further analyses.

2 Preliminary analyses

We report the outcome of four preliminary analyses that were run to rule out potential alternative explanations for the current results that are discussed in our manuscript.

2.1 Blocks in the arithmetic tasks

The results of arithmetic were based on the data of both blocks of each task, i.e. one without and one with metacognitive monitoring question. It could be that asking a metacognitive question after solving arithmetic items, had an influence on arithmetic performance. We therefore also analysed our data based only on the arithmetic block that did not include metacognitive questions. The interpretation of the results was identical to when the results were based on both blocks together. Consequently, the results presented in the manuscript are based on the full dataset (i.e. block with and without metacognitive monitoring questions).

2.2 Strategy use within the arithmetic tasks

The RTs of the arithmetic tasks indicated that it is likely that different strategies were used to solve the arithmetic items. We used different methods to estimate strategy use (i.e. procedural strategies vs. arithmetic fact retrieval). We calculated a Tau estimate which reflects the positive skew of the distribution of response times of a participant, with increases in Tau reflecting a slowing on some trials (Penner-Wilger & Lefevre, 2006). As a result, Tau can be used as an index of procedure use. On the other hand, we used the data of the small arithmetic items (i.e. items with an answer smaller or equal to 10 for the addition items, and items with an answer smaller or equal to 25 for the multiplication items) as an index of retrieval. This problem-size approach to investigate strategy use has been widely used in the literature (De Smedt, 2016) and the assumption that retrieval strategies are typically used more often to solve small problems has been confirmed by empirical data in children (e.g. Barrouillet, Mignon, & Thevenot, 2008; Imbo & Vandierendonck, 2008). We calculated average RTs of the small items and a Tau estimate for every participant, and correlated these two with the other key variables in our study (see Table 2.S1 and Table 2.S2 below). Importantly, as they are based on RT, these estimates of strategy use do not provide information about the associations between our key variables and arithmetic accuracy.

Table 2.S1
Correlational analyses of the response time (RT) of the small items as an index of retrieval and the Tau estimate as an index of procedural use in the addition task

Variable	Small			Tau		
	r	*p*	BF_{10}	*r*	*p*	BF_{10}
Executive functions						
Inhibition: Flanker (IE) [a, b]	-.051	.571	0.131	-.092	.310	0.187
Inhibition: Stroop (IE) [a, c]	.105	.243	0.219	.189	.035	1.026
Shifting (average # items needed to switch) [d]	-.006	.945	0.115	-.074	.423	0.157
Updating (accuracy)	-.120	.179	0.271	-.111	.215	0.238
Metacognition						
General metacognitive knowledge (# correct) [e]	-.218	.014	2.132	-.202	.023	1.393
Calibration of confidence – Addition [e]	-.228	.011	2.769	-.216	.016	1.978
Symbolic numerical magnitude processing (RT)[f]	.379	<.001	>100	.258	.005	5.854

Note. [a] IE = inverse efficiency of the incongruent condition, calculated by dividing the response time by the accuracy; the higher the score, the worse the performance [b]; Controlled for the Flanker task baseline condition (IE); [c] Controlled for the animal Stroop task baseline condition (IE); [d] The higher the score, the worse the performance; [e] The higher the score, the better the performance; [f] Controlled for performance on the motor speed task (RT).

Table 2.S2
Correlational analyses of the response time (RT) of the small items as an index of retrieval and the Tau estimate as an index of procedural use in the multiplication task

Variable	Small			Tau		
	r	*p*	BF_{10}	*r*	*p*	BF_{10}
Executive functions						
Inhibition: Flanker (IE) [a, b]	-.078	.386	0.162	-.086	.340	0.338
Inhibition: Stroop (IE) [a, c]	.041	.648	0.123	.069	.443	0.149
Shifting (average # items needed to switch) [d]	-.014	.879	0.116	-.052	.576	0.134
Updating (accuracy)	-.056	.530	0.135	-.013	.886	0.112
Metacognition						
General metacognitive knowledge (# correct) [e]	-.140	.117	0.371	-.137	.125	0.352
Calibration of confidence – Multiplication [e]	-.513	<.001	>100	-.434	<.001	>100
Symbolic numerical magnitude processing (RT)[f]	.288	.002	16.069	.360	<.001	>100

Note. [a] IE = inverse efficiency of the incongruent condition, calculated by dividing the response time by the accuracy; the higher the score, the worse the performance [b]; Controlled for the Flanker task baseline condition (IE); [c] Controlled for the animal Stroop task baseline condition (IE); [d] The higher the score, the worse the performance; [e] The higher the score, the better the performance; [f] Controlled for performance on the motor speed task (RT).

As can be seen in tables 2.S1 and 2.S2, the pattern of results is largely the same for both strategy indicators in both operations. Consequently, based on these estimates, we conclude that our results were not likely to be confounded by strategy use and, therefore, no distinction in strategy use was made in the results presented in the manuscript. However, the analyses using these estimates should be interpreted with caution, as they do not consider strategy use on a trial-by-trial basis (Siegler, 1987) and children could be using other strategies than assumed based on the assumptions made to calculate these estimates. Future studies should therefore measure the strategy that participants use to solve each problem via trial-by-trial strategy reports.

2

2.3 Task-specific metacognitive monitoring

The score of task-specific metacognitive monitoring included accuracy, but accuracy was very high in the arithmetic tasks. As children tend to be overconfident (Destan & Roebers, 2015) - indeed also in this study children often indicated that they thought they were right, which in most cases (i.e. combined with accurate arithmetic answers) resulted in the highest metacognitive monitoring score - our results might be biased. We therefore also analysed the metacognitive monitoring data based on the items with incorrect arithmetic answers only. The interpretation of the results did not change when only the incorrectly solved arithmetic items were considered. Therefore, all results in the paper are based on the full dataset.

2.4 Inhibition tasks

In the inhibition tasks, i.e. Flanker task and animal Stroop task, there were different conditions, namely a baseline or neutral condition (i.e. Flanker: only one arrow was presented in the centre of the screen; Stroop: animals occupied the same area on the screen), a congruent condition (i.e. Flanker: distracter arrows pointed in the same direction as the target; Stroop: larger animal in real life = larger image on the screen) and an incongruent condition (i.e. Flanker: distracter arrows pointed in the opposite direction as the target; Stroop: larger animal in real life = smaller image on the screen). Descriptive statistics for the different conditions are presented in Table 2.S3 below. To check if the task manipulation in the inhibition tasks worked, we verified if the performance (IE) on the incongruent condition differed from the performance (IE) on the baseline condition. This was the case for both the Flanker task ($t(125) = -9.03$, $p < .001$) and the animal Stroop task ($t(125) = 13.54$, $p < .001$). There was no facilitation-effect (i.e. better performance in the congruent condition than the baseline condition) in neither the Flanker task nor the animal Stroop task. Performance was even significantly poorer in the congruent condition in both the Flanker task ($t(125) = -9.772$, $p < .001$) and the animal Stroop task ($t(125) = -3.338$, $p < .001$) compared to the baseline condition. Adding another dimension to the task (i.e. different picture sizes) clearly made the task more difficult in both the congruent and incongruent conditions. This might be due to inhibitory demands in both conditions and therefore, we did not use the difference score (i.e. incongruent condition minus congruent condition) as index for inhibition. For all analyses presented in the manuscript, we used the IE scores of the incongruent condition controlled for the IE score of their respective baseline condition via residuals using partial correlation.

Table 2.S3

Descriptive statistics of the different conditions (i.e. neutral, congruent and incongruent) in the inhibition tasks

	n	*M*	*SD*	Range
Flanker task				
Baseline task				
Neutral condition (IE) [a,b]	216	689.36	101.26	[479.80-986.42]
Inhibitory task				
Congruent condition (IE) [a,c]	126	840.16	200.15	[493-1730.53]
Incongruent condition (IE) [a,d]	126	1115.01	550.80	[604.50-5047.67]
Animal Stroop task				
Baseline task				
Neutral condition (IE) [a,e]	126	895.11	185.83	[508.05-1651.02]
Inhibitory task				
Congruent condition (IE) [a,f]	126	942.49	208.66	[560.53-1790.44]
Incongruent condition (IE) [a,g]	126	1106.09	196.96	[709-1732.23]

Note. [a] IE = inverse efficiency, calculated by dividing the response time by the accuracy; the higher the scores, the worse the performance; [b] One arrow in the middle of the screen; [c] Distractor arrows in the same direction as the target; [d] Distractor arrows in the opposite direction as the target; [e] Both animals occupied the same area on the screen; [f] Larger animal in real life = larger image on the screen; [g] Larger animal in real life = smaller image on the screen.

CHAPTER 3

What does arithmetic count on?

A longitudinal panel study on the roles of numerical magnitude processing, executive functions and metacognition in primary school.

The content of this chapter is under revision as:

Bellon, E., Fias, W., & De Smedt, B. (Under Revision). What does arithmetic count on?: A longitudinal panel study on the roles of numerical magnitude processing, executive functions and metacognition in primary school. *Journal of numerical cognition.*

Chapter 3

What does arithmetic count on?

A longitudinal panel study on the roles of numerical magnitude processing, executive functions and metacognition in primary school.

Abstract

While there is a wealth of studies on the associations between arithmetic and individual (meta)cognitive functions such as numerical magnitude processing, executive functions and metacognition, far fewer studies investigate these diverse functions simultaneously to explore their unique roles in arithmetic. Such research is needed, because these three functions have been shown to be interrelated. Even fewer studies have examined these associations with arithmetic longitudinally, thereby hardly accounting for prior arithmetic performance in their design. The current study used a longitudinal panel design to simultaneously investigate the roles of numerical magnitude processing, executive functions and metacognition in arithmetic performance and development during an important developmental period in which all these functions develop (7-9 years-olds). Participants were 121 typically developing children, who were tested on the abovementioned functions and arithmetic in second and third grade of primary school. Our results demonstrate that symbolic numerical magnitude processing and metacognition have unique predictive roles for later arithmetic. The data indicate that executive functions are not so strong predictors of arithmetic, when different other important (meta)cognitive functions are considered. Prior arithmetic performance remains to be the most robust predictor for later arithmetic performance. The results of this study emphasize the need to investigate (meta)cognitive correlates of arithmetic simultaneously and with a longitudinal panel design to obtain a more thorough understanding of functions playing a role in individual differences in arithmetic performance and development.

Introduction

Decades of research on the correlates of mathematics development in children has led to the identification of several important (cognitive) functions that are involved in mathematics performance and its development. Important examples of these functions are numerical magnitude processing (e.g. De Smedt, Noël, Gilmore, & Ansari, 2013, for a review; Schneider et al., 2017, for a meta-analysis), executive functioning (e.g. Bull & Lee, 2014; Cragg & Gilmore, 2014; Friso-Van Den Bos, Van Der Ven, Kroesbergen, & Van Luit, 2013) and metacognition (e.g. Garofalo & Lester, 1985; Lucangeli & Cornoldi, 1997; Rinne & Mazzocco, 2014; Schneider & Artelt, 2010; Schoenfeld, 1987; van der Stel, Veenman, Deelen, & Haenen, 2010).

Firstly, numerical magnitude processing, defined as children's elementary intuitions about quantity and the ability to understand the meaning of numbers, has been reliably identified as correlating with and predictive of individual differences in mathematics performance and development (e.g. De Smedt et al., 2013; Schneider et al., 2017). These findings indicate that the better individuals' numerical magnitude processing skills are, the better their (concurrent) mathematical performance is. This was especially true for symbolic numerical magnitude processing compared to non-symbolic numerical magnitude processing (Schneider et al., 2017). The relations between the processing of non-symbolic and symbolic number and their development, constitute one of the most debated topics in numerical cognition, yet the existing body of evidence converge to the conclusion that symbolic abilities are the most critical for the development of mathematics (e.g., Merkley & Ansari, 2016). In (early) primary school, numerical magnitude processing skills substantially develop as children become faster and more accurate at numerical magnitude processing tasks, especially symbolic tasks (e.g. Matejko & Ansari, 2016; Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013).

Secondly, executive functions refer to top-down mental processes that allow us to respond flexibly to our environment and engage in deliberate, goal-directed, thought and action (Cragg & Gilmore, 2014). In accordance with Miyake et al. (2000), executive functioning is typically operationalised as consisting of the processes of inhibition, shifting and updating. Inhibition refers to one's ability to control one's attention, behaviour, and thoughts to override a strong internal predisposition or external lure and instead do what is more appropriate or necessary (Diamond, 2013). Shifting is defined as the disengagement from an irrelevant task set or strategy, and the subsequent initiation of a new, more appropriate set (van der Sluis, de Jong, & van der Leij, 2007). Updating involves holding information in memory and flexibly manipulating it (Baddeley & Hitch, 1994). As there is both unity and diversity in these executive functions (Miyake et al., 2000), and as different components of executive functions are differently related to mathematics (Bull & Lee, 2014), it is of utmost importance to include all three aspects of executive functioning when examining their role in other processes such as mathematics. Executive functions in general, and updating in particular, are found to be correlated with and predictive of mathematical performance and development (e.g. Bull & Lee, 2014; Cragg & Gilmore, 2014; Friso-Van Den Bos et al., 2013), indicating better executive functioning skills are associated with better mathematics skills. Examples of possible underlying mechanisms driving these associations are the importance of storage and retrieval of partial results during problem-solving processes or the suppression of inappropriate strategies or interfering information (e.g. Bull & Lee, 2014). From very early in life onwards, and throughout primary school, major advances in executive functioning occur (e.g. Carlson, Zelazo, & Faja, 2013; Diamond, 2013; Huizinga, Dolan, & van der Molen, 2006), providing children with better inhibition, shifting and updating skills.

Thirdly, metacognition, which was first introduced by Flavell (1979), is often broadly defined as 'thinking about your thinking'. It encompasses both knowledge about your cognition (e.g. Brown, 1978; Flavell, 1979) and how people monitor and control their cognition on-task (e.g. Nelson & Narens, 1990). The relation between metacognition and mathematics has been extensively studied (e.g. Rinne & Mazzocco, 2014; Schoenfeld, 1992; Stillman & Mevarech, 2010; van der Stel et al., 2010) and there is a long tradition of research investigating metacognition in mathematics education (e.g. De Corte, Verschaffel, & Op 't Eynde, 2000; Schneider & Artelt, 2010). These studies show that successful mathematics performance depends not only on having adequate knowledge, but also on having sufficient awareness, monitoring and control of that knowledge (e.g. Garofalo & Lester, 1985). Metacognitive knowledge and skills develop substantially in primary school (e.g. Lyons & Ghetti, 2010; Roebers & Spiess, 2017; Schneider, 2008, 2010), resulting in better general metacognitive knowledge (e.g. Schneider & Löffler, 2016) and improved monitoring and control skills (e.g. Garrett, Mazzocco, & Baker, 2006; Schneider & Lockl, 2008). Over its extended course of development in which metacognition becomes increasingly under the individual's conscious control, metacognition becomes more explicit, powerful and effective.

The role of these functions in mathematical performance has shown to vary depending on the mathematical domain under investigation, for example, for the roles of numerical magnitude processing (Schneider et al., 2017) and working memory (Peng et al. 2016). Consequently, when studying these functions, they should be investigated in relation to subcomponents of mathematics performance rather than to general mathematics achievement tests, which include diverse mathematical domains. Therefore, the current study specifically focusses on arithmetic.

Because a multitude of (meta)cognitive functions have an impact on arithmetic performance (Dowker, 2019c), such as numerical magnitude processing, executive functions and metacognition, it is essential to investigate such functions simultaneously. The majority of the abovementioned studies on the associations between these functions and arithmetic have mainly focused on associations with one (meta)cognitive function, and only very few studies investigated the role of each of these different functions *simultaneously* in arithmetic performance. For example, much less is known about whether, on the one hand, executive functioning continues to predict mathematics skills after taking numerical magnitude processing into account, and, on the other hand, to what extend numerical magnitude processing itself and its association with arithmetic is determined by more domain-general processes such as executive functioning and metacognition.

The limited research that did include different (cognitive) functions, found that this simultaneous investigation had an important impact on the results. For example, in their meta-analysis, Chen and Li (2014) found that the overall effect size of non-symbolic magnitude comparison in mathematical competence was significantly lower in studies controlling for general non-numerical cognitive abilities compared to studies not controlling for them. Likewise, Schneider et al. (2017) suggested that including other cognitive abilities (such as inhibition) in regression models might have a similar effect on the association between symbolic numerical magnitude processing and mathematics performance. Simanowski & Krajewski (2019) found that, after controlling for early numerical magnitude processing, executive functions in kindergarten were no longer predictive of mathematics skills in first and second grade. Furthermore, there is large theoretically overlap between executive functions and metacognition. Both are higher-order, control processes related to the regulation of behaviour, they follow a similar developmental trajectory, and studies investigating both functions simultaneously suggest that executive functions and metacognition are related (see Roebers & Feurer, 2015, for a short overview). Although there are only a small number of studies that investigate executive functions and metacognition simultaneously, these studies show that metacognitive skills are stronger predictors of academic performance (e.g. Bellon, Fias, & De Smedt, 2019; Bryce, Whitebread, & Szűcs, 2015; Roebers, Cimeli, Röthlisberger, & Neuenschwander, 2012). Despite these observations, studies including a variety of (cognitive) functions are scarce, such that for future studies it is essential include various (cognitive) abilities when investigating the associations between them and arithmetic performance.

In accordance with this suggestion, Bellon et al. (2019) simultaneously investigated symbolic numerical magnitude processing, executive functions and metacognition in 7-8-year olds. They found unique roles of symbolic magnitude skills, metacognitive monitoring and updating in arithmetic performance, in addition to each other. On the other hand, no associations between shifting or inhibition and arithmetic were found. This study thus showed that when considered together, symbolic numerical magnitude processing, some aspects of executive functioning, and metacognition each explain unique variance in arithmetic skills in young primary school children.

Yet, the study by Bellon et al. (2019), and by extension most other studies on associations between (meta)cognitive functions and arithmetic performance, only reports on concurrent relations, leaving important issues unresolved. First, it remains unknown whether the associations between arithmetic and these functions, when considered simultaneously, are stable over development. Second, it is unclear whether these (meta)cognitive functions (i.e. numerical magnitude processing, executive functions and metacognition) also predict later arithmetic performance, and, finally, going one step further, whether they do so when prior arithmetic performance is taken into account and thus whether they contribute to the development of arithmetic skills.

When the associations between arithmetic and numerical magnitude processing, executive functions and metacognition are studied in isolation from each other, longitudinal studies have confirmed the predictive power for later arithmetic performance of each of these (cognitive) functions separately. For example, several studies found evidence for the predictive value of numerical magnitude processing for later arithmetic (e.g. De Smedt, Verschaffel, & Ghesquière, 2009; Sasanguie, Van den Bussche, & Reynvoet, 2012; Vanbinst, Ghesquière, & De Smedt, 2015). In their meta-analysis, Schneider et al. (2017) showed that these associations are stronger for symbolic than for non-symbolic numerical magnitude processing. Within literature on the executive functions as well, there is evidence is for the predictive value of executive functions, especially updating skills, for later arithmetic performance (e.g. De Smedt, Janssen, et al., 2009; Lee & Bull, 2016; Mazzocco & Kover, 2007; Passolunghi, Mammarella, & Altoè, 2008; Van der Ven, Kroesbergen, Boom, & Leseman, 2012). While studies on the longitudinal associations between metacognition and arithmetic are scarce, the few available studies confirm the predictive power of metacognition for children's later arithmetic (e.g. Rinne & Mazzocco, 2014; van der Stel & Veenman, 2010).

As has been argued above when discussing the available cross-sectional studies on the associations between these functions and arithmetic, these longitudinal studies also focus merely on one (meta)cognitive function to predict later arithmetic performance. As such, they fail to identify the unique contributions of such functions when other critical (meta)cognitive functions that predict arithmetic are considered. Even more critical, most of these longitudinal studies fail to include prior arithmetic performance as an important predictor in their models, and hence do not investigate the importance of these functions relative to prior arithmetic performance. This is crucial, as extensive evidence has demonstrated that early academic performance is a robust indicator of later performance (e.g. Duncan et al., 2007). Moreover, including prior arithmetic performance importantly yields the possibility to investigate the predictive power of these functions for development in arithmetic (Duncan et al., 2007). Indeed, to thoroughly study whether a (meta)cognitive function at time point 1 predicts arithmetic at time point 2, it is necessary to control for arithmetic at time point 1, otherwise concurrent correlations between the (meta)cognitive function and arithmetic at both time points may confound the investigated predictive association. Namely, the predictive association between (meta)cognitive functions at time point 1 and arithmetic at time point 2 may then be entirely driven by the relationship between arithmetic at both time points (i.e. its stability).

The current study tackles these important issues in the extant literature by, on the one hand, simultaneously studying the role of symbolic numerical magnitude processing, executive functions and metacognition in arithmetic and, at the same time, providing a longitudinal follow up of children from second to third grade of primary school, of which the results of the first time point (i.e. second grade) have been published (Bellon et al., 2019). Using a longitudinal panel design, the current study aims to unravel (1) the stability of the associations found in early primary school over time and, importantly, (2) uncover the longitudinal associations of symbolic numerical magnitude processing, executive functions and metacognition with arithmetic, and (3) investigate these predictive associations when taking prior arithmetic performance into account.

By following up second graders one year later, we specifically investigated these issues in a crucial developmental period for the investigated (meta)cognitive functions – as was discussed above – as well as arithmetic. At this age, arithmetic development is characterized by the transition from initial effortful strategies to solve basic arithmetic items (e.g. counting strategies), to efficient arithmetic strategies (e.g. decomposition) or automation through arithmetic fact retrieval (e.g. Siegler, 1996). Substantial development in children of this age is also present in symbolic numerical magnitude processing (e.g. Matejko & Ansari, 2016), executive functions (e.g. Carlson et al., 2013; Diamond, 2013) and metacognitive skills (e.g. Schneider, 2010, 2015; Schneider & Lockl, 2008). Because of these developmental progressions, it is likely that the interrelations between arithmetic and symbolic numerical magnitude processing, executive functions and metacognition might change during this developmental period (e.g. Bull & Lee, 2014; Van der Ven et al., 2012). Investigating the interplay of these variables in this critical developmental period is essential to foster learning and develop targeted intervention programs.

Methods

1 Participants

Participants were 121 Flemish third graders (63 girls; M_{age} = 8 years 8 months, SD = 3 months, range = 8 years 2 months to 9 years 2 months). All children were participants of an ongoing longitudinal panel study (*n* = 127) of which the first time point has been published (Bellon et al., 2019); in the current study, the children were followed up one year later. All children were typically developing and they had no diagnosis of a developmental disorder. They all had a predominantly middle-to-high socioeconomic background. For every participant, written informed parental consent was obtained. The study was approved by the social and societal ethics committee of KU Leuven.

2 Procedure

The procedure and tasks in this longitudinal follow up were exactly the same as those at time point one, the findings of which were reported in Bellon et al. (2019). All participants were tested at their own school during regular school hours and all completed three sessions: an individual session including the arithmetic tasks, a session in small groups of 5 children including the computerized cognitive tasks, and a group-administered session including the metacognitive questionnaire and the measure of intellectual ability. The sessions took about 30, 40, and 60 min, respectively. All children completed the tasks in the same order.

3 Materials

3.1 Arithmetic

Arithmetic was assessed with two single-digit computerized production tasks, namely addition and multiplication, presented in separate tasks (i.e. 64 items for each task). Arithmetic items were presented on the computer screen and children were asked to answer verbally, as quickly and accurately as possible. For each task, stimuli were pseudo-randomly divided into two blocks (i.e. one of each commutative pair in each block). During the second block of each arithmetic task, a specific metacognitive monitoring measure was added to the task (see below). Performance measures were accuracy of the answers and average RT for correct responses, which were calculated for each operation separately.

3.2 Executive functions

Executive functioning was measured with computerized inhibition, shifting, and updating tasks. None of the executive functions tasks included numerical stimuli, which allowed us to investigate executive functions without numerical processing confounds.

3
3.2.1 Inhibition.

We used a classic arrow Flanker task (Eriksen & Eriksen, 1974; Huizinga et al., 2006), a speeded choice reaction time task where participants needed to respond to target stimuli (i.e. a left or right pointing arrow presented at the centre of the screen) flanked by distractors (i.e. two arrows on each side of the target). Items were part of the congruent condition if all arrows pointed in the same direction; items were part of the incongruent condition if the distractors pointed to the opposite direction as the target. Congruent and incongruent items were presented interchangeably within the task. In total 40 items were presented. Children needed to indicate the direction of the middle arrow while suppressing the direction of the distractors. In the baseline condition only one arrow was shown so the task was the same, but suppressing information was not necessary.

An animal Stroop task (based on Szűcs, Devine, Soltesz, Nobes, & Gabriel, 2013) was additionally administered. In this task children had to indicate which animal of two simultaneously presented images of coloured animals was larger in real life. One animal image was presented with an area on the screen four times larger than the other image, yielding two conditions (i.e. congruent condition in which the larger animal in real life was the larger image on the screen; incongruent in which the larger animal in real life was the smaller image on the screen). In total 40 items were administered. Children were required to ignore the size of the images on the screen and to respond based on the animal size in real life only. In the baseline condition all animals on screen had the same size so that the task was the same, but suppressing information was not necessary.

Because both accuracy and RT constitute essential parts of inhibition, we calculated inverse efficiency (IE) scores (i.e. average RT for correct answers divided by average accuracy) for both inhibition tasks. Based on the previous study with these tasks in this population (Bellon et al., 2019), the IE of items in the incongruent condition were used as performance measure. The IE of items in the baseline conditions were used as a control in the analyses.

3.2.2 Shifting.

To measure shifting skills, we used the Wisconsin Card Sorting Task (WCST; Grant & Berg, 1948), in which participants needed to determine how to sort cards on the basis of unspecified categories (i.e. shape and colour). Children were given no explicit instructions about the sorting rules; these needed to be inferred based on the feedback that was given after every item. Without notice, the sorting rule changed after a variable number (seven, eight, or nine) of consecutive correct responses. After this switch item, children needed to disengage from the previous sorting rule and discover and/or apply the other sorting rule. The sorting rule changed 20 times during the whole task. The performance measure was the average number of items a child needed to switch between rules.

3.2.3 Updating skills.

Updating was assessed by means of a standard 2-back task (Pelegrina et al., 2015). In this continuous recognition task, a sequence of items is shown. For each item, participants had to indicate whether the presented stimulus was identical to the stimulus presented 2 trials back. Two blocks of 20 items were administered. The performance measure was the accuracy of the answer.

3.3 Metacognition

Two aspects of metacognition, namely general metacognitive knowledge and metacognitive judgments incorporated into the arithmetic task, were included in this study.

3.3.1 General metacognitive knowledge.

To measure metacognitive abilities independent of arithmetic, we used a general metacognitive questionnaire (Haberkorn et al., 2014), in which 15 situations involving mental performance were described together with two possible strategies. Children indicated which of the two strategies they thought fitted the situation the best, or they could indicate that both solutions were equally good. The performance measure was the number of correct answers.

3.3.2 Metacognitive monitoring.

On a trial-by-trial basis, children were asked to judge the accuracy of their answer (e.g. Bellon et al., 2019; Rinne & Mazzocco, 2014) during the second block of both arithmetic tasks (32 items for each operation). After giving their answer to the arithmetic problem, children needed to indicate if they thought their answer was “Correct”, “Incorrect” or if they “Did not know”. The alignment between one’s judgment on the accuracy of the answer and the accuracy of the answer itself was calculated as a measure of metacognitive monitoring: A correct arithmetic answers yielded a score of 2 if children said they were correct, 1 if they said they did not know, or 0 if they said they were incorrect. This scale was reversed when the arithmetic answer was incorrect. The metacognitive monitoring score per child was the mean of all monitoring scores (i.e. monitoring score per item; n = 32) and was calculated for each arithmetic task separately.

3.4 Numerical magnitude processing

A symbolic numerical magnitude comparison task was used to assess children’s symbolic numerical magnitude processing skills. Participants had to indicate which of two simultaneously presented Arabic digits was numerically larger. The performance measure was average RT of correct responses on the 72 administered trials.

3.5 Control variables

Intellectual ability was assessed by means of the Raven’s Standard Progressive Matrices (Raven et al., 1992). The performance measure was the number of correctly solved patterns. A *motor speed* task was included as a control for children’s response speed on the keyboard (De Smedt & Boets, 2010).

Two shapes were simultaneously presented on either side of the screen, and children needed to indicate which of the two shapes was filled. The performance measure was average RT of correct responses in the 20 trials.

4 Data analysis

We ran frequentist and Bayesian analyses using both univariate and multivariate techniques. Hence, we were able to explore our data by means of a well-known method to gauge statistical support for the hypotheses of interest (i.e. frequentist statistics), while at the same time allowing us to test the degree of support for a hypothesis (i.e. degree of strength of evidence in favour of or against any given hypothesis; i.e. using Bayes Factors). Bayes factors were interpreted following the classification scheme in (Andraszewicz et al., 2015).

For the Pearson correlation analyses, we used a default prior width set to 1, provided by the statistical program JASP (JASP, 2019). In the regression analyses, to quantify the evidence in favour of our hypotheses, a $BF_{inclusion}$ was calculated for every predictor in the model. The $BF_{inclusion}$ represents the change from prior to posterior odds (i.e. BF_{10}), where the odds concern all the models with a predictor of interest to all models without that predictor (i.e. a Bayes factor for including a predictor averaged across the models under consideration). We used a default prior width provided by JASP of .354 (prior for *r* scale covariates) for all linear regression analyses.

Results

The results on time point 1 (T_1) have been reported in Bellon et al. (2019). Below the results of the study on time point 2 (T_2) and the results of the longitudinal analyses (i.e. combining data from T_1 and T_2) can be found.

1 Cross-sectional analyses at T_2

1.1 Descriptive statistics

The descriptive statistics of the key variables measured at T_2 are presented in Table 3.1. A table with descriptive statistics of all administered measures at T_2 – including control variables and baseline conditions – can be found in Appendix A.

Table 3.1
Descriptive statistics of the key variables at T_2

	n	*M*	*SD*	Range
Arithmetic				
Addition				
Accuracy	121	.97	.03	[.84-1.00]
Response time (ms)	121	2577.92	846.50	[1410.31-6102.02]
Multiplication				
Accuracy	121	.93	.07	[.66-1.00]
Response time (ms)	121	4341.55	1955.00	[1474.02-14792.50]
Numerical magnitude processing				
Symbolic numerical magnitude comparison task				
Response time (ms)	121	745.59	132.90	[520.17-1212.37]
Executive functions				
Inhibition				
Flanker task – incongruent condition				
Accuracy	121	.93	.08	[.58-1.00]
Response time (ms)	121	776.33	169.02	[505.74-1392.89]
Inverse efficiency (RT/accuracy)[a]	121	856.32	235.74	[548.68-1784.65]
Animal Stroop task – incongruent condition				
Accuracy	121	.93	.07	[.75-1.00]
Response time (ms)	121	926.27	179.71	[560.90-1758.65]
Inverse efficiency (RT/accuracy)[a]	121	1006.15	221.05	[623.22-2344.87]
Shifting				
WCST (average # items needed to switch)[b]	119	5.62	4.90	[2-46]
Updating				
2-back task (accuracy)	121	.75	.06	[.47-.87]
Metacognition				
General metacognitive knowledge				
General metacognitive questionnaire (# correct)[c]	121	11.01	2.27	[3-15]
Metacognitive monitoring [d]				
In addition task	121	1.88	0.13	[1.44-2.00]
In multiplication task	121	1.86	0.13	[1.34-2.00]

Note. All response time variables are average response time for the correct responses. [a] Inverse efficiency scores were calculated by dividing the response time by the accuracy; the higher the score, the worse the performance; [b] Score = total number of items needed to switch divided by number of blocks completed; [c] Number of correct answers; [d] Alignment between children's metacognitive judgment and the accuracy of their arithmetic answer, i.e. correct arithmetic answers yielded a score of 2 if children said they were *Correct*, 1 if they said *I am not sure*, and 0 if they said they were *Incorrect*; this scale was reversed when the arithmetic answer was incorrect. The higher the score, the better the metacognitive monitoring.

3

1.2 Correlational analyses

Pearson correlation coefficients were calculated to examine the associations between the different variables under study and addition (Table 3.2a), and multiplication performance (Table 3.2b). A full matrix of all intercorrelations is provided in Appendix B.

Table 3.2a
Correlation analyses of the accuracy and response time (RT) of addition

Variable	Addition accuracy			Addition RT		
	r	*p*	BF_{10}	*r*	*p*	BF_{10}
Symbolic numerical magnitude processing (RT)[a, b]	.07	.43	0.16	.46	<.001	>100
Executive functions						
Inhibition – Flanker (IE)[a,c]	-.12	.19	0.27	.09	.32	0.19
Inhibition – Stroop (IE) [a, d]	-.12	.18	0.28	.04	.67	0.13
Shifting [a] (average # items needed to switch)	.05	.63	0.13	-.01	.96	0.12
Updating (accuracy) [e]	.10	.30	0.19	-.25	.006	4.62
Metacognition						
General metacognitive knowledge (# correct) [e]	.14	.13	0.36	-.31	.001	36.27
Metacognitive monitoring – Addition [e]	.48	<.001	>100	-.38	<.001	>100

Note. [a] The higher the scores, the worse the performance; [b] Controlled for performance on the motor speed task (RT); [c] Controlled for the Flanker task baseline condition (IE); [d] Controlled for the animal Stroop task baseline condition (IE); [e] The higher the score, the better the performance.

3

Table 3.2b

Correlation analyses of the accuracy and response time (RT) of multiplication

Variable	Multiplication accuracy			Multiplication RT		
	r	*p*	BF_{10}	*r*	*p*	BF_{10}
Symbolic numerical magnitude processing (RT)[a, b]	-.22	.02	1.88	.35	<.001	>100
Executive functions						
Inhibition – Flanker (IE)[a,c]	-.18	.05	0.82	.04	.63	0.13
Inhibition – Stroop (IE) [a, d]	-.04	.64	0.13	.11	.24	0.23
Shifting [a] (average # items needed to switch)	-.15	.10	0.44	.13	.17	0.29
Updating (accuracy) [e]	.01	.95	0.11	-.09	.30	0.19
Metacognition						
General metacognitive knowledge (# correct) [e]	.14	.14	0.34	-.16	.08	0.51
Metacognitive monitoring – Multiplication [e]	.58	<.001	>100	-.36	<.001	>100

Note. [a] The higher the scores, the worse the performance; [b] Controlled for performance on the motor speed task (RT); [c] Controlled for the Flanker task baseline condition (IE); [d] Controlled for the animal Stroop task baseline condition (IE); [e] The higher the score, the better the performance.

1.2.1 Numerical magnitude processing.

Symbolic numerical magnitude processing was significantly related to addition and multiplication RT: Children with better symbolic numerical magnitude processing skills performed faster when doing arithmetic. Bayes factors indicated decisive evidence for these associations. There was anecdotal evidence for the association between symbolic numerical magnitude processing and multiplication accuracy. The null hypothesis was supported for the association between addition accuracy and symbolic numerical magnitude processing RT.

1.2.2 Executive functions.

Accuracy on the updating task was significantly correlated with addition RT, indicating that children with better updating skills were faster to give a correct response in the addition task. The Bayes factor indicated moderate evidence for this association. There were no other significant and supported correlations with the remaining executive functioning variables. Moreover, the Bayes factors indicated evidence in favour of the null hypotheses for all these associations.

1.2.3 Metacognition.

General metacognitive knowledge was significantly associated with addition RT, indicating that children with better global metacognitive knowledge performed faster on the addition task. The Bayes factor indicated very strong evidence in favour of this association. There was anecdotal evidence for the null hypothesis for the associations of addition and multiplication accuracy, and multiplication RT with general metacognitive knowledge.

Metacognitive monitoring skills were significantly related to performance (i.e. accuracy and RT) in both arithmetic tasks, with Bayes factors indicating decisive evidence. This suggests that children with better on-task metacognitive monitoring skills performed better (i.e. faster and more accurate) in the arithmetic tasks.

1.3 Regression analyses

Regression analyses were performed to assess the unique contribution of the different (meta)cognitive functions to concurrent arithmetic performance: All variables that were significantly related to addition and/or multiplication performance and for which Bayesian analyses indicated more than anecdotal evidence in favour of an association were entered simultaneously into the regression model. Table 3.3 presents the results of our regression analyses.

3

Table 3.3a
Regression analyses of arithmetic accuracy

Variable	Addition accuracy (R^2 = .26)				Multiplication accuracy (R^2 = .36)			
	β	*t*	*p*	$BF_{inclusion}$	*β*	*t*	*p*	$BF_{inclusion}$
Intelligence (# correct)[a]	.19	2.44	.02	5.52	.18	2.40	.02	4.36
Metacognitive monitoring [a, b]	.46	5.84	<.001	>100	.55	7.48	<.001	>100

Note. [a] The higher the score, the better the performance. [b] Metacognitive monitoring in the addition task for addition accuracy, Metacognitive monitoring in the multiplication task for multiplication accuracy.

Table 3.3b
Regression analyses of arithmetic RT

Variable	Addition RT (R^2 = .39)				Multiplication RT (R^2 = .27)			
	β	*t*	*p*	$BF_{inclusion}$	*β*	*t*	*p*	$BF_{inclusion}$
Motor speed task (RT)[a]	.04	0.44	.66	0.62	.14	1.65	.10	1.40
Symbolic numerical magnitude processing (RT)[a]	.44	5.63	<.001	>100	.34	3.95	<.001	>100
Updating (accuracy) [b]	-.10	-1.33	.19	1.12	.01	0.17	.86	0.51
General metacognitive knowledge [b]	-.14	-1.92	.06	2.55	-.04	-0.42	.67	0.56
Metacognitive monitoring [b, c]	-.29	-3.80	<.001	>100	-.32	-3.97	<.001	>100

Note. [a] The higher the scores, the worse the performance; [b] The higher the score, the better the performance; [c] Metacognitive monitoring in the addition task for addition RT, Metacognitive monitoring in the multiplication task for multiplication RT.

1.3.1 Arithmetic accuracy.

Accuracy in both addition and multiplication was significantly predicted by metacognitive monitoring and intelligence. The evidence for the contribution of metacognitive monitoring was decisive.

1.3.2 Arithmetic RT.

Response time in both addition and multiplication was significantly predicted by symbolic numerical magnitude processing and metacognitive monitoring. The evidence for the contribution of these predictors was decisive for both symbolic numerical magnitude processing and metacognitive monitoring. Anecdotal evidence was found for the predictive power of general metacognitive knowledge for addition RT. There was no evidence in favour or against an association between addition RT and updating. Anecdotal evidence for the null hypothesis was found concerning the associations of multiplication RT with updating and with general metacognitive knowledge.

1.4 Interim summary cross-sectional results

A comparison of the results at T_1 (Bellon et al., 2019), when the participants were in second grade (i.e. 7-8-years-old), with the results of the current follow up at T_2 (i.e. third grade, 8-9-years-old) revealed that the associations found between symbolic numerical magnitude processing and arithmetic, between addition RT and general metacognitive knowledge, and between metacognitive monitoring and arithmetic are stable effects, found at both time points. Less stable effects were found for the executive functions: Updating was no longer significantly correlated with addition accuracy, but was significantly correlated with addition RT at $T_{2,}$ although there was no evidence for the unique predictive power of updating for concurrent addition RT. Additionally, updating was no longer significantly correlated with multiplication performance at T_2. While at T_1 the result on the association of addition with performance on the animal Stroop task was indecisive (i.e. with a Bayes factor of 1 for the association), the results at T_2 clearly indicate evidence for the null hypothesis.

2 Longitudinal correlational analyses

Pearson correlation coefficients were calculated to examine the longitudinal associations between the different variables measured at T_1 (i.e. second grade) and addition (Table 3.4a) and multiplication performance (Table 3.4b) measured at T_2 (i.e. third grade). A full matrix of all longitudinal intercorrelations is provided in Appendix C.

Table 3.4a

Longitudinal correlations of the accuracy and response time (RT) of addition at T_2

Variable	Addition accuracy T_2			Addition RT T_2		
	r	*p*	BF_{10}	*r*	*p*	BF_{10}
Arithmetic T_1						
Prior addition performance [a]	.47	<.001	>100	.78	<.001	>100
Symbolic numerical magnitude processing (RT) T_1 [b, f]	.12	.19	0.27	.41	<.001	>100
Executive functions T_1						
Inhibition – Flanker (IE) [b, c]	-.11	.24	0.23	-.01	.93	0.12
Inhibition – Stroop (IE) [b, d]	-.11	.25	0.22	.08	.42	0.16
Shifting [b] (average # items needed to switch)	.06	.55	0.14	-.08	.39	0.17
Updating (accuracy) [e]	.17	.06	0.76	-.10	.25	0.23
Metacognition T_1						
General metacognitive knowledge (# correct) [e]	.01	.94	0.11	-.32	<.001	72.89
Metacognitive monitoring – Addition [e]	.33	<.001	76.80	.32	<.001	62.01

Note. [a] Addition accuracy T_1 for addition accuracy T_2, Addition RT T_1 for addition RT T_2; [b] The higher the scores, the worse the performance; [c] Controlled for the Flanker task baseline condition (IE); [d] Controlled for the animal Stroop task baseline condition (IE); [e] The higher the score, the better the performance; [f] Controlled for performance on the motor speed task (RT).

Table 3.4b

Longitudinal correlations of the accuracy and response time (RT) of multiplication at T_2

Variable	Multiplication accuracy T_2			Multiplication RT T_2		
	r	*p*	BF_{10}	*r*	*p*	BF_{10}
Arithmetic T_1						
Prior multiplication performance [a]	.54	<.001	>100	.64	<.001	>100
Symbolic numerical magnitude processing (RT) T_1 [b, f]	-.004	.97	0.11	.34	<.001	>100
Executive functions T_1						
Inhibition – Flanker (IE) [b, c]	-.01	.92	0.12	-.03	.74	0.12
Inhibition – Stroop (IE) [b, d]	-.14	.14	0.33	.04	.70	0.12
Shifting [b] (average # items needed to switch)	-.15	.11	0.42	.04	.64	0.13
Updating (accuracy) [e]	.19	.03	1.27	.01	.92	0.11
Metacognition T_1						
General metacognitive knowledge (# correct) [e]	.16	.08	0.52	-.30	.001	32.87
Metacognitive monitoring – Multiplication [e]	.40	<.001	>100	-.21	.02	1.55

Note. [a] Multiplication accuracy T_1 for multiplication accuracy T_2, Multiplication RT T_1 for multiplication RT T_2; [b] The higher the scores, the worse the performance; [c] Controlled for the Flanker task baseline condition (IE); [d] Controlled for the animal Stroop task baseline condition (IE); [e] The higher the score, the better the performance; [f] Controlled for performance on the motor speed task (RT).

3

2.1 Numerical magnitude processing

Symbolic numerical magnitude processing was significantly correlated with later RT in arithmetic, indicating that children with better symbolic numerical magnitude processing skills were faster to give a correct answer in an addition and in a multiplication task one year later. The null hypotheses were supported for the associations between symbolic numerical magnitude processing and later arithmetic accuracy.

2.2 Executive functions

Updating measured at T_1 was significantly related to multiplication accuracy at T_2, yet, the Bayes factor indicated there was only anecdotal evidence for this association. Associations with all other executive functions were non-significant and all Bayes factors indicated evidence for the null hypotheses.

2.3 Metacognition

General metacognitive knowledge was significantly related to later arithmetic RT, with Bayes factors indicating very strong evidence in favour of these associations. General metacognitive knowledge did not significantly correlate with arithmetic accuracy, with Bayes factors indicating strong to anecdotal evidence in favour of the null hypotheses.

Metacognitive monitoring was significantly correlated with arithmetic performance one year later, indicating that children with better metacognitive monitoring skills early in primary school, performed better (i.e. faster and more accurate) in arithmetic one year later. Bayes factors indicated there was strong evidence in favour of the correlation between metacognitive monitoring and later addition performance (i.e. accuracy and RT) and decisive evidence for later multiplication accuracy. There was only anecdotal evidence in favour of the association between metacognitive monitoring and later multiplication RT.

3 Longitudinal regression analyses

Longitudinal regression analyses were performed to assess the unique contribution of the different cognitive variables to arithmetic performance one year later: All variables that were significantly related to addition and/or multiplication performance and for which Bayesian analyses indicated more than anecdotal evidence in favour of an association, were entered simultaneously into a first regression model (Model 1). In a second regression model (Model 2), respective prior arithmetic performance was added to the model. Table 3.5 presents the results of these regression analyses.

Table 3.5a
Regression analyses of arithmetic accuracy

Variables T_1	Addition accuracy T_2				Multiplication accuracy T_2			
	β	t	p	$BF_{inclusion}$	B	t	p	$BF_{inclusion}$
Model 1 – Without prior arithmetic performance								
Intelligence (# correct) [a]	.27	3.24	.002	46.16	.25	3.11	.002	31.14
Metacognitive monitoring [a, b]	.29	3.43	.001	79.81	.37	4.52	<.001	>100
Model 2 – With prior arithmetic performance								
Intelligence (# correct) [a]	.19	2.28	.02	3.28	.18	2.24	.03	2.49
Metacognitive monitoring [a, b]	.12	1.30	.20	1.05	-.03	-.27	.79	0.45
Prior arithmetic accuracy [b]	.36	3.77	<.001	>100	.52	4.13	<.001	>100

Note. [a] The higher the score, the better the performance; [b] Metacognitive monitoring in the addition task for addition accuracy, Metacognitive monitoring in the multiplication task for multiplication accuracy.

Table 3.5b

Regression analyses of arithmetic RT

Variables T_1	Addition RT T_2				Multiplication RT T_2			
	β	t	p	$BF_{inclusion}$	β	t	p	$BF_{inclusion}$
Model 1 – Without prior arithmetic performance								
Motor speed task (RT)[a]	-.15	-1.82	.07	3.85	-.08	-.88	.38	1.17
Symbolic numerical magnitude processing (RT) [a]	.40	4.87	<.001	>100	.33	3.34	<.001	>100
General metacognitive knowledge [b]	-.24	-3.07	.003	32.35	-.23	-2.73	.01	11.52
Metacognitive monitoring [b, c]	-.31	-4.12	<.001	>100	-.20	2.44	.02	6.35
Model 2 – With prior arithmetic performance								
Motor speed task (RT)[a]	-.08	-1.24	.22	.33	-.06	-.80	.43	0.44
Symbolic numerical magnitude processing (RT) [a]	.16	2.40	.02	1.17	.17	2.24	.03	2.06
General metacognitive knowledge [b]	-.13	-2.12	.04	1.06	-.18	-2.53	.01	4.48
Metacognitive monitoring [b, c]	-.12	-1.94	.06	0.67	-.08	-1.17	.24	0.58
Prior arithmetic RT [c]	.65	9.85	<.001	>100	.55	7.43	<.001	>100

Note. [a] The higher the scores, the worse the performance; [b] The higher the score, the better the performance; [c] Metacognitive monitoring in the addition task for addition RT, Metacognitive monitoring in the multiplication task for multiplication RT.

3.1 Arithmetic accuracy

Arithmetic accuracy was significantly predicted by earlier metacognitive monitoring performance, in addition to the predictive power of intelligence. However, once prior arithmetic performance was taken into account, there was no evidence for or against the predictive power of metacognitive monitoring. Prior arithmetic accuracy significantly predicted later arithmetic accuracy, with Bayes factors indicating decisive evidence for the predictive power of earlier performance.

3.2 Arithmetic RT

Arithmetic response time was significantly predicted by earlier symbolic numerical magnitude processing, general metacognitive knowledge, and metacognitive monitoring performance, in addition to each other. Once prior arithmetic performance (i.e. RT) was additionally taken into account, there was anecdotal and moderate evidence for the predictive power for multiplication RT of symbolic numerical magnitude processing and general metacognitive knowledge, respectively, but no evidence for or against their predictive power for addition RT. There was no evidence for or against the predictive power of metacognitive monitoring for later arithmetic RT. Prior arithmetic RT significantly predicted later arithmetic RT, with Bayes factors indicating decisive evidence for the predictive power of earlier arithmetic performance.

Discussion

Numerous studies have highlighted the importance of symbolic numerical magnitude processing (see Schneider et al., 2017, for a meta-analysis), executive functions (see Friso-Van Den Bos et al., 2013, for a meta-analysis) and metacognition (e.g. Schneider & Artelt, 2010) for arithmetic. Yet, few studies have considered different functions simultaneously, and even fewer studies did so in a longitudinal way. As a result, there is a lack of a comprehensive understanding of the role of these functions in arithmetic. Therefore, the current study combined (a) a simultaneous investigation of these (meta)cognitive functions – which have been identified as important for arithmetic performance and development when investigated in isolation but have not been studied in concert, and (b) a longitudinal panel design – which allows us to investigate the predictive power of these functions for later arithmetic performance taking into account prior arithmetic performance. Importantly, this investigation was done in a crucial developmental period for all of the included functions (i.e. arithmetic, numerical magnitude processing, executive functions, metacognition). Our data show the important, unique roles of symbolic numerical magnitude processing and metacognition in arithmetic and indicate that executive functions are not so strong predictors of individual differences in arithmetic, when different other important functions are considered. On the other hand, our longitudinal data also indicate the strongest, most robust predictive role of prior arithmetic performance for later performance. In the remainder of this discussion, we first discuss the cross-sectional results exploring the stability of the associations over time. Second, we consider the longitudinal associations between the (meta)cognitive functions and arithmetic and examine the predictive power of these (meta)cognitive functions for arithmetic, in addition to each other. Next, we discuss the additional predictive value of the (meta)cognitive functions once prior arithmetic performance is taken into account. We conclude with suggestions for future research.

1 Cross-sectional results

Taken together with the results of the first time point (Bellon et al., 2019), the cross-sectional results of the second time point (i.e. in third graders) provide insight into the stability of the associations found in early primary school over time. As such, the current study provides, for each (meta)cognitive process, a more comprehensive image of its role throughout arithmetic development in early to middle primary school.

Symbolic numerical magnitude processing was reliably associated with arithmetic skills over development (i.e. both in second and third grade), which is in line with the meta-analytic findings of (Schneider et al., 2017). Several underlying mechanisms might be at play in the concurrent associations between symbolic numerical magnitude processing and arithmetic. A possible explanation of this association might be that the measures of arithmetic performance require the interpretation and transformation of information presented in symbolic form (i.e. Arabic numerals; Schneider et al., 2017).

Furthermore, a good understanding of the numbers in terms of their magnitudes can be helpful for choosing an efficient calculation strategy (e.g. Linsen et al., 2015).

The results shows a less clear role of executive functions in explaining individual differences in arithmetic in both grades, with even a slight decrease in the role of executive functions over development. This indicates that over development, executive functions, being effortful processes in nature (Diamond, 2013), take a less prominent place in individual differences in arithmetic, as over time arithmetic develops from a skill relying on effortful strategies to efficient, automatized problem solving or fact retrieval. A similar conclusion was made by Lee and Bull (2016), who found that the correlation between executive functions and mathematics performance varied across grade, with the strongest relations in early primary school compared to later in development. This could be due to the fact that single-digit arithmetic is fairly easy for children upward of middle primary school, requiring less executive resources at that time. It is plausible that a more prominent role for executive functions would have been found if more difficult arithmetic items (e.g. multi-digit arithmetic) were included.

The pattern of results concerning the associations with metacognition shows stable effects, as general metacognitive knowledge is reliably associated with addition RT (i.e. at both time points), and metacognitive monitoring is reliably associated with arithmetic performance at both time points. One reason why general metacognitive knowledge might be specifically associated with concurrent addition and not multiplication, could be that within multiplication the focused-on strategy is rote memorization of multiplication tables, whereas in addition metacognitive strategy selection (e.g. Geurten et al., 2018) may be more important. To fully understand the association between general metacognitive knowledge and multiplication RT, more research is needed, as Bayes factors indicated there was no evidence for or against these associations at both time points. The stable associations of metacognitive monitoring with concurrent arithmetic performance across development indicate that correctly appraising your performance is not only important in the early learning stages of arithmetic performance, but also later in development (in line with Rinne & Mazzocco, 2014), over and above other important cognitive functions.

2 Longitudinal results

The abovementioned concurrent results, however, do not answer critical questions on the longitudinal associations between these variables, and the unique, predictive value of these (meta)cognitive functions for later arithmetic performance, in addition to each other. Therefore, the current study also investigated the longitudinal associations.

The current results show clear evidence for a longitudinal association between symbolic numerical magnitude processing in early primary school and arithmetic RT one year later. In line with the existing literature (e.g. De Smedt et al., 2013; De Smedt, Verschaffel, et al., 2009; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013; Schneider et al., 2017; Vanbinst et al., 2015), our longitudinal analyses show that there is a clear predictive role of symbolic numerical magnitude processing for arithmetic RT. Importantly, in accordance with the critical need to include non-numerical functions when investigating the role of symbolic numerical magnitude processing in mathematics (Schneider et al., 2017), this study additionally confirms the predictive role of symbolic numerical magnitude processing over other important (meta)cognitive functions. Along with the abovementioned mechanisms underlying concurrent associations between symbolic numerical magnitude processing and arithmetic, proficient numerical magnitude processing skills might induce the transition to more efficient strategies based on the characteristics of the numbers in the problem (Booth & Siegler, 2008). Moreover, arithmetic facts might be stored in long-term memory in a meaningful way, i.e. according to their magnitude (e.g. Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Verguts & Fias, 2005) and meaningful facts might be easier to store and recall from memory (e.g. Robinson, Menchetti, & Torgesen, 2002).

While the results on the longitudinal association between updating and arithmetic accuracy provided no evidence for or against an association, it was clear that updating was not a strong predictor of later arithmetic performance. Thus, the predictive role of updating found in previous research (e.g. Bull & Lee, 2014; Friso-Van Den Bos et al., 2013; Van der Ven et al., 2012) was not confirmed in the current study, which could be due to the differences in the investigated mathematics subdomain (e.g. Best, Miller, & Naglieri, 2011; Van der Ven et al., 2012). The current study examined the predictive power of updating in arithmetic, whereas most studies use general mathematics tests, for which higher correlations with executive functions are consistently found (Friso-Van Den Bos et al., 2013). In line with results on the concurrent associations of arithmetic with executive functions, the current study shows clear evidence for the absence of longitudinal associations between inhibition and shifting, and arithmetic. Going one step further, the current study shows a lack of their predictive power for later arithmetic performance, which is in line with the existent literature (e.g. Bull & Lee, 2014; Van der Ven et al., 2012). In general, it is not unlikely that the role of executive functions is dependent on the learning stage and experience of children and thus decreases over development once arithmetic reaches an predominantly automatic level that comes with additional practice (Lee & Bull, 2016; Vanbinst & De Smedt, 2016a). As Holmes, Gathercole, and Dunning (2009) suggest, executive functions might be especially important for the acquisition of mathematical skills, and less so later in development of that skill.

In line with previous research, metacognitive monitoring was found to be an important predictor of later arithmetic (e.g. Carr, Alexander, & Folds-Bennett, 1994; Carr & Jessup, 1995; Rinne & Mazzocco, 2014). Importantly, this study is the first to show a clear predictive role of metacognition for arithmetic in younger primary school children, and, in particular, in addition to other important cognitive functions, thereby highlighting its importance for arithmetic. A possible underlying mechanism could be that better insight in one's own performance (e.g. better detection of errors) creates more learning moments in which incorrect arithmetic answers are detected and possibly replaced with new, correct answers. In turn, this could lead to better arithmetic performance.

Due to the longitudinal panel design of this study, we were also able to take into account prior arithmetic performance. This is of utmost importance, as the strongest and most robust predictor of a child's later academic performance is their earlier academic performance (Duncan et al., 2007). As a result, controlling for prior performance is necessary to obtain insight into how these (meta)cognitive variables predict the development of arithmetic between the time points.

The current study importantly shows that the evidence for added value of symbolic numerical magnitude processing in predicting development in arithmetic over prior arithmetic performance in the exact same task, is only anecdotal, as was the additional predictive power of general metacognitive knowledge for addition RT. The evidence for the predictive power of general metacognitive knowledge for multiplication RT, on the other hand, was moderate. Lastly, once prior arithmetic performance was taken into account, there was no evidence for or against the predictive power of metacognitive monitoring. The substantial decrease in predictive power of the (meta)cognitive functions for arithmetic once prior arithmetic performance is taken into account, is not surprising, given the very strong autoregressive effect (i.e. $r = .78$ for addition; $r = .64$ for multiplication). Therefore, the found anecdotal evidence for the predictive power of symbolic numerical magnitude processing is still meaningful. The results also emphasize the importance of metacognitive knowledge for arithmetic development, which was strongest for multiplication RT, possibly due to the larger development in multiplication RT compared to addition RT in this developmental period. The predictive role of general metacognitive knowledge for later arithmetic, over prior arithmetic performance, was also stronger compared to numerical magnitude processing and metacognitive monitoring, which could be due to a larger overlap of numerical magnitude processing and metacognitive monitoring with the arithmetic task compared to the metacognitive questionnaire. In general, however, these results indicate that the investigated (meta)cognitive functions play a larger role in concurrent performance than they do in development of arithmetic (i.e. change in arithmetic performance from second to third grade). The results confirm the strong, robust predictive value of prior arithmetic performance for later arithmetic performance.

3

The design of the current study proves to be a valuable direction for future research to yield a thorough understanding of the interplay between symbolic numerical magnitude processing, executive functions and metacognition in mathematics performance and development. Therefore, future studies should explore this developmental interplay in more advanced arithmetic tasks (e.g. multi-digit arithmetic) or in other mathematical subdomains (e.g. fractions). Because longitudinal studies are unable to establish causality, future studies should also consider experimental designs in which the current (meta)cognitive functions are manipulated to investigate their causal relation with arithmetic. Training or intervention studies should be used to yield a deeper understanding of the unique role of these (meta)cognitive functions, but also to possibly foster arithmetic performance and development.

Conclusion

In sum, this study yields a deeper understanding of the role of symbolic numerical magnitude processing, executive functions and metacognition in arithmetic performance and development in primary school children. Our data show the important, unique role of symbolic numerical magnitude processing and metacognition in arithmetic and indicate that executive functions are not so strong predictors of individual differences in arithmetic, when different other important functions are considered. Additionally, the strong, robust predictive value of prior arithmetic performance for later arithmetic performance was confirmed.

Appendixes

1 Appendix A – Descriptive statistics of all administered cross-sectional data in third graders

Table 3.A1

Descriptive statistics of all administered variables

	n	*M*	*SD*	Range
Arithmetic				
Addition				
Accuracy	121	.97	.03	[.84-1.00]
Response time (ms)	121	2577.92	846.50	[1410.31-6102.02]
Multiplication				
Accuracy	121	.93	.07	[.66-1.00]
Response time (ms)	121	4341.00	1955.00	[1474.02-14792.50]
Numerical magnitude processing				
Symbolic numerical magnitude comparison task				
Response time (ms)	121	745.59	132.90	[520.17-1212.37]
Executive functions				
Inhibition				
Flanker task – baseline				
Accuracy	121	.94	.08	[.6-1.00]
Response time (ms)	121	582.75	86.47	[393.7-811.2]
Inverse efficiency (RT/accuracy)[a]	121	619.14	88.41	[437.44-901.33]
Flanker task – incongruent condition				
Accuracy	121	.93	.08	[.58-1.00]
Response time (ms)	121	776.33	169.02	[505.74-1392.89]
Inverse efficiency (RT/accuracy)[a]	121	856.32	235.74	[548.68-1784.65]
Animal Stroop task – baseline				
Accuracy	121	.98	.04	[.85-1.00]
Response time (ms)	121	856.83	163.03	[485.90-1325.65]
Inverse efficiency (RT/accuracy)[a]	121	875.40	167.55	[485.90-1395.42]
Animal Stroop task – incongruent condition				
Accuracy	121	.93	.07	[.75-1.00]
Response time (ms)	121	926.27	179.71	[560.90-1758.65]
Inverse efficiency (RT/accuracy)[a]	121	1006.15	221.05	[623.22-2344.87]

(table continues on next page)

3

	n	*M*	*SD*	Range
Shifting				
WCST (average # items needed to switch)[b]	119	5.62	4.90	[2-46]
Updating				
2-back task (accuracy)	121	.75	.06	[.47-.87]
Metacognition				
General metacognitive knowledge				
General metacognitive questionnaire (# correct)[c]	121	11.01	2.27	[3-15]
Metacognitive monitoring [d]				
In addition task	121	1.88	0.13	[1.44-2.00]
In multiplication task	121	1.86	0.13	[1.34-2.00]
Control variables				
Intellectual ability				
Raven's Standard Progressive Matrices[c]	121	502.54	151.70	[329.95-1913.25]
Motor speed on keyboard				
Motor reaction time task Response time	121	38.72	6.96	[13-54]

Note. All Response time variables are average response time for the correct responses. [a] Inverse efficiency scores were calculated by dividing the response time by the accuracy; the higher the score, the worse the performance; [b] Score = total number of items needed to switch divided by number of blocks completed; [c] Number of correct answers; [d] Alignment between children's metacognitive judgment and the accuracy of their arithmetic answer, i.e. correct arithmetic answers yielded a score of 2 if children said they were *Correct*, 1 if they said *I am not sure*, and 0 if they said they were *Incorrect*; this scale was reversed when the arithmetic answer was incorrect. The higher the score, the better the metacognitive monitoring.

3

2 Appendix B – All intercorrelations of the administered cross-sectional data in third graders

Table 3.B1

Correlation analyses between the administered measures

				1a	1b	2a	2b	3	4	5	6	7	8	9	10	11
Arithmetic	Addition	1a. Addition ACC														
		1b. Addition RT														
			r	-.28	-											
			p	.002	-											
			BF_{10}	11.63	-											
	Multiplication	2a. Multiplication ACC														
			r	.54	-.39	-										
			p	<.001	<.001	-										
			BF_{10}	>100	>100	-										
		2b. Multiplication RT														
			r	-.03	.68	-.35	-									
			p	.74	<.001	<.001	-									
			BF_{10}	0.12	>100	>100	-									
Numerical magnitude processing	Symbolic NMP	3. Symbolic NMP RT														
			r	.07	.46	-.22	.35	-								
			p	.43	<.001	.02	<.001	-								
			BF_{10}	0.16	>100	1.91	>100	-								
Executive functions	Inhibition	4. Flanker task IE														
			r	-.12	.09	-.18	.04	.22	-							
			p	.19	.32	.05	.63	.02	-							
			BF_{10}	0.27	0.19	0.82	0.13	1.88	-							
		5. Stroop task IE														
			r	-.12	.04	-.04	.11	.08	.02	-						
			p	.18	.67	.64	.24	.41	.83	-						
			BF_{10}	0.28	0.13	0.13	0.23	.16	0.12	-						
	Shifting	6. WCST														
			r	.05	-.01	-.15	.13	.15	.12	.05	-					
			p	.63	.96	.10	.17	.11	.19	.61	-					
			BF_{10}	0.13	0.12	0.44	0.29	.42	0.27	0.13	-					
	Updating	7. 2-back task ACC														
			r	.10	-.25	.01	-.09	-.09	-.20	.04	-.19	-				
			p	.30	.006	.95	.30	.34	.03	.65	.04	-				
			BF_{10}	0.19	4.62	0.11	0.19	0.18	1.25	0.13	1.02	-				

(table continues on next page)

3

				1a	1b	2a	2b	3	4	5	6	7	8	9	10	11
Meta-cognition	General metacognitive knowledge	8. Metacognitive questionnaire														
			r	.14	-.31	.14	-.16	-.17	-.01	-.14	.04	.16	-			
			p	.13	.001	.14	.08	.06	.89	.13	.64	.09	-			
			BF_{10}	0.36	36.27	0.34	0.51	0.62	0.12	0.35	0.13	0.49	-			
	Task-specific metacognition	9. Metacognitive monitoring - Addition														
			r	.48	-.38	.28	-.14	.02	-.07	-.01	.04	.25	.13	-		
			p	<.001	<.001	.002	.13	.86	.45	.92	.69	.007	.15	-		
			BF_{10}	> 100	>100	15.56	0.35	0.12	0.15	.12	0.12	4.26	0.31	-		
		10. Metacognitive monitoring - Multiplication														
			r	.37	-.27	.58	-.36	-.04	.01	.02	-.10	.16	.06	.48	-	
			p	<.001	.003	<.001	<.001	.69	.94	.80	.28	.09	.52	<.001	-	
			BF_{10}	>100	8.72	>100	>100	0.12	0.11	.12	0.21	.48	0.14	>100	-	
Control Variables	Intelligence	11. Raven														
			r	.24	-.04	.27	.04	.07	-.26	-.01	.01	.17	-.004	.11	.16	-
			p	.007	.63	.003	.64	.42	.004	.92	.89	.07	.97	.24	.08	-
			BF_{10}	4.01	0.13	8.59	0.13	0.16	6.42	0.12	0.12	0.57	0.11	0.23	0.52	-
	Motor speed	12. Motor reaction time task RT														
			r	-.16	.31	-.07	.31	.35	.13	-.03	-.02	-.10	-.22	-.26	-.13	.008
			p	.08	.001	.43	.001	<.001	.17	.74	.86	.27	.02	.004	.16	.93
			BF_{10}	0.51	36.01	0.16	34.37	>100	0.30	0.12	0.12	0.21	1.99	7.04	0.30	0.11

Note. ACC = accuracy, RT = response time, NMP = numerical magnitude processing; All correlations with the Flanker task incongruent condition are controlled for the performance (IE) on the Flanker baseline condition; All correlations with the animal Stroop task incongruent condition are controlled for the performance (IE) on the animal Stroop baseline task; All correlations with the numerical magnitude processing task are controlled for performance (RT) on the motor speed task, except for the correlations with the inhibition tasks, which are controlled for their respective baseline (IE).

3 Appendix C – All longitudinal intercorrelations of the administered measures

Table 3.C1

Longitudinal correlations between the administered measures

		T₁		1a	1b	2a	2b	3	4	5	6	7	8	9	10	11	12
				T₂													
Arithmetic	Addition	1a. Addition ACC															
			r	.47	-.30	.31	-.04	.04	-.10	-.09	.01	.16	.12	.39	.16	.33	-.12
			p	<.001	.001	<.001	.66	.63	.26	.32	.88	.07	.21	<.001	.09	<.001	.20
			BF_{10}	>100	26.16	50.37	0.13	0.13	0.22	0.19	0.12	0.56	0.25	>100	0.48	94.35	0.25
		1b. Addition RT															
			r	-.30	.78	-.31	.54	.32	.05	-.01	.01	-.10	-.19	-.21	-.22	-.01	.33
			p	.001	<.001	<.001	<.001	<.001	.60	.91	.95	.28	.03	.02	.02	.95	<.001
			BF_{10}	31.88	>100	49.34	>100	71.90	0.13	0.11	0.12	0.20	1.05	1.75	2.15	0.11	>100
	Multiplication	2a. Multiplication ACC															
			r	.43	-.41	.54	-.25	-.06	.04	.05	-.04	.06	.09	.25	.38	.26	-.24
			p	<.001	<.001	<.001	.01	.50	.69	.58	.70	.53	.31	.01	<.001	.004	.01
			BF_{10}	>100	>100	>100	4.34	0.14	0.12	0.13	0.12	0.14	0.19	4.54	>100	7.56	3.46
		2b. Multiplication RT															
			r	-0.02	.48	-0.11	.64	0.22	-.12	.02	-0.09	-0.01	.04	-.03	-.23	.18	.18
			p	.79	<.001	0.25	<.001	.02	.21	.85	0.31	.96	.66	.72	.01	.05	.06
			BF_{10}	0.12	>100	0.22	>100	2.13	0.25	0.12	0.19	0.11	0.13	0.12	2.35	0.80	0.70
Numerical magnitude processing	Symbolic NMP	3. Symbolic NMP RT															
			r	.12	.41	-.004	.34	.49	.13	.10	.04	-.27	-.32	-.09	-.05	.12	.30
			p	.19	<.001	.97	<.001	<.001	.17	.30	.71	.003	<.001	.34	.62	.21	.001
			BF_{10}	0.27	>100	0.11	>100	>100	0.30	0.19	0.12	8.35	49.14	0.18	0.13	0.25	23.37
Executive functions	Inhibition	4. Flanker task IE															
			r	-.11	-.01	-.01	-.03	-.05	.59	.17	-.01	-.12	.02	-.10	.01	-0.35	-.01
			p	.24	.93	.92	.74	.60	<.001	.07	.94	.20	.83	.29	.90	<.001	.96
			BF_{10}	0.23	0.12	0.12	0.12	0.13	>100	0.59	.12	.26	.12	0.20	0.12	>100	0.11
		5. Stroop task IE															
			r	-.11	.08	-.14	.04	.10	.31	.25	.05	-.07	.13	-.11	-.13	-.11	-.11
			p	.25	.42	.14	.70	.26	.001	.01	.63	.43	.17	.25	.17	.25	.24
			BF_{10}	0.22	0.16	0.33	0.12	0.22	40.19	4.69	0.13	0.16	0.29	0.22	0.29	0.22	0.23
	Shifting	6. WCST															
			r	.06	-.08	-.15	.04	.07	.12	-.01	.15	-.01	.19	.24	-.02	-.17	.04
			p	.55	.39	.11	.64	.48	.22	.90	.10	.90	.05	.01	.80	.07	.69
			BF_{10}	0.14	0.17	0.42	0.13	0.15	0.25	0.12	0.43	0.12	0.84	2.87	0.12	0.56	0.13
	Updating	7. 2-back task ACC															
			r	.17	-.10	.19	.01	-.04	-.11	.01	.05	.22	.07	.27	.30	.28	-.10
			p	.06	.25	.03	.92	.63	.22	.94	.59	.02	.44	.003	.001	.002	.29
			BF_{10}	0.76	0.23	1.27	0.11	0.13	0.24	0.11	0.13	1.95	0.15	10.85	33.92	13.45	0.20

(table continues on next page)

3

		T₁		1a	1b	2a	2b	3	4	5	6	7	8	9	10	11	12
				T2													
Meta-cognition	General meta-cognitive knowledge	8. Metacognitive questionnaire															
			r	.01	-.32	.16	-.30	-.17	-.17	-.03	-.11	.19	.37	.03	.07	.01	-.15
			p	.94	<.001	.08	.001	.06	.06	.75	.23	.04	<.001	.76	.48	.92	.10
			BF10	0.11	72.89	0.52	32.87	0.65	0.65	0.12	0.24	1.02	>100	0.12	0.15	0.11	0.44
	Task-specific metacognition	9. Metacognitive monitoring - Addition															
			r	.33	-.32	.30	-.16	-.14	-.08	.01	-.12	.09	.06	.51	.36	.11	-.07
			p	<.001	<.001	.001	.08	.12	.40	.92	.18	.31	.52	<.001	<.001	.24	.43
			BF10	76.80	62.01	24.36	0.51	0.38	0.16	0.11	0.28	0.19	0.14	>100	>100	0.23	0.16
		10. Metacognitive monitoring - Multiplication															
			r	.24	-.31	.40	-.21	-.02	.03	.08	-.02	-.02	.01	.19	.34	.23	-.13
			p	.01	<.001	<.001	.02	.82	.75	.37	.85	.85	.91	.04	<.001	.01	.15
			BF10	3.52	47.43	>100	1.55	0.12	0.12	0.17	0.12	0.12	0.11	0.93	>100	2.99	0.32
Control Variables	Intelligence	11. Raven															
			r	.31	-.08	.30	-.01	.03	-.25	-.11	-.07	.18	.02	.15	.22	.65	.09
			p	.001	.37	.001	.94	.71	.01	.25	.46	.05	.82	.09	.02	<.001	.32
			BF10	42.82	0.17	30.74	0.11	0.12	5.64	0.22	0.15	0.76	0.12	0.46	1.88	>100	0.18
	Motor speed	12. Motor reaction time task RT															
			r	-.08	.09	-.13	.14	.14	.14	-.05	.01	-.11	-.06	-.21	-.21	-.07	.25
			p	.37	.36	.17	.13	.14	.14	.60	.96	.22	.53	.02	.02	.48	.01
			BF10	0.17	0.17	0.29	0.35	0.34	0.34	0.13	0.12	0.24	0.14	1.50	1.42	0.15	5.10

Note. ACC = accuracy, RT = response time, NMP = numerical magnitude processing; All correlations with the Flanker task incongruent condition are controlled for the performance (IE) on the Flanker baseline condition; All correlations with the animal Stroop task incongruent condition are controlled for the performance (IE) on the animal Stroop baseline task; All correlations with the numerical magnitude processing task are controlled for performance (RT) on the motor speed task, except for the correlations with the inhibition tasks, which are controlled for their respective baseline (IE)

CHAPTER 4

Too anxious to be confident?

A panel longitudinal study into the interplay of metacognition and mathematics anxiety in arithmetic achievement.

The content of this chapter is under revision as:

Bellon, E., Fias, W., & De Smedt, B. (Under Revision). Too anxious to be confident? A panel longitudinal study into the interplay of metacognition and mathematics anxiety in arithmetic achievement. *Journal of Educational Psychology*.

Chapter 4

Too anxious to be confident?

A panel longitudinal study into the interplay of metacognition and mathematics anxiety in arithmetic achievement.

Abstract

Both metacognitive monitoring and mathematics anxiety have been identified as associated with or predictive of individual differences in arithmetic achievement in primary school children. It is unclear, however, how these variables are interrelated and whether their interrelation impacts their respective associations with arithmetic achievement. Gaining insight into their interplay is of utmost importance for the design of targeted interventions. We used a panel longitudinal design to investigate arithmetic achievement, metacognitive monitoring skills and mathematics anxiety in 127 typically developing 7-8-year-olds (second grade) and followed them up one year later (in third grade). As such, participants were in the middle of an important developmental period for arithmetic, as well as metacognitive monitoring and mathematics anxiety. Our preregistered analyses showed that metacognitive monitoring and mathematics anxiety are correlated and that this association strengthened over development. However, this association was mediated by arithmetic achievement. In line with the deficit model of mathematics anxiety, our findings indicate an increasingly important role of mathematics anxiety in arithmetic achievement and metacognition. Furthermore, our results indicated that the association between arithmetic achievement and metacognitive monitoring was unique and specific, without mediation or moderation of mathematics anxiety. Arithmetic achievement was found to be a unique, powerful predictor over developmental time of both metacognitive monitoring and mathematics anxiety, over and above their respective autoregressive effects. These results emphasise the importance of arithmetic achievement in the development of mathematics anxiety, metacognitive monitoring and their interrelations.

Introduction

Individual differences in mathematical performance have been a frequently investigated topic in the existing educational and psychological literature. Most of this research has focused on cognitive skills, such as numerical magnitude processing (e.g. Schneider et al., 2017) or working memory (e.g. Peng et al., 2016), that underlie these individual differences. There has been, however, much less research on more affective and metacognitive variables, which also play a critical role in academic learning and its individual differences (e.g. Dowker, 2019c; Schneider, 2010). One such important affective factor is mathematics anxiety. An extensive body of research has shown that mathematics anxiety or "*a feeling of tension and anxiety that interferes with the manipulation of numbers and solving of mathematical problems in ordinary life and academic situations*" (Richardson & Suinn, 1972, p. 551), is systematically found to be moderately negatively related to mathematics performance (e.g. see Hembree, 1990; Ma, 1999; Namkung et al., 2019, for meta-analyses) and some studies have specifically shown this negative association with arithmetic (e.g. Ashcraft et al., 1998; Harari et al., 2013; Hunt et al., 2017; Sorvo et al., 2017). More recently, the metacognitive variable metacognitive monitoring has been found to be a unique predictor of individual differences in arithmetic achievement, in addition to important other predictors such as symbolic numerical magnitude processing or working memory (Bellon et al., 2019; Rinne & Mazzocco, 2014). Metacognitive monitoring is part of the broader concept 'metacognition' – which involves the ability to assess one's own cognitive knowledge and ability (e.g. Vo et al., 2014) and how people monitor and control their cognition on-task (e.g. Bryce et al., 2015). Metacognitive monitoring is defined as the subjective self-assessment of how well a cognitive task will be/is/has been performed (Morsanyi et al., 2019; Nelson & Narens, 1990).

Research on mathematics anxiety and metacognitive monitoring has been done in isolation from each other, making their interrelation and unique contribution to the development of mathematics unclear. Gaining insight into this issue is of utmost importance, as learning mathematics involves a complex interplay of diverse processes including cognitive, metacognitive and affective processes (e.g. Carey et al., 2016; Dowker, 2019e; Dowker et al., 2016; Hill et al., 2016; Mammarella et al., 2015). Moreover, it is likely that metacognitive monitoring and mathematics anxiety are associated to each other, because both are linked to thinking about your performance, and because they are both associated to individual differences in mathematics. The precise, developmental associations between mathematics anxiety and metacognitive monitoring, and whether this interplay influences their respective association with mathematical performance, remain unknown. Understanding this complex interplay of associations was exactly the aim of the current panel longitudinal study. In the remainder of this introduction, we discuss potential ways in which metacognitive monitoring, mathematics anxiety and mathematics performance might interact, and we outline the rationale for the current study.

One possibility is that mathematics anxiety influences metacognitive monitoring. For example, mathematics anxiety might hinder the efficiency of metacognitive monitoring (Morsanyi et al., 2019), as anxiety impairs the functioning of the goal-directed attentional system (Morsanyi et al., 2019); see the processing efficiency theory: Eysenck & Calvo, 1992; and see the attentional control theory: Eysenck et al., 2007). As a result, mathematics anxiety can lead to a biased interpretation of actual performance, i.e. an underestimation of success and an overestimation of failure. Yet, this lower confidence in one's own mathematical abilities as a result of math anxiousness (Hembree, 1990), might in turn might lead to a weaker tendency for overconfidence, which is observed in the general population (e.g. Finn & Metcalfe, 2014; Flavell et al., 1970; Yussen & Levy, 1975) and thus result in better calibration of confidence (i.e. lower confidence in lower achievers). On the other hand, mathematics anxiety might cause rumination and preoccupying thoughts that consume cognitive resources (e.g. Ashcraft et al., 1998; Eysenck & Calvo, 1992) that could otherwise be used for metacognitive monitoring. Mathematics anxiety might also result in a higher likelihood of rushing through or premature termination of mathematical tasks or items (Ashcraft & Faust, 1994), without thorough metacognitive processing of the task or item. Contrarily, mathematics anxiety might lead to exaggerated error monitoring, in which math anxious children may compensate for anxiety-related processing inefficiencies through increased cognitive effort. For example, they might engage in more stringent monitoring processes, such as double-checking answers (Morsanyi et al., 2019; Moser et al., 2013).

Another possibility is that metacognitive monitoring influences the development of mathematics anxiety. For example, with increasing experience in mathematics, the individual differences in mathematics achievement between children and their peers might become more apparent. When children become more aware of these individual differences in achievement, i.e. as a consequent of their developing metacognitive skills, this might increase feelings of pressure and anxiety, thereby strengthening the relation between metacognition and mathematics anxiety. Moreover, repeated experiences of failure in mathematics together with a growing awareness of this failure might lead to math anxiousness. For example, Tobias (1986) stated that poor performance is a function of the acquisition deficit and the observed elevation in anxiety is attributable to student's metacognitive awareness of their incomplete learning. Alternatively, when previously math anxious children who perform well become more aware of their own good performance, they might become less math anxious.

Although several hypotheses on the interplay between metacognitive monitoring and mathematics anxiety as well as their association with achievement are possible, this question has not been rigorously, empirically tested. The existing literature offers some suggestions as cross-sectional associations have been observed between test anxiety and metacognitive skilfulness in secondary school students (Veenman et al., 2000), between mathematics anxiety and metacognition in Chinese 10-year-olds and Turkish 12-year-olds word problem solving (Lai et al., 2015; Özcan & Gümüs, 2019), in university students in arithmetic (Legg & Locker, 2009) and in their general mathematics achievement (Erickson

& Heit, 2015). However, in view of the observation that the associations between metacognitive monitoring and mathematics achievement (e.g. Bellon et al., 2019), and between mathematics anxiety and mathematics achievement (e.g. Krinzinger et al., 2009; Ma & Kishor, 1997; Maloney & Beilock, 2012; Ramirez et al., 2013, 2016; Vukovic et al., 2013) are already observed in the early grades of primary school, it is critical to investigate the interplay of these variables at a much earlier age than in the abovementioned studies. To date, it remains unclear how metacognitive monitoring and mathematics anxiety are related to each other and to mathematical achievement in early primary school children. Even more critical, none of the abovementioned studies has collected longitudinal data, which makes making claims on the developmental dynamics of the associations problematic.

The current study will use a longitudinal panel design to thoroughly investigate the associations between metacognitive monitoring, mathematics anxiety and arithmetic achievement. Looking into this question in young primary school children is of utmost importance, as metacognitive monitoring skills are rapidly developing in this age range and (early signs of) mathematics anxiety emerge, while at the same time children's acquisition of arithmetic is rapidly developing at this age and constitutes a crucial building block for later development in more complex mathematical abilities (e.g. Kilpatrick et al., 2001). More specifically, critical developments in metacognitive monitoring, e.g. an increase in its accuracy, are observed during middle childhood (e.g. Ghetti, 2008; Lyons & Ghetti, 2010; Schneider, 2008, 2010; Schneider & Lockl, 2008; Schneider & Löffler, 2016). Over second and third grade, the 7-to-9 year-olds typically intensify their skills in addition and subtraction – which they learned in first grade. They are introduced to the multiplication tables and get intensive training with a focus on rote memorization of these arithmetic facts. Lastly, mathematics anxiety was found to develop at an early age (e.g. 5-6 years; Maloney & Beilock, 2012; Ramirez et al., 2016; Skemp, 1986; Wu et al., 2012) and has been linked to a child's first experiences to mathematics and to the use of rote memorization (e.g. Rossnan, 2006). These first years of formal mathematics instruction are thus of crucial importance in the development of mathematics anxiety. Because of this rapid and essential development of these three variables, it is critical to investigate the interrelations between arithmetic achievement, metacognitive monitoring and mathematics anxiety at this stage of development. Therefore, we specifically recruited second graders (7-8 year-olds) and followed them up one year later in third grade (8-9 year-olds).

To the best of our knowledge, the present study is the first to use a longitudinal panel design to understand the developmental dynamics between arithmetic achievement, metacognitive monitoring and mathematics anxiety. Additionally, such a longitudinal panel design allows us to gain further insight into these interrelations through the use of mediation and/or moderation analyses, which are difficult on the basis of cross-sectional data alone (e.g. Maxwell et al., 2011; Maxwell & Cole, 2007; Selig & Preacher, 2009). *Mediation* analyses investigate the mechanisms by which an effect operates, namely whether a variable's effect on an outcome variable can be partitioned into direct and indirect (via a mediator variable) effects. Mediation analysis thus answers the question of *how* the effect occurs (Hayes, 2018): Is the effect of metacognitive monitoring on arithmetic achievement a direct effect, or is this effect driven by the association of metacognitive monitoring with mathematics anxiety, which in turn is associated to arithmetic achievement? For example, Özcan and Gümüs (2019) found in seventh grade Turkish students that metacognitive experience mediated the effect of mathematics anxiety on problem solving. It remains to be determined whether this can also be observed at an earlier developmental stage. *Moderation* analyses, on the other hand, investigate the boundary conditions of an effect, namely whether a variable (i.e. the moderating variable) influences or is related to the size of one variable's effect on another. Moderation analysis thus answers the question of *when* the effect occurs (Hayes, 2018): Is there only an effect of metacognitive monitoring on arithmetic achievement for low math anxious children or is the effect present across the entire distribution of achievement? Legg and Locker (2009), for example, found that metacognition moderated the relation between mathematics anxiety and arithmetic performance in adults, predicting that performance would decrease as anxiety increased, except at high metacognition levels.

1 The current study

The current study is the first to simultaneously consider metacognitive monitoring, mathematics anxiety and arithmetic achievement in young primary school children. We investigated this issue with a longitudinal panel design, which allowed us to go beyond the traditional correlational analyses and regression models with mediation and moderation models. This design made it possible to thoroughly investigate the important outstanding conundrum on the developmental dynamics of arithmetic achievement, metacognitive monitoring and mathematics anxiety in young primary school children, which is critical to develop effective educational interventions. Our study was preregistered on the Open Science Framework (OSF; https://osf.io/4xgh9/?view_only=deac4754b67b4a30a98003680b3f9536). The participants of this study were part of a larger longitudinal project on executive functions and metacognition in arithmetic achievement, of which some data of the first data point has already been published (Bellon et al., 2019). Importantly, no data on mathematics anxiety were analysed or reported before.

Method

1 Participants

Participants were 127 second graders (64 girls; M_{age} = 7 years 11 months, SD = 4 months, range = 7 years 4 months to 8 years 5 months) at the first time point, and 121 of them were followed up one year later in third grade (63 girls; M_{age} = 8 years 8 months, SD = 3 months, range = 8 years 2 months to 9 years 2 months). They were all typically developing children from Flanders (Belgium) who had no diagnosis of a developmental disorder and had dominantly middle-to-high socioeconomic background. For every participant, written informed parental consent was obtained. The study was approved by the Social and Societal Ethics Committee of the KU Leuven.

2 Procedure

At both time points, all children participated in three test sessions, which took place at their own school during regular school hours. All children completed the tasks in the same order: Firstly, an individual session including the arithmetic achievement tasks and a task-specific metacognitive monitoring measure within the arithmetic tasks. Secondly, a session in small groups of five children containing cognitive computer tasks. Lastly, a group-administered session in the classroom containing the evaluation of arithmetic achievement, general metacognitive knowledge, mathematics anxiety and intellectual ability. The full cognitive testing battery is posted on the OSF page of this project. Below we describe the key variables and control variables for our preregistered research questions of the current study.

3 Materials

Materials consisted of standardized tests, paper-and-pencil tasks, and computer tasks designed with E-prime 2.0 (Schneider et al., 2002).

3.1 Arithmetic

Arithmetic skills were assessed with a custom computerized task and a standardized achievement test (i.e. Tempo Test Arithmetic; de Vos, 1992).

The custom computerized task was the same as in Bellon et al. (2019). A single-digit addition and a single-digit multiplication production task were administered, with 64 trials per operation. Children were presented with the item for 2000 ms and afterwards, a black screen appeared until response. The children were asked to answer verbally and as quickly and accurately as possible. Response times (RTs) and answers were registered. Each task was pseudo-randomly divided into two blocks and during the second block of each arithmetic task, a metacognitive monitoring measure was added to the task (see below). Performance measures were average RT for correct answers and accuracy of the answers, which were calculated for both tasks together, resulting in one average arithmetic RT and average arithmetic accuracy score per child.

Arithmetic fluency was assessed by the Tempo Test Arithmetic (TTA; de Vos, 1992); a standardized pen-and-paper achievement test of arithmetical fluency. This achievement test comprises five columns of arithmetic items (one column per operation and a mixed column), each increasing in difficulty. Participants got one minute per column to provide as many correct answers as possible. The performance measure was the total number of correctly solved items within the given time (i.e. total score over the five columns).

3.2 Metacognitive monitoring

In the second block of the computerized arithmetic tasks (*n* = 32 per operation), a metacognitive monitoring measure was added to the items. Children had to report their judgment on the accuracy of their answer to the arithmetic item on a trial-by-trial basis (e.g. Bellon et al., 2019, Rinne & Mazzocco, 2014). More specifically, after giving their answer to the arithmetic problem, children had to indicate if they thought their arithmetic answer was *Correct*, *Incorrect,* or if they *Did not know*. We used emoticons in combination with the options (e.g. ☺ and *Correct*) to make the task more attractive and feasible for children (see Figure 4.1).

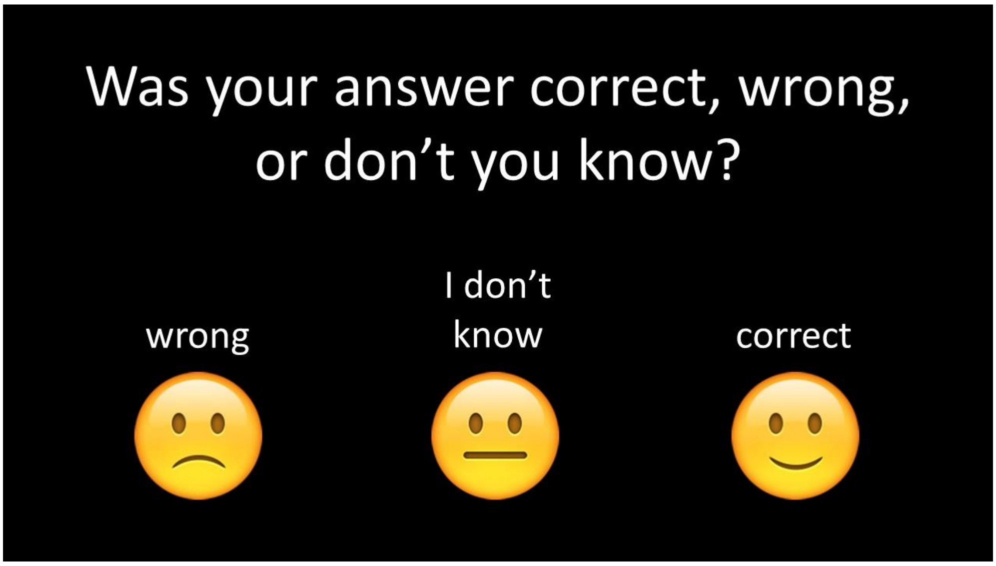

Figure 4.1. Example of metacognitive monitoring question after arithmetic item.

Metacognitive monitoring skills were operationalised as calibration of this judgment (i.e. the alignment between one’s judgment in the accuracy of their answer to a problem and the actual accuracy of the answer). A calibration score of 2 was obtained if their metacognitive judgment corresponded to their actual performance: metacognitively judged as *Correct* and indeed correct arithmetic answer; metacognitively judged as *Incorrect* and indeed incorrect arithmetic answer. A calibration score of 0 was obtained if children's metacognitive judgement did not correspond to their actual performance: metacognitively judged as *Correct* and in fact incorrect arithmetic answer; metacognitively judged as *Incorrect* and in fact correct arithmetic answer. A calibration score of 1 was obtained if children indicated they *Did not know*. The metacognitive monitoring score per child was the mean of all calibration scores (i.e. calibration score per arithmetic item) over both arithmetic tasks (n = 64). The higher the calibration scores, the better the metacognitive monitoring skills. To familiarize the children with the task, six practice items were presented.

3.3 Mathematics anxiety

To measure mathematics anxiety, we adapted the mathematics anxiety questionnaire developed by Suinn and Edwards (1982). In this questionnaire, 15 situations involving mathematics in daily life (e.g. *How anxious are you / would you be to do homework for math class?*) were described. Children had to indicate which of five emoticons best represented their feeling within that situation (i.e. ranging from "not at all anxious/stressed" with corresponding emoticon ☺ to "very anxious/stressed" with corresponding emoticon ☹). The researcher read the situations aloud one by one. Children were given a response form with the emoticons, such that they could follow each item and indicate their answer. The more anxious the children indicated they were, the higher the score (e.g. ranging from 1 point for "not at all anxious/stressed" to 5 points for "very anxious/stressed"). The performance measure was the total number of points over the 15 situations.

3.4 Intellectual ability

Intellectual ability was assessed through Raven's Standard Progressive Matrices (Raven et al., 1992). Children were given 60 multiple-choice items in which they had to complete a pattern. The performance measure was the number of correctly solved patterns.

4 Data analysis

A comprehensive analysis plan was preregistered on the OSF page of this project. We ran frequentist analyses using both uni- and multivariate techniques, as well as Bayesian analyses, using Bayes factors. Although Bayes factors provide a continuous measure of degree of evidence, we interpreted them according to some conventional approximate guidelines for interpretation (Andraszewicz et al., 2015, for a classification scheme): BF_{10} = 1 provides no evidence either way, BF_{10} > 1 anecdotal, BF_{10} > 3 moderate, BF_{10} > 10 strong, BF_{10} > 30 very strong and BF_{10} >100 decisive evidence for the alternative hypothesis.

To answer our research questions, we used correlational and regression analyses. For the Bayesian analyses, we used a default prior provided by the statistical program JASP (JASP, 2019): a Cauchy prior distribution centred at zero; prior width set to 1 for Pearson correlations and to .354 for the linear regression analyses. For every predictor in each regression model, a $BF_{inclusion}$ was calculated. This is a

Bayes factor for including the predictor averaged across the models under consideration (i.e. the models with versus the models without the predictor of interest).

Because the current study had a longitudinal panel design, it was possible to additionally perform both mediation and moderation analyses to investigate direct and indirect effects in the presented associations, and to investigate whether a certain variable influences or is related to the size of one variable's effect on another variable. We used the PROCESS computational toolbox for SPSS (Hayes, 2018) using Model 4 ('simple mediation model') for all mediation analyses, and Model 1 ('simple moderation model') for all moderation analyses. The strength and significance of the mediation models were tested using the bootstrapping method with 10 000 iterations (Preacher et al., 2007).

Results

1 Descriptive statistics

The descriptive statistics of all key measures are presented in Table 4.1.

Table 4.1
Descriptive statistics of the key variables

	n	*M*	*SD*	Range
Arithmetic				
Custom task				
Accuracy				
T_1	127	.89	.08	[.54-1.00]
T_2	121	.95	.05	[.77-1.00]
Response time (ms)				
T_1	127	5431.89	2554.31	[2242.60-20115.33]
T_2	121	3459.73	1304.24	[1520.84-10447.26]
Standardized task [a]				
T_1	126	67.68	16.51	[33-108]
T_2	121	85.51	19.18	[48-127]
Metacognition				
Metacognitive monitoring [b]				
T_1	127	1.79	0.14	[1.06-2.00]
T_2	121	1.87	0.11	[1.50-2.00]
Mathematics anxiety				
T_1	127	28.82	9.72	[15-59]
T_2	121	26.91	8.22	[15-54]
Control				
Intellectual ability				
Raven [a]				
T_1	127	34.50	8.35	[10-50]
T_2	121	38.72	6.96	[13-54]

Note. T_1 = time point 1 (Grade 2); T_2 = time point 2 (Grade 3). [a] Number of correct answers; [b] Alignment between children's metacognitive judgment and the accuracy of their arithmetic answer.

4

In the analyses reported below, the standardized achievement test of arithmetic (i.e. TTA) was used as performance measure for arithmetic achievement. This was done to have an assessment of arithmetic that was independent of the task in which metacognitive monitoring was measured. Moreover, the standardized achievement test was the most ecologically valid measure, assessing arithmetic achievement as it is usually assessed in the classroom. As a validity check, we verified whether the performance measures of the custom tasks were correlated with the widely-used standardized arithmetic achievement test that we utilized (i.e. TTA). This was the case for both the accuracy (T_1: $r = .43$, $p < .001$, $BF_{10} > 100$; T_2 : $r = .47$, $p =$, $BF_{10} > 100$) and the response time for correct answers (T_1: $r = -.63$, $p < .001$, $BF_{10} > 100$; T_2 : $r = -.73$, $p < .001$, $BF_{10} > 100$).

There was a significant increase in performance between the two time points in arithmetic achievement ($t(119) = -18.74$, $p < .001$, $BF_{10} > 100$), in metacognitive monitoring ($t(120) = -6.30$, $p < .001$, $BF_{10} > 100$) and in the raw scores of the intellectual ability measure ($t(120) = -7.10$, $p < .001$, $BF_{10} > 100$), but not in mathematics anxiety ($t(120) = 1.80$, $p = .07$, $BF_{10} = 0.48$).

1.1 Correlational analyses

Pearson correlation coefficients of the associations between arithmetic achievement, metacognitive monitoring and mathematics anxiety are presented in Table 4.2. A full correlation matrix that includes all variables reported on here is included in Appendix A. As correlations of our key variables with intellectual ability were not significant/supported, intellectual ability was not considered further.

Arithmetic achievement was significantly correlated with metacognitive monitoring at both time points and across time, with Bayes factors indicating decisive evidence in favour of all these associations. Arithmetic achievement was significantly correlated with mathematics anxiety at both time points and across time. An inspection of the Bayes factors provided a more nuanced interpretation of these associations, indicating decisive evidence for all associations with mathematics anxiety at T_2. The Bayes factors indicated moderate evidence in favour of the association between arithmetic achievement T_1 and mathematics anxiety T_1 and only anecdotal evidence for the association between mathematics anxiety T_1 and arithmetic achievement T_2.

The association between metacognitive monitoring and mathematics anxiety was significant at both time points, yet the Bayes factors only supported the association at T_2, while at T1 the evidence of this associations was only anecdotal. Across time points, the results show that mathematics anxiety at T_1 was significantly correlated with metacognitive monitoring at T_2, with a Bayes factor indicating there was moderate evidence in favour of the association. The association of metacognitive monitoring T_1 with mathematics anxiety T_2 was not significant, with a Bayes Factor indicating moderate evidence in favour of the null hypothesis.

Table 4.2
Correlation analyses of arithmetic achievement, metacognitive monitoring and mathematics anxiety at both time points

		Arithmetic achievement		Metacognitive monitoring		Mathematics anxiety
		T_1	T_2	T_1	T_2	T_1
Arithmetic achievement						
T_2						
	r	.85	-			
	p	< .001	-			
	BF_{10}	>100	-			
Metacognitive monitoring						
T_1						
	r	.41	.35	-		
	p	<.001	<.001	-		
	BF_{10}	>100	>100	-		
T_2						
	r	.38	.36	.45	-	
	p	<.001	<.001	<.001	-	
	BF_{10}	>100	>100	>100	-	
Mathematics anxiety						
T_1						
	r	-.25	-.20	-.21	-.24	-
	p	.004	.03	.02	.008	-
	BF_{10}	6.20	1.25	1.65	3.82	-
T_2						
	r	-.35	-.42	-.17	-.40	.35
	p	<.001	<.001	.07	<.001	<.001
	BF_{10}	>100	>100	0.58	>100	>100

Note. T_1 = time point 1 (Grade 2); T_2 = time point 2 (Grade 3).

1.2 Regression analyses

Regression analyses were used to predict arithmetic achievement, metacognitive monitoring performance, and mathematics anxiety in third grade (T_2), based on arithmetic achievement, metacognitive monitoring performance, and mathematics anxiety in second grade (T_1). Only variables that were significantly correlated with the outcome measure at the level of zero-order correlations were considered in the regression models. Firstly, arithmetic achievement at T_2 was predicted with multiple regression analyses using the autoregressor (arithmetic achievement T_1) and metacognitive monitoring T_1 (Table 4.3a) and mathematics anxiety T_1 (Table 4.3b); secondly, metacognitive monitoring T_2 was predicted with multiple regression analyses using the autoregressor (metacognitive monitoring T_1), and arithmetic achievement at T_1 (Table 4.3c), and mathematics anxiety T_1 (Table 4.3d) and both arithmetic achievement T1 and mathematics anxiety T1 simultaneously (Table 4.3e); lastly, mathematics anxiety T_2 was predicted with multiple regression analyses using the autoregressor (mathematics anxiety T_1) and arithmetic achievement T_1 (Table 4.3f).

Table 4.3
Regression analyses of arithmetic achievement *T_2, metacognitive monitoring T_2 and mathematics anxiety T_2*

	Arithmetic achievement T_2			
	β	t	p	$BF_{inclusion}$
(a) Adjusted $R^2 = .71$				
Autoregressor – Arithmetic achievement T_1	.84	15.63	<.001	>100
Metacognitive monitoring T_1	.02	0.29	.77	0.15
(b) Adjusted $R^2 = .71$				
Autoregressor – Arithmetic achievement T_1	.80	16.69	<.001	>100
Mathematics anxiety T_1	.01	0.10	.92	0.14
	Metacognitive monitoring T_2			
	β	t	p	$BF_{inclusion}$
(c) Adjusted $R^2 = .24$				
Autoregressor - Metacognitive monitoring T_1	.37	4.24	<.001	>100
Arithmetic achievement T_1	.23	2.61	.01	8.35
(d) Adjusted $R^2 = .22$				
Autoregressor - Metacognitive monitoring T_1	.42	5.11	<.001	>100
Mathematics anxiety T_1	-.16	-1.89	.06	1.93
(e) Adjusted $R^2 = .25$				
Autoregressor - Metacognitive monitoring T_1	.35	4.02	<.001	>100
Arithmetic achievement T_1	.21	2.38	.02	5.86
Mathematics anxiety T_1	-.10	-1.15	.25	1.07
	Mathematics anxiety T_2			
	β	t	p	$BF_{inclusion}$
(f) Adjusted $R^2 = .19$				
Autoregressor - Mathematics anxiety T_1	.28	3.33	.001	59.70
Arithmetic achievement T_1	-.28	-3.28	.001	52.75

Note. T_1 = time point 1 (Grade 2); T_2 = time point 2 (Grade 3).

All outcome measures (i.e. arithmetic achievement, metacognitive monitoring, and mathematics anxiety at T_2) were significantly predicted by their autoregressor on top of the other considered variables.

Arithmetic achievement was not significantly predicted by either metacognitive monitoring or mathematics anxiety on top of its autoregressor, with Bayes factors indicating moderate to strong evidence for the null hypotheses.

Metacognitive monitoring T_2 was significantly predicted by arithmetic achievement $T_{1,}$ in addition to both the autoregressor and to mathematics anxiety T_1, with Bayes factors indicating moderate evidence for the predictive value of arithmetic achievement T_1. Mathematics anxiety $T_{1,}$ on the other hand did not significantly predict metacognitive monitoring T_2 on top of the autoregressor, with Bayes factors indicating only anecdotal evidence for the influence of mathematics anxiety in predicting metacognitive monitoring.

Mathematics anxiety T_2 was significantly predicted by arithmetic achievement T_1, in addition to the autoregressor, with a Bayes Factor indicating very strong evidence in favour of the predictive value of arithmetic achievement T_1. Metacognitive monitoring T_1 was not significantly related to mathematics anxiety T_2, for which reason it was not included as a predictor variable in the regression analyses.

1.2.1 Mediation analyses.

These regression analyses might hint to underlying interrelations between arithmetic achievement, metacognition and mathematics anxiety that may explain some variance in their respective bivariate associations (e.g. Is the association between arithmetic achievement and metacognitive monitoring influenced by mathematics anxiety?). Yet the abovementioned regression analyses cannot provide evidence for those hypotheses. Therefore, we performed mediation analyses to investigate whether the associations that were found, were either direct effects of one variable on another, or whether the associations were mediated via a third variable. For example, we investigated whether metacognitive monitoring directly predicted arithmetic achievement, or whether its predictive power was indirect, via mathematics anxiety. As preregistered, only associations that were found significant/supported in the zero-order correlational analyses were considered for mediation and moderation analyses.

All mediation analyses are presented in Figure 4.2. We first investigated whether mathematics anxiety mediated the association between arithmetic achievement and metacognitive monitoring. Mathematics anxiety T_1 did not mediate the relation between metacognitive monitoring T_1 as antecedent and arithmetic achievement T_2 as consequent variable (bootstrapped 95% confidence interval of indirect path [-1.43; 10.47]; Figure 4.2A), nor did mathematics anxiety T_1 mediate the relation between arithmetic achievement T_1 as antecedent and metacognitive monitoring T_2 as consequent variable (bootstrapped 95% confidence interval of indirect path [0.000; 0.0007]; Figure 4.2B). Both direct effects (i.e. direct effect of metacognitive monitoring on arithmetic achievement and direct effect of arithmetic achievement on metacognitive monitoring) were significant.

Secondly, we tested whether metacognitive monitoring mediated the association between arithmetic achievement and mathematics anxiety. Metacognitive monitoring T_1 mediated the relation between mathematics anxiety T_1 as antecedent variable and arithmetic achievement T_2 as consequent (bootstrapped 95% confidence interval of indirect path [-.30; -.02]; Figure 4.2C). The direct effect of mathematics anxiety on arithmetic achievement was not significant. On the other hand, metacognitive monitoring T_1 did not mediate the relation between arithmetic achievement T_1 as antecedent variable and mathematics anxiety T_2 as consequent (bootstrapped 95% confidence interval of indirect path [-0.04; 0.03]; Figure 4.2D). The direct effect of arithmetic achievement on mathematics anxiety was significant.

Lastly, we tested whether arithmetic achievement mediated the association between metacognitive monitoring and mathematics anxiety. A mediation analysis showed that arithmetic achievement T_1 mediated the relation between mathematics anxiety T_1 as antecedent variable and metacognitive monitoring T_2 as consequent (bootstrapped 95% confidence interval of indirect path [-0.0020; -0.0003]; Figure 4.2E). The direct effect of mathematics anxiety on metacognition was not significant. As there was no significant association between mathematics anxiety T_2 and metacognitive monitoring T_1, no mediation analyses were performed on this association.

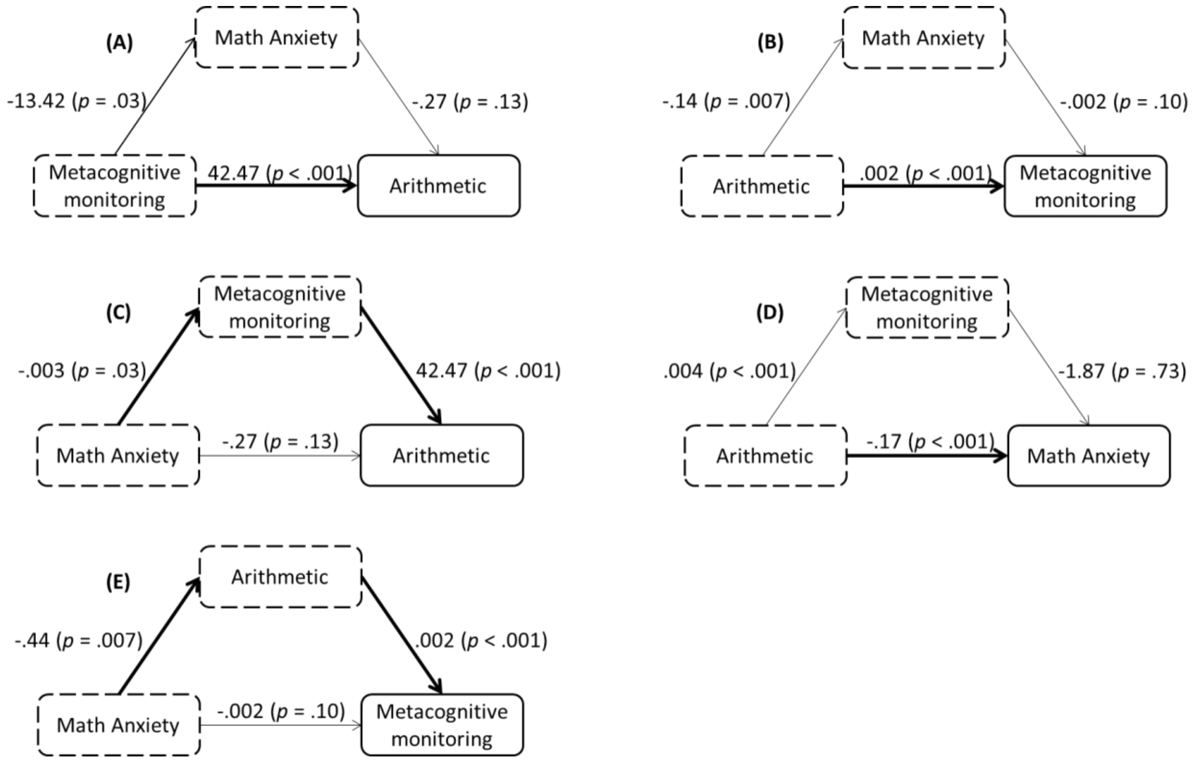

Figure 4.2. Mediation analyses between arithmetic achievement, metacognitive monitoring and mathematics anxiety. Antecedent variables and mediators (i.e. in box with dashed lines) are measured at T_1. Consequent variables (i.e. in box with solid line) are measured at T_2.

1.2.2 Moderation analyses.

While through mediation analyses, we investigated whether the found effects were direct or indirect within the entire population, we used moderation analyses to explore whether the (size of the) effect of one variable on another is dependent on a third variable. Namely, we wondered when or under what circumstances one variable exerts an effect on another. For example, we studied whether the relationship between arithmetic achievement and metacognitive monitoring might be different at different levels of mathematics anxiety.

Using moderation analyses with arithmetic achievement (T_2) as dependent variable, metacognitive monitoring T_1 did not moderate the effect of mathematics anxiety T_1, nor did mathematics anxiety T_1 moderate the effect of metacognitive monitoring T_1. The interaction between metacognitive monitoring T_1 and mathematics anxiety T_1 was not significant ($\beta = -2.14$, $p = .11$).

Using moderation analyses with metacognitive monitoring (T_2) as dependent variable, arithmetic achievement T_1 did not moderate the effect of mathematics anxiety T_1, nor did mathematics anxiety T_1 moderate the effect of arithmetic achievement T_1. The interaction between mathematics anxiety T_1 and arithmetic achievement T_1 was not significant (β = -0.00, p = .80).

Using moderation analyses with mathematics anxiety (T_2) as dependent variable, arithmetic achievement T_1 did not moderate the effect of metacognitive monitoring T_1, nor did metacognitive monitoring T_1 moderate the effect of arithmetic achievement T_1. The interaction between metacognitive monitoring T_1 and arithmetic achievement T_1 was not significant (β = -0.12, p = .67).

Discussion

Throughout the literature on arithmetic achievement, metacognitive monitoring and mathematics anxiety have been identified as associated with arithmetic achievement in children (e.g. Bellon et al., 2019; Hembree, 1990; Rinne & Mazzocco, 2014; Shrager & Siegler, 1998). As metacognitive monitoring and mathematics anxiety have been studied in relative isolation, to date, it is unclear whether or not considering both variables simultaneously has an influence on their associations with arithmetic achievement. Some indications on the interplay between these variables can be found in the existing literature on adults (e.g. Erickson & Heit, 2015; Legg & Locker, 2009), yet, studies on this specific interplay in children are scarce. Importantly, the developmental dynamics between these variables have never been studied. The current study is the first to tackle this important outstanding issue in young primary school children by including metacognitive monitoring and mathematics anxiety simultaneously to investigate their interrelation and interplay in arithmetic achievement in second to third grade, using a longitudinal panel design. This design allowed us to thoroughly investigate the developmental dynamics and interrelations of arithmetic achievement, metacognitive monitoring and mathematics anxiety in early primary school, a crucial developmental period for arithmetic achievement, as well as metacognitive monitoring and mathematics anxiety.

The results of this study provide essential new insights into this issue, while on the other hand replicating some important results from the existing literature in the isolated fields of metacognitive monitoring and mathematics anxiety. As an initial step, the intercorrelations between arithmetic achievement, metacognitive monitoring and mathematics anxiety were studied longitudinally. Firstly, we found significant and supported associations between the standardized arithmetic achievement test and metacognitive monitoring at both time points (in line with Bellon et al., 2019; Rinne & Mazzocco, 2014). Importantly, this study is the first to longitudinally confirm this relation across time points. Secondly, significant and supported associations between arithmetic achievement and mathematics anxiety were found on both time points, and across time, in line with the meta-analytic results of Hembree (1990), Ma (1999) and Namkung et al. (2019), who found associations of a similar size (i.e.

r = [-.27 – -.34]). Going beyond the existing body of evidence, the current study is the first to show that metacognitive monitoring and mathematics anxiety are associated at different time points and across time in second and third grade primary school children.

Due to its longitudinal panel design, the current study was able to provide empirical evidence on the developmental dynamics of arithmetic achievement, metacognitive monitoring and mathematics anxiety in young primary school children and to uncover whether underlying interrelations between arithmetic achievement, metacognition and mathematics anxiety explain some variance in their respective bivariate associations. As such, the current study provided empirical evidence pointing to the direction of the association between arithmetic achievement and metacognitive monitoring. The results of the regression analyses suggest that the longitudinal association between arithmetic achievement and metacognitive monitoring is mostly driven by the influence of arithmetic achievement on metacognitive monitoring skills rather than the other way around, as arithmetic achievement was found to be a unique predictor of later metacognitive monitoring (over and above its autoregressive effect and mathematics anxiety), while on the other hand metacognitive monitoring did not predict later arithmetic achievement on top of its autoregressive effect. This longitudinal relation between arithmetic achievement and metacognitive monitoring cannot be explained by mathematics anxiety, as we did not observe any mediation or moderation effects of mathematics anxiety.

In line with the existing literature, our results suggest that mathematics anxiety, and its negative association with academic achievement, is already present in young primary school children (e.g. Maloney & Beilock, 2012; Ramirez et al., 2016; Wu et al., 2012). The Bayes factors in the current study provided important insight into the strength of the associations with arithmetic achievement, showing that the negative association between arithmetic achievement and mathematics anxiety increases across development, a finding that is in line with existing research (e.g. Dowker, 2019e; Dowker et al., 2012, 2016; Ma & Kishor, 1997; Wood et al., 2012). These changes in the association between arithmetic achievement and mathematics anxiety may be the result of the increasing experience of mathematical success and failure. Mathematics anxiety may increase in those children whose poor performance results in repeated failure experiences, in contrast to children who experience greater success in mathematics (Dowker, 2019). As according to behaviouristic models, anxiety emerges as an obligatory response to an aversive stimulus (Watson & Rayner, 1920), it is plausible that frequent poor mathematics performance leads to negative emotions such as mathematics anxiety (e.g. Krinzinger et al., 2009).

Another possible underlying mechanism that might be at play in the increased association between arithmetic achievement and mathematics anxiety is the metacognitive awareness of one's own performance. Specifically, primary school children develop a better notion of their own performance over time and, consequently, the individual differences in arithmetic achievement between themselves and their peers might become more apparent. This may increase feelings of pressure and anxiety and strengthen the relation between arithmetic achievement and mathematics anxiety. This hypothesis is in line with Tobias (1986), who stated that poor performance is a function of an acquisition deficit and the observed elevation in anxiety is attributable to student's metacognitive awareness of their incomplete learning.

We further tested these associations via regression analyses on the longitudinal data. The results showed that while mathematics anxiety did not predict later arithmetic achievement, arithmetic achievement uniquely predicted mathematics anxiety on top of its autoregressive effect and without mediation or moderation of metacognitive monitoring. These results are in line with the deficit model of mathematics anxiety, stating that poor performance leads to higher anxiety about that situation in the future (Ma & Xu, 2004; Tobias, 1986). Concerning mathematics anxiety, this means that recollection of poor mathematics performance may generate mathematics anxiety (e.g. Carey et al., 2016; Dowker et al., 2016; Hembree, 1990; Ma & Xu, 2004; Maloney et al., 2015; Sorvo et al., 2019). On the other hand, our study partially confirms the hypothesis made by Tobias (1986) that metacognitive monitoring mediates the association between arithmetic achievement and mathematics anxiety. While our results indicate a strong, unique predictive power of arithmetic achievement for mathematics anxiety, without mediation or moderation by metacognitive monitoring, mathematics anxiety does not directly predict arithmetic achievement. Its effect is indirect: the longitudinal predictive power of mathematics anxiety on arithmetic achievement is mediated by metacognitive monitoring. This indicates that mathematics anxiety is only predictive of arithmetic achievement through metacognitive awareness. This result is in line with Lai and colleagues (2015), who found the same mediation process in 10-year-old Chinese children in word problem solving, and with Özcan and Gümüs (2019), who found this in 11-12-year-old Turkish children. Our results go beyond the previous ones by showing this mediation already occurs in earlier grades of primary school, and by using longitudinal data that allowed us to more properly investigate potential mediation processes.

The current study shows that the negative association between metacognitive monitoring and mathematics anxiety increases over development. This might be due to the age of the participants, as both metacognitive monitoring and mathematics anxiety are at a crucial point in development in early primary school. On the one hand, metacognitive monitoring accuracy is increasing at this age (e.g. Ghetti, 2008; Lyons & Ghetti, 2010; Schneider, 2008, 2010; Schneider & Lockl, 2008; Schneider & Löffler, 2016). With children getting better at correctly evaluating their performance, repeated failure in arithmetic may have a stronger impact and lead to an increasing association between metacognitive monitoring and mathematics anxiety. Moreover, children might become more aware of the individual differences in arithmetic achievement between them and their peers and thereby increase feelings of pressure and anxiety, which may strengthen the relation between metacognition and mathematics anxiety. On the other hand, as mathematics anxiety might develop very early in primary school (e.g. Maloney & Beilock, 2012; Ramirez et al., 2013), earlier mathematics anxiety may limit one's ability to correctly monitor their performance later in development and thereby strengthening the association between them (Morsanyi et al., 2019). The results of our longitudinal associations are most in line with this last hypothesis, as only the association between mathematics anxiety at T_1 and metacognitive monitoring at T_2 was significant and supported.

To conscientiously investigate these associations between metacognitive monitoring and mathematics anxiety, we also examined whether performance in arithmetic achievement might affect this relation between metacognitive monitoring and mathematics anxiety. When early arithmetic achievement was included as a predictor on top of mathematics anxiety and the autoregressor of metacognitive monitoring, there was no evidence for the predictive power of mathematics anxiety for later metacognitive monitoring. This was confirmed by the mediation analyses, which showed that across time points, mathematics anxiety did not directly predict metacognitive monitoring, but that this relation was mediated via arithmetic achievement. These results suggest that the reason for a correlation between metacognitive monitoring and mathematics anxiety across time points is because mathematics anxiety is related to arithmetic achievement, which in turn correlates to later metacognitive monitoring skills.

Integrating these different results, a similar pattern of increased associations of mathematics anxiety with arithmetic achievement and metacognitive monitoring is observed over development. This might suggest an increasingly important role of mathematics anxiety in the development of primary school children at different levels of performance, i.e. academic as well as metacognitive performance. It is therefore important to make sure early signs of mathematics anxiety are detected or, ideally, that mathematics anxiety is prevented. This can be done by, for example, modelling positive attitudes to mathematics and avoiding expressing negative attitudes towards children, so that a vicious spiral in which mathematics anxiety and difficulties with mathematics reinforce one another is prevented (e.g. Dowker et al., 2016).

The predictive power of arithmetic achievement for metacognitive monitoring as well as mathematics anxiety was strong and independent of their autoregressors and, respectively, mathematics anxiety and metacognitive monitoring. The same was not true for the predictive power of either metacognitive monitoring or mathematics anxiety for arithmetic achievement. It is important to acknowledge that the autoregressive effect of arithmetic achievement was very high (i.e. $r = .85$, $p < .001$), which makes the possibility to explain additional variance in arithmetic achievement on top of this autoregressor very difficult. On the other hand, our arithmetic achievement measure might already capture individual differences in metacognitive monitoring and mathematics anxiety, which are both processes that co-occur with performance. This may have also been reflected in the autoregressive effect. To further investigate this possibility, future studies should test the effects of interventions targeted at both metacognitive monitoring and mathematics anxiety to explore their impact on arithmetic achievement and as such bypass the potential measurement overlap with arithmetic.

No moderation effects were found within the associations between arithmetic achievement, metacognitive monitoring and mathematics anxiety in primary school children, which indicates that the associations that were found were equal across different levels of arithmetic achievement, metacognitive monitoring and mathematics anxiety within the entire sample. These results add important insight to the existing literature, by contradicting hypotheses saying the association between arithmetic achievement and mathematics anxiety might be stronger in highly metacognitive competent children, or hypotheses saying the association between metacognitive monitoring and mathematics anxiety might be stronger in low arithmetic achievers. Our results are not in line with the study on 56 adults by Legg and Locker (2009), who found that metacognition (as measured using a metacognitive inventory assessing planning, checking, monitoring and evaluating behaviours) moderated the relation between arithmetic achievement and mathematics anxiety, predicting that performance decreased as anxiety increased, except at high metacognition levels. These differences in results indicate that the interplay between arithmetic achievement, metacognition and mathematics anxiety may be different at different ages, and in different operationalisations of metacognition. Future studies should therefore further clarify at what ages and for which aspects of metacognition moderation of the association between arithmetic achievement and mathematics anxiety occurs.

A limitation of the current study might be that the mathematics anxiety measure used in the current study focusses mainly on the affective dimension of mathematics anxiety (i.e. emotions of fear, nervousness and tension with their associated physiological reactions, which occur in the presence of numerical stimuli, whether or not there is a threat of failure or evaluation; Wigfield & Meece, 1988). Mathematics anxiety indeed consists of two different dimensions (Dowker, 2019e), namely the affective and the cognitive dimension. It might be the case that the cognitive dimension (i.e. concerns about how one is performing and the fear of failure) is even more highly correlated with performance monitoring because of the overlap in 'thinking about your performance'. However, most research focusing on the affective dimension has typically shown a relation between mathematics anxiety and mathematics achievement in primary school children (e.g. Vukovic et al., 2013; Wu et al., 2012), while studies focusing on the cognitive dimension have tended not to show such an association in young children (e.g. Dowker et al., 2012; Krinzinger et al., 2009; Wood et al., 2012). Studies which included both dimensions of mathematics anxiety have suggested that performance in young children is related to the affective but not the cognitive dimension (e.g. Dowker, 2019; Sorvo et al., 2017).

Future research should build on this first empirical study into the interrelations between arithmetic achievement, metacognition and mathematics anxiety. Such research should further examine the role of executive functions within the present intercorrelations, as these executive functions may contribute to the correlation between skills at different time points. Indeed, executive functions have been associated with mathematics ability (e.g. Bull & Lee, 2014), with mathematics anxiety (e.g. Ashcraft & Kirk, 2001) and, mostly on a theoretical level, with metacognitive monitoring (e.g. Roebers & Feurer, 2016).

However, the report of the first wave of this longitudinal study (Bellon et al., 2019) revealed a lack of associations between executive functions, arithmetic achievement and our current measure of metacognitive monitoring, making it rather unlikely that these executive functioning measures would have an important impact on the current findings.

In view of the observation of an association between general anxiety and enhanced amplitude of error-related negativity (see Moser et al. 2013 for a meta-analysis) there is a need for brain imaging studies to look at the relation of mathematics anxiety with monitoring using EEG, particularly in developmental populations. Other possibilities would be to investigate the connectivity between prefrontal regions – which are known for their involvement in metacognitive processes (e.g. Fleming & Dolan, 2014) – and the amygdala (i.e. mathematics anxiety is associated to hyperactivity in the right amygdala regions, which are important for processing negative emotions; Young et al., 2012), or to investigate differences in activation patterns between high vs low mathematics anxious children during monitoring tasks using fMRI. As most of the existing neuro-imaging studies are on adults, it is important to specifically study this association between (mathematics) anxiety and metacognitive monitoring in children.

Conclusion

To conclude, this study further unravels the interplay between arithmetic achievement, metacognitive monitoring and mathematics anxiety in early primary school children in second and third grade. The current study shows that, while metacognitive monitoring and mathematics anxiety are indeed correlated, the association between metacognitive monitoring and arithmetic achievement is a unique, specific one, without mediation or moderation of mathematics anxiety. The results of this study also clearly indicate that arithmetic achievement in itself is an important, unique predictor of both metacognitive monitoring and mathematics anxiety later in development, emphasizing the importance of skill development in the development of metacognitive monitoring, mathematics anxiety and their interrelations. Strengthening children’s arithmetic achievement through targeted interventions potentially increases their metacognitive monitoring and reduces mathematics anxiety, a possibility that should be tested in future research.

Appendixes

1 Appendix A – Correlational analyses between the administered measures

Table 4.A1

Correlation analyses between the administered measures

			1a		1b		1c		2		3		4
			Arithmetic ACC		Arithmetic RT		TTA		Metacognitive monitoring		Mathematics anxiety		Intellectual ability
			T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1
Arithmetic - Custom task	1a. Arithmetic ACC												
	T_2												
		r	.60	-									
		p	<.001	-									
		BF_{10}	>100	-									
	1b. Arithmetic RT												
	T_1												
		r	-.21	-.19	-								
		p	.02	.04	-								
		BF_{10}	1.74	0.97	-								
	T_2												
		r	-.30	-.34	.72	-							
		p	.001	<.001	<.001	-							
		BF_{10}	18.99	>100	>100	-							
Arithmetic - Standard task	1c. TTA												
	T_1												
		r	.43	.38	-63	-.66	-						
		p	<.001	<.001	<.001	<.001	-						
		BF_{10}	>100	>100	>100	>100	-						
	T_2												
		r	.39	.47	-.64	-.34	.85	-					
		p	<.001	<.001	<.001	<.001	<.001	-					
		BF_{10}	>100	>100	>100	>100	>100	-					
Metacognitive monitoring	2. Metacognitive monitoring												
	T_1												
		r	.72	.42	-.27	-.28	.41	.35	-				
		p	<.001	<.001	.002	.002	<.001	<.001	-				
		BF_{10}	>100	>100	12.89	13.19	>100	>100	-				
	T_2												
		r	.41	.56	-.21	-.34	.38	.36	.45	-			
		p	<.001	<.001	.02	<.001	<.001	<.001	<.001	-			
		BF_{10}	>100	>100	1.52	>100	>100	>100	>100	-			

(table continues on the next page)

			1a		1b		1c		2		3		4
			Arithmetic ACC		Arithmetic RT		TTA		Metacognitive monitoring		Mathematics anxiety		Intellectual ability
			T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1
Mathematics anxiety	3. Mathematics anxiety questionnaire												
	T_1												
		r	-.23	-.17	.16	.31	-.25	-.20	-.21	-.24	-		
		p	.01	.06	.07	.001	.004	0.3	.02	.01	-		
		BF_{10}	2.70	0.69	0.56	40.10	6.20	1.25	1.65	3.82	-		
	T_2												
		r	-.22	-.40	.19	.41	-.35	-.42	-.17	-.40	.35	-	
		p	.02	<.001	.03	<.001	<.001	<.001	.07	<.001	<.001	-	
		BF_{10}	2.12	>100	1.05	>100	>100	>100	0.58	>100	>100	-	
Control variable – Intellectual ability	4. Intellectual ability - Raven												
	T_1												
		r	.28	.34	.18	.03	-.01	.02	.13	.22	-.09	-.03	-
		p	.001	<.001	.04	.73	.88	.82	.14	.02	.29	.75	-
		BF_{10}	19.07	>100	0.92	0.12	0.11	0.12	0.33	1.86	0.19	0.12	-
	T_2												
		r	.34	.29	.14	.02	-.04	-.002	.20	.16	-.10	-.06	.65
		p	<.001	.001	.13	.85	.63	.98	.03	0.9	.27	.53	<.001
		BF_{10}	>100	19.21	0.36	0.12	0.13	0.11	1.27	0.49	0.21	0.14	>100

Note. ACC = accuracy; RT = response time for the correct answers.

CHAPTER 5

Metacognition across domains

Is the association between arithmetic and metacognitive monitoring domain-specific?

The content of this chapter has been published as:

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), e0229932. doi: 10.1371/journal.pone.0229932

Chapter 5

Metacognition across domains.

Is the association between arithmetic and metacognitive monitoring domain-specific?

Abstract

Metacognitive monitoring is a critical predictor of arithmetic in primary school. One outstanding question is whether this metacognitive monitoring is domain-specific or whether it reflects a more general performance monitoring process. To answer this conundrum, we investigated metacognitive monitoring in two related, yet distinct academic domains: arithmetic and spelling. This allowed us to investigate whether monitoring in one domain correlated with monitoring in the other domain, and whether monitoring in one domain was predictive of performance in the other, and vice versa. Participants were 147 typically developing 8-9-year-old children (Study 1) and 77 typically developing 7-8-year-old children (Study 2), who were in the middle of an important developmental period for both metacognitive monitoring and academic skills. Pre-registered analyses revealed that within-domain metacognitive monitoring was an important predictor of arithmetic and spelling at both ages. In 8-9-year-olds the metacognitive monitoring measures in different academic domains were predictive of each other, even after taking into account academic performance in these domains. Monitoring in arithmetic was an important predictor of spelling performance, even when arithmetic performance was controlled for. Likewise, monitoring in spelling was an important predictor of arithmetic performance, even when spelling performance was controlled for. In 7-8-year-olds metacognitive monitoring was domain-specific, with neither correlations between the monitoring measures, nor correlations between monitoring in one domain and performance in the other. Taken together, these findings indicate that more domain-general metacognitive monitoring processes emerge over the ages from 7 to 9.

Introduction

"*Learn from your mistakes*" is an old saying that (grand)parents teach their children. This goes back to the premise that making mistakes is associated with learning. Noticing your mistakes is an example of monitoring your cognition. This monitoring of cognition is a facet of metacognition, a concept first introduced by Flavell (1979). One critical component of metacognition is procedural metacognition. This is a self-reflecting, higher-order cognitive process, which indicates how people monitor and control their cognition during ongoing cognitive processes (Flavell, 1999; Nelson & Narens, 1990). Metacognitive monitoring is an important aspect of procedural metacognition and is defined as the subjective self-assessment of how well a cognitive task will be/is/has been performed (Morsanyi et al., 2019; Nelson & Narens, 1990).

Two recent studies found evidence for metacognitive monitoring as an important predictor of arithmetic performance (Bellon et al., 2019; Rinne & Mazzocco, 2014). To determine the role of metacognitive monitoring, these authors asked children on a trial-by-trial basis to report their judgement of the accuracy of their answers during an arithmetic task. Both studies found that successful appraisal of the accuracy of one's arithmetic judgement is a powerful predictor of arithmetic performance in primary school children. To date, however, it is unclear whether the results regarding the strength of the role of metacognitive monitoring in arithmetic are specific to the arithmetic domain, or whether they are reflective of a more general metacognitive monitoring process in academic performance; an outstanding question on which this study will focus.

Metacognition has been regarded as a fundamental skill influencing cognitive performance and learning in domains as diverse as arithmetic, memory, reading, perception, and many others (e.g. Annevirta et al., 2007; Block & Peskowitz, 1990; Efklides & Misailidi, 2010; Freeman et al., 2017; Kuhn, 2000; Lyons & Ghetti, 2013; Özsoy, 2011; Rinne & Mazzocco, 2014; Schneider, 1998; Schneider & Artelt, 2010; Schraw et al., 2006; Veenman et al., 2006, 2004). The importance of metacognition that was found in existing research in different (cognitive) domains is not surprising, as metacognitive aspects, such as knowing the limits of your own knowledge and being able to regulate that knowledge, are essential components of self-regulated and successful learning (Schraw et al., 2006), enabling learners to improve their cognitive performance. For example, good metacognition allows learners to correctly allocate study-time, check answers when they feel unsure about the correctness of the answer or provide a learning moment when an error is detected. Besides being considered a global ability playing a role in a large range of domains, metacognition, and consequently metacognitive monitoring, is usually considered to be a domain-general cognitive process that is correlated across content domains. This suggests that people who are good at evaluating their performance for one task, also tend to be good at evaluating their performance for another task (e.g. Geurten et al., 2018; Schraw et al., 1995). There is, however, evidence suggesting that this domain-generality only emerges over development.

Geurten and colleagues (2018) recently observed that metacognition is first domain-specific and then generalizes across domains as children mature. They found a gradual shift from domain-specific towards domain-general metacognition across the arithmetic and memory domains in children aged between 8 and 13. In adults, more evidence for the domain-generality has been observed. Veenman and colleagues (1997) and Schraw and colleagues (Schraw et al., 1995; Schraw & Nietfeld, 1998) found that metacognitive measures are correlated across unrelated (cognitive) tasks. More specifically, Schraw and colleagues (1995) found significant correlations between metacognitive measures across eight different domains ranging from historic knowledge to knowledge of caloric values of food. This domain-general hypothesis in adults is also supported by brain imaging data that show that adults' metacognitive abilities for different types of tasks partially depend on common neurobiological structures, such as the prefrontal cortex and precuneus (Fleming & Dolan, 2014).

However, domain-specific knowledge and skills also seem to be important for metacognitive monitoring. For example, in young children (ages 5 to 8 years), Vo and colleagues (2014) showed that metacognition in the numerical domain was unrelated to metacognition in the emotional domain, suggesting young children's metacognition is domain-specific. Based on their empirical findings, Schraw and colleagues (1995) suggested that in adults metacognitive monitoring within a specific domain is governed by general metacognitive processes in addition to domain-specific knowledge. Löffler, Von Der Linden and Schneider (2016) documented a twofold effect of expertise on monitoring in soccer: Although domain-specific knowledge enhances monitoring performance in some situations, more optimistic estimates (presumably due to the application of a familiarity heuristic) typically reduce monitoring accuracy in experts. Likewise, in mathematics, metacognitive monitoring has been found to be a function of domain-specific ability (e.g. Garcia et al., 2016; Lingel et al., 2019). Taken together, the existing research also illustrates the importance of domain-specific knowledge and skills for metacognitive monitoring.

This issue of domain-specificity is a longstanding debate within the metacognitive literature (e.g. Geurten et al., 2018; Schraw et al., 1995; Schraw & Nietfeld, 1998; van Bon & Kuijpers, 2016; Veenman et al., 1997), both at the behavioural and brain-imaging level. Yet, in children, the results are scarce and rather inconclusive, with different results for various age groups as well as metacognitive measures.

Firstly, age-related differences in the results on domain-specificity of metacognition in children are not surprising, as a critical development in monitoring is observed during early to late childhood (e.g. Geurten et al., 2018; Lyons & Ghetti, 2010). For example, in (early) primary school, metacognitive monitoring accuracy is found to increase (e.g. Ghetti, 2008; Lyons & Ghetti, 2010; Schneider, 2008, 2010; Schneider & Lockl, 2008; Schneider & Löffler, 2016). In the same developmental time period of these age-related improvements in monitoring of cognition, there are also important age-related improvements in academic skills, such as arithmetic and spelling. The age-related metacognitive improvements are recognized to underlie several aspects of cognitive development in various domains (e.g. improvements in accuracy; e.g. Lyons & Ghetti, 2010). Furthermore, based on their empirical findings, Geurten and colleagues (2018) conclude that a gradual shift toward domain-general metacognition occurs in children between 8 and 13 years, and that metacognition is no more bound by task content and domain knowledge after the age of 10. Against this background and to thoroughly investigate the domain-specificity question in children, the current research specifically recruited 8-9 year-olds (third grade; Study 1) and 7-8 year-olds (second grade; Study 2), who are in the middle of this important developmental period for both metacognitive monitoring and academic skills.

Secondly, the different results on domain-specificity of metacognition in children for different metacognitive measures may in part be due to different aspects of metacognition being investigated. Metacognition includes both declarative and procedural metacognition. As metacognition encompasses different aspects, it is not surprising that these different aspects of metacognition follow different developmental paths (Schneider & Löffler, 2016) and that they are differently associated with domain-specific knowledge and skills. A recent study by Bellon and colleagues (2019), for example, found that within-domain metacognitive monitoring was associated with arithmetic performance, while declarative metacognitive knowledge was not. The authors suggest this might indicate that children's metacognition is more domain-specific than it is domain-general. Yet, the authors based their suggestion on results on different aspects of metacognition, which were measured fundamentally differently (i.e. online, trial-by-trial reports for metacognitive monitoring vs. general questionnaire for declarative metacognitive knowledge), making testing the domain-specificity hypothesis as well as making strong claims about domain-specificity of metacognition troublesome.

To overcome these issues, the current research specifically focused on the monitoring aspect of metacognition. Extending the existing body of data, we included, in addition to the metacognitive monitoring measure in arithmetic, the same metacognitive monitoring measure in another domain of academic learning, i.e. spelling. By including metacognitive monitoring measures in two domains, and, importantly, by using the exact same paradigm to measure it, the current study was able to investigate the question of domain-specificity more thoroughly. The paradigm to measure metacognitive monitoring was the same as in Bellon et al. (2019) and Rinne and Mazzocco (2014). Spelling was included as a second domain to maximize the comparability of the two tasks in which metacognitive monitoring was measured. Arithmetic and spelling are quintessential domains in primary school and in both domains primary school children go through crucial developmental steps. Based on the children's curriculum, we were able to select age-appropriate items. This allowed us to thoroughly investigate whether the results on the role of metacognitive monitoring in arithmetic are specific to the arithmetic domain or not.

Based on the outstanding issues outlined above, this study aims to extend and deepen our knowledge on the domain-specificity of the role of metacognition in different academic domains in middle childhood. Specifically, this study will investigate whether metacognitive monitoring is domain-specific or not by investigating (a) the associations between within-domain metacognitive monitoring and arithmetic and spelling; (b) whether metacognitive monitoring in one domain is associated with and/or predicted by metacognitive monitoring in the other domain; (c) whether performance in one domain is associated with and/or predicted by metacognitive monitoring in the other domain, and (d) these questions in two different age groups in primary school to fully grasp potential transitional periods in the domain-specificity of metacognitive monitoring.

If, on the one hand, metacognition is highly domain-general, then metacognitive monitoring in the arithmetic and spelling tasks will be correlated and predictive of each other, even when controlled for academic performance – as arithmetic and spelling are highly related domains; and metacognitive monitoring in one domain will be associated with and predictive of academic performance in the other domain. If, on the other hand, metacognition is highly domain-specific, then the associations described above will be non-significant (frequentist statistics) and Bayes factors will be close to zero (Bayesian statistics; see below). These questions are investigated in two different age groups for which, based on the existing literature, different predictions can be made on the extent to which metacognitive monitoring is domain-general. By selecting participants in these two age groups, we aimed to capture an important period in the development of (the domain-generality of) metacognitive monitoring. In Study 1, we investigated these questions in 8-9-year-olds, for which domain-generality of metacognitive monitoring was predicted (third grade). Study 2 investigated these questions in younger children, namely 7-8-year-olds, for which more domain-specificity of metacognitive monitoring was predicted (second grade).

Study 1: Metacognitive monitoring in arithmetic and spelling in 8-9-year-olds (third grade)

Method

1 Participants

Participants were 147 typically developing Flemish 8-9 year-olds (third grade; 69 girls; M_{age} = 8 years, 10 months; SD = 3 months; [8 years 4 months - 9 years 4 months]), without a diagnosis of a developmental disorder, and who came from a dominantly middle-to-high socio-economic background. This study was approved by the social and societal ethics committee of KU Leuven. For every participant, written informed parental consent was obtained.

2 Procedure

All participants participated in four test sessions, which took place at their own school during regular school hours. They all completed the tasks in the same order. In the context of a larger project, all children first participated in an individual session of which the data are not included in the current manuscript. Second, a session in small groups of eight children took place, including the computerized spelling task and motor speed task. Third, a second session in small groups took place, including the computerized arithmetic task and motor speed task. Fourth, in a group session in the classroom, the standardized arithmetic and spelling tests and the test of intellectual ability were administered. Sessions were separated by one to three days on average; they were never adjacent. Below we describe the key variables and control variables used to answer our research questions. The full cognitive testing battery is posted on the Open Science Framework (OSF) page of this project (https://osf.io/ypue4/?view_only=ce9f97af0e3149c28942a43499eafd32).

3 Materials

Materials consisted of written standardized tests and computer tasks designed with Open Sesame (Mathôt et al., 2012). Arithmetic and spelling skills were assessed with both a custom computerized task and a standardized test (i.e. Arithmetic: Tempo Test Arithmetic, de Vos, 1992; Spelling: standardized dictation, Moelands & Rymenans, 2003). The computerized tasks for arithmetic and spelling were specifically designed to be as similar as possible, to minimize the possibility that the results on domain-specificity of metacognition were due to differences in paradigm. Both tasks were multiple choice tasks with specifically selected age-appropriate items (i.e. single digit addition and multiplication for arithmetic; three specific Dutch spelling rules for spelling). After a first introductory block, in the second block of each task, participants had to report their judgment on the accuracy of their academic answer after each trial, using the same metacognitive monitoring measure in both tasks.

3.1 Arithmetic

3.1.1 Custom computerized arithmetic task.

This single-digit task included addition and multiplication items, and comprised all combinations of the numbers 2 to 9 for each operation (n = 36). The task consisted of two blocks, i.e. one introductory block without (n = 12) and one with (n = 60) a metacognitive monitoring measure (see below). Stimuli were pseudo-randomly divided into the two blocks and children were given a short break between blocks. Each block was preceded by four practice trials to familiarize the child with the task requirements. Performance on the practice items was not included in the performance measures. In both blocks, addition items were presented first (n = 6 in the first block; n = 30 in the second block). After a short instruction slide indicating multiplication items would follow, the multiplication items were presented (n = 6 in the first block; n = 30 in the second block). The position of the numerically largest operand was balanced. Each item was presented with two possible solutions, one on the left and one on the right side of the screen. In half of the items, the correct solution was presented on the left side of the screen. Incorrect solutions for the addition items were created by adding or subtracting 1 or 2 to the solution (*n* = 7 for every category), or by using the answer to the corresponding multiplication item (e.g. 6 + 3 with incorrect solution 18; *n* = 8). The incorrect solutions for the multiplication items were table related, i.e. solution -1 or +1 one of the operands (e.g. 6 × 3 with incorrect solution 24; *n* = 7 for every category), or the answer to the corresponding addition (e.g. 8 × 2 with incorrect solution 10; *n* = 8). Each trial started with a 250 ms fixation point in the centre of the screen and after 750 ms the stimulus appeared in white on a black background. The stimuli remained visible until response. The children had to indicate which of the presented solutions for the problem was correct (by pressing the corresponding key). The response time and answer were registered via the computer. Performance measures were both accuracy and the response time for correct answers in the second block (*n* = 60).

3.1.2 Standardized arithmetic task.

Arithmetic fluency was assessed by the Tempo Test Arithmetic (TTA; de Vos, 1992); a standardized pen-and-paper test of arithmetical fluency, which comprises five columns of arithmetic items (one column per operation and a mixed column), each increasing in difficulty. Participants got one minute per column to provide as many correct answers as possible. The performance measure was the total number of correctly solved items within the given time (i.e. total score over the five columns).

3.2 Spelling

3.2.1 Custom computerized spelling task.

Spelling performance was measured with a computerized task consisting of two blocks, i.e. one introductory block without (*n* = 12) and one with (*n* = 60) a metacognitive monitoring measure (see below). Stimuli were pseudo-randomly divided into the two blocks and children were given a short break between blocks. Each block was preceded by six practice trials to familiarize the child with the task requirements. Performance on the practice items was not included in the performance measures. The items consisted of a Dutch word with a missing part, that was replaced by an underscore (e.g. 'ko_ie' for 'koffie'), presented with two possible solutions, one on the left and one on the right side of the screen. We used three specific Dutch spelling rules, which were the focus of spelling instruction at the participants' age. Firstly, the rule of open and closed syllables was used, on the basis of which one can figure out if one or two vowels or consonants have to be written. Secondly, the extension rule was used, on the basis of which one can figure out if words with a [t] sound at the end of the word are written with a 't' or a 'd'. To correctly spell these two types of words, children can either use these rules, or when they have extensive experience with these words, retrieve the correct spelling from their memory. Flemish third graders have the most experience with the extension rule, and are in the learning phase for the open and closed syllables rule. Stepwise, they go from learning the rule and using the procedure to spell the words, towards automatization of the correct spelling and thus retrieving it from memory. This spelling development is analogous to arithmetic development in third grade (i.e. from procedure use to retrieval). A third category of words was added for which no rule is available, but only retrieval from long-term memory is possible (i.e. au/ou-words; ei/ij-words). The diphthongs in these words have the same pronunciation, but are spelt differently (e.g. ‘r*ei*s’ vs. ‘w*ij*s’ have both the [ɛi] sound) – there is no rule to determine whether one or the other diphthong should be used and children have to learn this by heart. All items were selected from curriculum-based glossaries. Incorrect solutions were created by using the related distractor (*n* = 14 for each category), namely one or two vowels or consonants for the open and closed syllables rule (e.g. koffie: ‘ko_ie’ with options ‘f’ or ‘ff’), ‘t’ or ‘d’ for the extension rule (e.g. kast: ‘kas_’ with options ‘t’ or ‘d’), and the related diphthong for the to-be-retrieved words (e.g. konijn: ‘kon_n’ with options ‘ei’ or ‘ij’). In half of the items, the correct solution was presented on the left side of the screen. Each trial started with a 250 ms fixation point in the centre of the screen and after 750 ms children were presented on audiotape with the word. Then, the visual stimulus appeared in white on a black background. The stimuli remained visible until response. The children had to indicate which of the presented solutions for the problem was correct (by pressing the corresponding key; i.e. left/right key). The response time and answer were registered via the computer. Performance measures were both accuracy and the response time for correct answers in the second block (*n* = 60).

3.2.2 Standardized spelling task.

Spelling ability was also measured with a standardized dictation (Moelands & Rymenans, 2003). We administered the subtest for children at the end of third grade, which includes age-appropriate, curriculum-based items. The experimenter read aloud 43 sentences and the participants had to write one word down that was repeated two times after the sentence was read. The performance measure was the total number of correctly written words.

3.3 Metacognitive monitoring

5

In the second block of the arithmetic and the spelling tasks (*n* = 60 for each task), a metacognitive monitoring measure was added to the items. Children had to report their judgment on the accuracy of their answer to the academic item on a trial-by-trial basis (e.g. Bellon et al., 2019; Rinne & Mazzocco, 2014). More specifically, after giving their answer to the arithmetic/spelling problem, children had to indicate if they thought their answer was *Correct*, *Incorrect*, or if they *Did not know*. We used emoticons in combination with the options (e.g. ☺ and *Correct*) to make the task more attractive and feasible for children (see figure 5.1). Children had to respond by pressing the key corresponding to their metacognitive judgment (i.e. indicated with corresponding emoticon stickers). Metacognitive monitoring skills were operationalised as calibration of this judgment (i.e. the alignment between one’s judgment in the accuracy of their answer to a problem and the actual accuracy of the answer). Namely, a calibration score of 2 was obtained if their metacognitive judgment corresponded to their actual performance (i.e. metacognitively judged as *Correct* and indeed correct academic answer; metacognitively judged as *Incorrect* and indeed incorrect academic answer), a calibration score of 0 if their metacognitive judgement did not correspond to their actual performance (i.e. metacognitively judged as *Correct* and in fact incorrect academic answer; metacognitively judged as *Incorrect* and in fact correct academic answer), and a calibration score of 1 if children indicated they *Did not know* about their academic answer. The metacognitive monitoring score per child was the mean of all calibration scores (i.e. calibration score per arithmetic/spelling item; n = 60 per domain) and was calculated for each task separately. The higher the calibration scores, the better the metacognitive monitoring skills. To familiarize the children with the task, practice items were presented in each task.

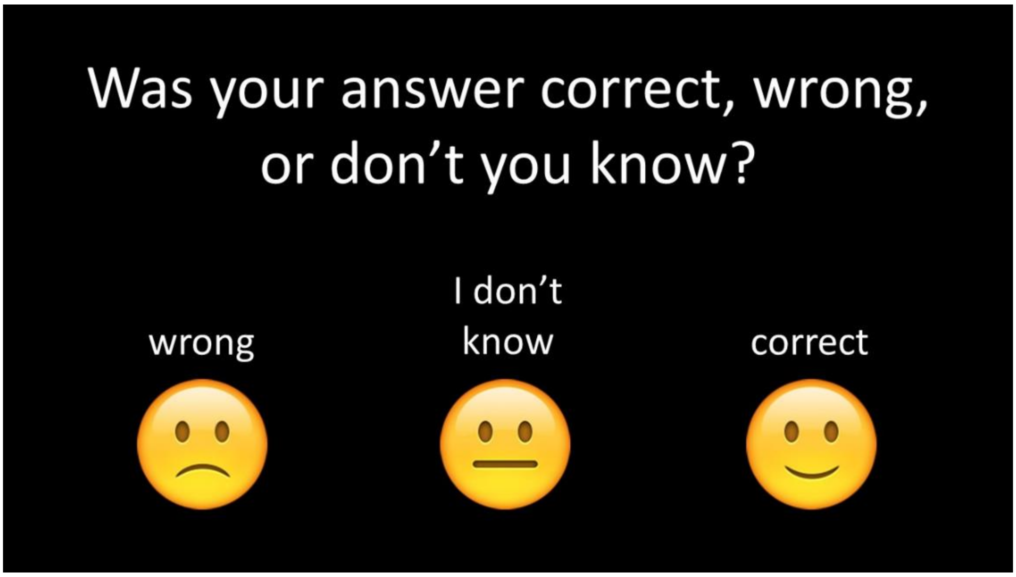

Figure 5.1. Example of metacognitive monitoring question after arithmetic/spelling item.

3.4 Control variables

3.4.1 Intellectual ability.

Intellectual ability was assessed through Raven's Standard Progressive Matrices (Raven et al., 1992). Children were given 60 multiple-choice items in which they had to complete a pattern. The performance measure was the number of correctly solved patterns.

3.4.2 Motor speed.

A motor speed task was included as a control for children's response speed on the keyboard (Bellon et al., 2019). Two shapes were simultaneously presented on either side of the screen and children had to indicate which of the two shapes was filled by pressing the corresponding key (i.e. left/right key). All shapes were similar in size and each shape occurred four times filled and four times non-filled, yielding 20 trials. The position of the filled shape was balanced. After fixation, stimuli appeared until response. Three practice trials were included to familiarize the children with the task. The performance measure was the average response time for correct responses.

4 Data analysis

A comprehensive analyses plan was preregistered on the OSF page of this project (https://osf.io/ypue4/?view_only=ce9f97af0e3149c28942a43499eafd32). The key analyses to answer our research questions are presented below; the results of the remaining preregistered analyses can be found in the supplementary materials.

We ran frequentist analyses using both uni- and multivariate techniques, as well as Bayesian analyses. Frequentist analyses allowed us to explore our data by means of a well-known method to gauge statistical support for the hypotheses of interest. Bayesian statistics allowed us to test the degree of support for a hypothesis (i.e. degree of strength of evidence in favour of or against any given hypothesis), expressed as the Bayes factor (BF_{10}; the ratio between the evidence in support of the alternative hypothesis over the null hypothesis). Although Bayes factors provide a continuous measure of degree of evidence, there are some conventional approximate guidelines for interpretation (Andraszewicz et al., 2015, for a classification scheme): $BF_{10} = 1$ provides no evidence either way, $BF_{10} > 1$ anecdotal, $BF_{10} > 3$ moderate, $BF_{10} > 10$ strong, $BF_{10} > 30$ very strong and $BF_{10} > 100$ decisive evidence for the alternative hypothesis; $BF_{10} < 1$ anecdotal, $BF_{10} < 0.33$ moderate, $BF_{10} < 0.10$ strong, $BF_{10} < 0.03$ very strong and $BF_{10} < 0.01$ decisive evidence for the null hypothesis. By adding these Bayesian analyses, we deepened our findings from the traditional analyses, as we were able to identify evidence in favour of the null hypothesis, consequently, identify which hypothesis is most plausible (i.e. alternative hypothesis vs. null hypothesis) and which predictors are the strongest. This is particularly relevant for the current study because we can compare the strength of evidence in favour of the domain-general hypothesis (i.e. association between metacognitive monitoring measures in different domains; association between performance and metacognitive monitoring across domains) versus the domain-specific hypothesis (i.e. no association between metacognitive monitoring measures in different domains; no association between performance and metacognitive monitoring across domains).

5

To answer our research questions, we used correlation and regression analyses. For the Bayesian analyses, we used a default prior with prior width set to 1 for Pearson correlations and to .354 for the linear regression analyses. For Bayesian regressions, a $BF_{inclusion}$ was calculated for every predictor in the model, which represents the change from prior to posterior odds (i.e. BF_{10}), where the odds concern all the models with a predictor of interest to all models without that predictor (i.e. a Bayes factor for including a predictor averaged across the models under consideration).

As planned in the preregistration, we excluded a child's performance on a task if this performance was more than three standard deviations from the mean of the task (i.e. ≤ 3% of the data per task). Due to unforeseen circumstances during data collection (e.g. school bell ringing), we additionally excluded some data at the item level (i.e. < 0.57% of items per task) that were considered to be measurement errors, i.e. when the data point was an outlier (i.e. more than three standard deviations from the mean) at both the item level (i.e. compared to the general mean of the item) and at the subject level (i.e. compared to the personal mean of the subject).

Results

The descriptive statistics of all measures are presented in Appendix A. Additionally, Pearson correlation coefficients of all variables under study were calculated (Appendix B). Although not originally pre-registered, we additionally re-calculated all analyses below with chronological age as an additional control variable. Considering chronological age within grade in the analyses reported below did not change the interpretation of the results (Appendix C).

1 The role of metacognitive monitoring in arithmetic and spelling performance

Pearson correlation coefficients of the associations between metacognitive monitoring and the academic skills are presented in Table 5.1.

Table 5.1
Correlation analyses of metacognitive monitoring and academic performance measures in 8-9-year-olds (Grade 3)

	Arithmetic			Spelling		
	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (TTA) [a]	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (dictation) [a]
Metacognitive monitoring						
Arithmetic						
r	.84	-.08	.38	.45	.11	.26
p	<.001	.38	<.001	<.001	.20	.003
BF_{10}	>100	0.16	>100	>100	0.24	9.65
Spelling						
r	.48	-.19	.33	.91	-.02	.66
p	<.001	0.03	<.001	<.001	.79	<.001
BF_{10}	>100	1.18	>100	>100	0.11	>100

Note. [a] Controlled for intellectual ability; [b] Controlled for intellectual ability and motor speed on the keyboard.

5

Metacognitive monitoring in the arithmetic task (MM_{arith}) was significantly correlated with arithmetic accuracy ($Arithmetic_{acc}$) and the tempo test arithmetic (TTA), with Bayes factors indicating decisive evidence in favour of the associations, even when controlling for intellectual ability. There was no significant correlation with response time for correct arithmetic answers ($Arithmetic_{rt}$) and the Bayes factor indicated moderate evidence in favour of no association.

Metacognitive monitoring in the spelling task (MM_{spell}) was significantly correlated with spelling accuracy ($Spelling_{acc}$) and dictation, with Bayes factors indicating decisive evidence in favour of the associations, even when controlling for intellectual ability. There was no significant correlation with response time for correct spelling answers ($Spelling_{rt}$) and the Bayes factor indicated moderate evidence in favour of no association.

Based on the absence of significant (frequentist statistics) and supported (Bayesian statistics) associations with our response time performance measures ($Arithmetic_{rt}$ and $Spelling_{rt}$), and because these measures only take into account data for correct answers, losing important information on performance and possibly overestimating performance, the response time performance measures will not be considered in further analyses.

2 Domain-specificity of the role of metacognitive monitoring

To examine domain-specificity of the role of metacognition, we first investigated the association between MM_{arith} and MM_{spell} with correlation and regression analyses. Specifically, we investigated whether MM_{arith} and MM_{spell} were correlated, even when controlling for intellectual ability and academic performance in both domains. Controlling for intellectual ability and performance in both standardized academic tasks was necessary, to make sure the observed associations between MM_{arith} and MM_{spell} were not (entirely) driven by their shared reliance on intellectual ability or by the high correlation between both academic domains.

Secondly, we studied the role of MM_{spell} in arithmetic performance and MM_{arith} in spelling performance with correlation and regression analyses. In other words, cross-domain correlations between academic performance in one domain and metacognitive monitoring in the other domain were calculated. As performance in the arithmetic and spelling tasks was highly correlated, the cross-domain associations of metacognitive monitoring and academic performance might rely on the correlation between the academic tasks. Therefore, we used regression models to investigate whether metacognitive monitoring in arithmetic uniquely predicted spelling performance on top of arithmetic performance, and vice versa.

5

In a final step, we investigated the unique contribution of cross-domain metacognitive monitoring to performance over within-domain metacognitive monitoring using regression models including metacognitive monitoring in both domains as predictors for academic performance.

2.1 Associations between metacognitive monitoring in different domains

MM_{arith} and MM_{spell} were significantly correlated, even when controlling for intellectual ability, and arithmetic and spelling performance on the standardized tasks (r = .42; p < .001; BF_{10} > 100). Regression analyses confirmed that metacognitive monitoring in one domain was uniquely predicted by metacognitive monitoring in the other domain, even when simultaneously considered with intellectual ability and performance on the standardized tasks in both academic domains (see Table 5.2). Additional post-hoc analyses that were not preregistered indicated that the results were the same when including academic achievement as measured with accuracy in the computerized academic tasks instead of academic achievement as measured with the standardized academic tasks.

Table 5.2

Regression analyses of MM_{arith} and MM_{spell} performance with metacognitive monitoring in the other domain and standardized task performance in both domains as predictors

	MM_{arith}			
	β	t	p	$BF_{inclusion}$
Intellectual ability	.16	2.12	.04	2.90
TTA	.26	3.62	<.001	72.57
Dictation	-.14	-1.50	.14	1.46
MM_{spell}	.51	5.26	<.001	>100
	MM_{spell}			
	β	t	p	$BF_{inclusion}$
Intellectual ability	.07	1.07	.29	0.38
Dictation	.55	8.77	<.001	>100
TTA	.01	0.13	.90	0.25
MM_{arith}	.34	5.26	<.001	>100

Note. MM_{arith} = metacognitive monitoring in the arithmetic task; MM_{spell} = metacognitive monitoring in the spelling task; TTA = Tempo Test Arithmetic.

2.2 Cross-domain performance associations of metacognitive monitoring

Table 5.1 shows cross-domain correlations between academic performance and metacognitive monitoring in the other domain. MM_{arith} was significantly correlated with both spelling performance measures (i.e. $Spelling_{acc}$ and dictation), with a Bayes factor indicating moderate to decisive evidence. MM_{spell} was significantly correlated with both arithmetic performance measures (i.e. $Arithmetic_{acc}$ and TTA), with a Bayes factor indicating decisive evidence.

We further investigated whether metacognitive monitoring in arithmetic uniquely predicted spelling performance on top of arithmetic performance; and vice versa. Namely, we predicted arithmetic performance based on MM_{spell} and dictation, and spelling performance based on MM_{arith} and TTA (Table 5.3). These regression analyses showed that, even when performance in the academic domain was taken into account, metacognitive monitoring in that domain remained a significant and supported predictor of academic performance in the other domain (all ps < .05; all BFs_{10} >5).

When metacognitive monitoring scores in both domains were considered simultaneously to predict academic performance (using regression analyses), only the role of metacognitive monitoring within the domain itself remained significant (frequentist statistics) and supported (Bayesian statistics). Namely, when MM_{arith} and MM_{spell} were used to predict arithmetic performance, only MM_{arith} was a significant and supported predictor ($Arithmetic_{acc}$: p < .001; $BF_{inclusion}$ > 100; TTA: p = .001; $BF_{inclusion}$ > 100), not MM_{spell} ($Arithmetic_{acc}$: p = .41; $BF_{inclusion}$ = 0.18; TTA: p = .10; $BF_{inclusion}$ = 1.36). On the other hand, when MM_{arith} and MM_{spell} were used to predict spelling performance, only MM_{spell} was a significant and supported predictor ($Spelling_{acc}$: p <.001; $BF_{inclusion}$ > 100; Dictation: p < .001; $BF_{inclusion}$ > 100), not MM_{arith} ($Spelling_{acc}$: p = .38; $BF_{inclusion}$ = .06; Dictation: p = .61; $BF_{inclusion}$ = .24).

Table 5.3

Regression analyses of arithmetic performance (i.e. arithmetic$_{acc}$ and TTA) and spelling performance (i.e. spelling$_{acc}$ and dictation) with metacognitive monitoring in the other domain and standardized task performance in the other domain as predictors

	Arithmetic							
	Arithmetic$_{acc}$				TTA			
	β	*t*	*p*	BF$_{inclusion}$	*β*	*t*	*p*	BF$_{inclusion}$
MM$_{spell}$	.54	5.18	<.001	5.03	.24	2.11	.04	>100
Dictation	-.06	-.54	.59	2.07	.19	1.73	.09	0.37
	Spelling							
	Spelling$_{acc}$				Dictation			
	β	*t*	*p*	BF$_{inclusion}$	*β*	*t*	*p*	BF$_{inclusion}$
MM$_{arith}$	.47	5.89	<.001	>100	.23	2.66	.009	10.84
TTA	.12	1.46	.15	0.86	.25	2.95	.004	23.59

Note. MM$_{arith}$ = metacognitive monitoring in the arithmetic task; MM$_{spell}$ = metacognitive monitoring in the spelling task; TTA = Tempo Test Arithmetic.

Interim discussion

The results of Study 1 reveal that within-domain metacognitive monitoring was an important predictor of both arithmetic and spelling performance. Monitoring measures in both domains were highly correlated and predictive of one another, even after controlling for intellectual ability and performance on both academic tasks. Both monitoring measures correlated with performance in the other academic domain, ever after controlling for performance within the domain (e.g. significant correlation of MM$_{arith}$ with spelling performance, controlled for arithmetic performance). When monitoring within the domain was added above monitoring across-domain, only monitoring within the domain remained a significant predictor of academic performance. Taken together, these results provide substantial evidence for domain-generality of metacognitive monitoring in academic domains in 8-9-year-olds, in addition to the importance of some degree of domain-specificity in monitoring skills.

These results leave the question of whether this domain-generality is the result of a shift (e.g. Geurten et al., 2018) in early primary school unanswered. One possibility is that the 8-9-year-olds already went through an important transition regarding domain-generality of metacognitive monitoring, but that such domain-generality is not observed at younger ages. On the other hand, it is possible that no shift to domain-generality has occurred because also at a younger age, domain-generality can be observed. To test this, we additionally recruited a new sample of children that were one year younger, i.e. 7-8-year-olds (Study 2). The same research questions as in Study 1 were studied using the exact same paradigm. This allowed us to test whether domain-generality is already observed at younger ages or not.

Study 2: Metacognitive monitoring in arithmetic and spelling in 7-8-year-olds (second grade)

Methods

1 Participants

Participants were 77 typically developing Flemish 7-8 year-olds (second grade; 49 girls; M_{age} = 7 years, 8 months; SD = 4 months; [7 years 1 month - 8 years 1 month]), without a diagnosis of a developmental disorder, and who came from a dominantly middle-to-high socio-economic background. For every participant, written informed parental consent was obtained.

2 Procedure

The procedure was the same as in Study 1.

3 Materials

Materials were the same as in Study 1. The items in the custom arithmetic and spelling tasks were adapted from Study 1 to be age appropriate for second graders. Namely, for arithmetic, only single-digit addition was administered (*n* = 30); for spelling only two specific Dutch spelling rules were used (i.e. extension rule and to be retrieved words with diphthongs; *n* = 30). The standardized arithmetic task was exactly the same as in Study 1. As for the standardized dictation, we administered the subtest for children in the middle of second grade, which includes age-appropriate, curriculum-based items (Moelands & Rymenans, 2003) (*n* = 42).

4 Data analysis

For this follow-up study, we carried out the same analyses as preregistered for Study 1 (https://osf.io/ypue4/?view_only=ce9f97af0e3149c28942a43499eafd32). The same exclusion criteria for data as in Study 1 were applied. Less than 4 % of the data per task was excluded as an outlier; less than 0.90% of the items per task were excluded as a measurement error.

Results

The descriptive statistics of all measures are presented in Appendix A. Additionally, Pearson correlation coefficients of all variables under study were calculated (Appendix B). Although not originally pre-registered, we additionally re-calculated all analyses below with chronological age as an additional control variable. Considering chronological age within grade in the analyses reported below did not change the interpretation of the results (Appendix C).

1 The role of metacognitive monitoring in arithmetic and spelling performance

Pearson correlation coefficients of the associations between metacognitive monitoring and academic performance are presented in Table 5.4.

Table 5.4
Correlation analyses of metacognitive monitoring and academic performance measures in 7-8-year-olds (Grade 2)

	Arithmetic			Spelling		
	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (TTA) [a]	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (dictation) [a]
Metacognitive monitoring						
Arithmetic						
r	.74	.30	.47	.11	.06	.20
p	<.001	.02	<.001	.38	.66	.11
BF_{10}	>100	2.60	>100	0.23	0.17	0.53
Spelling						
r	.03	.11	.05	.89	.03	.32
p	.84	.40	.69	<.001	.82	.01
BF_{10}	0.16	0.11	0.17	>100	0.16	4.12

Note. [a] Controlled for intellectual ability; [b] Controlled for intellectual ability and motor speed on the keyboard.

MM_{arith} was significantly correlated with all three arithmetic performance measures. Bayes factors indicate that the evidence for an association with $Arithmetic_{acc}$ and the TTA is decisive, while there is only anecdotal evidence for an association with $Arithmetic_{rt}$.

MM_{spell} was significantly correlated with both $Spelling_{acc}$ and dictation, with Bayes factors indicating moderate to decisive evidence for an association. There was no significant correlation with $Spelling_{rt}$ and the Bayes factor indicated moderate evidence in favour of no association.

Based on the same rationale as Study 1, the response time performance measures were not considered in further analyses.

2 Domain-specificity of the role of metacognitive monitoring

MM_{arith} and MM_{spell} were not significantly correlated after controlling for intellectual ability (r = .14, p = .28). The Bayes factor indicated there was moderate evidence in favour for no association (BF_{10} = 0.28). Hence, further control analyses (i.e. in line with Study 1 in which the correlation between MM_{arith} and MM_{spell} was also controlled for performance on the TTA and dictation) were not performed.

Table 5.4 shows cross-domain correlations between academic performance and metacognitive monitoring in the other domain. MM_{arith} was not significantly correlated with any of the spelling performance measures. Bayes factors indicated moderate evidence in favour of no association. MM_{spell} was not significantly related to any of the arithmetic measures. Bayes factors indicated anecdotal to moderate evidence in favour of no association.

Interim discussion

The results of Study 2 revealed that within-domain metacognitive monitoring was an important predictor of both arithmetic and spelling performance. Monitoring measures in both domains were not correlated, and both monitoring measures did not correlate with performance in the other academic domain. These results provide substantial evidence for domain-specificity of metacognitive monitoring in academic domains in 7-8-year-olds (second graders). No domain-general effect of metacognitive monitoring was observed, in contrast to the 8-9-year-olds (third grade children; Study 1).

General discussion

Two recent studies found evidence for within-domain metacognitive monitoring as an important predictor of arithmetic (Bellon et al., 2019; Rinne & Mazzocco, 2014). One outstanding question is whether these results regarding the role of metacognitive monitoring in arithmetic are specific to the arithmetic domain, or whether they are reflective of a more general role of metacognitive monitoring in academic performance. This study adds to the existing literature in an important way by (a) investigating metacognitive monitoring in two related, yet distinct academic domains, (b) studying whether monitoring in one domain was associated with and predictive of monitoring in the other domain (and vice versa), and (c) studying whether monitoring in one domain was associated with and predictive of performance in the other domain (and vice versa), and importantly by (d) doing this in two important age groups, namely children aged 8-9 (Study 1) and 7-8 (Study 2) who are in an important developmental phase for both academic performance and metacognition, and using the exact same paradigm in both age groups and both domains.

Our results reveal that within-domain metacognitive monitoring was an important predictor of both arithmetic and spelling performance in both 8-9-year-olds (Study 1) and 7-8-year-olds (Study 2). Although metacognitive monitoring in arithmetic and spelling were highly correlated and predictive of one another in 8-9-year-olds (Study 1), they were not in younger 7-8-year-old children (Study 2). In 8-9-year-olds, but not in 7-8-year-olds, both monitoring measures correlated with performance in the other academic domain, even after controlling for performance within the domain (e.g. significant correlation of MM_{arith} with spelling performance, controlled for arithmetic performance). These results provide evidence for the emergence of domain-generality of metacognitive monitoring between second and third grade (i.e. 7-9-year-olds).

Our results nicely replicate associations between metacognitive monitoring and academic performance (e.g. Bellon et al., 2019; Freeman et al., 2017; Rinne & Mazzocco, 2014; Roebers et al., 2014; Roebers & Spiess, 2017). Combining the data of both studies, we are able to confirm the theoretically assumed development of metacognition from highly domain- and situation-specific to more flexible and domain-general with practice and experience (Borkowski et al., 2000). Our results regarding a possible underlying domain-general element of metacognitive monitoring in middle primary school children (8-9-year-olds) are in line with the existing literature in older ages and/or other domains (e.g. Geurten et al., 2018; Schraw et al., 1995; Veenman et al., 1997). For example, Schraw and colleagues (Schraw et al., 1995; Schraw & Nietfeld, 1998) and Veenman and colleagues (1997) found evidence for domain-generality of metacognitive monitoring in adults; Geurten et al. (2018) observed a shift to domain-general metacognition between 8 and 13 across the arithmetic and memory domain. Our results also show the importance of domain-specific knowledge for metacognitive performance, as was previously found in non-academic domains (i.e. soccer) by for example Löffler and colleagues (2016), in very young children by Vo and colleagues (2014), and in 12-year-olds in mathematics by Lingel and colleagues (2019). Our results add to this body of research that domain-generality of metacognitive monitoring emerges between the ages of 7-to-9, yet that domain-specific knowledge and skills remain important for metacognitive monitoring, even in highly related academic domains.

Schraw and colleagues (1995) note that when performance is correlated among domains (i.e. as they were in Study 1), correlated metacognitive monitoring scores (i.e. as they were in Study 1) pose no serious threat to the assumption that monitoring is domain-specific. However, when they are correlated after removing the variation attributable to performance scores, as we did using partial correlations and regression analyses, this outcome cannot be explained on the basis of domain-specific knowledge and a domain-general argument needs to be invoked. As both monitoring performances remained significantly correlated after removing the variation attributable to performance scores, our results indicate that in 8-9-year-olds (Study 1) there might be an underlying domain-general element of metacognition within both metacognitive monitoring scores. This was not observed in 7-8-year-olds (Study 2). All in all, these results point to the emergence of domain-generality of metacognitive monitoring in between second (7-8-year-olds) and third (8-9-year-olds) grade of primary school.

Our results still provide some evidence for a domain-specific element of metacognitive monitoring in 8-9-year-olds. Although metacognitive monitoring across-domain was an important predictor of performance, the associations with monitoring within-domain were significantly larger than with monitoring across-domain. Once monitoring within a domain was considered, the predictive power of monitoring across-domain was no longer significant/supported. These results suggest the continuing importance of domain-specific knowledge and skills. This domain-specific element could explain the additional predictive power of monitoring within-domain in addition to metacognitive monitoring across-domain.

Based on the important role that metacognitive monitoring was found to have in arithmetic performance (Bellon et al., 2019; Rinne & Mazzocco, 2014), the current study investigated the domain-specificity question of metacognition by also including spelling performance. We deliberately included a different, yet correlated skill within the academic domain to thoroughly investigate the extent to which metacognition might be domain-specific. This is different from existing research, where the domain-specificity question was investigated in very distant domains. For example, Vo and colleagues (2014) investigated domain-specificity in the numerical domain versus emotion discrimination. The use of spelling alongside arithmetic made it possible to use the exact same paradigm to measure metacognitive monitoring and maximize the comparability of the two tasks. The fact that the computerized tasks for arithmetic and spelling were specifically designed to be as similar as possible, minimized the possibility that the results on domain-specificity of metacognition were due to differences in paradigms. By including standardized arithmetic and spelling tasks, which are not as similar to each other and measure performance in an ecologically valid way, we minimized the possibility that the results on domain-generality of metacognition were due to similarities in paradigms. While there is substantial evidence in the current studies for the emergence of domain-general metacognitive monitoring processes, the results also indicate that, even in highly related domains, domain-specific knowledge and skills are important for metacognitive monitoring in primary school children.

Although the custom arithmetic and spelling tasks were designed with age-appropriate items, a slight difference in task difficulty was present, with the computerized spelling tasks being more difficult than the arithmetic tasks. Schraw and colleagues (1995) pointed out that task difficulty, as a characteristic of the test environment, might have an important influence on metacognitive monitoring. They found that, with different task difficulty levels, metacognitive monitoring in adults was mostly domain-specific, yet, once tests were matched on test characteristics, monitoring was mostly domain-general. To make sure our results were not influenced by this slight difference in task difficulty, we selected, post-hoc, a subset of items per task (n = 40 for Study 1; n = 20 for Study 2) that were matched on task difficulty (i.e. t-test comparing accuracy in arithmetic and spelling selection: Study 1: $t(138) = 0.12$, $p = .91$; Study 2: $t(71) = 0.36$, $p = .72$). These post-hoc exploratory results showed that our findings on metacognitive monitoring and its specificity did not change when restricting the analyses to those items that were matched in task difficulty.

Performance measures of arithmetic and spelling were accuracy in the computerized tasks, and widely-used, standardized pen-and-paper tasks. As accuracy data were a fundamental part of our metacognitive monitoring scoring, in the interpretation of the results, the largest focus should be on the standardized measures, as metacognitive monitoring was measured independently from these measures. The computerized and the standardized tasks were both age-appropriate measures, yet the standardized tasks focused less on specific items of the curriculum (i.e. only single-digit arithmetic in the computerized arithmetic task; only specific Dutch spelling rules in the computerized spelling task), for which reason they were more wide-ranged and valid measures of children’s arithmetic and spelling skills. The standardized tasks were the most ecologically valid measures, assessing arithmetic and spelling performance as they are assessed in the classroom. Including these standardized tasks in the design is an essential asset of this study compared to the existing literature (e.g. Bellon et al., 2019; Rinne & Mazzocco, 2014), as we were able to generalize our results from the role of metacognitive monitoring within the task, to within the domain, independently from the task in which monitoring was measured.

Although the driving mechanisms for the gradual development from domain-specificity to domain-generality of metacognitive monitoring cannot be determined on the basis of the current study, it is important to reflect on why metacognition shifts to being more domain-general around the ages 8-9. The existing literature offers some theoretical possibilities, albeit speculatively, that should be investigated in future research.

The development from more domain-specificity of metacognitive monitoring towards more domain-generality in this age group is likely reflective of a gradual transition that occurs in the development of primary school children (e.g. Schneider, 2010). In early stages of this development, children's metacognitive monitoring might still be highly dependent on the (characteristics of the) specific stimuli, while over development, through experiences of failure and success, and with practice in assessing one's performance as well as in (academic) tasks, monitoring might become more generic. These hypotheses and our results can also be interpreted within the dual-process framework of metacognition (e.g. Koriat, 2007; Koriat & Ackerman, 2010; Koriat & Levy-sadot, 1999), which Geurten et al. (2018) used to interpret their findings. According to this dual-process framework of metacognition (Koriat, 2007; Koriat & Ackerman, 2010; Koriat & Levy-sadot, 1999), metacognitive judgments can, on the one hand, be experience-based, i.e. based on fast and automatic inferences made from a variety of cues that reside from immediate feedback from the task and that are then heuristically used to guide decisions. As such, these metacognitive judgments are task-dependent and probably difficult to generalize across domains. On the other hand, metacognitive judgments can be information-based, i.e. based on conscious and deliberate inferences, in which various pieces of information retrieved from memory are consulted and weighted in order to reach an advised judgment. These conscious and effortful judgments are more likely to generalize to other domains. Taken together with the current results, this dual-processing model of metacognition may suggest that 7-8 year-old (second grade) children preferentially rely on automatic inferences when making judgments, while improvements of metacognitive abilities may enable 8-9 year-old children (third grade) to rely more often on conscious and deliberate information-based processes.

Another explanation for the gradual shift from domain-specificity to domain-generality of metacognition could be that this development might be associated with the development in other general cognitive functions, such as working memory capacity or intellectual ability. For example, Veenman and colleagues (2005) found that metacognitive skills develop alongside, but not entirely as part of intellectual ability. Growth in these other general cognitive functions might enable a shift from domain-specificity to domain-generality of metacognition.

Finally, the development from domain-specificity towards domain-generality might also be driven by education, as teachers instruct children on assessing their own performance, which is at first very focused on specific tasks. Over development, children might internalise this into a semantic network of their own abilities, which in turn might generalise to other tasks and thus become more general.

It is essential to note that none of the above-mentioned hypotheses can be empirically evaluated within the current study. The focus of the current study was on *whether* a development toward domain-generality in metacognitive monitoring occurs in primary school children, in related academic domains, and, secondly *when* this occurs. The question on *how*, i.e. what mechanisms lie behind this, and *why* this is the case at this age, are important questions for future research.

Future research should also examine the question of domain-specificity of metacognition longitudinally, investigating the potential shift from domain-specificity to domain-generality in the same group of primary school children. Such a research design will allow one to investigate the directions of the associations between metacognition and academic performance and how these associations evolve over time. Furthermore, brain-imaging research in children could be very useful to investigate the question of domain-specificity of metacognition, by, for example, testing whether metacognitive abilities for different types of tasks (partially) depend on common neurobiological structures such as the prefrontal cortex, as has been observed in adults (e.g. Fleming & Dolan, 2014).

Conclusion

To conclude, the results of this study show that metacognitive monitoring of performance is an important predictor of academic skills in primary school children. While in young primary school children (7-8-year-olds), this process is domain-specific, in slightly older children (8-9-year-olds), this is a predominantly domain-general process, in which metacognitive monitoring of performance is an important predictor of academic skills independently of the academic task and domain it is measured in, even in highly related domains. Besides depending on domain-general metacognitive processes, metacognitive monitoring remains to be dependent of domain-specific performance and knowledge. Knowing whether metacognition is rather domain-specific or domain-general, and when domain-generality emerges, is of importance for educators, as this might impact on how they provide instructions in metacognitive monitoring, namely for each task or domain separately (i.e. domain-specific metacognition) or concurrently in different tasks and domains (expecting it to transfer to new domains; domain-general metacognition).

Appendixes

1 Appendix A – Descriptive statistics

Table 5.A1

Descriptive statistics of the key variables in 8-9-year olds (Grade 3; Study 1)

	n	*M*	*SD*	Range
Arithmetic				
Custom task				
Accuracy	141	.94	.06	[.75-1.00]
Response time (ms)	141	3833	1479	[1537-8543]
Standardized task				
Total score	144	77.70	19.20	[36.00-133.00]
Spelling				
Custom task				
Accuracy	145	.78	.12	[.45-.98]
Response time (ms)	145	2434	729	[1188-4586]
Standardized task				
Total score	144	30.00	8.28	[9.00-43.00]
Metacognitive monitoring				
In arithmetic task	141	1.88	0.12	[1.53-2.00]
In spelling task	145	1.53	0.25	[0.88-1.97]
Control				
Intellectual ability				
Raven	143	36.10	7.43	[13.00-52.00]
Motor Speed				
Accuracy	143	.98	.03	[.90-1.00]
Response time (ms)	143	593	135	[390-1044]

Table 5.A2

Descriptive statistics of the key variables in 7-8-year-olds (Grade 2; Study 2)

	n	*M*	*SD*	Range
Arithmetic				
Custom task				
Accuracy	73	.89	.12	[.50-1.00]
Response time (ms)	73	4384	1541	[1606-8273]
Standardized task				
Total score	68	27.24	6.74	[14-43]
Spelling				
Custom task				
Accuracy	76	.70	.13	[.43-1.00]
Response time (ms)	76	2994	1063	[1052-7368]
Standardized task				
Total score	68	32.65	5.66	[17-42]
Metacognitive monitoring				
In arithmetic task	73	1.74	0.25	[0.97-2.00]
In spelling task	76	1.38	0.26	[0.97-1.93]
Control				
Intellectual ability				
Raven	68	29.16	8.39	[11-48]
Motor Speed				
Accuracy	74	.99	.03	[.95-1.00]
Response time (ms)	74	634	138	[413-1073]

2 Appendix B – All intercorrelations

Table 5.B1

Correlation analyses between the administered measures in 8-9-year-olds (Grade 3; Study 1)

				1a	1b	1c	2a	2b	2c	3a	3b	4
Arithmetic	Custom task	1a. Arithmetic ACC		-								
		1b. Arithmetic RT										
			r	.12								
			p	.14	-							
			BF_{10}	0.31	-							
	Standardized task	1c. TTA										
			r	.26	-.71	-						
			p	.001	<.001	-						
			BF_{10}	16.85	>100	-						
Spelling	Custom task	2a. Spelling ACC										
			r	.44	-.18	.33	-					
			p	<.001	.03	<.001	-					
			BF_{10}	>100	1.05	>100	-					
		2b. Spelling RT										
			r	.22	.52	-.33	-.04	-				
			p	.005	<.001	<.001	.60	-				
			BF_{10}	3.49	>100	>100	0.12	-				
	Standardized task	2c. Dictation										
			r	.33	-.29	.39	.75	-.24	-			
			p	<.001	<.001	<.001	<.001	.004	-			
			BF_{10}	>100	42.30	>100	>100	6.45	-			
Metacognitive monitoring		3a. MM_{arith}										
			r	.89	.00	.40	.47	.15	.37	-		
			p	<.001	.99	<.001	<.001	.08	<.001	-		
			BF_{10}	>100	0.10	>100	>100	0.46	>100	-		
		3b. MM_{spell}										
			r	.52	-.18	.37	.92	-.05	.73	.58	-	
			p	<.001	.03	<.001	<.001	.56	<.001	<.001	-	
			BF_{10}	>100	1.14	>100	>100	0.12	>100	>100	-	
Control variables	Intellectual ability	4. Raven										
			r	.19	-.02	.16	.34	-.01	.28	.25	.31	-
			p	.03	.81	.05	<.001	.91	<.001	.003	<.001	-
			BF_{10}	1.20	0.11	0.66	>100	0.11	29.26	8.22	>100	-
	Motor speed	5. Motor speed task RT										
			r	.20	.47	-.32	.00	.33	-.08	.13	.01	-.08
			p	.02	<.001	<.001	.99	<.001	.33	.13	.94	.35
			BF_{10}	1.59	>100	>100	0.10	>100	0.17	0.33	0.10	0.16

Note. ACC = accuracy; RT = response time for the correct answers; TTA = Tempo Test Arithmetic; MM_{arith} = metacognitive monitoring in the arithmetic task; MM_{spell} = metacognitive monitoring in the spelling task.

5

Table 5.B2

Correlation analyses between the administered measures in 7-8-year-olds (Grade 2; Study 2)

				1a	1b	1c	2a	2b	2c	3a	3b	4
Arithmetic	Custom task	1a. Arithmetic ACC		-								
		1b. Arithmetic RT										
			r	.46	-							
			p	<.001	-							
			BF_{10}	>100	-							
	Standardized task	1c. TTA										
			r	.45	-.05	-						
			p	<.001	.70	-						
			BF_{10}	>100	0.17	-						
Spelling	Custom task	2a. Spelling ACC										
			r	.15	.05	.13	-					
			p	.22	.70	.28	-					
			BF_{10}	0.31	0.16	0.27	-					
		2b. Spelling RT										
			r	.01	.33	-.20	-.06	-				
			p	.92	.004	.11	.62	-				
			BF_{10}	0.15	7.75	0.55	.16	-				
	Standardized task	2c. Dictation										
			r	.28	.20	.24	.28	-.05	-			
			p	.02	.11	.05	.02	.69	-			
			BF_{10}	1.91	0.55	0.96	2.16	.17	-			
Metacognitive monitoring		3a. MM_{arith}										
			r	.80	.37	.47	.19	.09	.28	-		
			p	<.001	.001	<.001	.11	.46	.02	-		
			BF_{10}	>100	20.90	>100	0.53	0.19	1.87	-		
		3b. MM_{spell}										
			r	.18	.19	.08	.90	.02	.36	.26	-	
			p	.12	.11	.51	<.001	.84	.003	.03	-	
			BF_{10}	0.48	0.52	0.19	>100	0.15	13.07	1.64	-	
Control variables	Intellectual ability	4. Raven										
			r	.46	.27	.12	.21	-.05	.24	.44	.28	-
			p	<.001	.03	.33	.10	.67	.05	<.001	.02	-
			BF_{10}	>100	1.49	0.24	0.60	0.17	1.06	131.5	2.06	-
	Motor speed	5. Motor speed task RT										
			r	.02	.25	-.43	-.17	.35	-.13	.05	-.13	-.31
			p	.87	.04	<.001	.14	.002	.29	.69	.26	.01
			BF_{10}	0.15	1.16	73.64	0.42	15.17	0.27	0.16	0.27	3.32

Note. ACC = accuracy; RT = response time for the correct answers; TTA = Tempo Test Arithmetic; MM_{arith} = metacognitive monitoring in the arithmetic task; MM_{spell} = metacognitive monitoring in the spelling task.

3 Appendix C – Analyses with chronological age

Although not originally pre-registered, we additionally calculated the analyses presented in the manuscript with chronological age as an additional control variable. Pearson correlation coefficients were calculated between age and academic and metacognitive performance measures in both grades (see table 5.C1 below). The associations between age and the other metrics were not statistically significant, and Bayes factors were all below 0.43, consequently pointing to evidence for the null hypotheses of no association between age and the variables under investigation. In line with the lack of significant correlations with age, post-hoc defined partial correlations and regression models to control for shared variance across age and the other metrics (see Tables C2-C6) indicate that including chronological age in the analyses does not change the interpretation of the current results.

Table 5.C1

Correlation analyses of chronological age and academic and metacognitive performance measures in both grades

		Study 1 - Grade 3	Study 2 - Grade 2
		Age	Age
Arithmetic performance			
Custom task			
Accuracy			
	r	-.07	.15
	p	.44	.22
	BF_{10}	0.14	0.31
Response time			
	r	-.09	-.09
	p	.28	.46
	BF_{10}	0.19	0.19
Standardized task			
	r	.06	.18
	p	.46	.15
	BF_{10}	0.13	0.42
Spelling performance			
Custom task			
Accuracy		.07	-.14
	r	.43	.26
	p	0.15	0.28
	BF_{10}		
Response time			
	r	-.13	-.13
	p	.12	.29
	BF_{10}	0.36	0.25
Standardized task			
	r	.06	.01
	p	.46	.94
	BF_{10}	0.14	0.15
Metacognitive Monitoring			
Arithmetic			
	r	-.14	.07
	p	.11	.59
	BF_{10}	0.37	0.17
Spelling			
	r	-.01	-.13
	p	.87	.28
	BF_{10}	0.11	0.26

Table 5.C2

Partial correlations of metacognitive monitoring and academic performance measures in 8-9-year-olds (Grade 3)

	Arithmetic			Spelling		
	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (TTA) [a]	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (dictation) [a]
Metacognitive monitoring						
Arithmetic						
r	.86	-.05	.43	.53	.11	.35
p	<.001	.53	<.001	<.001	.22	<.001
BF_{10}	>100	0.13	>100	>100	0.23	>100
Spelling						
r	.53	-.15	.38	.93	-.04	.71
p	<.001	.09	<.001	<.001	.68	<.001
BF_{10}	>100	0.45	>100	>100	0.12	>100

Note. All correlations are additionally controlled for chronological age. [a] Controlled for intellectual ability; [b] Controlled for intellectual ability and motor speed on the keyboard.

Table 5.C3

Partial correlations of metacognitive monitoring and academic performance measures in 7-8-year-olds (Grade 2)

	Arithmetic			Spelling		
	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (TTA) [a]	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (dictation) [a]
Metacognitive monitoring						
Arithmetic						
r	.80	.37	.46	.16	.08	.17
p	<.001	.001	<.001	.23	.52	.18
BF_{10}	>100	20.31	>100	0.32	0.20	0.38
Spelling						
r	.06	.11	.11	.89	-.01	.36
p	.66	.42	.39	<.001	.92	.003
BF_{10}	0.17	0.22	0.22	>100	0.16	11.93

Note. All correlations are additionally controlled for chronological age. [a] Controlled for intellectual ability; [b] Controlled for intellectual ability and motor speed on the keyboard.

5

Table 5.C4

Partial correlations of metacognitive monitoring measures

	Study 1 – Grade 3	Study 2 – Grade 2
	Metacognitive monitoring Spelling	Metacognitive monitoring Spelling
Metacognitive monitoring Arithmetic		
r	.41 [a]	.17 [b]
p	<.001	.19
BF_{10}	>100	0.37

Note. [a] Partial correlation controlled for intellectual ability, arithmetic and spelling performance on the standardized tasks and chronological age; [b] Partial correlation controlled for intellectual ability and chronological age.

Table 5.C5

Regression analyses of MM_{arith} and MM_{spell} performance with metacognitive monitoring in the other domain, standardized task performance in both domains and chronological age as predictors (Grade 3)

	MM_{arith}			
	β	*t*	*p*	$BF_{inclusion}$
Age	-.12	-1.77	.08	2.04
Intellectual ability	.14	1.91	.06	2.12
TTA	.27	3.68	<.001	84.62
Dictation	-.12	-1.18	.24	1.06
MM_{spell}	.49	4.93	<.001	>100
	MM_{spell}			
	β	*t*	*p*	$BF_{inclusion}$
Age	.01	0.21	.84	0.19
Intellectual ability	.08	1.28	.20	0.36
Dictation	.55	8.49	<.001	>100
TTA	-.001	-0.01	.99	0.19
MM_{arith}	.34	4.93	<.001	>100

Note. TTA = Tempo Test Arithmetic; MM_{arith} = metacognitive monitoring in the arithmetic task; MM_{spell} = metacognitive monitoring in the spelling task.

Table 5.C6

Regression analyses of arithmetic performance (i.e. arithmetic$_{acc}$ and TTA) and spelling performance (i.e. spelling$_{acc}$ and dictation) with metacognitive monitoring in the other domain, standardized task performance in the other domain and chronological age as predictors (Grade 3)

	Arithmetic							
	Arithmetic$_{acc}$				TTA			
	β	t	p	BF$_{inclusion}$	β	t	p	BF$_{inclusion}$
Age	-.04	-0.52	.61	0.29	.04	0.48	.63	0.42
MM$_{spell}$	.53	4.99	<.001	>100	.22	1.95	.05	3.19
Dictation	-.07	-0.64	.53	0.30	.19	1.68	.10	1.83
	Spelling							
	Spelling$_{acc}$				Dictation			
	β	t	p	BF$_{inclusion}$	β	t	p	BF$_{inclusion}$
Age	.13	1.75	.08	1.24	.10	1.18	.24	1.14
MM$_{arith}$	.50	6.06	<.001	>100	.25	2.80	.006	10.16
TTA	.09	1.15	.25	0.68	.23	2.63	.01	11.67

Note. MM$_{arith}$ = metacognitive monitoring in the arithmetic task; MM$_{spell}$ = metacognitive monitoring in the spelling task; TTA = Tempo Test Arithmetic.

Supplemental Materials

In these supplementary materials, all remaining preregistered analyses for both studies that were not reported on in the main body of the manuscript are presented.

1 Study 1

1.1 Preliminary correlation analyses

As a validity check, we first verified whether the performance measures of the custom tasks that we made for this study were correlated with the widely-used standardized arithmetic and spelling tasks that we utilized (i.e. TTA and dictation). This was the case for both the accuracy (arithmetic: $r = .257$, $p = .002$; spelling: $r = .748$, $p < .001$) and the response time for correct answers (arithmetic: $r = -.709$, $p < .001$; spelling: $r = -.248$, $p = .003$).

1.2 General metacognitive knowledge

1.2.1 Method.

To measure metacognitive abilities independently of arithmetic and spelling, we used a general metacognitive questionnaire (adapted from Haberkorn, Lockl, Pohl, Ebert, & Weinert, 2014). In this questionnaire, 15 situations involving mental performance (e.g. *Which strategy do you think is better to make sure you won't forget to take your skates to school the next day?*) were described and three possible answers (e.g. *a. Write a note on a piece of paper; b. Think strongly about the skates; c. Both proposed strategies are equally good/bad*) were presented. The researcher read the situations and the corresponding options aloud one by one. Children were given a response form with pictures of the three possible answers, so they could follow each item and indicate their answer. The performance measure was the number of correct answers. The mean score on the general metacognitive knowledge questionnaire was 10.70 (SD = 2.35; range [4.00-15.00]).

5

The preregistered analyses involving the general metacognitive knowledge questionnaire are presented below.

1.2.2 Associations of academic performance and general metacognitive knowledge.

The general metacognitive knowledge questionnaire (MC_{know}) was not significantly associated with the arithmetic and spelling measures. For the arithmetic measures, the Bayes factors indicated moderate evidence in favour of no association with $Arithmetic_{rt}$ and the TTA; there was no considerable evidence in favour of or against an association with $Arithmetic_{acc}$. For the spelling measures, the Bayes factors indicated moderate evidence in favour of no association with $Spelling_{acc}$ and $Spelling_{rt}$; there was no considerable evidence in favour of or against an association with dictation.

Table 5.S1

Correlation analyses of metacognition measures and academic performance measures in 8-9-year-olds (Grade 3)

	Arithmetic			Spelling		
	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (TTA) [a]	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (dictation) [a]
Metacognitive knowledge						
r	.14	-.02	.06	.12	.03	.15
p	.12	.81	.50	.17	.75	.08
BF_{10}	0.36	0.11	0.13	0.27	0.11	0.49

Note. [a] Controlled for intellectual ability; [b] Controlled for intellectual ability and motor speed.

1.2.3 The unique role in academic performance of metacognitive monitoring within- and across-domain and metacognitive knowledge.

The correlation analyses show significant associations between our academic performance measures and metacognitive monitoring within the domain, in contrast to no significant correlations and (moderate) evidence in favour of no associations of the academic performance measures and general metacognitive knowledge. Comparing the strength of the associations confirms this pattern: For every performance measure, the strength of the association with metacognitive monitoring within the domain was significantly larger than the strength of the association with general metacognitive knowledge (William-Steiger tests; all p's < .001).

Regression analyses were performed to assess the unique contribution of our different metacognitive measures (i.e. metacognitive monitoring within- and across-domain and general metacognitive knowledge) to arithmetic and spelling performance. Therefore, all metacognitive measures were simultaneously entered into the regression models, together with intellectual ability as a control measure (Table 5.S2). For every model, all variance inflation factors (VIF) were smaller than 1.60, indicating no issues with collinearity among predictors.

Table 5.S2

Regression analyses of arithmetic and spelling performance with metacognitive monitoring within- and cross-domain, MC_{know} and intellectual ability as predictors

	Arithmetic								Spelling							
	$Arithmetic_{acc}$				Standardized task (TTA)				$Spelling_{acc}$				Standardized task (dictation)			
	β	t	p	BF inclusion	β	t	p	BF inclusion	β	t	p	BF inclusion	β	t	p	BF inclusion
MM_{arith}	.83	15.06	<.001	>100	.33	3.36	.001	97.81	-.04	-0.98	.33	0.06	-.05	-0.67	.51	0.18
MM_{spell}	.04	0.69	.50	0.09	.15	1.52	.13	0.70	.93	23.11	<.001	>100	.69	8.91	<.001	>100
MC_{know}	.08	1.88	.06	0.33	-.01	-0.17	.87	0.28	.02	0.76	.45	0.06	.10	1.63	.11	0.43
Intellectual ability	-.01	-0.12	.90	0.07	.02	0.26	.80	0.30	.04	1.23	.22	0.08	.09	1.38	.17	0.30

Note. MM_{arith} = metacognitive monitoring in the arithmetic task; MM_{spell} = metacognitive monitoring in the spelling task.

These results show that when metacognitive monitoring within- and across-domain, and general metacognitive knowledge were considered simultaneously, only the role of metacognitive monitoring within the domain itself remained significant (frequentist statistics) or supported (Bayesian statistics) for each academic performance measure (i.e. $Arithmetic_{acc}$, $Spelling_{acc}$, TTA, dictation).

1.2.4 Discussion.

Whereas the lack of a significant/supported association of academic performance with MC_{know} suggests that domain-generality of metacognition might be limited to metacognitive monitoring within the academic domains, it is important to note some characteristics of this study that might limit this interpretation of the specificity of metacognition. The current study included two aspects of metacognition, namely declarative, general metacognitive knowledge (MC_{know}) and an aspect of procedural metacognition (i.e. metacognitive monitoring; MM_{arith} and MM_{spell}). In all above-mentioned results, the role of metacognitive monitoring surpasses the role of general metacognitive knowledge in academic performance. Firstly, this difference in results could be due to a difference in the cognitive process, namely declarative vs. procedural metacognition. It is not surprising that different aspects of metacognition follow different developmental paths (Schneider & Löffler, 2016) and are differently associated with domain-specific skills. To further investigate whether this difference in results is indeed due to a difference in metacognitive processes, future studies should not only include domain-general declarative metacognitive knowledge, but also measures of domain-specific, declarative metacognitive knowledge. Secondly, the difference in results between metacognitive monitoring and declarative metacognitive knowledge could be due to a difference in performance measures that were used to measure these skills. The general metacognitive knowledge questionnaire (used to measure declarative, general metacognitive knowledge), while being a validated measure, differs in two considerable ways from our monitoring measure, which was online and more detailed (i.e. on a trial-by-trial basis). These aspects of the current study could limit the interpretation of the specificity of metacognition outside the academic domains and outside of metacognitive monitoring. Drawing on our results, future research should include measures of both metacognitive monitoring and control (i.e. both aspects of procedural metacognition) and general as well as domain-specific declarative metacognitive knowledge (e.g. Neuenhaus, Artelt, Lingel, & Schneider, 2011) to further investigate the domain-specificity question of metacognition.

1.3 Academic-domain-related differences in the role of metacognitive monitoring

We investigated whether there were academic-domain-related differences (i.e. differences between arithmetic and spelling) in the (strength of the) role of metacognition (i.e. task-specific metacognitive monitoring on the one hand and general metacognitive knowledge on the other).

Our results show that there were differences in the strength of the role of metacognitive monitoring depending on the domain. Using the Fisher r-to-z transformation, we found that the association between MM_{spell} and $Spelling_{acc}$ was significantly larger than the association between MM_{arith} and $Arithmetic_{acc}$ (p = .014). This was also the case for the standardized task, with the association between MM_{spell} and dictation being significantly larger than the association between MM_{arith} and the TTA (p = .001). There were no significant differences between the domains in the strength of the associations with MC_{know}.

2 Study 2

2.1 Preliminary correlation analyses

The same validity tests were performed as in Study 1, showing that for accuracy (arithmetic: r = .46, p < .001, BF_{10} > 100; spelling: r = .28, p = .02, BF_{10} = 2.16) performance on the custom task was correlated with the standardized task. For response time for correct answers (arithmetic: r = -.05, p = .70, BF_{10} = 0.17; spelling: r = -.05, p = .69, BF_{10} = 0.17) this was not the case.

2.2 General metacognitive knowledge

The preregistered analyses involving the general metacognitive knowledge questionnaire in Grade 2 are presented below. The mean score on the general metacognitive knowledge questionnaire was 6.65 (SD = 2.44; range [1.00-11.00]).

2.2.1 Associations of academic performance and general metacognitive knowledge.

The general metacognitive knowledge questionnaire (MC_{know}) was not significantly associated with the arithmetic and spelling measures. Bayes factors indicated moderated evidence in favour of no association with both response time measures, and for $Spelling_{acc}$ and dictation. There was only anecdotal evidence for no association with $Arithmetic_{acc}$ and TTA.

Based on the lack of significant/supported associations, no further analyses were performed.

Table 5.S3
Correlation analyses of metacognition measures and academic performance measures in 7-8-year-olds (Grade 2)

	Arithmetic			Spelling		
	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (TTA) [a]	Custom task – Accuracy [a]	Custom task - RT [b]	Standardized task (dictation) [a]
Metacognitive knowledge						
r	.21	.05	.23	.10	.03	.05
p	.09	.68	.07	.43	.84	.71
BF_{10}	0.62	0.16	0.79	0.21	0.16	0.16

2.3 Academic-domain-related differences in the role of metacognition

We investigated whether there were academic-domain-related differences (i.e. differences between arithmetic and spelling) in the (strength of the) role of metacognition (i.e. task-specific metacognitive monitoring on the one hand and general metacognitive knowledge on the other).

Our results show that there were almost no differences in the strength of the role of metacognitive monitoring depending on the domain. Using the Fisher r-to-z transformation, we found that only the association between MM_{spell} and $Spelling_{acc}$ was significantly larger than the association between MM_{arith} and $Arithmetic_{acc}$ (p = .005). There were no other significant differences between the domains in the strength of the associations with metacognition.

CHAPTER 6

Metacognition in children's brains

The neurobiological basis of metacognitive monitoring during arithmetic in the developing brain.

The content of this chapter is under revision as:

Bellon, E., Fias, W., & De Smedt, B. (Under Revision). Metacognition in children's brains. The neurobiological basis of metacognitive monitoring during arithmetic in the developing brain. *Human Brain Mapping*.

Chapter 6

Metacognition in children's brains

The neurobiological basis of metacognitive monitoring during arithmetic in the developing brain.

Abstract

In contrast to a substantial body of research on the neurobiological basis of cognitive performance in several academic domains, less is known about how the brain generates metacognitive awareness of such performance. The existing work on the neurobiological underpinnings of metacognition has almost exclusively been done in adults and has largely focused on lower level cognitive processing domains, such as perceptual decision making. Extending this body of evidence, we investigated metacognitive monitoring by asking children to solve arithmetic problems, an educationally relevant higher-order process, while providing concurrent metacognitive reports during fMRI acquisition. This was done in a sample of 55 primary school children aged 9-10-years-old. The current study is the first to demonstrate that brain activity during metacognitive monitoring, relative to the control task, increased in the left inferior frontal gyrus in children. This brain activity further correlated with children's arithmetic development over a three year time period. These data are in line with the frequently suggested, yet never empirically tested, hypothesis that activity in the prefrontal cortex during arithmetic is related to the higher-order process of metacognitive monitoring.

Introduction

Cognitive neuroscience has made considerable progress in understanding the neurobiological basis of cognitive performance in several academic domains, such as arithmetic. Much less is known, however, about how the brain generates metacognitive awareness of task performance (Fleming & Dolan, 2012) during academic performance. Understanding the neurobiological basis of metacognition is essential, as this higher-order process supports reflection upon and control of other cognitive processes, and occupies a central role in human cognition (Flavell, 1979). Age-related improvements in children's ability to monitor and regulate their mental operations are widely recognized to be a driving force in cognitive development, underlying age-related improvements in accuracy on a wide variety of tasks (e.g. Lyons & Ghetti, 2010), such as arithmetic (Rinne & Mazzocco, 2014). In view of the extensive behavioural work on the importance of metacognition in academic performance (e.g. Roebers et al., 2012; Schneider & Artelt, 2010; Schraw et al., 2006), there is a need to further our understanding of metacognitive processes in the context of academic skills at the level of the brain.

Metacognition is considered a higher brain function that strongly depends on the prefrontal cortex or PFC (see Pannu & Kaszniak, 2005, and Shimamura, 2000, for reviews). Adult studies on the neurobiological correlates of metacognitive judgments across different tasks have pointed to a consistent involvement of a fronto-parietal network (e.g. Fleming & Dolan, 2014; see Vaccaro & Fleming, 2018, for a meta-analysis). There are, however, three critical limitations in the current literature on the neurobiological underpinnings of metacognition that motivated the current study. Firstly, and to the best of our knowledge, the existing body of data is solely based on adult studies. Therefore, the results cannot be generalized to the neurobiological basis of metacognition in children without thorough empirical investigation. Secondly, this adult work has almost exclusively been done in lower level cognitive processing domains, such as perceptual decision making (e.g. Fleming et al., 2012; Fleming & Dolan, 2014; Shimamura, 2000; Vaccaro & Fleming, 2018). Yet, there is evidence to suggest that there is specificity, i.e. regional specialization within the PFC, concerning the neurobiological basis of metacognition with respect to metacognitive processes in different tasks and domains (e.g. Baird et al., 2013; McCurdy et al., 2013). Hitherto, it remains unknown what the neurobiological correlates of metacognition on high-level cognitive processing are. Thirdly, Vaccaro and Fleming (2018) indicated that some aspects of the neurobiological basis of metacognition have been overlooked. Most research has focused on brain activity related to metacognitive confidence judgements in task performance or related to the extent to which a metacognitive judgment effectively tracks task performance (i.e. metacognitive ability). Yet, the fundamental question of which brain regions are involved in engaging in a metacognitive monitoring task regardless of participants' behavioural performance (in other words, the level of confidence that participants indicate or their metacognitive ability) has been neglected. Answering this question is crucial to understand the underlying neurocognitive architecture supporting metacognitive abilities. This was precisely the aim of the current study. We therefore examined metacognitive judgment-related activity in itself, namely activation that results from contrasts comparing the requirement of metacognitive judgment against a control condition.

The current study tackles these important issues by investigating them for the first time in children. We investigated which brain region(s) are active when engaging in a metacognitive monitoring task through the use of retrospective metacognitive monitoring judgments in a higher-level cognitive process, namely arithmetic.

Investigating the brain activity during metacognitive monitoring of arithmetic also adds to the existing body of developmental brain imaging studies that have studied brain activity during arithmetic (Arsalidou et al., 2018, for a meta-analysis; Peters and De Smedt, 2017, for a systematic review), as this might lead to a better understanding of the activity in prefrontal regions which has been consistently observed during arithmetic. Indeed, it has been frequently suggested that this prefrontal activation during arithmetic reflects metacognitive monitoring as well as working memory load or goal-directed problem solving (e.g. Ansari et al., 2005; Arsalidou et al., 2018; Houdé et al., 2010; Kaufmann et al., 2006, 2011; Kucian et al., 2008; Menon, 2015; Rivera et al., 2005). However, this suggestion that the control networks that are active during arithmetic might point, at least partially, to the involvement of metacognitive processes, has never been empirically tested.

This suggestion is not far-fetched as behavioural work has revealed that metacognitive monitoring is a unique predictor of individual differences in arithmetic in children (Bellon et al., 2019; Rinne and Mazzocco, 2014). Interestingly, Ansari et al. (2011) showed in adults that medial and lateral regions of the PFC were correlated with the detection of arithmetic errors and deployment of control following an arithmetic error. These authors suggested that activation of these regions might suggest greater awareness of mistakes during calculation, pointing to the role of metacognition.

In sum, the current study is the first to empirically investigate which brain regions are involved in engaging in metacognitive monitoring within a higher-order cognitive processing domain (i.e. arithmetic), and to do so in primary school children. Investigating this also sheds light on the frequently suggested, but never empirically tested hypothesis that metacognitive monitoring processes, which were found to be an important predictor of arithmetic skills in behavioural research, could partially explain the increases in prefrontal activation that are often observed when doing arithmetic.

We examined these questions in primary school children aged 9-10 as they are in the midst of an important developmental period of both arithmetic (e.g. Vanbinst, Ceulemans, et al., 2015) and metacognition (e.g. Schneider, 2010). Children participated in an fMRI experiment in which they were asked to solve arithmetic problems and to answer either metacognitive questions (i.e. experimental condition) or to make a colour judgement (i.e. control condition) while they were in the scanner. To further explore the association between brain activity during metacognitive monitoring and children's arithmetic development, we specifically recruited children that took part in a larger longitudinal behavioural project in which developmental arithmetic data were collected. This allowed us to explore associations between children's brain activity during metacognitive monitoring and their arithmetic development.

Method

1 Participants

Participants were 55 children (30 girls; 2 left-handed), aged 9 to 10 years old (M_{age} = 10 years 2 months, SD = 3 months, [9 years 7 months - 10 years 7 months]). After correction for movement in the scanner (see below), the final sample consisted of 50 participants. All children were recruited from an ongoing 3-year-longitudinal study on the role of metacognitive monitoring in arithmetic (Bellon et al., 2019). They were all typically developing children, who had no diagnosis of a developmental disorder, nor reported a history of psychiatric or neurological illness. They had normal or corrected-to-normal vision, and a dominantly middle-to-high socioeconomic background. For every participant, written informed parental consent was obtained. In return for participating, all children were given a financial compensation. The study was approved by the Medical Ethical Committee of KU Leuven (S59167).

2 Procedure

Each child participated in two sessions. During the first session, children were extensively informed about the scanning procedure. They were familiarized with the MRI environment and procedures using a mock scanner in which every step of the MRI procedure was practiced while the noise of the scanner was simulated. They also completed an arithmetic fluency test (see below). Additionally, an extensive cognitive test battery was administered, as part of an ongoing longitudinal study in which these children participated, including executive functioning, numerical magnitude processing, reading ability and mathematics anxiety. The data from this behavioural test battery were not considered for the current study. During the second session, brain imaging data were collected. Both functional data (during an arithmetic task) and structural data were acquired (for scanning parameters see below). The full MRI-protocol lasted approximately 50 minutes.

3 Imaging task

An arithmetic task was performed by the children in the scanner. This task was specifically designed to tap into both arithmetic and metacognitive processes, using a specific protocol adapted from recent behavioural research (Bellon et al., 2019; Rinne & Mazzocco, 2014). Similar metacognitive protocols have also been used in adult neuro-imaging research (e.g. Chua et al., 2009; Fleming et al., 2012; Hilgenstock et al., 2014). An overview of the arithmetic task, including its timing is illustrated in Figure 6.1. The task was presented across five functional runs in a block fMRI design. In each run, 30 multiplication items were presented in which children were asked to indicate which of the two presented solutions (i.e. one on either side of the screen) was correct. Two conditions were administered (i.e. experimental condition and control condition, see Figure 6.2) within each run. Each run was divided into six blocks: experimental (*n* = 3) and control (*n* = 3) blocks were alternated. A block comprised of a long fixation (15s), an indication of which condition would follow (1000 ms), five arithmetic trials of the same condition (35s) and an end fixation (15s); see Figure 6.3. Each arithmetic trial consisted of a short fixation (200 ms), a presentation of the multiplication item and a response screen (in total 4300 ms), a short black screen (100 ms) and an additional question depending on the condition (2500 ms). A multiplication item consisted of the presentation of the arithmetic problem (2000 ms), the presentation of a white equality sign (100 ms), the presentation of a coloured equality sign and two solutions to the arithmetic problem (i.e. one lure and one correct solution; 2100 ms), and a black screen (100 ms). Children answered using buttons on a response box corresponding to the location of the response options on the screen. The duration of each run was approximately 5 minutes.

Each participant was presented with a set of 150 multiplication items. Multiplication was chosen as arithmetic operation of interest to ensure considerable inter- and intra-individual variability in performance by using items of different difficulty levels, while still using a task with which children were very familiar, and which was as ecologically valid as possible. To maximize variability in both arithmetic performance and metacognition processes (experimental condition, see below) a wide range of multiplication items was included, ranging from easy items (*n* = 50; i.e. single-digit multiplications items with 0-1 and 2-9 as operands, and single times double digit items with 0-1 or 10-11 and 12-19 or 2-9, respectively, as operands) over standard multiplication tables (*n* = 50; i.e. single-digit multiplications with 2-9 as operands) to hard items (*n* = 50; i.e. single- times double-digit multiplications with 2-9 and 12-19 as operands). We did not include ties, standard single-digit items that were considered 'too easy' (i.e. 2x3, 2x4, 3x4 and their commutative pairs), and hard items that were considered 'too difficult' (i.e. operands 17-19 combined with operands 7-9). In each run, the same number of single-digit items as well as single- times double-digit items was presented. The number of times a specific operand was presented in one run was equally distributed across runs. Commutative pairs were never presented within the same run.

All multiplication items were presented horizontally, in white (Calibri, font size 80) on a black background and in Arabic digits. On presentation of the two solutions to the arithmetic problem, the children were asked to indicate where the correct solution was presented by pressing the leftmost or rightmost button on the response boxes for the left or right response alternatives, respectively. Lure solutions were one of five possible categories, namely the correct solution plus or minus the value of the largest operand, the correct solution plus or minus the value of the smallest operand or the solution to the corresponding addition. As a result, most of the proposed incorrect solutions were table related products. Lures from each category were evenly distributed over blocks and conditions. The position of the correct answer was balanced.

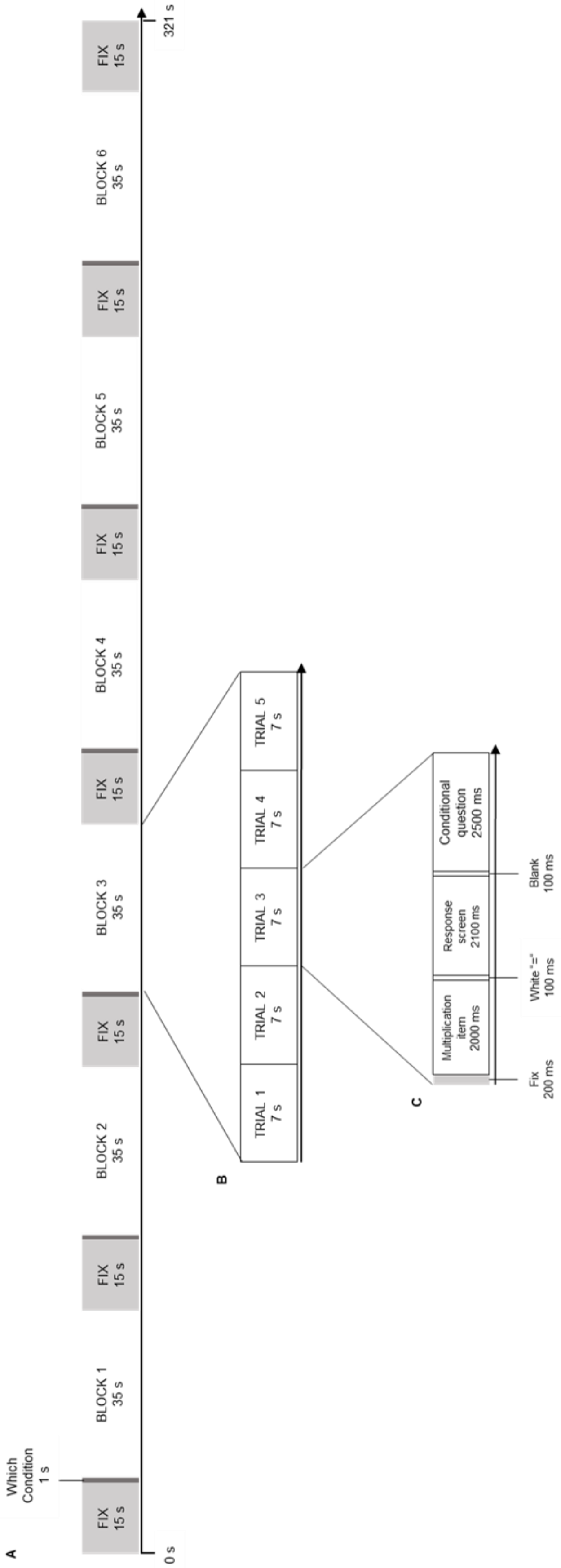

Figure 6.1 Schematic overview of the arithmetic task. (A) Overview of run; (B) Overview of block; (C) Overview of trial.

To truly isolate the act of engaging in metacognitive processes, two conditions were created, namely an experimental condition in which a metacognitive question was asked after the arithmetic item, and a control condition, in which every aspect of the arithmetic task was identical, and only the nature of the question that was asked after the arithmetic item differed. In this control condition, a question on colour was asked.

3.1 Experimental condition: Metacognitive question

In the metacognitive condition, after each arithmetic item children were asked to report their judgment on the accuracy of their arithmetic answer, by indicating whether they thought their answer was "*Correct*", "*Incorrect*" or whether they "*Did not know*". We used emoticons in combination with the options to make the task more attractive and feasible for the children (see Figure 6.2 left panel). The participating children were very familiar with this task, as they already participated in an ongoing longitudinal study in which this protocol to assess metacognitive monitoring was used (Bellon et al., 2019).

3.2 Control condition: colour question

In the control condition, after each arithmetic item, children were asked which of three colours the equality sign (presented simultaneously with the two solutions) had. Importantly, the equal sign was coloured in both conditions, to make conditions as similar as possible. Only in the control condition children were asked to report on the colour (see Figure 6.2 right panel). This specific control condition was used to engage similar memory processes as during the metacognitive judgment (i.e. both involve thinking back), yet the content of the cognitive process was entirely different, as in the metacognitive condition the children think back to their own performance, while in the colour condition they have to remember the colour they saw.

Taken together, the two conditions were exactly the same in terms of timing, nature of the stimuli and arithmetic task. The only difference between them was that in the experimental condition they had to make a judgement on their own performance on the item, while in the control condition they had to make a judgement about colour of the item.

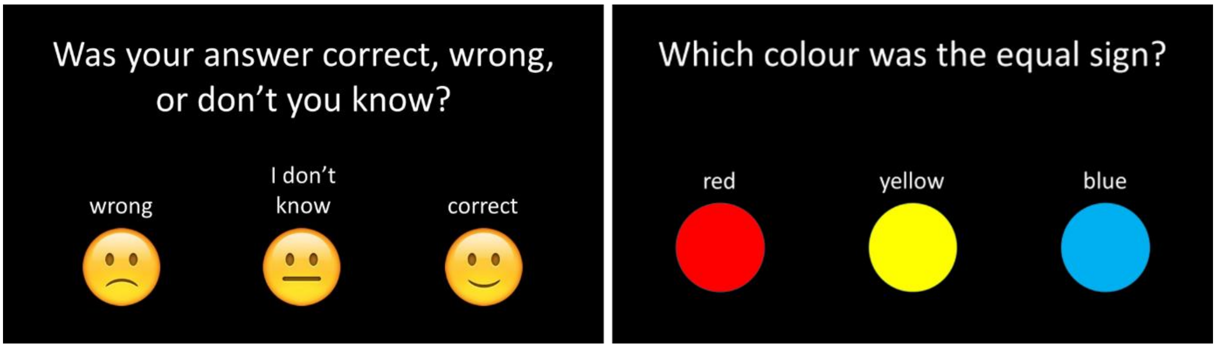

Figure 6.2. Screen presented for the experimental condition: Metacognitive question (left); Screen presented for the control condition: Colour question (right).

6

In Figure 6.3, an overview of a block in both conditions is presented, in which detailed information of the course of an arithmetic item can be found.

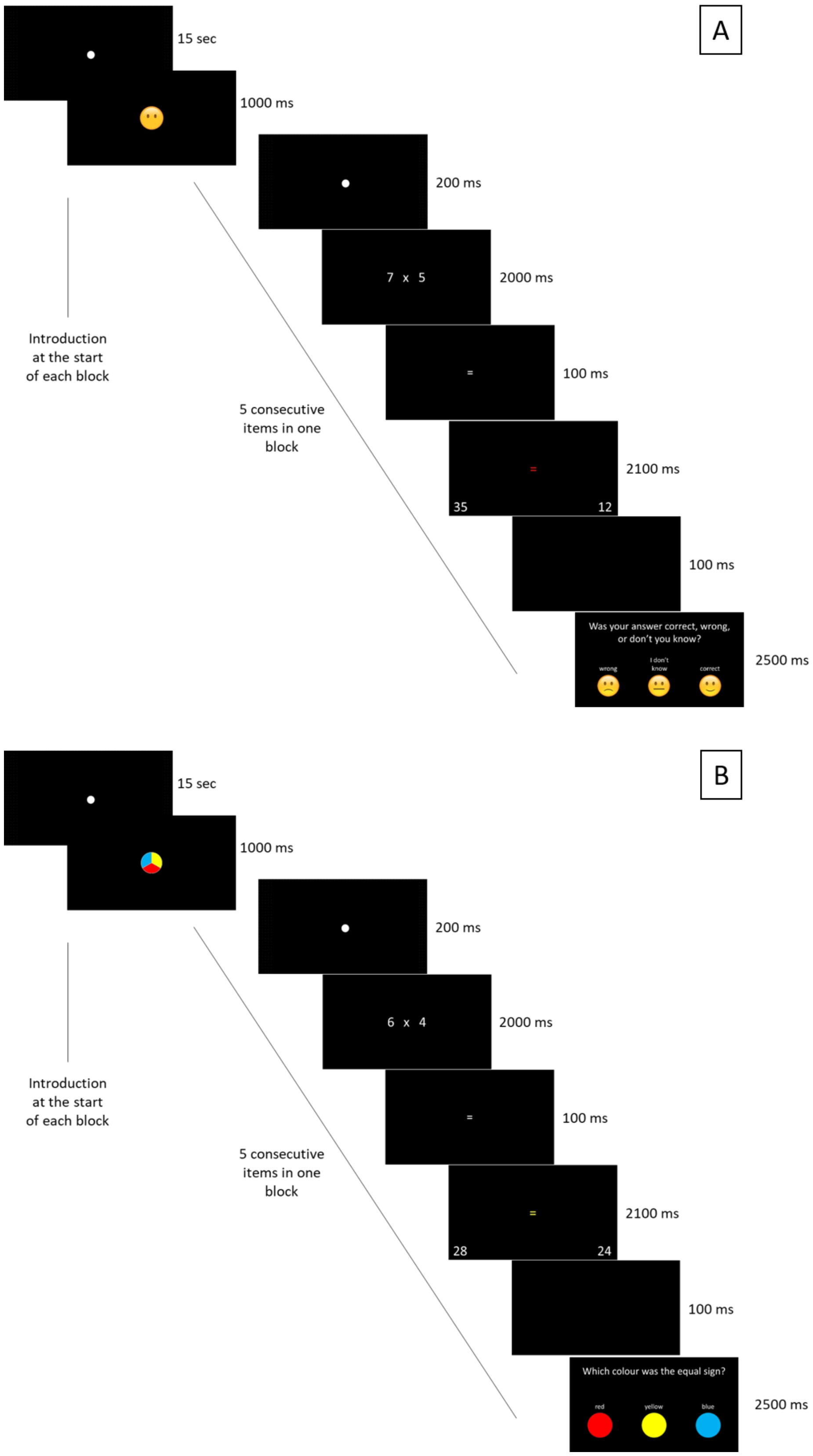

Figure 6.3. Overview of a block in the experimental (A) and control (B) condition.

Stimuli were presented using a script written in MATLAB (The MathWorks Inc., 2018), displayed using PsychToolbox 3 (Brainard, 1997), via a projector (NEC Display Solutions) onto a screen, which was made visible through a mirror attached to the head coil, located approximately 46cm behind the participants' eyes.

3.3 Scanning parameters

Structural and functional images were collected via a 3.0T Philips Ingenia CX MRI Scanner with a SENSE 32-channel head coil, located at the Department of Radiology of the University Hospital in Leuven, Belgium. Soft padding was used to stabilize the children's heads in order to minimize head motion. For the fMRI data, slices were recorded in ascending order, using a EPI sequence (52 slices, 2.19 x 2.19 x 2.2mm voxel size, 2.2 mm slice thickness, 0.3mm interslice gap, TR = 3000ms, TE = 29.8ms, 90° flip angle, 96 x 96 acquisition matrix) and covered the whole brain (field of view: 210 x 210 x 130 mm). Each run consisted of 107 measurements. Furthermore, a high-resolution T1-weighted anatomical image (MPRAGE sequence, 182 slices, resolution 0.98 x 0.98 x 1.2mm^3, TE = 4.6ms, 256 x 256 acquisition matrix, 8° flip angle, 250 x 250 x 218 mm field of view) was acquired for each participant.

4 Behavioural task outside the scanner

Arithmetic fluency was assessed by the Tempo Test Arithmetic (TTA; de Vos, 1992); a standardized pen-and-paper test of arithmetical fluency which comprises five columns of arithmetic items (one column per operation and a mixed column), each increasing in difficulty. Participants got one minute per column to provide as many correct answers as possible. The performance measure was the total number of correctly solved items within the given time (i.e. total score over the five columns).

Because all participants were enrolled in a longitudinal study (Bellon et al., 2019), performance on the TTA was not only available from the behavioural session that accompanied the MRI session, but also from when these participants were in second and third grade (i.e. 7-8 and 8-9-years-old, respectively). These data were further included in the current study.

5 Data analysis

All pre-processing was conducted with the Statistical Parametric Mapping (SPM) software package for MATLAB (SPM12, Wellcome Department of Cognitive Neurology, London). Functional images were corrected for slice-timing differences and for head motion artefacts by realigning all images to the mean image, and were co-registered to the high-resolution anatomical image. Both functional and anatomical images were normalized to the standard Montreal Neurological 152-brain average template. As a final pre-processing step, functional images were spatially smoothed using a Gaussian kernel of 6 mm full-width at half maximum (FWHM).

To avoid a decrease in data quality due to movement during scanning, two motion criteria (see also Peters et al., 2018) were used to identify excessive movement during functional runs. Firstly, all runs in which participants moved more than one voxel size (2.2 mm) in the x-, y- or z-direction on two consecutive images, were discarded. Secondly, runs in which an Euclidean distance measure (i.e. an additive measure of the amount of motion in all directions from one time point to another), exceeded one voxel size, were also removed. Participants with less than three runs without excessive movement, were discarded in all analyses on both the imaging and behavioural data. This criterion led to the discarding of 5 participants, leading to a final sample of 50 children. Of these remaining participants, 7% of the runs were discarded from the analyses due to excessive motion.

After pre-processing, as a part of the first level analysis, the effect of the experimental condition per voxel was estimated by creating a general linear model per participant. Onset and duration of each block of each condition were modelled. These regressors were convolved with a canonical hemodynamic response function (HRF). The six motion realignment parameters for each subject were included as regressors of no interest in the general linear models, to further control for variation due to movement artefacts.

To measure the neurobiological correlates of metacognitive monitoring, a 'metacognition contrast' was created in the first-level analysis by subtracting the average BOLD response of the control condition (i.e. colour task) from the experimental condition (i.e. metacognitive question), resulting in voxel-wise *t*-statistics maps for each participant.

Finally, a second-level group analysis was performed on the first level contrast images of the 'metacognition contrast' using a one-sample t-test to identify brain regions with higher activity during metacognitive judgment than during the control condition. We studied activation at a whole brain level, threshold of $p < .05$ after family wise error or FWE correction, to control for multiple comparisons. Anatomical labels of results were defined using the xjView toolbox for SPM (https://www.alivelearn.net/xjview).

To further understand the results of the metacognitive contrast, functionally defined region(s) of interest (ROI) were generated from significantly activated cluster(s) in this contrast, using the Marsbar toolbox for MATLAB (Brett et al., 2002). From the ROI(s), we extracted the contrast estimates of the metacognitive contrast, also using Marsbar. High values indicated a large difference between the activation in the metacognitive condition versus the control condition. These contrast estimates were then used for examining brain-behaviour correlations.

As in the adult literature specific regions were found depending on the studied metacognitive aspect (e.g. judgment-related activity, judgment level or metacognitive monitoring ability; Vaccaro & Fleming, 2018), we firstly explored whether the activation found for engaging in metacognitive thought (i.e. judgment-related activity) was correlated with these other metacognitive aspects (i.e. absolute metacognitive judgment and metacognitive monitoring ability), which were inferred from the behavioural data of in-scanner performance (for scoring see Table 6.1). Pearson correlations were calculated between the contrast estimates of the metacognitive contrast and the different metacognitive aspects, i.e. absolute metacognitive judgment and metacognitive monitoring ability.

Secondly, against the background of behavioural research in which metacognitive monitoring was an important predictor of arithmetic performance, we further explored whether the activation found for engaging in metacognitive thought was associated with children's arithmetic and its development. Therefore, we used developmental behavioural data from the longitudinal study in which these children were enrolled (Bellon et al., 2019). Specifically, children's score on the TTA was used as an indicator of their arithmetic fluency, which were collected at each time point (grade 2, grade 3 and grade 4). From these data, a linear regression was calculated to predict their arithmetic fluency. For each individual we derived an intercept and slope, which reflected the starting level and the change over time, respectively. These behavioural measures were subsequently correlated with the extracted contrast estimates of the metacognitive contrast.

Results

1 In-scanner behavioural results

In-scanner behavioural results were only analysed for runs that were included in the imaging analyses. Descriptive statistics of the in-scanner behavioural results are displayed in Table 6.1.

To verify whether the two conditions of the arithmetic task in the scanner (i.e. metacognitive (MC) condition and colour (C) condition) differed in task difficulty level, we compared (a) whether or not participants were able to provide an answer to the arithmetic item within the given time frame (i.e. 2000 ms), independent of the accuracy of that answer (i.e. a score of 0 was given if participants failed to answer within the time limit; a score of 1 when they were able to answer within the time frame) and (b) the number of correct arithmetic responses that were given within the time limit (i.e. a score of 0 when participants chose the incorrect solution to the arithmetic item; a score of 1 when they chose the correct solution). Importantly, trials in which participants did not respond, or responded too late due to the time limit, were excluded from the correct responses scores. No differences between the conditions were found on either of the arithmetic performance measures (independent sample t-test arithmetic response rate: M_{MC} = 0.85 , SD_{MC} = 0.10 ; M_C = 0.83, SD_C = 0.13; $t(98) = 0.83$, $p = .41$; independent sample t-test correct arithmetic responses: M_{MC} = 0.81 , SD_{MC} = 0.09 ; M_C = 0.83, SD_C = 0.8; $t(98) = -1$, $p = .32$). Bayes factors for these analyses indicated evidence for the null hypothesis of no difference between the conditions (both BF_{10}'s < 0.33). This equivalence indicates there was no difference in degree of cognitive demand in the arithmetic task between the two conditions, and thus ensures that differences in brain activity between these conditions are not due to variation in arithmetic task performance.

Table 6.1
Arithmetic and metacognitive performance in the scanner

	n	*M*	*SD*	Range	Theoretical maximum
In-scanner arithmetic performance					
Arithmetic response rate [a]	50	0.84	0.11	[0.59-1.00]	1.00
Arithmetic correct responses [b, c]	50	0.82	0.08	[0.62-1.00]	1.00
In-scanner absolute metacognitive judgment					
Absolute accuracy judgment [b, d]	50	2.62	0.17	[2.08-2.94]	3.00
In-scanner metacognitive monitoring ability					
Monitoring ability [b, e]	50	1.65	0.15	[1.26-1.95]	2.00

Note. [a] A score of 0 was given if participants failed to answer the arithmetic item within the time limit of 2000 ms; a score of 1 when they were able to answer within the time frame. [b] Only items on which participants were able to provide an arithmetic answer within the time frame were included in this measure. [c] A score of 0 was obtained if the arithmetic answer given was incorrect, a score of 1 if the arithmetic answer was correct. [d] A score of 3 was given if children indicated they were certain their arithmetic answer was correct, a score of 2 if they indicated they were unsure about their arithmetic answer, a score of 1 if they thought their arithmetic answer was incorrect. [e] A score of 2 was obtained if their metacognitive judgment corresponded to their actual performance (i.e. metacognitively judged as *Correct* and indeed correct academic answer; metacognitively judged as *Incorrect* and indeed incorrect academic answer), a score of 0 if their metacognitive judgement did not correspond to their actual performance (i.e. metacognitively judged as *Correct* and in fact incorrect academic answer; metacognitively judged as *Incorrect* and in fact correct academic answer), and a score of 1 if children indicated they *Did not know* about their academic answer.

2 Imaging results

To isolate areas of functional significance during which participants metacognitively judged the accuracy of their arithmetic answer, we examined the difference in neural activation between the metacognitive condition and the control (i.e. colour) condition, i.e. the metacognition contrast. An overview of the clusters that were more active during the metacognition than during the colour condition can be found in Table 6.2. A visualisation of this contrast is displayed in Figure 6.4. These differences were FWE corrected at $p < .05$. Our findings revealed that engaging in a metacognitive task was associated with stronger activation in the left inferior frontal gyrus (IFG). There were no other clusters that showed increased activity during the metacognition as compared to the control condition.

Table 6.2
Region, coordinates of the peak voxel, number of voxels (k) and t-value of the activation clusters elicited by the metacognitive contrast. Voxel coordinates are presented in MNI space.

	Peak coordinates				
Cluster	*x*	*y*	*z*	*k*	*t*
Metacognition > Control condition					
Left inferior frontal gyrus	-47	30	-5	75	7.04
	-56	21	13	10	4.94

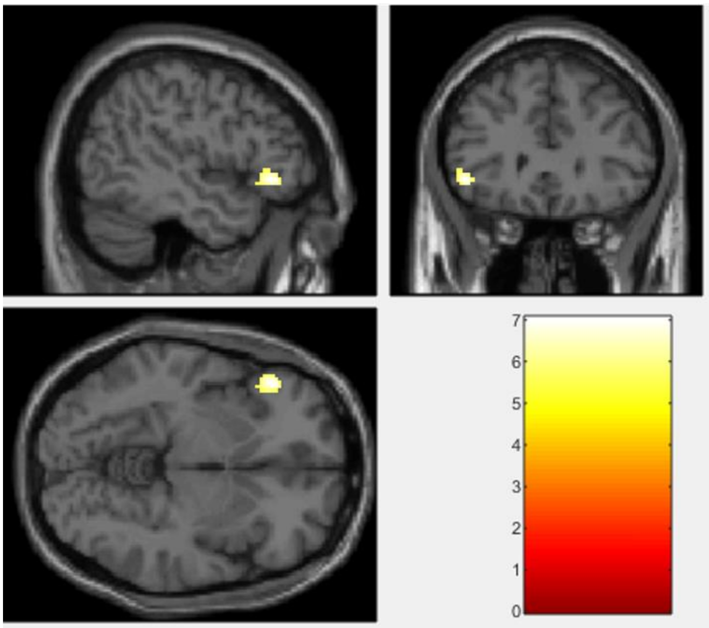

Figure 6.4. Results from the whole brain analysis of the metacognitive contrast. The activation map was corrected for multiple comparisons through a family wise error (FWE) correction with a $p < 0.05$ threshold.

3 Brain-behaviour correlations

The significant cluster found in the left IFG was used as ROI to further understand the results of the metacognitive contrast. From this ROI, the contrast estimates of the metacognitive contrast were extracted. These beta-values, which represent the activation difference between the metacognitive and the control condition, were correlated with metacognitive and arithmetic performance indices (see below).

3.1 Absolute metacognitive judgement & metacognitive monitoring ability

We explored whether the activation found for engaging in metacognitive thought (i.e. activation in the left IFG) was also significantly correlated with other metacognitive aspects (i.e. metacognitive judgment level and metacognitive monitoring ability; Figure 6.5). No significant correlations were found between brain activation for engaging in a metacognitive monitoring task and metacognitive judgment level or metacognitive monitoring ability. Bayes factors pointed to evidence for the null hypotheses.

3.2 Arithmetic

The results of the TTA on three time points are displayed in Table 6.3. Significant age-related changes in TTA score were found, with performance in each time point significantly differing from the other time points ($F(2, 147) = 29.80$, $p < .001$; Post hoc tests using Bonferroni correction: all p's < .02). The intercept and slope of that change over time were calculated, indicating that on average children started with a performance of around 60 arithmetic items solved in 5 minutes, and each year, they were able to solve on average 14 items more.

Table 6.3
Performance on the TTA on three time points and development operationalised as intercept and slope of the regression line between the three time points

	n	*M*	*SD*	Range
TTA T_1 (grade 2)	50	72.62	16.37	[41-108]
TTA T_2 (grade 3)	50	90.32	19.34	[52-127]
TTA T_3 (grade 4)	50	100.72	19.34	[65-142]
Intercept	50	59.79	18.31	[21.33-105.33]
Slope	50	14.05	5.83	[2.5-26.0]

We further explored whether the activation found for engaging in metacognitive thought was associated with arithmetic development, as measured by intercept and slope of the regression line of TTA performance on three time points (Figure 6.5). A significant correlation was found between brain activation for engaging in a metacognitive monitoring task and the intercept of arithmetic development. There was no significant correlation with slope in arithmetic development and Bayes factors pointed to evidence for the null hypothesis.

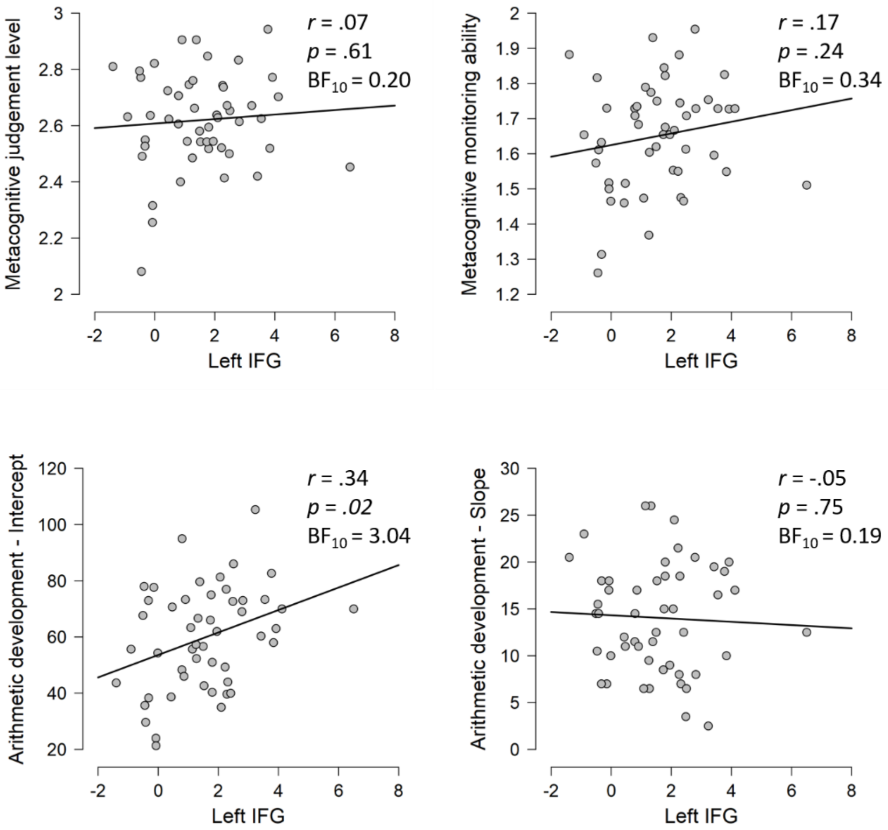

Figure 6.5. Scatterplots with fit lines of the associations and Pearson correlation coefficients between the behavioural measures of metacognition and arithmetic and brain activation in the left IFG.

Discussion

The current study tackled an important gap in the existing literature on how the brain generates metacognitive awareness of task performance. While there is already some evidence on this ability in adults (Fleming and Dolan, 2012), there are no brain imaging data available on this issue in children. Moreover, research focused predominantly on lower level cognitive processing and has mostly neglected particular aspects of the neurobiological basis of metacognition, namely which brain regions are involved in engaging in a metacognitive monitoring task.

Addressing these gaps in the literature, the current study was the first to explicitly investigate the brain activation underlying the engagement in metacognitive monitoring in children, and during metacognitive monitoring in an academic task. We observed increased activation in the left IFG relative to the control task. No other increases in brain activity during metacognitive monitoring were observed. Brain-behavior correlations indicated that brain activity related to engaging in metacognitive monitoring and behavioral arithmetic performance were associated. These data are in line with the suggestion that prefrontal activation in the arithmetic brain network may be, at least partially, related to metacognition.

A comparison of the existing literature, which is exclusively based on adults, and the current data in children, demonstrates both similarities and differences in the neurobiological basis of engaging in metacognitive monitoring. Our results are in line with Chua et al. (2009), who found greater activity in the left inferior frontal region (BA 47) in adults for retrospective metacognitive monitoring compared to a prospective feeling-of-knowing. Our data are also in accordance with the in adults consistently observed activation increases in prefrontal regions during metacognitive monitoring. However, the exact location where this increased activation in the prefrontal cortex is found, differs depending on the very diverse study characteristics in the existing literature. These include operationalisation of metacognitive monitoring and the metacognitive aspect under study (e.g. confidence versus metacognitive ability), used contrasts (e.g. monitoring versus fixation or task performance), and the domain in which metacognitive monitoring was studied (e.g. perceptual decision making versus memory domain). For example, using low versus high confidence as metacognitive measure compared to fixation Chua et al. (2006) found activation differences in the PFC including anterior, dorsolateral and posterior regions of the bilateral IFG. Yet, when comparing confidence rating and a recognition task instead of fixation, they found different activation patterns (e.g. right orbitofrontal regions). Yokoyama et al. (2010) found that, in adults who were good at predicting the correctness of their recognition memory performance (i.e. as measured by a significantly positive gamma), brain regions exhibiting higher activity during confidence rating compared to a perceptual task included bilateral superior frontal regions.

Using a similar design as in the current study, a small number of studies in adults have examined the brain activity of engaging in metacognitive monitoring independent of participants actual behavioural task performance; i.e. the brain activity regardless of which metacognitive judgment (e.g. “I think I’m (in)correct”) is given and regardless of whether one’s metacognitive judgment is aligned with the actual task performance. Specifically, Fleming and colleagues (2012) found that, in adults, in a perceptual decision making task, the right rostrolateral PFC showed greater activity during self-report compared to a matched control condition.

Because of this large variability in characteristics of the studies that investigated the neurobiological basis of metacognitive monitoring, it is desirable to follow a meta-analytic approach to obtain a reference to which the results of the current study can be compared. The activation likelihood estimation (ALE) composite meta-analysis of metacognition-related activity by Vaccaro and Fleming (2018) revealed a consistent involvement of a frontoparietal network, including a cluster in the left IFG (peak coordinate in MNI: -36 28 -6; volume in mm^3 = 1432; maximum ALE value = 0.0318). The current results in children are in line with this observation. Our results also align with their meta-analysis investigating retrospective metacognitive judgments and revealing consistent activation in the left IFG (Vaccaro and Fleming, 2018). It is worth noting that both meta-analyses also revealed other significant clusters in metacognitive monitoring in adults (e.g. bilateral parahippocampal), which were not found in the current study in children.

To more quantitatively compare metacognitive monitoring related activation found in our study to those associated with monitoring in the broader, existing adult literature, we obtained the association test maps for the term 'monitoring' and the term 'judgement' from Neurosynth (www.neurosynth.org; (Yarkoni et al., 2011); accessed November 2019), a platform for automatically synthesizing the results of many different neuroimaging studies using text-mining and meta-analyses to generate mappings between neurobiological and cognitive states. Data on the term 'metacognition' were not available in Neurosynth. The meta-analytical map associated with the term 'monitoring' describes the likelihood that a region will be activated if the study contains the term 'monitoring' over and above other terms in the database including 1335 terms, 507 891 activations reported in 14 371 studies. The automated meta-analysis of 465 studies containing 'monitoring' revealed a map that contained a cluster in the left IFG (false discovery rate criterion of .01), of which the peak value was -34 24 -4. This suggests some overlap between the current result (i.e. peak value -47 30 -5, k = 75, voxel size 2.2) and the Neurosynth data for 'monitoring'. The automated meta-analysis of 290 studies containing 'judgement' also revealed a map that contained a cluster in the left IFG, which included the peak value found in the current study. Taken together, the existing meta-analytic data are thus overlapping with our results on the neurobiological basis of metacognitive monitoring using retrospective metacognitive judgments.

The current study adds to the existing literature, as we explicitly investigated the neurobiological basis of metacognitive monitoring in children of a narrow age range and in higher order cognitive processing. Research with such a specific focus is of utmost importance to functionally specify brain activation associated with metacognitive processes. This furthers our understanding of the underlying neurocognitive architecture supporting metacognitive abilities. Investigating the activation tracking the requirement for a metacognitive judgment in particular, is an essential area of research, as a detailed meta-analysis of research in this area (Vaccaro and Fleming, 2018) demonstrated a lack of studies investigating this, even in the adult population. The current study addressed that lacuna.

Because we specifically isolated the brain regions involved in metacognitive monitoring in arithmetic in children, the current study yields a unique opportunity to explore the overlap between metacognitive monitoring processes and arithmetic in children. During arithmetic, children are known to activate various parietal and frontal areas (Peters & De Smedt, 2017), a network that also includes the left IFG. Kucian et al. (2008) and Kawashima et al. (2004) also found significant activation increases during exact calculation and multiplication, respectively, in the left IFG. The current results, identifying the left IFG as the neurobiological basis for engaging in metacognitive monitoring, are in line with the frequently suggested hypothesis (e.g. Ansari et al., 2005; Arsalidou et al., 2018; Houdé et al., 2010; Kaufmann et al., 2011, 2006; Kucian et al., 2008; Menon, 2015; Rivera et al., 2005), that part of this prefrontal activation that is consistently found during arithmetic in children points to metacognitive awareness.

The exploratory brain-behaviour correlations further reveal an association between brain activity related to engaging in metacognitive monitoring and arithmetic performance: Higher activation in the left IFG while engaging in metacognitive monitoring on arithmetic performance, is associated with better arithmetic performance. This aligns with Peters et al. (2018), who found higher activation during arithmetic in the left IFG for children with better arithmetic performance. Importantly, our data are not reflective of individual differences in error making or post-error responses, as such an association would reveal a negative correlation between arithmetic performance and left IFG activation, instead of the currently found positive association. Moreover, there was no difference in arithmetic accuracy between the metacognitive and the control condition, making it unlikely that activation related to errors would be captured in the metacognitive contrast estimates.

Future research should build on our results to deepen our understanding of how the brain generates metacognitive awareness of task performance. Such studies should examine age-related changes in the neurobiological basis of metacognition in higher order processes, via comparing different age groups or by using longitudinal data. This is particularly relevant as metacognitive monitoring gradually shifts from being a more domain-dependent ability to a more domain-general process (Geurten et al., 2018). As such, an interesting avenue for future research is to study whether the current results are specific for arithmetic or whether the left IFG is generally involved in metacognitive monitoring in other domains. Additional research is also needed to investigate whether brain activation differs between different metacognitive judgments (e.g. “I think I’m correct” vs. “I think I’m incorrect”) and between correct vs. incorrect arithmetic trials, a possibility that we could not examine given the design of the current study. By building on this first empirical study on the neurobiological basis of metacognitive monitoring in children in arithmetic, subsequent studies might further clarify the role of metacognitive monitoring in arithmetic that was found in earlier behavioural research, at the level of the brain.

Conclusion

To conclude, this study is the first to reveal the neurobiological basis of metacognitive monitoring in children during an educationally relevant higher order process in the left IFG. The current design yielded a unique opportunity to explore the overlap between the neurobiological basis of metacognitive monitoring and arithmetic performance in children, as it has been frequently suggested, but was never empirically tested, that prefrontal activation during arithmetic performance pointed to control mechanisms such as metacognition. Our results are in line with this suggestion.

CHAPTER 7

General discussion & Perspectives

Chapter 7
General discussion & Perspectives

The important thing is to never stop questioning.
Curiosity has its own reason for existing.
– A. Einstein

What child's play is for some, is a Sisyphean task for others. This certainly also applies to arithmetic, a domain in which large individual differences are present, already early in life (e.g. Berch et al., 2016; Dowker, 2005, 2019b). Through the different studies within this dissertation, I aimed to contribute to the knowledge on the cognitive, metacognitive and affective correlates of arithmetic performance and development, which may eventually help to develop effective instructional approaches and might contribute to designing scientifically validated remediation programs for children at risk for or with difficulties in arithmetic.

The first two empirical studies in the current dissertation (Chapter 2 and Chapter 3) simultaneously investigated different cognitive and metacognitive processes and their associations with arithmetic. These chapters provided the groundwork for the current dissertation. Throughout these studies, metacognitive monitoring emerged as an important, unique process related to arithmetical performance and development. Because of this promising role of metacognitive monitoring in arithmetic, the subsequent chapters (Chapters 4 to 6) of this dissertation fleshed out the role of metacognitive monitoring in more detail.

Although the current dissertation tackles important gaps in the existing body of research, the interest in metacognition is not new. In line with Descartes' (1628) *"Cogito ergo sum",* many scholars contend that the ability to reflect upon our thoughts and behaviour, upon our cognition in general, constitutes the core of what makes us human. Rightfully so, countless philosophers and scientists have been fascinated by this self-reflective nature of human thought. Furthermore, metacognition has everyday relevance, not only for patients with neurological or psychiatric disorders (e.g. David et al., 2012), or the healthy aged with deficits in metamemory (Souchay et al., 2007), but also for learners and for educators who want to promote learning, for researchers interested in how the mind works (Schneider & Artelt, 2010), and for people who swear they left their keys by the front door only to find them on the kitchen table. Learning is difficult in a world that changes ceaselessly (Meyniel & Dehaene, 2017), and good metacognitive knowledge and skills are an essential part of acquiring a new skill. For example, good metacognitive monitoring enables one to discriminate between (arithmetic) errors and correct responses, and as such may improve (arithmetic) accuracy by facilitating the implementation of control (Rinne & Mazzocco, 2014). While people are capable of robust evaluation of their cognition (Yeung & Summerfield, 2012), even at a young age (e.g. primary school, Lyons & Ghetti, 2010), one's belief about one's own knowledge and/or performance is rarely entirely valid, and even often spurious. Nevertheless, people frequently act upon such beliefs and they may redirect behaviour. Bad self-judgment and/or overconfidence in one's own judgment may lead to the failure to take advantage of learning opportunities, due to, for example, not seeking more information when performance is suboptimal (Desender et al., 2018). Furthermore, bad self-judgment and/or overconfidence may have disastrous consequences in high-risk situations. Dramatic world events, such as high rates of entrepreneurial failure, global stock market crashes, the explosion of the Space Shuttle Challenger and the nuclear accident at Chernobyl have all been blamed on overconfidence (Molenberghs et al., 2016; Moore & Healy, 2008).

An important part of the scientific literature on metacognition has evaluated the role of metacognition in learning, memory, thinking, problem solving, and decision making (e.g. Metcalfe & Shimamura, 1994). Hence, an extensive body of research on the role of metacognition in different domains of performance is available that provides the indispensable foundation on which the research in this dissertation is built. The contribution of the current dissertation to this literature is the specific angle through which the role of metacognition was investigated. Firstly, I specifically focused on a subdomain of mathematics, namely arithmetic performance and development, to more carefully unravel the previously observed associations between metacognition and general mathematical performance and development. Secondly, I investigated the role of metacognition in addition to other cognitive and affective processes that are widely recognized as important for arithmetic. Thirdly, I further focused on one specific aspect of metacognition, that is the metacognitive monitoring of accuracy. To that end, I used a more experimental approach, asking children on a trial-by-trail basis to report their judgment on the accuracy of their arithmetic answer.

To summarize, even though a vast body of literature exists on, on the one hand, children's arithmetic skills and development, and on the other hand, metacognition in mathematics, the association between children's arithmetic performance and development with their metacognitive abilities is not completely understood. The overall aim of this doctoral project was thus to broaden our knowledge on the correlates of arithmetic performance and development by focusing on the role of metacognition, thereby complementing detected gaps in the literature using different methodological frameworks and investigating different cognitive, metacognitive and affective processes.

In the remainder of this general discussion, I will discuss the main findings of the current dissertation, followed by a methodological and theoretical discussion, standing still at the strengths and weaknesses of the current dissertation and providing some suggestions for future research. Next, I will further discuss some educational considerations related to the current research. This chapter ends with a general conclusion.

Integrative Summary

1 The unique roles of numerical magnitude processing, executive functions and metacognition in arithmetic performance and development in primary school children

Chapter 2 in the present dissertation provided the groundwork for all subsequent studies. This empirical study focused on simultaneously investigating numerical magnitude processing, executive functions (inhibition, shifting and updating) and metacognition, to unravel their unique roles in arithmetic performance in young primary school children (i.e. second graders). Importantly, these second graders were in the middle of an important developmental period not only for arithmetic, but also for the other cognitive and metacognitive processes under investigation.

The approach of investigating domain-specific and domain-general processes in concert is an important contribution to the existing literature on mathematics in general and arithmetic in specific, as within this body of research, domain-specific and domain-general processes have been frequently studied in isolation (e.g. Fias, 2016, for a critical discussion). A crucial and novel characteristic of the study reported in Chapter 2 is the inclusion of metacognition on top of the other cognitive processes. The consideration of metacognition is especially important, because of its close link to executive functions (see Roebers, 2017; Roebers & Feurer, 2016).

The results indicated that symbolic numerical magnitude processing was uniquely associated with arithmetic performance, even when other important processes, namely executive functions and metacognition, were considered simultaneously. As such, it was demonstrated that this association does not merely arise as a result of a common reliance of numerical magnitude processing and arithmetic on these other, domain-general processes, which has been suggested in the existing literature. Moreover, Chapter 2 established that, when all three components of executive functions were considered together, there was only evidence for the unique role of updating in explaining individual differences in arithmetic performance. This emphasises the need to include all three components of executive functions when investigating their role in primary school children. Concerning metacognition, the results demonstrated that declarative, general metacognitive knowledge was a significant predictor of addition speed, on top of numerical magnitude processing and executive functions. Although this novel result needs to be replicated, this may suggest that good metacognitive knowledge is important for arithmetic when a child is already somewhat proficient in the area. A critical finding in Chapter 2 is that metacognitive monitoring was a strong, stable and unique predictor of concurrent arithmetic performance. This study was the first to demonstrate this importance of metacognitive monitoring in arithmetic, over and above the role of numerical magnitude processing and executive functions.

In sum, the results of Chapter 2 showed that numerical magnitude processing, updating, and metacognition were all significantly and uniquely associated with aspects of arithmetic. Importantly, this study emphasized the need to include metacognition in cognitive and developmental research on mathematical cognition. Indeed, this research field has not paid a lot of attention to the study of metacognition as is evidenced by the fact that this process has scarcely been mentioned in some of its major overview works (e.g. Campbell, 2005; Cohen Kadosh & Dowker, 2015; Dowker, 2019c; David C. Geary, 1994; Gilmore et al., 2018a; Henik & Fias, 2018). However, this is in sharp contrast with the educational research on mathematical learning, as metacognition has been studied quite extensively in the domain of mathematics education (e.g. Schneider & Artelt, 2010).

Because Chapter 2 only reported on concurrent associations, we were not able to assess the stability of the found associations over development, to examine whether these processes explain individual differences in later arithmetic performance, and to investigate the predictive power of these processes for the development of arithmetic while accounting for prior arithmetic performance. Indeed, besides the need for the simultaneous investigation of numerical magnitude processing, executive functions and metacognition, an important gap in the current literature was the lack of longitudinal investigation on this topic. Such research is critical as it allows us to better understand the developmental dynamics between the investigated processes (i.e. predictors vs. consequences, the possibility of bi-directional associations and predictors of change). Therefore, in **Chapter 3**, I used a longitudinal panel design to thoroughly investigate the stability of the found associations, the longitudinal associations between the investigated processes (i.e. numerical magnitude processing, executive functions and metacognition) and arithmetic and the predictive power of these processes over and above prior arithmetic performance.

Chapter 3 established the stability of the concurrent associations of symbolic numerical magnitude processing, general metacognitive knowledge and metacognitive monitoring with different aspects of arithmetic performance. Our findings confirm the unique role of each of these processes throughout arithmetic development in early to middle primary school. Unexpectedly, no strong evidence was found for the role of executive functioning in explaining individual differences in arithmetic performance in third grade. The lack of strong associations with measures of executive functioning might be explained by the fact that arithmetic becomes a more automatized, less effortful skill, in which controlled processing, as measured by the executive functioning tasks, might not be powerful predictor of individual differences anymore.

Chapter 3 also investigated longitudinal associations to determine whether numerical magnitude processing, executive functions and metacognition each explain individual differences in later arithmetic performance. These longitudinal associations mirrored the concurrent associations found in third grade. We observed a unique association of symbolic numerical magnitude processing, general metacognitive knowledge and metacognitive monitoring in second grade, with arithmetic performance in third grade. No associations between executive functions in second grade and arithmetic in third grade were observed.

The most critical element in Chapter 3 was the inclusion of the autoregressive effect of arithmetic. After taking into account prior arithmetic performance, only moderate evidence for the predictive role of general metacognitive knowledge and anecdotal evidence for numerical magnitude processing was found. Although it needs to be acknowledged that the control for autoregressive effects is a very stringent one, no strong claims can be made about the predictive power of metacognitive monitoring for arithmetic development. Importantly, our results emphasize the crucial, robust predictive role of prior arithmetic performance for later performance and development. This emphasises a major issue in the existing mathematical cognition literature. Many longitudinal studies in the field of arithmetic have failed to include prior arithmetic performance as an important predictor of individual differences in mathematical performance in their models (e.g. Sasanguie et al., 2012; Vanbinst, Ghesquière, et al., 2015). Hence these studies do not investigate the importance of (meta)cognitive processes relative to prior arithmetic performance. Including prior arithmetic performance yields the possibility to examine the predictive power of the investigated (meta)cognitive processes for development in arithmetic, by ensuring that the concurrent correlations between these (meta)cognitive processes and arithmetic do not confound the investigated predictive association. Hence, the current results urge for the inclusion of prior performance in future longitudinal studies on mathematical cognition.

Chapter 2 and Chapter 3 significantly added to our understanding of the developmental dynamics between several cognitive and metacognitive processes and arithmetic in a crucial developmental period in primary school (i.e. second to third grade). Building on our results, further research on these developmental dynamics earlier in development (e.g. first grade) is very promising. Such an investigation could provide us with a deeper understanding of these dynamics in children at the start of their arithmetic learning process.

2 The interplay of metacognitive monitoring and mathematics anxiety in arithmetic in primary school children

Learning arithmetic involves a complex interplay of diverse processes, including metacognitive and affective processes. When investigating the association between metacognitive monitoring and arithmetic, their interplay with mathematics anxiety is of particular interest. This is because, on the one hand, both metacognitive monitoring and mathematics anxiety have been shown to be associated with arithmetic performance and development. On the other hand, metacognitive monitoring as well as mathematics anxiety are processes that are linked to thinking about your performance. **Chapter 4** therefore focused on the interplay of metacognitive monitoring and mathematics anxiety in arithmetic.

Specifically, I aimed to investigate whether metacognitive monitoring and mathematics anxiety were indeed associated in primary school children and whether their respective associations with arithmetic were influenced by this interrelation. I also investigated the role of arithmetic performance itself within this interplay. To reach this goal, a longitudinal panel design was applied, using correlations and multiple regression analysis, as well as moderation and mediation models. Mediation analyses enabled us to investigate potential mechanisms by which certain effects operated, while moderation analyses enabled us to investigate when, i.e. at which level of another process, an effect occurred between two processes.

The findings of our preregistered analyses confirmed the concurrent and longitudinal associations between, on the one hand, arithmetic and metacognitive monitoring, and, on the other hand, arithmetic and mathematics anxiety. It should be noted that our results mainly showed associations of arithmetic with mathematics anxiety in third grade, and much less with mathematics anxiety in second grade. The same pattern of results was observed for the association between metacognitive monitoring and mathematics anxiety: The findings revealed that metacognitive monitoring and mathematics anxiety were significantly correlated concurrently and that this association was more pronounced in third grade than in second grade. These results might suggest an increasing importance of mathematics anxiety in the development of primary school children at different levels of performance, i.e. academic as well as metacognitive. While this pattern of results needs to be further investigated in future research, these results urge to tackle (signs of) mathematics anxiety already early in primary school.

The current longitudinal study provided further insights into the developmental dynamics of the association between metacognitive monitoring and mathematics anxiety: Earlier mathematics anxiety was associated with later metacognitive monitoring but not vice versa. This suggests that mathematics anxiety might hinder one's ability to correctly monitor one's performance, rather than that metacognitive awareness of performance leads to mathematics anxiety. However, no conclusive evidence was found for a predictive effect of mathematics anxiety for metacognitive monitoring once prior metacognitive monitoring was taken into account. The results further suggest that the predictive power of mathematics anxiety for metacognitive monitoring was mediated through arithmetic performance, which emphasises the pivotal role of arithmetic performance in the development of metacognitive monitoring and mathematics anxiety.

This pivotal role of performance was also gleaned from the observation of a strong predictive role of arithmetic for both metacognitive monitoring and mathematics anxiety, on top of their respective autoregressors. Furthermore, mediation analyses demonstrated that this predictive role was a direct effect, without mediation or moderation by metacognitive monitoring or mathematics anxiety. These results are thus in line with the deficit model of mathematics anxiety (Carey et al., 2016), stating that poor arithmetic performance leads to higher mathematics anxiety in the future.

When prior arithmetic performance was considered, neither metacognitive monitoring nor mathematics anxiety predicted later arithmetic performance. This might be explained by the extremely strong autoregressive effect of arithmetic, which makes explaining additional variance very difficult. Another explanation is that the effects of metacognitive monitoring and mathematics anxiety are inherently intertwined with the measure of arithmetic, as both metacognitive monitoring and mathematics anxiety are processes that co-occur with performance. These two hypotheses are discussed in more detail section below (See Section 4 of the General Considerations).

As was documented in the discussion to Chapter 3, controlling for the autoregressive effect of prior arithmetic performance was a very stringent control. Therefore, Chapter 4 additionally examined the longitudinal associations between metacognitive monitoring, mathematics anxiety and arithmetic without the inclusion of the autoregressor. On the one hand, the predictive power of metacognitive monitoring for later arithmetic performance was found to be a direct effect, without mediation or moderation by mathematics anxiety. On the other hand, the predictive power of mathematics anxiety for later arithmetic performance was an indirect effect, mediated – but not moderated – by metacognitive monitoring. This may suggest that mathematics anxiety is only predictive of later arithmetic performance through metacognitive awareness. Taken together, both of these findings emphasize the importance of metacognitive monitoring for later arithmetic. However, cautious interpretation of these findings is needed, because these results were only found when prior arithmetic performance was not considered.

A further interesting investigation of the interrelations between metacognitive monitoring and mathematics anxiety could be to examine the association between different categories of metacognitive judgements and mathematics anxiety. In the current study, no distinction in the performance measure of metacognitive monitoring was made between a correct metacognitive judgment in the context of a correct arithmetic response (i.e. correct arithmetic answer was and child indicated “I think my answer was correct”) versus in the context of an arithmetic error (i.e. arithmetic answer was incorrect and child indicated “I think my answer was incorrect”). Both situations resulted in a high metacognitive monitoring score. A distinction between correct metacognitive judgments depending on the accuracy of the arithmetic answer may potentially unravel other interrelations between metacognitive monitoring and mathematics anxiety. It is plausible that children who are frequently correct and know they are correct are less math anxious than children who frequently make arithmetic errors and realize this. To thoroughly investigate this, a careful design is needed in which there is a substantial number of observations in each category (e.g. both correct and erroneous responses; different metacognitive judgments), which may be realized by adaptively adjusting difficulty level of the arithmetic items.

In light of the overarching topic of the current dissertation, a crucial finding in Chapter 4 is that the association that was found between arithmetic and metacognitive monitoring is unique, and that it cannot be explained by mathematics anxiety. The results suggest that the longitudinal association between arithmetic and metacognitive monitoring is mostly driven by the influence of arithmetic on metacognitive monitoring rather than the other way around. This hypothesis should be addressed in further research that bypasses the abovementioned drawbacks of autoregressive effects, for example, by using training studies, as will be discussed below (Section 4 of the General Considerations).

3 Metacognitive monitoring across domains

Chapter 5 specifically addressed the question on the domain-specificity versus domain-generality of metacognitive monitoring. Metacognitive monitoring has been investigated in and associated with a wide range of cognitive domains, but the extent to which metacognitive monitoring is domain-general, is debated (e.g. Schraw et al., 1995). In other words, it is unclear whether metacognitive monitoring in a certain domain is reliant on domain-specific components or whether it reflects a general performance monitoring process that is recruited to evaluate performance on a variety of tasks.

In Chapter 5, two studies in primary school children were described to thoroughly investigate the domain-specificity versus domain-generality hypothesis. Two different age groups were selected because it has been suggested that in primary school there is a shift from domain-specificity to domain-generality occurring between the age of 8 and 13 years (Geurten et al., 2018). One group was specifically chosen just under this age-range, i.e. a 7-8-year-old group, for which children's metacognitive monitoring was expected to be domain-specific. The other age group, i.e. 8-9-year-olds, was exactly at the beginning of the age range for which domain-generality is assumed to start emerging. The fact that these two groups only differed in one grade is an important strength of the design. This allowed us to use the same arithmetic, spelling and metacognitive monitoring paradigm, to maximize comparability between age groups. On the other hand, the additional use of standardized academic performance tasks ensured that the results on domain-generality of metacognitive monitoring could not be merely explained by similarities in paradigms to measure academic performance.

Another critical strength of the two studies in Chapter 5 is that these included different but highly related academic domains, i.e. arithmetic and spelling. A stringent test of the possibility of domain-specificity requires the examination of metacognitive monitoring in related domains. Indeed, investigating this in distant domains (e.g. Vo et al., 2014) would, by design, increase the likelihood of revealing evidence for such domain-specificity. When investigating the question of domain-specificity in highly related domains, it is, however, of the utmost importance to remove variation attributable to performance scores, because associations between performance on the cognitive tasks can confound the results on the domain-specificity or domain-generality of metacognitive monitoring in those domains (Schraw et al., 1995). Therefore, another crucial strength of Chapter 5 is that performance on the cognitive tasks was controlled for in our models.

Specifically, Chapter 5 investigated in two age groups whether metacognitive monitoring was associated with performance in both domains. Critically, it was examined whether metacognitive monitoring in one domain correlated with metacognitive monitoring in the other domain as well as whether monitoring in one domain was predictive of performance in the other domain, and vice versa. The findings indicate that within-domain metacognitive monitoring was indeed associated with arithmetic and spelling performance at both ages. Taking into account the stringent empirical test of domain-specificity, the results showed domain-specificity of metacognitive monitoring in the youngest age group and aspects of domain-specificity in the oldest group. However, in the oldest group, the critical findings were that metacognitive monitoring measures in different domains were predictive of each other and monitoring in one domain predicated performance in the other and vice versa, even after taking into account performance. Taken together, these findings thus suggest the emergence of more domain-general metacognitive monitoring processes between the ages of 7 and 9, in line with what has been proposed by Geurten et al. (2018). The findings of Chapter 5 significantly added to our understanding of the development of metacognitive monitoring in primary school, which can benefit the refinement of theoretical frameworks of the gradual shift from domain-specificity of metacognitive monitoring towards more domain-generality of metacognitive monitoring (see Section 1.4 of the general considerations).

4 A developmental cognitive neuroscience perspective on metacognitive monitoring during arithmetic in children

Chapter 6 in the present dissertation included behavioural as well as brain-imaging measures, i.e. using fMRI, to examine the role of metacognitive monitoring in arithmetic at the neurobiological level. Based on the existing brain imaging literature on both arithmetic (e.g. Peters & De Smedt, 2017) and metacognitive monitoring (e.g. Fleming & Dolan, 2014), investigating this was particularly interesting because of the overlap in brain networks that are associated with both processes. Research on the neurobiological basis of arithmetic has systematically observed increases in activity in prefrontal regions during arithmetic (e.g. Arsalidou et al., 2018; Peters & De Smedt, 2017). On the other hand, metacognition is considered a higher-order brain function that strongly depends on the prefrontal cortex (e.g., Shimamura, 2000). Research that examined the overlap in brain activity between arithmetic and metacognition in the prefrontal cortex was, however, lacking. Furthermore, no studies on the neurobiological basis of metacognitive monitoring were available in children. In view of the massive changes in brain structure and function throughout childhood, the mere generalization of adult findings to developmental populations could be problematic (e.g. De Smedt, 2018).

The study presented in Chapter 6 tackled these issues by, firstly, examining the neurobiological basis of metacognitive monitoring during arithmetic in children. Secondly, the association between brain activation during metacognitive monitoring, as measured via fMRI, and arithmetic outside the scanner was investigated. It is important to note that, to enhance ecological validity, the paradigm that was administered in the MRI scanner explicitly addressed arithmetical skills that are formally taught in school. Because measures of performance in the scanner are obtained in a very strict and unnatural setting, I deliberately chose to correlate brain activity during metacognitive monitoring with an ecologically valid, widely used standardized measure of children’s arithmetic performance in the classroom (i.e. TTA; see De Smedt et al., 2010, for a discussion). We therefore specifically recruited children who participated in the studies in Chapter 2 and Chapter 3. As a result, we were able to combine the longitudinal behavioural data on arithmetic performance and development over a three-year period with the brain imaging data in Chapter 5. Another way to enhance the ecological validity that we used in this neuro-imaging study was that, for the measurement of metacognitive monitoring, we used exactly the same method of assessment that was used throughout the behavioural studies presented in this doctoral dissertation.

A critical challenge in fMRI studies is to develop an appropriate control condition to isolate the specific process under investigation. To do so for metacognitive monitoring, a stringent control condition, i.e. performing the same arithmetic task with a judgment on colour instead of a metacognitive judgment, was used. As a result, the only difference between the experimental and the control condition was in metacognitive monitoring. Our findings demonstrated that metacognitive monitoring in children is reflected in brain activity in the left inferior frontal gyrus (IFG). The observed region of activation is in line with the expectations for the neurobiological basis of metacognitive monitoring, based on the adult literature (Vaccaro & Fleming, 2018). Activity in the left IFG has also been consistently found during arithmetic (Peters & De Smedt, 2017). The results thus provided evidence in line with the hypothesis that part of the prefrontal activation during arithmetic in children might reflect metacognitive monitoring processes. Going one step further, correlations between our unique set of longitudinal behavioural data and brain imaging data further pointed toward an association between brain activity related to engaging in metacognitive monitoring and arithmetic performance.

fMRI measurements require participants to lie very still in a noisy environment, which is quite different from what happens in the classroom. By combining these neuro-imaging data with the more ecologically valid longitudinal behavioural data, we furthered our understanding of the neurobiological processes that play a role when children learn school-relevant skills. Neuro-imaging data on metacognitive monitoring can add to our understanding of what happens when children engage in metacognitive monitoring processes. Especially in children, this way of measuring metacognition has been neglected. Yet, this is a promising avenue for further research as these neurobiological measures expand the methodological toolbox of behavioural researchers (e.g. de Smedt & Verschaffel, 2010, for a critical overview). Enhancing our understanding of how the brain works can further our understanding of arithmetic performance and development and the processes associated with individual differences herein.

General theoretical and methodological considerations

When contemplating the current dissertation and its results, it is important to critically reflect upon its strengths and limitations, which are theoretical as well as methodological in nature. Within this doctoral dissertation, several cognitive, metacognitive and affective processes were examined in concert. Throughout these studies, metacognitive monitoring was found to be an important, unique process related to arithmetic, hence, its role was studied in more detail over several studies. As a result, the current dissertation provided further insight into the interplay of different correlates of arithmetic and presented novel findings on the role of metacognition in arithmetic in primary school children, investigated with both behavioural and neuro-imaging methods.

Against this background, I will start with some critical considerations regarding the study of metacognition. Secondly, I will discuss the interrelations between metacognition and executive functions. Thirdly, I will elaborate on measurement challenges, such as the issue of task impurity, related to the study of the various cognitive, affective and metacognitive processes within this dissertation. Fourthly, I will review important considerations with regard to our correlational, individual differences approach, including the directionality of the investigated associations. Finally, I will discuss the complexity of the investigated processes themselves and the studied interrelations between cognitive, metacognitive and affective processes.

1 A meta-perspective on metacognition

Across the studies within this dissertation, the role of metacognition was investigated in concert with different cognitive and affective processes, at different ages, using both measures of declarative as well as procedural metacognition, and within both behavioural and neuro-imaging frameworks. In this section, I discuss several theoretical and methodological issues related to our measure of metacognition and to our results on the role of metacognition in arithmetic.

1.1 Focused investigation of a broad construct

Metacognition is a very broad concept, encompassing many different aspects that differ substantially (Flavell, 1979). There is a long tradition of research on metacognition in general and metacognition in mathematics (education) in particular, and, as a result, there exists an extensive body of research on this topic. The current dissertation contributes to this body of research by focusing on particular aspects of metacognition in a particular subdomain of mathematics.

Specifically, this doctoral dissertation included two aspects of metacognition, namely metacognitive monitoring of accuracy, and declarative metacognitive knowledge, both central aspects of metacognition. This focused investigation enabled us to functionally specify the association between mathematics and metacognition and as such build up a more profound understanding of critical components of metacognition for mathematics. One advantage of simultaneously investigating different aspects of metacognition is that it allows to examine their differential effects on, in this dissertation, arithmetic performance. On the one hand, metacognitive monitoring of accuracy was found to be a stable, unique concurrent correlate of arithmetic performance, and predictive of later performance, but not of development. On the other hand, declarative domain-general metacognitive knowledge was found to be a stronger longitudinal predictor of performance and development of specific aspects of arithmetic and not so much a concurrent correlate. It should also be noted that our measures of metacognitive monitoring and declarative metacognitive knowledge were not significantly correlated. This is in line with the abovementioned argument that different aspects of metacognition, while all related to 'thinking about your thinking', differ substantially.

It is important to acknowledge that our specific operationalisation of metacognition is limited, as it does not capture the richness of the many aspects of metacognition. For example, one aspect of metacognition that I did not investigate are metacognitive control processes. These are defined as the individual's executive activities enabling the use and adaptation of different cognitive operations with the aim to increase learning behavior or test performance (Roebers et al., 2014) and include, for example, allocation of study time or correction of errors. Because of the intertwined nature of metacognitive monitoring and metacognitive control (Nelson & Narens, 1990), it would be interesting for future studies to also include measures of metacognitive control. Individual differences in metacognitive control could play an important role in arithmetic, and in the interrelation between arithmetic and metacognitive monitoring. As stated by (Begg et al., 1992), p. 195) "*Knowing an item is inadequate is of little value in the absence of skills that will remediate the inadequacy.*", indicating that monitoring alone may not be sufficient for good performance. It may thus be especially conducive for arithmetic performance and development to not only have good metacognitive monitoring skills, as was investigated in the current dissertation, but to have a combination of good metacognitive monitoring skills and good metacognitive control skills, and as such having good metacognitive *regulation.* This metacognitive regulation may provide a solid basis for performance enhancement: "*A system that monitors itself (even imperfectly) may use its own introspections as input to alter the system's behavior*" (Nelson & Narens, 1990), p. 128). Future research should investigate this hypothesis including both measures of metacognitive monitoring and metacognitive control.

1.2 How to measure metacognition?

There is substantial heterogeneity in the approaches that studies use to measure metacognition, yet they all share the fact that they elicit subjective beliefs about cognition (Fleming & Dolan, 2012). Many methods for measuring metacognition have been used, such as questionnaires (e.g. Haberkorn et al., 2014), thinking-aloud protocols (e.g. van der Stel & Veenman, 2010), observations (e.g. Veenman & Spaans, 2005) or on-task trial-by-trial metacognitive judgments (e.g. Rinne & Mazzocco, 2014). Within this dissertation, I used two different measures of metacognition: A validated questionnaire for children's declarative metacognitive knowledge (Haberkorn et al., 2014) and a trial-by-trial metacognitive monitoring protocol (e.g. Rinne & Mazzocco, 2014). Because of its central role within this dissertation, I will first discuss this metacognitive monitoring paradigm in detail, commenting on both its strengths and weaknesses.

The metacognitive monitoring paradigm used in this dissertation was embedded in a custom-made, computerized arithmetic task. Hence, it measured metacognitive monitoring applied to the arithmetic domain, which accentuates the importance of the specific items on which metacognitive monitoring is measured. Within the studies reported on in Chapter 2 to 5, this arithmetic task included single-digit addition and multiplication items. The same metacognitive paradigm was also used in our neuro-imaging study (Chapter 6), embedded in a multiplication task that included not only single-digit, but also multi-digit items. As this study encompassed the last wave of our longitudinal follow up, children's arithmetic skills had already substantially further developed. Therefore, this modification of the difficulty of the arithmetic task was introduced to ensure considerable inter- and intra-individual variability in arithmetic performance and metacognitive monitoring of accuracy. The research reported in Chapter 4 additionally used the same metacognitive monitoring paradigm embedded in a spelling task. In sum, all items used in the tasks in which our metacognitive monitoring paradigm was embedded, were age-appropriate and educationally relevant, as they were important topics of instruction at the age of our participants. This correspondence of the items on which metacognitive monitoring was measured with what is taught in school enhanced the practical and educational relevance of our findings and, as such, is an important strength of the current approach.

A relevant distinction that is made in the existing literature on metacognitive monitoring is that of granularity of the judgment: Metacognitive judgements can occur both on a global (i.e. across multiple items of a task) or on a local (i.e. specific on one item) scale (e.g. Pieschl, 2009; Schraw, 1994). Our measure of metacognitive monitoring measure consisted of trial-by-trial reports of children's judgement of the accuracy of their arithmetic answer. These trial-by-trial measures thus provided us with a precise, local measure of children's metacognitive monitoring. As a result, it was possible to verify at the item level whether or not children were aware of when they were correct and when they made an arithmetic error. Additionally, in contrast to most global measures of metacognitive monitoring, our measure of monitoring consisted of metacognitive judgments on a multitude of items. This provided a more nuanced measure of monitoring compared to one global metacognitive judgment.

Based on these trial-by-trial reports, the performance measure of metacognitive monitoring was the alignment between the child's metacognitive judgement of accuracy and the actual accuracy of the arithmetic answer. Investigating trial-by-trial correspondence between accuracy and metacognitive judgment on accuracy is a widely used measure of metacognitive monitoring (e.g. Fleming & Lau, 2014), yet the way in which this correspondence is operationalised, differs substantially between studies. In all studies within the current dissertation, it was operationalised as follows: Correct academic answers yielded the highest metacognitive score if children said they were correct, and the lowest metacognitive score if they said they were incorrect. This scale was reversed when the academic answer was incorrect. As such our operationalisation focused on children's metacognitive accuracy regardless of the accuracy of their academic answer itself. It is important to note that this measure is not the same as a measure of confidence. Naturally, metacognitive monitoring and confidence are closely linked, because confidence is involved when children make metacognitive judgments. For example, when children have low confidence in their arithmetic performance, they will likely indicate that they think their arithmetic answer is wrong. However, what is measured by the concepts of metacognitive accuracy and of confidences critically differs. Our measure of metacognitive monitoring, which is a measure of metacognitive accuracy, encompassed two aspects. On the one hand, actual performance on an academic item and, on the other hand, the child's judgment of the accuracy of that performance. The correspondence between these two defined the metacognitive monitoring score that was consistently used in this doctoral dissertation. Hence, the way in which I operationalised metacognitive monitoring critically differs from plain measures of confidence, in which no correspondence with actual performance is made.

It needs to be emphasized that other operationalisations of the correspondence between one's metacognitive judgment and one's accuracy on a task have been reported in the existing literature. These include the statistical correlation between accuracy and confidence over trials (e.g. Pearson's *r*) or measures based on signal detection theory (e.g. type 2 d') and receiver operating characteristics (ROC) analysis (see Fleming & Lau, 2014, for a critical overview). All these measures differ in how accurately they reflect a person's metacognitive monitoring skills. A distinction that is often made in this context is the difference between 'metacognitive bias' and 'metacognitive sensitivity' (e.g. Fleming & Lau, 2014). Metacognitive bias is the overall level of confidence expressed, independent of whether the answer is correct or incorrect, or in other words, the difference in subjective confidence despite constant task performance. Metacognitive sensitivity or metacognitive accuracy is the extent to which a metacognitive judgment discriminates between correctly and incorrectly solved trials. In this dissertation, I have focused on metacognitive sensitivity as our measure of metacognitive monitoring.

It is important to note that metacognitive sensitivity is often affected by task performance itself. One will have higher metacognitive accuracy in an easy task compared to a hard task (Fleming & Lau, 2014). On the other hand, most measures of metacognitive sensitivity may be subject to metacognitive bias. For example, it is known that children have the tendency to indicate high confidence in their performance, and as such tend to be overconfident (Destan & Roebers, 2015). Yet, theoretically, these aspects of metacognitive monitoring are independent: A child with high overall confidence (i.e. metacognitive bias), may still be sensitive to trial-by-trial variation in performance (i.e. high metacognitive accuracy). It is important to consider the effects of task performance and metacognitive bias in light of our results on metacognitive accuracy. For example, although age-appropriate items were used and substantial inter-individual variability in arithmetic and metacognitive monitoring performance was found, the average academic performance in our studies was fairly high in the tasks in which metacognitive monitoring was measured (i.e. range of average academic performance = [0.70-0.97]). As children tend to be overconfident (Destan & Roebers, 2015), combined with high accuracy rate, this may result in an overestimation of children's metacognitive monitoring skills. Yet, several results in the current studies demonstrate that high task performance or metacognitive bias did not drive our results on the association between arithmetic and metacognitive monitoring. Firstly, when the metacognitive monitoring data were analysed based on the items with incorrect arithmetic answers only, the interpretation of the results did not change (e.g. Supplementary materials Chapter 2). Secondly, the interpretation of our results did not change when performance measures of metacognitive monitoring were based on a subset of items for which academic task performance was lower (e.g. general discussion section Chapter 5). Thirdly, when task difficulty was increased, for example by including double-digit arithmetic items in Chapter 6, our results on the association between metacognitive monitoring and arithmetic remained consistent.

The second measure of metacognition that was used within this dissertation, albeit less prominently, was the general metacognitive knowledge questionnaire. I deliberately chose a validated questionnaire for primary school children that did not include questions on mathematics or arithmetic. Firstly, this allowed to preliminarily explore the domain-specificity hypothesis of metacognition by comparing a metacognitive measure within the arithmetic task versus a metacognitive measure without mathematical content (see below for a critical discussion). Because this measure did not tap into domain-specific mathematical knowledge, it allows to investigate the role of declarative metacognitive knowledge without mathematical confounds. While such independent investigation is certainly important, it is critical to note that such an approach does not allow one to examine the plausibly important role of declarative metacognitive knowledge related to the mathematical domain. Secondly, it is important to acknowledge that the declarative metacognitive knowledge questionnaire that was used, predominantly focused on the strategy component of declarative metacognitive knowledge, with less focus on the person and task categories of metacognitive knowledge (Flavell, 1979). Although a focus on metacognition on strategy is in line with the emphasize on strategy knowledge in the existing literature (e.g. Carr et al., 1994), it is important to keep this in mind when interpreting the results on declarative metacognitive knowledge.

1.3 *The complex association between metacognition and cognition*

Many researchers claimed that metacognition enhances learning (e.g. Dehaene, 2020; Flavell, 1979; Rinne & Mazzocco, 2014; Schoenfeld, 1992; Veenman et al., 2006; Wang et al., 1990). In line with this idea, Kuhn (2000) and Lyons and Ghetti (2010) contended that metacognition, as a higher-order process, regulates and drives development in cognitive domains. On the other hand, Begg (1992), concluded based on two experiments in university students, that metacognitive monitoring had no value for later performance, thus suggesting it was epiphenomenal. Already in 1978, Ann Brown questioned whether metacognition was just an epiphenomenon (Brown, 1978), having no causal influence on cognition. This debate illustrates the complex relation between metacognition and cognition. Metacognition draws on cognition and as such it is very hard to have adequate metacognitive knowledge of and skills in a domain without substantial domain-specific knowledge (Veenman et al., 2006). The question one can ask is then: Is good metacognitive monitoring just merely a reflection of arithmetic accuracy or, else, does good metacognitive monitoring in fact contribute to arithmetic accuracy? (Rinne & Mazzocco, 2014).

Although Brown put forward the question whether metacognition is epiphenomenal, she was convinced otherwise and stated that metacognition is not just cognition itself, but that the concept reflects a real change of emphasis towards cognition about one's own cognitions, which has value in itself. In line with Brown's idea, Kuhn (2000) argued that metacognition does not appear abruptly as an epiphenomenon in relation to first-order cognition, but emerges early in life, and follows an extended developmental course during which it becomes more explicit and effective. In the existing literature, it has been amply indicated that metacognition is not just epiphenomenal. Examples of this include that metacognitive monitoring is critical for providing input to metacognitive control processes (e.g. allocation of study time; Nelson & Narens, 1990), metacognitive monitoring prepares individuals to make effective use of feedback (e.g. Butler et al., 2008) and metacognitive knowledge provides learners with adequate learning strategies, such as rehearsal of newly learned materials (Flavell, 1979). In line with the idea that metacognition is not purely epiphenomenal, several findings in the current dissertation pointed to the importance of metacognition for arithmetic. A robust, uniquely predictive power of both declarative metacognitive knowledge and metacognitive monitoring for arithmetic performance was found on top of other important processes for arithmetic, such as numerical magnitude processing, mathematics anxiety and intellectual ability. Furthermore, declarative metacognitive knowledge uniquely predicted development in arithmetic performance.

The complex relation between metacognition and cognition is not only illustrated in the debate on whether or not metacognition is an epiphenomenon, but also in the question how people know what they know. Numerous scholars have theorized on and empirically investigated this topic (e.g. Koriat, 2007, and Kuhn, 2001, for critical overviews), thereby discussing the intertwined nature of cognition and metacognition.

One important facet of this debate is the question of whether metacognition and cognitive performance draw upon the same information. Examples of two different accounts regarding this question are the direct-access view and the cue-utilization approach (e.g. Busey et al., 2000; Fleming & Dolan, 2012; Koriat, 2007; Koriat & Levy-sadot, 1999). In the direct access view, metacognitive judgments are based upon a direct survey of, for example, memory contents, whereas in the cue-utilization approach various mnemonic cues, such as retrieval fluency, are used (e.g. Fleming & Dolan, 2012; Koriat, 1997). In the current body of literature, most empirical evidence is found for the cue-utilization approach (see Koriat, 2007, for an overview), indicating that metacognitive judgments are not solely based on the same information as first-level task performance.

What then underlies metacognitive judgments, making them separable from cognitive judgments? Busey et al. (2000) stated that, for example, retrospective metacognitive judgements, are not only based on the information that determines accuracy, i.e. on the basis of a direct access to information in memory, but that they are also formed through the analytic consideration of aspects of the study and test conditions. This is in line with the cue-utilization view, which argues that metacognitive judgments are inferential in origin. They are based on a variety of cues and heuristics (Benjamin & Bjork, 1996). Hence, the accuracy of metacognitive judgments depends on the validity of these cues. A common distinction that is made in the cue-utilization view is the distinction between experience-based and information-based (or theory-based) judgements (e.g. Koriat, 2007; Schneider, 2015a). Experience-based judgements are based on fast and automatic inferences made from a variety of cues that reside from immediate feedback from the task, such as familiarity of the task and the ease of processing. On the other hand, information-based metacognitive judgments are based on conscious and deliberate inferences. In information-based judgements, various pieces of information retrieved from memory are consulted and weighted in order to reach an advised judgment. These may include perceived difficulty of the task, the conditions of learning of the information in the task and perceived self-efficacy in the domain of the task. In sum, metacognitive judgments are thus affected conjointly by the content of declarative information retrieved from long-term memory and by, for example, the ease with which an answer is produced.

In light of the results of the current dissertation, this cue-utilization approach may provide insight into the improvements of metacognition over development and the developmental shift from domain-specificity of metacognition towards more domain-generality. This development may be related to improvements of metacognitive abilities that enable a shift from preferential reliance on automatic inferences to a more frequent reliance on conscious and deliberate information-based processes. It may thus be interesting to train children to rely more on effective information-based judgments to enhance their metacognitive monitoring (see Koriat & Bjork, 2006, for a similar idea in the context of judgments of learning).

Although metacognitive accuracy and accuracy in a cognitive task may be based on partially overlapping information (see Busey et al., 2000, for a review), and are often positively correlated (e.g. Rinne & Mazzocco, 2014), dissociations between, for example, judgments on cognitive performance and cognitive performance itself, have been consistently found in the existing literature. For instance, people often report high confidence in incorrectly solved items or in memories that never happened (e.g. Roediger & McDermott, 1995), indicating metacognition and cognition are dissociable constructs, that do not necessarily rely on the same information.

In the discussion on the separability of cognition and metacognition, brain imaging research proved to be an important source of evidence, as it can compare neurobiological substrates of both processes. Such studies indeed indicated that performance on a cognitive task can be distinguished from metacognitive monitoring in that task (e.g. Chua et al., 2004; Fleming & Dolan, 2014). This was specifically found for retrospective confidence judgements (e.g. Chua et al., 2006; Fleming & Dolan, 2014), which is how metacognitive monitoring was operationalised in the current dissertation. For example, fMRI studies comparing brain activity during task performance versus retrospective metacognitive judgements found important brain activation differences between task performance and metacognitive monitoring (e.g. Chua, Schacter, Rand-Giovannetti, & Sperling, 2006; Chua, Schacter, & Sperling, 2009). Lesion studies have also confirmed the separability between task performance and metacognitive performance by showing, for example, that patients with parietal lesions may have impairments in retrospective metacognitive performance despite little or no impairment in accompanied task performance (e.g. Berryhill, 2012; Davidson et al., 2008; Simons, Peers, Mazuz, Berryhill, & Olson, 2010).

1.4 Metacognition: Domain-specificity in domain-generality?

Metacognition has commonly been regarded as a domain-general process that is associated to cognitive performance and learning in various domains (e.g. Efklides & Misailidi, 2010). However, especially in young children, domain-specificity of metacognition has been emphasized (e.g. Kelemen et al., 2000; Vo et al., 2014). Several studies within this dissertation dealt in more or less concrete ways with the question of the domain-specificity versus domain-generality of metacognition.

Firstly, the studies reported in Chapter 2 and 3 provided some first evidence for domain-specificity. These findings demonstrated that within-domain metacognitive monitoring was a consistent, unique correlate of concurrent arithmetic performance, while domain-general declarative metacognition was not. Strong claims on domain-specificity, however, cannot be made based on these results. That is because these measures of metacognition did not only differ on their domain-specificity, but also on the metacognitive aspect they measured (i.e. metacognitive knowledge versus procedural metacognition), and there were large operationalisation differences between the tasks (e.g. within task measurement versus task-independent questionnaire). Secondly, in contrast to the more incidental role of domain-general declarative metacognition in concurrent arithmetic, a more prominent role of this declarative metacognition was found in arithmetic development. This may have reflected a gradual shift towards domain-generality of metacognition over development (e.g. Geurten et al., 2018).

While providing some indications on the domain-specificity versus domain-generality of metacognition, the design of the studies in Chapters 2 and 3 was not suitable to thoroughly investigate this issue. Therefore, Chapter 5 deliberately investigated this question of domain-specificity by contrasting metacognitive monitoring in two different academic domains. Using this more careful design, evidence was found for the emergence of more domain-generality of metacognitive monitoring between second and third grade, although evidence was also found for the continuing importance of domain-specific components. It needs to be acknowledged that this study only focused on the metacognitive monitoring aspect of metacognition and not on declarative metacognitive knowledge. One possibility to further our knowledge on the domain-specificity versus domain-generality of metacognition would be to investigate this question by focusing on declarative metacognition. By comparing domain-specific declarative metacognition measures, e.g. declarative metacognition questionnaires on mathematics versus on spelling, one could unravel whether the gradual transition from domain-specificity towards more domain-generality in declarative metacognitive knowledge can also be found.

Brain imaging studies might further add to our answer to the question of domain-specificity vs. domain-generality with a useful additional analytical toolbox. More specifically, it remains unclear whether a shift from domain-specificity to domain-generality of metacognitive monitoring in different domains relies on shift from distinct neural recourses to a shared resource, e.g. the prefrontal cortex, that supports metacognitive monitoring across domains. One way to investigate this question, building on the findings of Chapter 6, is to examine whether the left IFG is specific for metacognitive monitoring in arithmetic or also supports metacognitive monitoring in other domains, such as spelling. An even more advanced way to investigate this could be through multivariate pattern analyses (MVPA) in which the similarity or dissimilarity of neural activation patterns elicited by monitoring in arithmetic versus monitoring in spelling can be inspected (e.g. Martens et al., 2018, for an example in the domain of perception). This can be done by training a model to distinguish between the brain activity patterns of tasks (monitoring in arithmetic vs. monitoring in spelling) based on a subgroup of the participants and subsequently test if the model then can accurately distinguish between the brain activity patterns in the two tasks of the remaining participants. These analyses allow one to examine whether the brain activation patterns elicited in different tasks are distinct or rather similar. This question could then be investigated at different ages, in view of the evidence for a developmental shift that was found in Chapter 5. As such, neuro-imaging research can provide a complementary perspective to the behavioural data on this topic, by uncovering whether neurobiological signatures of metacognitive monitoring across tasks are common or distinct.

2 Metacognition and executive functions: The same, only different?

This dissertation included both metacognition and executive functions as processes of investigation. This is in line with the recently increasing interest to connect both processes theoretically and empirically (see Roebers, 2017, for an extensive review). It is important to note that the aim of our simultaneous investigating was not to make strong claims on the interrelation of both concepts in itself, but to investigate their unique role in arithmetic performance and development in addition to each other. This is particularly interesting because of the large similarities between these two processes, as both metacognition and executive functions are higher-order, control processes related to the regulation of behaviour. In the existing literature, the definitions of these processes also overlap substantially. For example, executive functions have been described as "*Skills required to monitor and control thought and action"* (Cragg & Gilmore, 2014, p. 63). Lee et al. (2013, p. 1933) describe executive functions as "*Executive functioning encompasses a large range of top-down control and monitoring processes*". Very similar definitions are used for procedural metacognition, as is shown in the following two examples: "*Metacognition research is focused on […] how people monitor and control their cognition on-task*" (Bryce et al., 2015, p. 182); "*Metacognition […] includes executive skills related to monitoring and self-regulation of one's own cognitive activities*" (Schneider & Lockl, 2008, p. 391). Besides overlapping definitions that can be found in the literature, metacognition and executive functions also follow a similar developmental trajectory and are associated with activation in anatomically similar brain regions, such as the dorsal-lateral prefrontal cortex (e.g. Roebers, 2017; Roebers & Feurer, 2016).

Based on the overlap between both concepts, empirical studies including both concepts and investigating their interplay are especially interesting. The interrelation between metacognition and executive functions might be rooted in various mechanisms. On the one hand, metacognition can play a critical role in executive functioning processes because it allows for top-down control of behaviour and an accurate determination of when which type of control is needed. For example, when through metacognitive monitoring of performance, one is uncertain about his/her performance level, executive functions could be triggered for adaptation by switching between strategies to perform a task or by inhibiting interfering information to better focus on the task at hand. On the other hand, executive functions might be necessary to make metacognitive processes possible, as, for example, metacognitive monitoring may rely on working memory skills. Studies investigating both functions simultaneously indicate that executive functions and metacognition, while being dissociable, are related (see Roebers & Feurer, 2016, for a short overview). This was also observed in this dissertation (Chapter 2 and 3), although we only observed an association with updating and not with other executive functions. When considered in concert to predict educational achievement, there is evidence that metacognitive skills are more important than executive functions (Bryce et al., 2015), as was also observed in the current dissertation (Chapter 2 and 3).

It is, however, important to note that both executive functions and metacognition are broad concepts and that specific conceptualisations and operationalisations between studies can thus differ substantially. These differences in operationalisation also affect the amount of overlap between the concepts and the results of investigations on their interrelation. Within the current dissertation, both metacognition and executive functions were operationalised via tasks that have been widely used in other studies. Importantly, for metacognition, I specifically focused on declarative metacognitive knowledge and procedural metacognitive monitoring, but not on metacognitive control. This might explain why no strong associations were found between our measures of metacognition and executive function, as the strongest associations between metacognition and executive function have been observed at the level of metacognitive control (see Roebers & Feurer, 2016, for a short overview).

3 General measurement challenges

Several measurement challenges related to the current dissertation have already been discussed both within the different chapters of this dissertation and throughout this general discussion. In the following section, I further elaborate on three measurement challenges in more detail, which emerged over the different studies. These relate to the use of performance measures, task impurity and self-ratings.

Within the current dissertation, several processes were investigated in which both accuracy and speed are important indicators of performance. For example, in arithmetic, people may strive for arithmetic fluency, that is, being both fast and accurate. Hence, in the different tasks used to measure different processes in the current dissertation (e.g. arithmetic, numerical magnitude processing), participants were asked to perform both accurately and fast. Accuracy of performance and speed of performance are thus inherently intertwined and investigating them separately might not always be ideal. On the one hand, the separate investigation of accuracy versus speed may be of interest to examine specific associations of either accuracy of performance or speed of performance with other processes. For example, in Chapter 2, we found that accuracy, but not speed of addition performance was associated with metacognitive knowledge, which may indicate that general metacognitive knowledge on the effectiveness of strategies is especially important to accurately solve an arithmetic problem and perhaps less important to quickly solve the problem. On the other hand, the use of a performance measure that includes both accuracy and response time simultaneously is theoretically appealing. There is much debate, however, on how to combine measures of speed and accuracy (e.g. Bruyer & Brysbaert, 2011; Liesefeld & Janczyk, 2019). One example is using inverse efficiency scores, i.e. dividing the average response time of correct answers by the overall accuracy, which we have used as performance measures for our inhibition tasks in Chapter 2 and 3. However, such scores are very sensitive to speed-accuracy trade-offs, which need to be evaluated before a combined performance measure is considered. Within this dissertation, exploratory analyses indicated that using separate or combined measures of accuracy and speed did not change the interpretation of our results on arithmetic, numerical magnitude processing, executive functioning and metacognition. This was also illustrated by the fact that, for example, an association between arithmetic and metacognitive monitoring was found independent of whether separate measures of accuracy and speed for arithmetic were used (e.g. Chapter 2) or whether a combined measure was used (e.g. TTA in Chapter 5).

A pivotal point in the measurement of cognitive, metacognitive and affective processes is the issue of task impurity. This measurement problem has received most attention in the context of executive functions (e.g. Bull & Lee, 2014; Miyake et al., 2000), but the same rationale holds for the measurement of metacognitive processes. Task impurity concerns the problem that a single task assesses different components (e.g. both executive and non-executive function processes), making a pure interpretation of what the task measures troublesome. In research on executive functions, this is particularly problematic, because executive functions are higher-order processes that manifest themselves only in operating on other processes. By nature, tasks measuring executive functions insurmountably also tap into other (cognitive) processes that are not necessarily relevant to the targeted executive functions, but, for example related to the domain they are measured in. For example, when executive functions are measured with tasks that use numerical stimuli, as is the case in the often-used digit span backward task or in a numerical Stroop task, it can be problematic to interpret whether performance is related to the executive functioning skills, to processing of numerical stimuli, or to both. To overcome this problem, I deliberately decided to not include numerical stimuli in the executive functioning tasks.

In the existing literature, task impurity is often addressed by using multiple tasks to measure each executive functioning component to obtain a measure of the latent ability level (e.g. Miyake et al., 2000). While I administered widely used executive functioning tasks, an important limitation of the current dissertation is that I did not include multiple measures for each executive functioning component. Consequently, I was not able to adopt such a latent variable approach. This limitation should be considered in light of the already large battery of tasks that I administered in primary school children within a school setting to measure cognitive, metacognitive and affective processes.

Related to the issue of task-impurity, the use of self-rating scales to measure mathematics anxiety, although the most commonly used method (Dowker et al., 2016), is not without limitations. Self-rating scales require self-awareness and functioning self-inspection: One needs to adopt a meta-perspective on his/her own cognition, behaviour and/or affect. As a result, self-rating scales of mathematics anxiety also include a metacognitive component and this may be troublesome when investigating the interrelations between mathematics anxiety and metacognition, as was the case in Chapter 5. However, it is critical to note that, in Chapter 5, I specifically focused on the metacognitive monitoring component of metacognition, using a measurement on a local scale (i.e. at the item level; Pieschl, 2009) with trial-by-trial metacognitive judgments on accuracy. The metacognitive ability that is needed to fill in self-rating scales may be much more related to global aspects of metacognition that reflect on general performance (e.g. on a whole task). It has been shown that local and global measures of metacognition represent distinct abilities that are not necessarily correlated (Pieschl, 2009). Future research could bypass this limitation of self-rating scales to measure mathematics anxiety by using physiological measures (e.g. heart rate or cortisol level), which do not include a self-reflective component (e.g. Avancini & Szűcs, 2019).

4 Correlational, individual differences approach

All studies presented within this dissertation were correlational in nature, using both cross-sectional and longitudinal designs. In the next section, I will elaborate on strengths and weaknesses of this approach, focussing on the use of a longitudinal design, the directionality of the investigated associations and the importance of the autoregressive effects, and on the critical difference between individual differences and prerequisites for performance when investigating the importance of other processes for arithmetic performance.

The longitudinal nature of the investigation of the cognitive, metacognitive and affective processes in arithmetic in the current dissertation is an important strength. Longitudinal studies provide a unique insight into the development of different processes and thus are especially valuable when processes are examined in the midst of development, as were the studied processes in the current dissertation. By investigating the interrelations during this crucial developmental time, I was able to unravel developmental dynamics of the interplay between these processes in the context of arithmetic. By using a longitudinal panel design, the current dissertation was able, firstly, to unravel the stability over time of the associations found in early primary school; secondly, to uncover the longitudinal associations between several cognitive, metacognitive and affective processes and arithmetic; and critically, to investigate these predictive associations while taking prior performance (e.g. prior arithmetic performance) into account.

Longitudinal studies are indeed important to investigate predictive associations between performance at earlier time points and later outcome. The dominant approach to study arithmetic performance and development within the existing literature, is investigating how other processes, such as executive functions, predict later arithmetic. Within this dissertation, I also predominantly set up our studies from this perspective. This approach is certainly worthwhile, because arithmetic is crucial in children's educational development and, more generally, in many aspects of modern life. It is thus essential to investigate what is associated with or predictive of arithmetic performance and development. However, to thoroughly study arithmetic performance and development, considering bi-directionality of the associations between arithmetic and its correlates is also important (e.g. Peng & Kievit, 2020, for a more general discussion of the issue of bidirectionality in predicting academic performance). The longitudinal study reported on in Chapter 4 also illustrates this. In this study, I not only investigated whether metacognitive monitoring and mathematics anxiety predicted later arithmetic performance, I also investigated whether arithmetic performance predicted later metacognitive monitoring and mathematics anxiety. The results indeed demonstrated the predictive power of arithmetic performance for these processes later in development, which was also confirmed in Chapter 5. Only by considering the bi-directionality of the associations, I was able to establish this pivotal role of academic performance in processes that are generally considered to be 'support processes'. On the other hand, academic performance can also play an important role in the interrelation between different processes, as was shown by the mediation by arithmetic performance of the association between early mathematics anxiety and later metacognitive monitoring reported in Chapter 4.

While longitudinal designs provide a wealth of information, especially on predictive associations over development, such studies do not provide evidence on underlying causal relationships per se. To inform us to some extent on the predictive power of different processes for arithmetic performance and development, and vice versa, a *panel* longitudinal design is essential, as such designs allow for the inclusion of autoregressive effects. Autoregressive effects are the effects a process has on itself measured at a later time, and thus describe the stability of individual differences in a process from one time point to the next (Selig & Little, 2012). The inclusion of the autoregressive effects means that the variance in a process at a later time point (e.g. T_2) that can be predicted by another process at $T_{1,}$ is *residual variance* controlling for previous levels of the outcome measure. Controlling for prior performance is necessary to avoid that concurrent correlations between processes at both time points confound the investigated predictive associations. Given the high stability of the processes investigated in the current dissertation, including autoregressive effects was an important strength of our design. This was nicely demonstrated in the context of the predictive association between metacognitive monitoring and arithmetic. Without considering autoregressive effects, both processes were predictive of each other at a later time point. When considering autoregressive effects, only strong evidence was found for a predictive role of prior arithmetic performance for later metacognitive monitoring, while the predictive power of metacognitive monitoring for later arithmetic performance was unclear. These results further confirmed the pivotal role of academic performance.

It is important to note that using autoregressive effects to investigate directionality of predictive associations also comes with certain limitations. For example, when autoregressive effects are very high, as was the case in this dissertation, explaining residual variance is very difficult. Furthermore, when processes co-occur, autoregressive effects may account for the variance explained by the co-occurring process. For example, metacognitive monitoring processes and mathematics anxiety co-occur with solving arithmetic problems. A measure of arithmetic might thus already capture individual differences in metacognitive monitoring and mathematics anxiety. These individual differences in metacognitive monitoring and mathematics anxiety may thus be reflected in the autoregressive effect of arithmetic, which confound the investigated predictive associations. Taken together, in order to make strong claims on causality, even longitudinal panel designs do not provide conclusive evidence. In that case experimental studies are required, such as intervention and/or training studies.

Future experimental studies can build on the presented correlational studies which demonstrated the interplay between and the importance of the different investigated processes. For example, instead of examining the extent to which metacognitive monitoring is related to arithmetic performance, future studies should investigate the question as to whether and how metacognitive monitoring causes gains in arithmetic performance and learning, by investigating the effect of fostering metacognitive monitoring on performance. Broad interventions focused on improving metacognitive knowledge and skills more generally in the context of mathematics education have already shown promising effects (see Schneider & Artelt, 2010, for a short overview; Dowker, 2019b), yet, it is not always clear which aspects of such programs have been successful and "*more research is needed on exactly which aspects of metacognition are important here*" (Dowker, 2019b, p. 299). The findings of the current dissertation plea to specifically investigate how metacognitive monitoring may improve arithmetic. For example, researchers can include a forced metacognitive monitoring component within a training paradigm in which arithmetic items are presented and feedback on the correctness of the answer is given. The effect of including a metacognitive monitoring component in an arithmetic training can then be investigated. Based on the findings of Butler et al. (2008) that feedback is especially helpful for low confidence responses, requiring children to explicitly give metacognitive judgments may efficiently promote learning. In the existing literature, promising evidence in adults is found that training of metacognitive monitoring results improves learning (e.g. Dunlosky et al., 2003). It remains to be determined whether such findings can be generalized to children and to the learning of arithmetic.

Another important consideration in the context of our research on the correlates of individual differences in arithmetic performance and development, is the difference between processes that are a prerequisite for doing arithmetic versus processes that drive individual differences. Indeed, children rely on several (cognitive) processes when solving arithmetical problems, and, as such, a certain skill level in these processes is needed to perform that task. Once that skill level is reached, which, in typically developing children, may be the case for several processes that are important for arithmetic, it is plausible that these processes may not explain additional variance in arithmetic performance. For example, our results indicated that inhibition, shifting, and (to a lesser extent) updating did not predictively explain individual differences in arithmetic. Yet, this does not imply that these processes are unimportant for arithmetic. It is likely that a certain level of executive functioning skills is a prerequisite or cognitive support for good mathematics learning. As the participants in our studies were all typically developing children in primary school, this level was probably reached for most of them. The role of support processes in arithmetic performance and development may become more clear either by using a more experimental approach, for example, by using dual-task designs that manipulate working memory (e.g. De Rammelaere et al., 2001) or by studying children with deficits in these support processes, such as children with ADHD, who show deficits in executive function.

5 Multiple processes, complex interactions

The current dissertation studied arithmetic performance and development in the context of different cognitive, metacognitive and affective processes, and at different ages in primary school. It is important to acknowledge the complexity of on the one hand, the processes in themselves, and on the other hand, their interrelations.

A critical strength of the current dissertation was bringing together different cognitive, metacognitive and affective processes in the investigation of performance and development in arithmetic. It is important to note that a long tradition of research has studied these different processes in isolation and as a result, extensive bodies of research exist on each of these different processes. Within each of these separate bodies of research, the complexity of the studied processes has been amply demonstrated. For example, most of the studied processes in the current dissertation are multi-faceted constructs, covering different aspects of an overarching process. For instance, executive functions encompass different components, such as inhibition. In turn, inhibition encompasses different aspects including cognitive and behavioural inhibition (Diamond, 2013). The same is true for metacognition, a broad concept covering declarative metacognitive knowledge as well as procedural metacognition. In turn, declarative knowledge encompasses knowledge on person, task and strategy aspects (Flavell, 1979) and procedural metacognition encompasses both monitoring and control processes (Nelson & Narens, 1990). In light of the topic of this dissertation, it is thus important to keep in mind that different aspects of the studied processes within this dissertation may have varying levels of involvement in different aspects of arithmetic.

Adding to this complexity, several of the studied processes within this dissertation have been investigated from the viewpoint of different research traditions, such as cognitive psychology or educational sciences. Their approaches to investigate these processes, namely the conceptualisation of the processes and the operationalisation of the processes in performance measures, substantially differ between different research traditions. For example, metacognition has been studied from the viewpoint of not only educational research traditions (e.g. Schneider & Artelt, 2010; Schoenfeld, 1992), but also in cognitive experimental research (e.g. Rahnev & Fleming, 2019), resulting in different research approaches such as observational methods, thinking-aloud protocols, questionnaires or strict experimental tasks investigating feeling of correctness.

The complexity of the investigated processes within this dissertation is also apparent in the complicacy of their interplay in the context of arithmetic performance and development. As stated by Ann Dowker (Dowker, 2019c, p. 4) "*Any statement that arithmetic ability is purely the product of a single factor is oversimplified*". In line with the approach within the current dissertation, it is thus essential to investigate different processes in concert when investigating arithmetic performance and development, because a multitude of processes have an impact on arithmetic. The findings of this dissertation indeed highlight the importance of such an approach, as I was able to isolate the unique contribution of different processes to arithmetic performance and development over and above each other. The uniqueness of such contributions can only be established if a multitude of processes is investigated.

In sum, performance in and development of arithmetic is rooted in a complex interplay of processes, reaching further than the specific processes investigated in the current doctoral thesis. Furthermore, these processes themselves are also complex and multi-faceted. Studies focussing on specific aspects of this complex puzzle, such as the ones presented in the current dissertation, are quintessential. They functionally unravel specific associations and underlying mechanisms, and, as such, contribute important pieces to the puzzle. Yet, one should not forget these pieces are only one part of the complex puzzle.

Educational considerations

Research on the correlates of arithmetic performance and development has the longstanding aim to help to develop effective instructional approaches and to contribute to designing evidence-informed remediation programs for children at risk for or with difficulties in arithmetic. In line with this aim, the findings of the current dissertation do not only help to refine theoretical frameworks on the correlates of arithmetic, as was discussed above, but might also be useful for the educational practice. Both science and education benefit from a good understanding of the processes that underlie the development of arithmetic in primary school children. Such an understanding may provide insight into processes as targets for instructional approaches as well as for training and intervention studies. Because all the included studies in this dissertation investigated the interplay of several processes in arithmetic in a crucial developmental period, they may provide a step towards (further research on) such evidence-informed approaches in education in a crucial period for fostering learning.

Based on the findings of this dissertation, a possible process to focus on in interventions is metacognition. On the one hand, metacognitive knowledge might be a fruitful target to enhance, even in young primary school children, as it emerged as the strongest predictor of arithmetic development in our studies. Teaching children about what factors (inter)act to affect their academic performance, for example by teaching them the difference between recognizing versus remembering as a learning strategy, may enhance later performance (Flavell, 1979). On the other hand, throughout all studies within this dissertation, metacognitive monitoring was found to be a strong, unique correlate and predictor of arithmetic performance. Based on this consistent finding, it might be relevant to include a metacognitive monitoring element in educational technology and serious games (e.g. Rekentuin; Van der Maas et al., 2014). For example, this could be done by, after solving an arithmetic item in such a game, asking children to indicate their confidence in their answer in a similar way to what was done in the current dissertation. Such a manipulation of the arithmetic learning process might be interesting in the light of our findings on the role of metacognitive monitoring in arithmetic and the fact that feedback after performance is especially helpful for low confidence responses (Butler et al., 2008).

Our findings concerning the suggested emergence of more domain-general metacognitive monitoring in academic domains over second to third grade of primary school may be relevant for thinking about how to stimulate metacognitive monitoring at different ages. More specifically, when educators want to promote metacognitive monitoring in children in itself or promote metacognitive monitoring to enhance academic learning, knowing whether metacognition is rather domain-specific or domain-general, and when domain-generality emerges, is of importance. This might impact how educators best provide instructions in metacognitive monitoring, namely for each task or domain separately (i.e. domain-specific metacognition) or concurrently in different tasks and domains (expecting it to transfer to new domains; domain-general metacognition). The findings of Chapter 5 show that there might be a tipping point between second and third grade. It is important to note that in the development of metacognition, parents and teachers may have a pivotal role, both by means of explicit instruction as well as modelling of behaviour. For example, Veenman et al. (2006) argue that the vast majority of students spontaneously pick up metacognitive knowledge and skills from their peers, parents and teachers. Roebers and Feurer (2016) state that parents' support to actively construct interpretations of mental states may foster procedural metacognitive development. In a large meta-analysis, Hattie (2009) referred to modelling metacognitive knowledge and skills as one of the most important skills in teachers. Furthermore, the British Education Endowment Foundation (n.d.) ranked the teaching of metacognition as one of the most successful educational interventions. In line with the results of Chapter 5, enhancing metacognitive monitoring in young primary school children may best be done in a domain-specific manner, hence in different courses separately, and transfer of these metacognitive monitoring skills is not to be expected at this young age. Yet, starting in third grade, this enhancing of metacognitive monitoring may be more domain-general and transfer may be expected. Further research is needed to investigate these suggestions.

It is also relevant for the educational practice to note that our findings demonstrate a pivotal role of academic performance, not only for later academic performance and development, as demonstrated in the very large autoregressive effects, but also for other metacognitive and affective processes. For example, we found that early arithmetic performance predicted both later metacognitive monitoring skills and later mathematics anxiety. Our results thus suggest that it is relevant to take performance level into account in development of metacognitive monitoring and mathematics anxiety. Enhancement of performance and domain-specific skills are likely crucial to further development of other processes such as metacognition, or to prevent (further) development of mathematics anxiety. Importantly, the current dissertation demonstrated an increasing role of mathematics anxiety in both arithmetic performance and metacognitive monitoring. As such, our results plea for careful detection of (early) signs of mathematics anxiety to prevent the potential vicious spiral of reinforcement between mathematics anxiety and difficulties with mathematics achievement and/or metacognitive monitoring. This might be especially important given society's increasing reliance on standardized testing and the detrimental impact of mathematics anxiety in such test situations.

The findings of the current dissertation might be particularly relevant for struggling learners, as they experience severe difficulties with arithmetic, and metacognitive abilities may have an important role in this learning process. It is important to note that the current dissertation and most of the existing body of research that investigated the role of metacognition in arithmetic has focused on typically developing children (but see Desoete & Roeyers, 2002). However, a substantial proportion of children present with dyscalculia, a neurodevelopmental learning disorder that is characterized by life-long difficulties in calculation that are not merely explained by intellectual disabilities, uncorrected sensory problems, mental or neurological disorders or inadequate instruction (American Psychiatric Association, 2013). Given the relative high prevalence of dyscalculia (i.e. around 7%; Butterworth, 2011; Peters & Ansari, 2019), it is important to determine whether the findings of the current dissertation also hold for these struggling learners. In one of the few studies on this topic, Rinne and Mazzocco (2014) demonstrated poor metacognitive monitoring in children with dyscalculia. However, they used an extremely small sample size of children with dyscalculia (n = 16), and the authors critically emphasized that replication of this finding was crucial. Future studies should thus investigate the role of metacognition in arithmetic in children with dyscalculia. Understanding these children's metacognitive knowledge and skills, which could either act as a risk or protective factor, and their interrelation with arithmetic, might provide a critical ground for the development of remedial interventions.

It is important to note that, in Flemish primary school, substantial gender gaps are found for mathematics performance (Kabinet Vlaams minister van Onderwijs, 2017), an observation that is applicable to many countries around the globe (Stoet & Geary, 2018). Such gender differences are also found in research on mathematics anxiety, even in primary school (Hill et al., 2016) and in self-confidence in mathematics, as was demonstrated in an extensive research report on the development of math capabilities and confidence in primary school (Nunes et al., 2009). Furthermore, lack of confidence has often been suggested to be the reason for the continued gender segregation in (higher) mathematics education and careers in mathematics-related fields (e.g. Hackett, 1985). This is especially worrisome when low confidence is not in accordance with actual performance, which consequently reflects poor metacognitive monitoring. As mathematics anxiety, self-confidence in mathematics and metacognition all involve thinking about one's own performance, gender differences are also likely present in, for example, children's metacognitive monitoring. Against this background, future studies should further examine whether gender plays a role in metacognitive knowledge and skills and whether these potential differences moderate the findings of the current dissertation.

While the current dissertation focused on characteristics that are more inherent to the child, the environment also influences a child's arithmetic performance and development. This is often overlooked when interpreting research findings, as we compare and build on the research findings of studies conducted all around the world, abstracting from, for example, the educational context in which children are learning arithmetic or the values of the society children grow up in. It is thus important to note that all studies in this doctoral dissertation were performed in participants in Flanders who all received formal schooling in the Flemish educational system. This may impact on, for example, the way arithmetic strategies are introduced and learned. A cross-cultural design, contrasting different educational approaches, maybe useful to examine whether cultural differences are found in the interplay of processes investigated in the current dissertation.

Within this dissertation, I have focused on arithmetic, a crucial aspect of mathematics. However, it is important to note that "*There is no such thing as arithmetic ability, only arithmetic abilities*" (Dowker, 2019c, p1): Arithmetic is made up of many components. I have mostly investigated arithmetic fluency by accuracy and response time measures and by using the Tempo-test Arithmetic, which combines the two. However, learning arithmetic also includes components such as understanding of arithmetic principles and computational estimation. Future research should further specify our findings by examining the interrelations between these other components of arithmetic and the investigated cognitive, metacognitive and affective processes. It may be especially interesting to investigate the association between metacognitive monitoring and computational estimation (e.g. Dowker, 2019a), as higher uncertainty level in the arithmetic task may have an impact on the association between performance on that task and metacognitive monitoring.

General Conclusion

Performance in and development of arithmetic is rooted in a complex interplay of processes. This dissertation added one piece to the complex puzzle that performance and development of arithmetic in children consists of by addressing important gaps in the existing literature, emphasizing the role of metacognition. In short, the main findings of this dissertation were 1) Numerical magnitude processing, executive functions and metacognition are each uniquely associated to arithmetic performance in primary school children; 2) Numerical magnitude processing and metacognition have unique predictive roles for later arithmetic, yet prior arithmetic performance remains the most robust predictor for later arithmetic performance; 3) The association found between metacognitive monitoring and arithmetic in primary school children is not affected by mathematics anxiety; 4) A transition from domain-specificity of metacognitive monitoring towards domain-generality of metacognitive monitoring in academic performance seems to occur between the age of 7 to 9 years; 5) Metacognitive monitoring in children is related to the left inferior frontal gyrus (IFG), suggesting an overlap with the arithmetic brain network. It is essential to recognize that we are still left with many outstanding questions in need of answers. By building on the existing literature, including the current dissertation, and through future research investigating multiple processes simultaneously, combining multiple methods such as behavioural and brain-imaging techniques, using different study designs such as cross-sectional, longitudinal and experimental methods, and studying children across ages, cultures, educational systems and demographic groups, we can further our understanding of the complexity of individual differences in this critical skill that is arithmetic. This understanding is not solely essential for academic purposes, but has crucial practical implications as well, paving the way for improving future perspectives for the learning child.

Bibliography

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Association.

Ancker, J. S., & Kaufman, D. (2007). Rethinking health numeracy: A multidisciplinary literature review. *Journal of the American Medical Informatics Association*, *14*(6), 713–721. https://doi.org/10.1197/jamia.M2464

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, *41*(2), 521–543. https://doi.org/10.1177/0149206314560412

Annevirta, T., Laakkonen, E., Kinnunen, R., & Vauras, M. (2007). Developmental dynamics of metacognitive knowledge and text comprehension skill in the first primary school years. *Metacognition Learning*, *2*, 21–39. https://doi.org/10.1007/s11409-007-9005-x

Ansari, D. (2010). Neurocognitive approaches to developmental disorders of numerical and mathematical cognition: The perils of neglecting the role of development. *Learning and Individual Differences*, *20*(2), 123–129. https://doi.org/10.1016/j.lindif.2009.06.001

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *NeuroReport*, *16*(16), 1769–1773. https://doi.org/10.1097/01.wnr.0000183905.23396.f1

Ansari, D., Grabner, R. H., Koschutnig, K., Reishofer, G., & Ebner, F. (2011). Individual differences in mathematical competence modulate brain responses to arithmetic errors: An fMRI study. *Learning and Individual Differences*, *21*(6), 636–643. https://doi.org/10.1016/j.lindif.2011.07.013

Ansari, D., & Lyons, I. M. (2016). Cognitive neuroscience and mathematics learning: how far have we come? Where do we need to go? *ZDM - Mathematics Education*, *48*, 379–383. https://doi.org/10.1007/s11858-016-0782-z

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, *8*(4), 170–177. https://doi.org/10.1016/j.tics.2004.02.010

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

Arsalidou, M., Pawliw-Levac, M., Sadeghi, M., & Pascual-Leone, J. (2018). Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies. *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2017.08.002

Arthurs, O. J., & Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *TRENDS in Neurosciences*, *25*(1), 27–31. https://doi.org/http://doi.org/10.1016/S0166-2236(00)01995-0

Ashcraft, M. H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In C. J. Brainerd, R. Kail, & J. Bisanz (Eds.), *Formal methods in developmental research* (pp. 302–338). Springer. https://doi.org/10.1007/978-1-4612-4694-7_9

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psycholocial Science*, 181–185. https://journals.sagepub.com/doi/pdf/10.1111/1467-8721.00196

Ashcraft, M. H. (2019). Models of math anxiety. In I. C. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics anxiety. What is known and what is still to be understood* (pp. 1–19). Routledge.

Ashcraft, M. H., & Faust, M. W. (1994). Mathematics anxiety and mental arithmetic performance: An exploratory investigation. *Cognition and Emotion*, *8*(2), 97–125. https://doi.org/10.1080/02699939408408931

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224–237. https://doi.org/10.1037/0096-3445.130.2.224

Ashcraft, M. H., Kirk, E. P., & Hopko, D. (1998). On the cognitive consequences of mathematics anxiety. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 175–196). Psychology Press.

Ashcraft, M. H., Krause, J. A., & Hopko, D. R. (2007). Is math anxiety a mathematical learning disability? In D. B. Berch & M. M. Mazzocco (Eds.), *Why is math so hard for some children* (pp. 329–348). Paul H. Brookes Publishing Co.

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment*, *27*(3), 197–205. https://doi.org/10.1177/0734282908330580

Avancini, C., & Szűcs, D. (2019). Psychophysiological correlates of mathematics anxiety. In I. C. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics anxiety. What is known and what is still to be understood* (pp. 42–61). Routledge.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485–493.

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education*, *10*(1), 10–33. https://doi.org/https://doi.org/10.1111/mbe.12100

Bailey, D. H., Littlefield, A., & Geary, D. C. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grades. *Journal of Experimental Child Psychology*, *113*, 78–92. https://doi.org/10.1016/j.jecp.2012.04.014

Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *Journal of Neuroscience*, *33*(42), 16657–16665. https://doi.org/10.1523/JNEUROSCI.0786-13.2013

Barrouillet, P., & Lépine, R. (2005). Working memory and children’ s use of retrieval to solve addition problems. *Journal of Experimental Child Psychology*, *91*, 183–204. https://doi.org/10.1016/j.jecp.2005.03.002

Barrouillet, P., Mignon, M., & Thevenot, C. (2008). Strategies in subtraction problem solving in children. *Journal of Experimental Child Psychology*, *99*, 233–251. https://doi.org/10.1016/j.jecp.2007.12.001

Batchelor, S., Torbeyns, J., & Verschaffel, L. (2019). Affect and mathematics in young children: an introduction. *Educational Studies in Mathematics*, *100*(3), 201–209. https://doi.org/10.1007/s10649-018-9864-x

Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology*, *4*(3), 195–218. https://doi.org/10.1080/09541449208406182

Bellon, E., Fias, W., & De Smedt, B. (2016). Are individual differences in arithmetic fact retrieval in children related to inhibition? *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00825

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.2019.01.012

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Lawrence Erlbaum associates.

Berch, D. B., Geary, D. C., & Mann Koepke, K. (Eds.). (2016). *Development of mathematical cognition: Neural substrates and genetic influences.* (Volume 2). Academic Press.

Berryhill, M. E. (2012). Insights from neuropsychology: pinpointing the role of the posterior parietal cortex in episodic and working memory. *Frontiers in Integrative Neuroscience*, *6*. https://doi.org/10.3389/fnint.2012.00031

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

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Block, K. K., & Peskowitz, N. B. (1990). Metacognition in spelling: Using writing and reading to self-check spellings. *The Elementary School Journal*, *91*(2), 151–164.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*(4), 1016–1031. https://doi.org/10.1111/j.1467-8624.2008.01173.x

Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities*, *43*(6), 541–552. https://doi.org/10.1177/0022219410371676

Borkowski, J. G., Chan, L. K. S., & Muthukrishna, N. (2000). A process-oriented model of metacognition: Links between motivation and executive functioning. *Issues in the Measurement of Metacognition*, 1–41.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1017/CBO9781107415324.004

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). *Region of interest analysis using an SPM toolbox*. https://doi.org/10.1201/b14650-28

British Education Endowment Foundation. (n.d.). https://educationendowmentfoundation.org.uk/

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, *26*(2), 571–593. https://doi.org/10.1207/s15326942dn2602

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 77–165). Lawrence Erlbaum associates. https://doi.org/10.3109/10826086909062003

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*(1), 5–13. https://doi.org/10.5334/pb-51-1-5

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

Bugden, S., & Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition*, *118*, 32–44. https://doi.org/10.1016/j.cognition.2010.09.005

Bull, R., Johnston, R. S., & Roy, J. a. (1999). Exploring the roles of the visual-spatial sketch pad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*, *15*(3), 421–442. https://doi.org/10.1080/87565649909540759

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, *8*(1), 36–41. https://doi.org/10.1111/cdep.12059

Bull, R., & Scerif, G. (2001). Executive functionin as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273–293. https://doi.org/10.1207/S15326942DN1903

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin and Review*, *7*(1), 26–48. https://doi.org/10.3758/BF03210724

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butterworth, B. (2011). Dyscalculia: From brain to education. *Science*, *332*, 1049–1053. https://doi.org/10.1126/science.334.6057.761-b

Butterworth, Brian, Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *Quarterly Journal of Experimental Psychology*, *54A*(4), 1005–1029. https://doi.org/10.1080/713756007

Campbell, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, *1*(2), 121–164.

Campbell, J. I. D. (2005). *Handbook of mathematical cognition*. Taylor & Francis Group.

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, *130*(2), 299–315. https://doi.org/10.1037/0096-3445.130.2.299

Cappelletti, M., & Fias, W. (Eds.). (2016). *The mathematical brain across the lifespan*. Progress in brain research. https://doi.org/10.1016/S0079-6123(16)30081-4

Carey, E., Hill, F., Devine, A., & Szűcs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01987

Carlson, S. M., Zelazo, P. D., & Faja, S. (2013). Executive Function. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology* (pp. 706–743). https://doi.org/10.1093/oxfordhb/9780199958450.013.0025

Carr, M., Alexander, J., & Folds-Bennett, T. (1994). Metacognition and mathematics strategy use. *Applied Cognitive Psychology*, *8*, 583–595.

Carr, M., & Jessup, D. L. (1995). Cognitive and metacognitive predictors of mathematics strategy use. *Learning and Individual Differences*, *7*(3), 235–247. https://doi.org/10.1016/1041-6080(95)90012-8

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Chinn, S. (2012). Beliefs, anxiety, and avoiding failure in mathematics. *Child Development Research*, 1–8. https://doi.org/10.1155/2012/396071

Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2003). Schooling, literacy, numeracy and labour market success. *The Economic Record*, *79*(245), 165–181. https://doi.org/10.1111/1475-4932.t01-1-00096

Chua, E. F., Pergolizzi, D., & Weintraub, R. R. (2014). The cognitive neuroscience of metamemory monitoring: Understanding metamemory processes, subjective levels expressed, and metacognitive accuracy. In S. M. Fleming & C. D. Frith (Eds.), *The cognitive neuroscience of metacognition* (pp. 267–191). Springer-Verlag.

Chua, E. F., Rand-Giovannetti, E., Schacter, D. L., Albert, M. S., & Sperling, R. A. (2004). Dissociating confidence and accuracy: Functional magnetic resonance imaging shows origins of the subjective memory experience. *Journal of Cognitive Neuroscience*, *16*(7), 1131–1142. https://doi.org/10.1162/0898929041920568

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. *NeuroImage*, *29*(4), 1150–1160. https://doi.org/10.1016/j.neuroimage.2005.09.058

Chua, E. F., Schacter, D. L., & Sperling, R. (2009). Neural correlates of metamemory. *Journal of Cognitive Neuroscience*, *21*(9), 1751–1765. https://doi.org/10.1162/jocn.2009.21123.Neural

Cohen Kadosh, R., & Dowker, A. (Eds.). (2015). *The Oxford handbook of numerical cognition*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.001.0001

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*(2), 63–68. https://doi.org/10.1016/j.tine.2013.12.001

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

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

Crone, E. A., Wendelken, C., Donohue, S. E., & Bunge, S. A. (2006). Neural evidence for dissociable components of task-switching. *Cerebral Cortex*, *16*, 475–486. https://doi.org/10.1093/cercor/bhi127

David, A. S., Bedford, N., Wiffen, B., & Gilleen, J. (2012). Failures of metacognition and lack of insight in neuropsychiatric disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1379–1390. https://doi.org/10.1098/rstb.2012.0002

Davidson, P. S. R., Anaki, D., Ciaramelli, E., Cohn, M., Kim, A. S. N., Murphy, K. J., Troyer, A. K., Moscovitch, M., & Levine, B. (2008). Does lateral parietal cortex support episodic memory?. Evidence from focal lesion patients. *Neuropsychologia*, *46*, 1743–1755. https://doi.org/10.1016/j.neuropsychologia.2008.01.011

De Corte, E., Verschaffel, L., & Op 't Eynde, P. (2000). Self-regulation. A characteristic and a goal of mathematics education. In P. Pintrich, M. Boekaerts, & M. Zeidner (Eds.), *Self-regulation: Theory, research and applications* (pp. 687–726). Lawrence Erlbaum associates.

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved? *Memory and Cognition*, *29*(2), 267–273. https://doi.org/10.3758/BF03194920

De Smedt, B. (2016). Individual differences in arithmetic fact retrieval. In D.C. Geary & K. Mann-Koepke (Eds.), *Development of mathematical cognition: Neural substrates and genetic influences* (pp. 219–243). Elsevier Academic Press.

De Smedt, B. (2018). Applications of cognitive neuroscience in educational research. In *Oxford Research Encyclopedia of Education*. https://doi.org/10.1093/acrefore/9780190264093.013.69

De Smedt, B., Ansari, D., Grabner, R. H., Hannula, M. M., Schneider, M., & Verschaffel, L. (2010). Cognitive neuroscience meets mathematics education. *Educational Research Review*, *5*(1), 97–105. https://doi.org/10.1016/j.edurev.2009.11.001

De Smedt, B., & Boets, B. (2010). Phonological processing and arithmetic fact retrieval: Evidence from developmental dyslexia. *Neuropsychologia*, *48*(14), 3973–3981. https://doi.org/10.1016/j.neuropsychologia.2010.10.018

De Smedt, B., & Grabner, R. H. (2015). Applications of neuroscience to mathematics education. In R. Cohen Kadosh & A. Dowker (Eds.), *The oxford handbook of numerical cognition*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.48

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology*, *103*(2), 186–201. https://doi.org/10.1016/j.jecp.2009.01.004

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

De Smedt, B., & Verschaffel, L. (2010). Traveling down the road: From cognitive neuroscience to mathematics education ... and back. *ZDM Mathematics Education*, *42*, 649–654. https://doi.org/10.1007/s11858-010-0282-5

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, *103*(4), 469–479. https://doi.org/10.1016/j.jecp.2009.01.010

De Visscher, A., & Noël, M.-P. (2014a). Arithmetic facts storage deficit: The hypersensitivity-to-interference in memory hypothesis. *Developmental Science*, *17*(3), 434–442. https://doi.org/10.1111/desc.12135

De Visscher, A., & Noël, M.-P. (2014b). The detrimental effect of interference in multiplication facts storing: Typical development and individual differences. *Journal of Experimental Psychology: General*, *143*(6), 2380–2400. https://doi.org/10.1037/xge0000029 de Vos, T. (1992). *Tempo-Test-Rekenen.* Berkhout.

Dehaene, S. (2020). *How we learn. Why brains learn better than any machine...for now*. Viking.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771

Desender, K., Van Opstal, F., & Van den Bussche, E. (2014). Feeling the conflict: The crucial role of conflict experience in adaptation. *Psychological Science*, *25*(3), 675–683. https://doi.org/10.1177/0956797613511468

Desoete, A., & Roeyers, H. (2002). Off-line metacognition - A domain-specific retardation in young children with learning disabilities? *Learning Disability Quarterly*, *25*, 123–139. https://doi.org/10.2307/1511279

Destan, N., Hembacher, E., Ghetti, S., & Roebers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5- to 7-year-old children. *Journal of Experimental Child Psychology*, *126,* 213–228. https://doi.org/10.1016/j.jecp.2014.04.001

Destan, N., & Roebers, C. M. (2015). What are the metacognitive costs of young children's overconfidence? *Metacognition Learning*, *10,* 347–374. https://doi.org/10.1007/s11409-014-9133-z

DeStefano, D., & LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*(3), 353–386. https://doi.org/10.1080/09541440244000328

Devine, A., Fawcett, K., Szucs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, *8*(33), 1–9. https://doi.org/10.1186/1744-9081-8-33

Dew, K. H., Galassi, J. P., & Galassi, M. D. (1984). Math anxiety: Relation with situational test anxiety, performance, physiological arousal, and math avoidance behavior. *Journal of Counseling Psychology*, *31*(4), 580–583. https://doi.org/10.1037/0022-0167.31.4.580

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). Oxford University Press. https://doi.org/10.1093/acprof

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

Dowker, A. (2005). *Individual differences in arithmetic. Implications for psychology, neuroscience and education.* Psychology Press.

Dowker, A. (2019a). A good guess. Estimation and individual differences. In A. Dowker (Ed.), *Individual differences in arithmetic. Implications for psychology, neuroscience and education* (pp. 161–183). Routledge.

Dowker, A. (2019b). Implications for helping children with their arithmetical difficulties. In *Individual differences in arithmetic. Implications for psychology, neuroscience and education* (pp. 286–323). Routledge.

Dowker, A. (2019c). *Individual differences in arithmetic. Implications for psychology, neuroscience and education* (Second edi). Routledge.

Dowker, A. (2019d). "Math doesn't like me anymore". The role of attitudes and emotions. In A. Dowker (Ed.), *Individual differences in arithmetic. Implications for psychology, neuroscience and education* (pp. 260–285). Routledge.

Dowker, A. (2019e). Mathematics anxiety and performance. In I. C. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics anxiety. What is known and what is still to be understood* (pp. 62–76). Routledge.

Dowker, A., Bennett, K., & Smith, L. (2012). Attitudes to mathematics in primary school children. *Child Development Research*, Article ID 124939. https://doi.org/10.1155/2012/124939

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*, 1–16. https://doi.org/10.3389/fpsyg.2016.00508

Dreger, R. M., & Aiken, L. R. (1957). The identification of number anxiety in a college population. *Journal of Educational Psychology*, *48*, 344–351. https://doi.org/10.1037/h0045894

Duncan, G. J., Claessens, A., Magnuson, K., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Dunlosky, John, Kubat-Silman, A. K., & Hertzog, C. (2003). Training monitoring skills improves older adults' self-paced associative learning. *Psychology and Aging*, *18*(2), 340–345. https://doi.org/10.1037/0882-7974.18.2.340

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology*, *91*, 113–136. https://doi.org/10.1016/j.jecp.2005.01.003

Efklides, A., & Misailidi, P. (Eds.). (2010). *Trends and prospects in metacognition research*. Springer.

Erickson, S., & Heit, E. (2015). Metacognition and confidence: comparing math to other academic subjects. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00742

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

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition and Emotion*, *6*(6), 409–434. https://doi.org/10.1080/02699939208409696

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336–353.

Fias, W. (2016). Neurocognitive components of mathematical skills and dyscalculia. In D. B. Berch, D. C. Geary, & K. Mann Koepke (Eds.), *Development of mathematical cognition: Neural substrates and genetic influences* (pp. 195–217). Academic Press. https://doi.org/10.1016/b978-0-12-801871-2.00008-3

Fias, W., Menon, V., & Szűcs, D. (2013). Multiple components of developmental dyscalculia. *Trends in Neuroscience and Education*, *2*(2), 43–47. https://doi.org/10.1016/j.tine.2013.06.006

Finn, B., & Metcalfe, J. (2014). Overconfidence in children's multi-trial judgments of learning. *Learning and Instruction*, *32*, 1–9. https://doi.org/10.1016/j.learninstruc.2014.01.001

Finnie, R., & Meng, R. (2001). Cognitive skills and the youth labour market. *Applied Economics Letters*, *8*(10), 675–679. https://doi.org/10.1080/13504850110037877

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

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annu. Rev. Psychol*, *50*, 21–45.

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology*, *1*, 324–340. https://doi.org/10.1016/0010-0285(70)90019-8

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, S. M., & Dolan, R. J. (2014). The Neural Basis of Metacognitive Ability. In S. M. Fleming & C. D. Frith (Eds.), *The cognitive neuroscience of metacognition* (pp. 245–265). Springer-Verlag. https://doi.org/10.1007/978-3-642-45190

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience*, *32*(18), 6117–6125. https://doi.org/10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 1–9. https://doi.org/10.3389/fnhum.2014.00443

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association*, *105*(2), 203–206. https://doi.org/10.5195/jmla.2017.88

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. https://doi.org/10.1016/j.lindif.2017.06.003

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

Friso-Van Den Bos, I., Van Der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, *10*, 29–44. https://doi.org/10.1016/j.edurev.2013.05.003

Fuhs, M. W., & Mcneil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science*, *16*(1), 136–148. https://doi.org/10.1111/desc.12013

Garcia, T., Rodríguez, C., González-Castro, P., González-Pienda, J., & Torrance, M. (2016). Elementary students' metacognitive processes and post-performance calibration on mathematical problem-solving tasks. *Metacognition Learning*, *11*, 139–170.

Garofalo, J., & Lester, F. K. (1985). Metacognition, cognitive monitoring , and mathematical performance. *Journal of Research in Mathematics Education*, *16*(3), 163–176.

Garrett, A. J., Mazzocco, M. M. M., & Baker, L. (2006). Development of the metacognitive skills of prediction and evaluation in children with or without math disability. *Learning Disabilities Research and Practice*, *21*(2), 77–88. https://doi.org/10.1111/j.1540-5826.2006.00208.x.

Geary, David C. (Ed.). (1994). *Children's mathematical development. Research and practical applications.* American Psychological Association.

Geary, David C., Hoard, M. K., & Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities*, *45*(4), 291–307. https://doi.org/10.1177/0022219410392046

Geary, David C., & Moore, A. M. (2016). Cognitive and brain systems underlying early mathematical development. *Progress in Brain Research*, *227*, 75–103. https://doi.org/10.1016/bs.pbr.2016.03.008

Geary, David C. (2011). Cognitive predictors of achievement growth in mathematics: A five year longitudinal study. *Developmental Psychology*, *47*(6), 1539–1552. https://doi.org/10.1037/a0025510.Cognitive

Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11267–11271. https://doi.org/10.1073/pnas.1220568110

Geurten, M., & Lemaire, P. (2017). Age-related differences in strategic monitoring during arithmetic problem solving. *Acta Psychologica*, *180*, 105–116. https://doi.org/10.1016/j.actpsy.2017.09.005

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. https://doi.org/10.1016/j.cogdev.2018.08.002

Ghetti, S. (2008). Rejection of false events in childhood. *Current Directions in Psychological Science*, *17*(1), 16–20.

Gierl, M. J., & Bisanz, J. (1995). Anxieties and attitudes related to mathematics in grades 3 and 6. *The Journal of Experimental Education*, *63*(2), 139–158.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE*, *8*(6), 1–9. https://doi.org/10.1371/journal.pone.0067374

Gilmore, C., Göbel, S. M., & Inglis, M. (2018a). *An introduction to mathematical cognition*. Routledge.

Gilmore, C., Göbel, S. M., & Inglis, M. (2018b). The development of arithmetic skills. In C. Gilmore, S. M. Göbel, & M. Inglis (Eds.), *An introduction to mathematical cognition* (pp. 51–72). Routledge.

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM Mathematics Education*, *47*, 771–782. https://doi.org/10.1007/s11858-014-0659-y

Goldfarb, L. (2018). Cognitive interferences and their development in the context of numerical tasks: Review and implications. In A. Henik & W. Fias (Eds.), *Heterogeneity of function in numerical cognition* (pp. 245–259). Academic Press.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/10.1037/h0059831

Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development*, *19*(1), 21–46. https://doi.org/10.1080/15248372.2017.1421538

Haberkorn, K., Lockl, K., Pohl, S., Ebert, S., & Weinert, S. (2014). Metacognitive knowledge in children at early elementary school. *Metacognition and Learning*, *9*(3), 239–263. https://doi.org/10.1007/s11409-014-9115-1

Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Counseling Psychology*, *32*(1), 47–56. https://doi.org/10.1037/0022-0167.32.1.47

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668. https://doi.org/10.1038/nature07246

Harari, R. R., Vukovic, R. K., & Bailey, S. P. (2013). Mathematics anxiety in young children: An exploratory study. *The Journal of Experimental Education*, *81*(4), 538–555. https://doi.org/10.1080/00220973.2012.727888

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis : A regression-based approach* (2nd ed.). The Guilford Press.

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1–21. https://doi.org/10.3758/s13428-017-0935-1

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33–46. https://doi.org/doi:10.2307/749455

Henik, A., & Fias, W. (Eds.). (2018). *Heterogeneity of function in numerical cognition*. Academic Press.

Hilgenstock, R., Weiss, T., & Witte, O. W. (2014). You'd better think twice: Post-decision perceptual confidence. *NeuroImage*, *99*, 323–331. https://doi.org/10.1016/j.neuroimage.2014.05.049

Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2016). *Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity*. *48*, 45–53. https://doi.org/10.1016/j.lindif.2016.02.006

Ho, H., Senturk, D., Lam, A. G., Zimmer, J. M., Hong, S., Okamoto, Y., Chiu, S.-Y., Nakazawa, Y., & Wang, C.-P. (2000). The affective and cognitive dimensions of math anxiety : A cross-national study. *Journal of Research in Mathematics Education*, *31*(3), 362–379.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*, 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), 9–15. https://doi.org/10.1111/j.1467-7687.2009.00848.x

Houdé, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: An fMRI meta-analysis of 52 studies including 842 children. *Developmental Science*, *13*(6), 876–885. https://doi.org/10.1111/j.1467-7687.2009.00938.x

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

Hunt, T. E., Bhardwa, J., & Sheffield, D. (2017). Mental arithmetic performance, physiological reactivity and mathematics anxiety amongst U.K. primary school children. *Learning and Individual Differences*, *57*, 129–132. https://doi.org/10.1016/J.LINDIF.2017.03.016

Imbo, I., & Vandierendonck, A. (2008). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychological Research*, *72*, 331–346. https://doi.org/10.1007/s00426-007-0112-8

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin and Review*, *18*(6), 1222–1229. https://doi.org/10.3758/s13423-011-0154-1

JASP. (2019). *JASP* (Version 0.11.1 [Computer software]). https://jasp-stats.org/

Kabinet Vlaams minister van Onderwijs, . (2017). *Dalende trend resultaten wiskunde basisonderwijs vraagt om verder onderzoek [Decreases in children's mathematical performance in primary school requires further investigation]*. https://onderwijs.vlaanderen.be/nl/dalende-trend-resultaten-wiskunde-basisonderwijs-vraagt-om-verder-onderzoek.

Kaufmann, L., Koppelstaetter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimmerhackl, L. B., Felber, S., & Ischebeck, A. (2006). Neural correlates of the number-size interference task in children. *NeuroReport*, *17*(6), 587–591. https://doi.org/10.1097/00001756-200604240-00007

Kaufmann, L., Wood, G., Rubinsten, O., & Henik, A. (2011). Meta-analyses of developmental fMRI studies investigating typical and atypical trajectories of number processing and calculation. *Developmental Neuropsychology*, *36*(6), 763–787. https://doi.org/10.1080/87565641.2010.549884

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., Sasaki, T., Sugiura, M., Watanabe, J., & Fukuda, H. (2004). A functional MRI study of simple arithmetic - A comparison between children and adults. *Cognitive Brain Research*, *18*, 227–233. https://doi.org/10.1016/j.cogbrainres.2003.10.009

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

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. National Academy Press. https://doi.org/10.17226/9822

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction*, *25*, 95–103. https://doi.org/10.1016/j.learninstruc.2012.12.001

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press. https://doi.org/10.2307/1422432

Koriat, A., & Ackerman, R. (2010). Choice latency as a cue for children's subjective confidence in the correctness of their answers. *Developmental Science*, *13*(3), 441–453. https://doi.org/10.1111/j.1467-7687.2009.00907.x

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(5), 1133–1145. https://doi.org/10.1037/0278-7393.32.5.1133

Koriat, A., & Levy-sadot, R. (1999). Processes underlying metacognitive judgments: Information- based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 483–502). Guilford.

Krinzinger, H., Kaufmann, L., & Willmes, K. (2009). Math anxiety and math ability in early primary school years. *Journal of Psychoeduc Assessment*, *27*(3), 206–225. https://doi.org/10.1177/0734282908330583

Kucian, K., Von Aster, M., Loenneker, T., Dietrich, T., & Martin, E. (2008). Development of neural networks for exact and approximate calculation: A fMRI study. *Developmental Neuropsychology*, *33*(4), 447–473. https://doi.org/10.1080/87565640802101474

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

Kuhn, D. (2001). How do people know? *Psychological Science*, *12*(1), 1–8. https://doi.org/10.1111/1467-9280.00302

Lai, Y., Zhu, X., Chen, Y., & Li, Y. (2015). Effects of mathematics anxiety and mathematical metacognition on word problem solving in children with and without mathematical learning difficulties. *PLOS ONE*, *10*(6), e0130570. https://doi.org/10.1371/journal.pone.0130570

Laski, E. V, & Dulaney, A. (2015). When prior knowledge interferes, inhibitory control matters for learning: The case of numerical magnitude representations. *Journal of Educational Psychology*, *107*(4), 1035–1050. https://doi.org/http://dx.doi.org/10.1037/edu0000034

Lee, K., & Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *Journal of Educational Psychology*, *108*(6), 869–882. https://doi.org/10.1037/edu0000090

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development*, *84*(6). https://doi.org/10.1111/cdev.12096

Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology*, *82*, 82–99. https://doi.org/10.1111/j.2044-8279.2010.02016.x

Legg, A. M., & Locker, L. (2009). Math performance and its relationship to math anxiety and metacognition. *North American Journal of Psychology*, *11*(3), 471–486.

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy trade-offs(?). *Behavior Research Methods*, *51*(1), 40–60. https://doi.org/10.3758/s13428-018-1076-x

Lingel, K., Lenhart, J., & Schneider, W. (2019). Metacognition in mathematics: do different metacognitive monitoring measures make a difference? *ZDM*, *51*, 587–600. https://doi.org/10.1007/s11858-019-01062-8

Linsen, S., Verschaffel, L., Reynvoet, B., & De Smedt, B. (2015). The association between numerical magnitude processing and mental versus algorithmic multi-digit subtraction in children. *Learning and Instruction*, *35*, 42–50. https://doi.org/10.1016/j.learninstruc.2014.09.003

Löffler, E., Von Der Linden, N., & Schneider, W. (2016). Influence of domain knowledge on monitoring performance across the life span. *Journal of Cognition and Development*, *17*(5), 765–785. https://doi.org/10.1080/15248372.2016.1208204

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *Journal of Neurolinguistics*, *24*(5), 583–591. https://doi.org/10.1016/j.jneuroling.2011.02.004

Lucangeli, D., & Cornoldi, C. (1997). Mathematics and metacognition: What is the nature of the relationship? *Mathematical Cognition*, *3*(2), 121–139. https://doi.org/10.1080/135467997387443

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General*, *141*(4), 635–641. https://doi.org/10.1037/a0027248

Lyons, K. E., & Ghetti, S. (2010). Metacognitive development in early childhood: New questions about old assumptions. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in metacognition research* (pp. 259–278). Springer, US. https://doi.org/10.1007/978-1-4419-6546-2_11

Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development*, *84*(2), 726–736. https://doi.org/10.1111/cdev.12004

Lyons, K. E., & Zelazo, P. D. (2011). Monitoring, metacognition, and executive function: Elucidating the role of self-reflection in the development of self-regulation. In B. B. Janette (Ed.), *Advances in child development and behovior* (pp. 379–412). Elsevier.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*(5), 520–540.

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, *28*(1), 26–47.

Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence*, *27*, 165–179. https://doi.org/10.1016/j.adolescence.2003.11.003

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, *16*(8), 404–406. https://doi.org/10.1016/j.tics.2012.06.008

Maloney, E. A., Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2015). Intergenerational effects of parents' math anxiety on children's math achievement and anxiety. *Psychological Science*, *26*(9), 1480–1488. https://doi.org/10.1177/0956797615592630

Mammarella, I. C., Caviola, S., & Dowker, A. (Eds.). (2019). *Mathematics anxiety. What is known and what is still to be understood*. Routledge.

Mammarella, I. C., Hill, F., Devine, A., Caviola, S., & Szűcs, D. (2015). Math anxiety and developmental dyscalculia: A study on working memory processes. *Journal of Clinical and Experimental Neuropsychology*, *37*(8), 878–887. https://doi.org/10.1080/13803395.2015.1066759

Martens, F., Bulthé, J., van Vliet, C., & Op de Beeck, H. (2018). Domain-general and domain-specific neural changes underlying visual expertise. *NeuroImage*, *169*, 80–93. https://doi.org/10.1016/j.neuroimage.2017.12.013

Matejko, A. A., & Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. *PLoS ONE*, *11*(3), 1–15. https://doi.org/10.1371/journal.pone.0149863

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324. https://doi.org/10.3758/s13428-011-0168-7

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23–44. https://doi.org/10.1037/1082-989X.12.1.23

Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model. *Multivariate Behavioural Research*, *46*(5), 816–841.

Mazzocco, M. M. M., & Kover, S. T. (2007). A longitudinal assessment of executive function skills and their association with math performance. *Child Neuropsychology*, *13*(1), 18–45. https://doi.org/10.1080/09297040600611346

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Yuet Liu, K., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *The Journal of Neuroscience*, *35*(5), 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744

Menon, V. (2015). Arithmetic in the Child and Adult Brain. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford Handbook of Numerical Cognition*. https://doi.org/10.1093/oxfordhb/9780199642342.013.041

Merkley, R., Matejko, A. A., & Ansari, D. (2017). Strong causal claims require strong evidence: A commentary on Wang and colleagues. *Journal of Experimental Child Psychology*, *153*, 163–167. https://doi.org/10.1016/j.jecp.2016.07.008

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition*. The MIT Press.

Meyniel, F., & Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *PNAS, 114*(9), E3859–E3868. https://doi.org/10.1073/pnas.1615773114

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Moelands, F., & Rymenans, R. (2003). *Schaal vorderingen in spellingvaardigheid voor Vlaanderen (SVS-V): handleiding*. Citogroep.

Molenberghs, P., Trautwein, F.-M., Böckler, A., Singer, T., & Kanske, P. (2016). Neural correlates of metacognitive ability and of feeling confident: A large-scale fMRI study. *Social Cognitive and Affective Neuroscience*, *11*(12), 1942–1951. https://doi.org/10.1093/scan/nsw093

Moore, D., & Healy, P. J. (2008). The trouble with overconfidence. *Psychiatric Review*, *115*(2), 502–517. https://doi.org/10.1192/pb.33.10.396a

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326.

Morsanyi, K., Cheallaigh, N. N., & Ackerman, R. (2019). Mathematics anxiety and metacognitive processes: Proposal for a new line of inquiry. *Psychological Topics*, *28*(1), 147–169. https://doi.org/10.31820/pt.28.1.8

Moser, J. S., Moran, T. P., Schroder, H. S., Donnellan, M. B., & Yeung, N. (2013). On the relationship between anxiety and error monitoring: a meta-analysis and conceptual framework. *Frontiers in Human Neuroscience*, *7*, 466. https://doi.org/10.3389/fnhum.2013.00466

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*, 490–502. https://doi.org/10.1016/j.jecp.2009.02.003

Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research*, *89*(3), 459–496. https://doi.org/10.3102/0034654319843494

Narayanan, N. S., Prabhakaran, V., Bunge, S. A., Christoff, K., Fine, E. M., & Gabrieli, J. D. E. (2005). The role of the prefrontal cortex in the maintenance of verbal working memory: An event-related fMRI analysis. *Neuropsychology*, *19*(2), 223–232. https://doi.org/10.1037/0894-4105.19.2.223

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

Neuenhaus, N., Artelt, C., Lingel, K., & Schneider, W. (2011). Fifth graders metacognitive knowledge: General or domain-specific? *European Journal of Psychology of Education*, *26*(2), 163–178. https://doi.org/10.1007/s10212-010-0040-7

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE*, *8*(7), e67918. https://doi.org/10.1371/journal.pone.0067918

Nunes, T., Bryant, P., & Sylva, K. (2009). Development of Maths Capabilities and Primary School. In *DCSF Research report RR118*.

Onderwijsdoelen Vlaanderen. (2018). https://www.onderwijsdoelen.be/wiskunde-lager-onderwijs

Opdenakker, M.-C., & Van Damme, J. (2007). Do school context, student composition and school leadership affect school practice and outcomes in secondary education? *British Educational Research Journal*, *33*(2), 179–206. https://doi.org/10.1080/01411920701208233

Özcan, Z. C., & Gümüs, A. E. (2019). A modeling study to explain mathematical problem-solving performance through metacognition, self-efficacy, motivation, and anxiety. *Australian Journal of Education*, *63*(1), 116–134.

Özsoy, G. (2011). An investigation of the relationship between metacognition and mathematics achievement. *Asia Pacific Education Review*, *12*(2), 227–235. https://doi.org/10.1007/s12564-010-9129-6

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, *15*(3), 105–130. https://doi.org/10.1007/s11065-005-7091-6

Park, D., Ramirez, G., & Beilock, S. L. (2014). The role of expressive writing in math anxiety. *Journal of Experimental Psychology: Applied*, *20*(2), 103–111. https://doi.org/10.1037/xap0000013

Passolunghi, M. C., Caviola, S., De Agostini, R., Perin, C., & Mammarella, I. C. (2016). Mathematics anxiety, working memory, and mathematics performance in secondary-school children. *Frontiers in Psychology*, *7*(42), 1–8. https://doi.org/10.3389/fpsyg.2016.00042

Passolunghi, M. C., Mammarella, I. C., & Altoè, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Developmental Neuropsychology*, *33*(3), 229–250. https://doi.org/10.1080/87565640801982320

Pelegrina, S., Lechuga, M. T., García-Madruga, J. A., Elosúa, M. R., Macizo, P., Carreiras, M., Fuentes, L. J., & Bajo, M. T. (2015). Normative data on the n-back task for children and young adolescents. *Frontiers in Psychology*, *6*, 1–11. https://doi.org/10.3389/fpsyg.2015.01544

Peng, P., & Kievit, R. A. (2020). The development of academic achievement and cognitive abilities: A bidirectional perspective. *Child Development Perspectives*. https://doi.org/10.1111/cdep.12352

Peng, P., Namkung, J., Barnes, M. A., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*(4), 455–473. https://doi.org/10.1037/edu0000079

Penner-Wilger, M., & Lefevre, J.-A. (2006). Decomposing the m: Using distributional analyses to provide a detailed description of addition and multiplication latencies. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *28*(28), 1944–1949.

Peters, G., De Smedt, B., Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2014). Subtraction by addition in children with mathematical learning disabilities. *Learning and Instruction*, *30*, 1–8. https://doi.org/10.1016/j.learninstruc.2013.11.001

Peters, L., & Ansari, D. (2019). Are specific learning disorders truly specific, and are they disorders? *Trends in Neuroscience and Education*, *17*, 100115. https://doi.org/10.1016/j.tine.2019.100115

Peters, L., Bulthé, J., Daniels, N., Op de Beeck, H., & De Smedt, B. (2018). Dyscalculia and dyslexia: Different behavioral, yet similar brain activity profiles during arithmetic. *NeuroImage: Clinical*, *18*, 663–674. https://doi.org/10.1016/j.nicl.2018.03.003

Peters, L., & De Smedt, B. (2017). Arithmetic in the developing brain: A review of brain imaging studies. *Developmental Cognitive Neuroscience*, *30*, 265–279. https://doi.org/10.1016/j.dcn.2017.05.002

Petronzi, D., Staples, P., Sheffield, D., Hunt, T. E., & Fitton-Wilde, S. (2019). Further development of the Children’s Mathematics Anxiety Scale UK (CMAS-UK) for ages 4–7 years. *Educational Studies in Mathematics*, *100*, 231–249. https://doi.org/10.1007/s10649-018-9860-1

Pieschl, S. (2009). Metacognitive calibration - An extended conceptualization and potential applications. *Metacognition and Learning*, *4*, 3–31. https://doi.org/10.1007/s11409-008-9030-4

Poldrack, R. A., Nichols, T., & Mumford, J. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge University Press. https://doi.org/10.1017/cbo9780511895029

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*(1), 185–227. https://doi.org/10.1080/00273170701341316

Punaro, L., & Reeve, R. (2012). Relationships between 9-year-olds’ math and literacy worries and academic abilities. *Child Development Research*. https://doi.org/10.1155/2012/359089

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*, 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Rahnev, D., & Fleming, S. M. (2019). How experimental procedures influence estimates of metacognitive ability. *Neuroscience of Consciousness*, *5*(1), 1–9. https://doi.org/10.1093/nc/niz009

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. *Journal of Experimental Child Psychology*, *141*, 83–100. https://doi.org/10.1016/j.jecp.2015.07.014

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, *14*(2), 187–202. https://doi.org/10.1080/15248372.2012.664593

Ramirez, G., Shaw, S. T., & Maloney, E. A. (2018). Math anxiety: Past research, promising interventions, and a new interpretation framework. *Educational Psychologist*, *53*(3), 145–164. https://doi.org/10.1080/00461520.2018.1447384

Raven, C. J., Court, J. H., & Raven, J. (1992). *Standard progressive matrices.* Oxford Psychologist Press.

Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, *19*(6), 551–554. https://doi.org/10.1037/h0033456

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), e98663. https://doi.org/10.1371/journal.pone.0098663

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. https://doi.org/10.1177/0956797612466268

Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, *15*(11), 1779–1790. https://doi.org/10.1093/cercor/bhi055

Robinson, C. S., Menchetti, B. M., & Torgesen, J. K. (2002). Toward a two-factor theory of one type of mathematics disabilities. *Learning Disabilities Research and Practice*, *17*(2), 81–89. https://doi.org/10.1111/1540-5826.00035

Roebers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45*, 31–51. https://doi.org/https://doi.org/10.1016/j.dr.2017.04.001

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

Roebers, C. M., & Feurer, E. (2016). Linking executive functions and procedural metacognition. *Child Development Perspectives*, *10*(1), 39–44. https://doi.org/10.1111/cdep.12159

Roebers, C. M., Krebs, S. S., & Roderer, T. (2014). Metacognitive monitoring and control in elementary school children: The interrelations and their role for test performance. *Learning and Individual Differences*, *29*, 141–149. https://doi.org/10.1016/j.lindif.2012.12.003

Roebers, C. M., & Spiess, M. A. (2017). The development of metacognitive monitoring and control in second graders: A short-term longitudinal study. *Journal of Cognition and Development*, *18*(1), 110–128. https://doi.org/10.1080/15248372.2016.1157079

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Rossnan, S. (2006). Overcoming math anxiety. *Mathitudes*, *1*(1), 1–4. https://doi.org/10.1090/mbk/062/04

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*(3), 418–431. https://doi.org/10.1016/j.jecp.2012.10.012

Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind, Brain, and Education*, *6*(3), 119–128. https://doi.org/10.1111/j.1751-228X.2012.01147.x

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science, 20*(3), 1–16. https://doi.org/10.1111/desc.12372

Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, *1*(2), 291–297. https://doi.org/10.1111/1467-7687.00044

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education, 2*(3), 114–121. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Schneider, W. (2010). The development of metacognitive competences. In B. M. Glatzeder, A. von Müller, & V. Goel (Eds.), *Towards a theory of thinking* (pp. 203–214). Springer-Verlag.

Schneider, W. (2015a). *Memory development from early childhood through emerging adulthood*. Springer. https://doi.org/10.1007/978-3-319-09611-7

Schneider, W. (2015b). Metacognitive development: Educational implications. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (Second Edi, Vol. 15). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.92011-7

Schneider, W., & Artelt, C. (2010). Metacognition and mathematics education. *ZDM - International Journal on Mathematics Education*, *42*, 149–161. https://doi.org/10.1007/s11858-010-0240-2

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime computer software and manual*. Psychology Software Tools.

Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: Evidence for developmental trends. In J. D. R. A. Bjork (Ed.), *Handbook of metamemory and memory* (pp. 391–409). Psychology Press.

Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 491–518). Oxford University Press.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 189–215). Erlbaum.

Schoenfeld, A. H. (1992). Learning to think mathematically. Problem solving, metacognition and sense making in mathematics. In *Handbook for research on mathematics teaching and learning* (pp. 334–370). Macmillan.

Schoenfeld, A. H. (2004). The Math Wars. *Educational Policy*, *18*(1), 253–286. https://doi.org/10.1177/0895904803260042

Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology*, *19*, 143–154. https://doi.org/10.1006/ceps.1994.1013

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, *36*, 111–139. https://doi.org/10.1007/s11165-005-3917-8

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. https://doi.org/DOI:10.1037/0022-0663.87.3.433

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

Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology*, *90*(2), 236–248. https://doi.org/10.1037/0022-0663.90.2.236 PDF Download PDF Cite Cite Email Print All Options Full text Full text - PDF Abstract/Details References 52

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, *48*(2), 630–633. https://doi.org/https://doi.org/10.2307/1128664

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of Developmental Research Methods.* (pp. 265–278).

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, *6*(2–3), 144–164.

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition*, *9*, 313–323. https://doi.org/10.1006/ccog.2000.0450

Shrager, J., & Siegler, R. S. (1998). A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*(5), 405–410.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*(3), 250–264.

Siegler, R. S. (1996). *Emerging Minds. The process of change in children's thinking*. Oxford University Press.

Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, *19*(3). https://doi.org/10.1111/desc.12395

Simanowski, S., & Krajewski, K. (2019). Specific preschool executive functions predict unique aspects of mathematics development: A 3-year longitudinal study. *Child Development*, *90*(2), 544–561. https://doi.org/10.1111/cdev.12909

Simons, J. S., Peers, P. V, Mazuz, Y. S., Berryhill, M. E., & Olson, I. R. (2010). Dissociation between memory accuracy and memory confidence following bilateral parietal lesions. *Cerebral Cortex*, *20*, 479–485. https://doi.org/10.1093/cercor/bhp116

Skemp, R. R. (1986). *The psychology of learning mathematics*. Penguin books.

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Dowker, A., & Aro, M. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *British Journal of Educational Psychology*, *87*(3), 309–327. https://doi.org/10.1111/bjep.12151

Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., Tolvanen, A., & Aro, M. (2019). Development of math anxiety and its longitudinal relationships with arithmetic achievement among primary school children. *Learning and Individual Differences*, *69*, 173–181. https://doi.org/10.1016/j.lindif.2018.12.005

Souchay, C., Moulin, C. J. A., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition*, *16*(4), 769–784. https://doi.org/10.1016/j.concog.2006.11.002

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2009). Affect 4.0. *Experimental Psychology*, *57*(1), 36–45. https://doi.org/10.1027/1618-3169/a000005

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

Stillman, G., & Mevarech, Z. (2010). Metacognition research in mathematics education: From hot topic to mature field. *ZDM - International Journal on Mathematics Education*, *42*(2), 145–148. https://doi.org/10.1007/s11858-010-0245-x

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, *29*(4), 581–593. https://doi.org/10.1177/0956797617741719

Suinn, R. M., & Edwards, R. (1982). The measurement of mathematics anxiety: The mathematics anxiety rating scale for adolescents—MARS-A. *Journal of Clinical Psychology*. https://doi.org/10.1002/1097-4679(198207)38:3<576::AID-JCLP2270380317>3.0.CO;2-V

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

Szűcs, D., & Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: A systematic review. *Trends in Neuroscience and Education*, *6*, 187–203. https://doi.org/10.1016/j.tine.2016.11.002

Tobias, S. (1986). Anxiety and cognitive processing of instruction. In R. Schwarzer (Ed.), *Self-related cognitions in anxiety and motivation* (pp. 35–54). Lawrence Erlbaum associates.

Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances*, *2*, 1–14. https://doi.org/10.1177/2398212818810591

van Bon, W., & Kuijpers, C. (2016). Beter leren rekenen gaat samen met grotere zekerheid, beter leren spellen met meer twijfel. *Pedagogiek, 36*(1), 49–70. https://doi.org/10.5117/PED2016.1.VBON

Van der Maas, H. L. J., van der Ven, S., & van der Molen, V. (2014). Oefenen op niveau: Het cijferspel in de rekentuin. *Volgens Bartjens*, *3*, 12–15.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*(5), 427–449. https://doi.org/10.1016/j.intell.2006.09.001

van der Stel, M., & Veenman, M. V. J. (2010). Development of metacognitive skillfulness: A longitudinal study. *Learning and Individual Differences*, *20*(3), 220–224. https://doi.org/10.1016/J.LINDIF.2009.11.005

van der Stel, M., Veenman, M. V. J., Deelen, K., & Haenen, J. (2010). The increasing role of metacognitive skills in math: A cross-sectional study from a developmental perspective. *ZDM - International Journal on Mathematics Education*, *42*(2), 219–229. https://doi.org/10.1007/s11858-009-0224-2

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics : A dynamic relationship. *British Journal of Educational Psychology*, *82*, 100–119. https://doi.org/10.1111/j.2044-8279.2011.02035.x

van Kraayenoord, C. E., & Schneider, W. E. (1999). Reading achievement, metacognition, reading self-concept and interest: A study of German students in grades 3 and 4. *European Journal of Psychology of Education*, *14*(3), 305–324. https://doi.org/10.1007/BF03173117

Vanbinst, K., Ceulemans, E., Ghesquière, P., & De Smedt, B. (2015). Profiles of children's arithmetic fact development: A model-based clustering approach. *Journal of Experimental Child Psychology*, *133*, 29–46. https://doi.org/10.1016/j.jecp.2015.01.003

Vanbinst, K., Ceulemans, E., Peters, L., Ghesquière, P., & De Smedt, B. (2018). Developmental trajectories of children's symbolic numerical magnitude processing skills and associated cognitive competencies. *Journal of Experimental Child Psychology*, *166*, 232–250. https://doi.org/10.1016/j.jecp.2017.08.008

Vanbinst, K., & De Smedt, B. (2016a). Individual differences in children's mathematics achievement: The roles of symbolic numerical magnitude processing and domain-general cognitive functions. *Progress in Brain Research*, *227*, 105–130. https://doi.org/10.1016/bs.pbr.2016.04.001

Vanbinst, K., & De Smedt, B. (2016b). Individual differences in children's mathematics achievement: The roles of symbolic numerical magnitude processing and domain-general cognitive functions. In M. Cappelletti & W. Fias (Eds.), *The mathematical brain across the lifespan* (pp. 105–130). Elsevier.

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain, and Education*, *6*(3), 129–136. https://doi.org/10.1111/j.1751-228X.2012.01148.x

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2015). Does numerical processing uniquely predict first graders' future development of single-digit arithmetic? *Learning and Individual Differences*, *37*, 153–160. https://doi.org/10.1016/j.lindif.2014.12.004

Vanbinst, K., Ghesquière, P., & De Smedt, B. (2019). Is the long-term association between symbolic numerical magnitude processing and arithmetic bi-directional? *Journal of Numerical Cognition*, *5*(3), 358–370. https://doi.org/https://doi.org/10.5964/jnc.v5i3.202

Vanbinst, K., van Bergen, E., Ghesquière, P., & De Smedt, B. (2020). Cross-domain associations of key cognitive correlates of early reading and early arithmetic in 5-year-olds. *Early Childhood Research Quarterly*, *51*, 144–152. https://doi.org/10.1016/j.ecresq.2019.10.009

Veenman, M. V. J., Elshout, J. J., & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction*, *7*(2), 187–209.

Veenman, M. V. J., Kerseboom, L., & Imthorn, C. (2000). Test anxiety and metacognitive skillfulness: Availability versus production deficiencies. *Anxiety, Stress and Coping*, *13*(4), 391–412. https://doi.org/https://doi.org/10.1080/10615800008248343

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

Veenman, M. V. J., & Van Cleef, D. (2019). Measuring metacognitive skills for mathematics: students' self-reports versus on-line assessment methods. *ZDM*, *51*, 691–701. https://doi.org/10.1007/s11858-018-1006-5

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*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills in early adolescence. *Instructional Science*, *14,* 89–109. https://doi.org/10.1016/j.learninstruc.2003.10.004

Verguts, T., & Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory & Cognition*, *33*(1), 1–16. https://doi.org/10.3758/BF03195293

Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). Information Age Publishing.

Vo, V. A., Li, R., Kornell, N., Pouget, A., & Cantlon, J. F. (2014). Young children bet on their numerical skills: Metacognition in the numerical domain. *Psychological Science*, *25*(9), 1712–1721. https://doi.org/10.1177/0956797614538458

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. *Contemporary Educational Psychology*, *38*, 1–10. https://doi.org/10.1016/J.CEDPSYCH.2012.09.001

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research*, *84*(1), 30–43.

Ward, J. (2015). *The Student's Guide to Cognitive Neuroscience* (Third edit). Psychology Press. https://doi.org/10.4324/9781315742397

Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(1), 1–14.

Watson, T. D., Azizian, A., & Squires, N. K. (2006). Event-related potential correlates of extradimensional and intradimensional set-shifts in a modified Wisconsin Card Sorting Test. *Brain Research*, *1092*(1), 138–151. https://doi.org/10.1016/j.brainres.2006.03.098

Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secundary school students. *Journal of Educational Psychology*, *80*(2), 210–216.

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment*, *24*(1), 226–239. https://doi.org/10.1037/a0025361

Wood, G., Pinheiro-Chagas, P., Júlio-Costa, A., Micheli, L. R., Krinzinger, H., Kaufmann, L., Willmes, K., & Haase, V. G. (2012). Math anxiety questionnaire: Similar latent structure in Brazilian and German school children. *Child Development Research*. https://doi.org/10.1155/2012/610192

Wu, S. S., Barth, M., Amin, H., Malcarne, V., & Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematics achievement. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00162

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670. https://doi.org/10.1038/nmeth.1635

Yeniad, N., Malda, M., Mesman, J., Van Ijzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, *23*, 1–9. https://doi.org/10.1016/j.lindif.2012.10.004

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Yokoyama, O., Miura, N., Watanabe, J., Takemoto, A., Uchida, S., Sugiura, M., Horie, K., Sato, S., Kawashima, R., & Nakamura, K. (2010). Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neuroscience Research*, *68*(3), 199–206. https://doi.org/10.1016/j.neures.2010.07.2041

Young, C. B., Wu, S. S., & Menon, V. (2012). The neurodevelopmental basis of math anxiety. *Psychological Science*, *23*(5), 492–501.

Yussen, S. R., & Levy, V. M. (1975). Developmental changes in predicting one's own span of short-term memory. *Journal of Experimental Child Psychology*, *19*, 502–508. https://doi.org/10.1016/0022-0965(75)90079-X

List of Publications

List of publications

Bibliography

Journal articles (peer reviewed)

Accepted Journal articles (peer reviewed)

Bellon, E., Fias, W., De Smedt, B. (2020). Metacognition across domains: Is the association between arithmetic and metacognitive monitoring domain-specific? *Plos One*. doi: 10.1371/journal.pone.0229932 (Impact factor: 2.78)

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. doi: 10.1016/j.jecp.2019.01.012 (citations: 1) (Impact factor: 2.98) Open Access

Bellon, E., Fias, W., De Smedt, B. (2016). Are Individual Differences in Arithmetic Fact Retrieval in Children Related to Inhibition? *Frontiers in Psychology*, *7*, Art.No. 825. doi: 10.3389/fpsyg.2016.00825 (citations: 7) (Impact factor: 2.13) Open Access

Under revision - Journal articles (peer reviewed)

Bellon, E., Fias, W., Ansari, D. & De Smedt, B. (Under Revision). The neurobiological basis of metacognitive monitoring during arithmetic in the developing brain. *Human Brain Mapping.* (Impact factor: 4.554)

Bellon, E., Fias, W., & De Smedt, B. (Under Revision). Too anxious to be confident? A panel longitudinal study into the interplay of metacognition and mathematics anxiety in arithmetic. *Journal of Educational Psychology*. (Impact factor: 5.178)

Bellon, E., Fias, W., & De Smedt, B. (Under Revision). What does arithmetic count on?: A longitudinal panel study on the roles of numerical magnitude processing, executive functions and metacognition in primary school. *Journal of Numerical Cognition* (No impact factor yet)

Vanbinst, K., Bellon, E., & Dowker, A. (Under Revision). Mathematics anxiety: An intergenerational approach. *Frontiers in Psychology*. (Impact factor: 2.13)

Abstracts/Presentations/Posters

Bellon, E., Fias, W., De Smedt, B. (2019). Online metacognition in children's arithmetic. Presented at the 2nd annual meeting of the Mathematical cognition and learning society (MCLS), Ottawa, Ontario, Canada, 16 Jun 2019-18 Jun 2019.

Bellon, E. (2019). Online metacognition in children's arithmetic. Presented at Research meeting Numerical cognition - Brain and Mind institute (BMI), Western University, London, Ontario, Canada, 13 Jun 2019.

Bellon, E., Fias, W., De Smedt, B. (2019). Neural correlates of metacognitive monitoring in arithmetic. Presented at the Meeting on Heterogeneous Contributions to Numerical Cognition, Ghent, Belgium.

Bellon, E., Fias, W., De Smedt, B. (2019). Online metacognition in children's arithmetic. Presented at the Annual meeting of the Belgian Association for Psychological Sciences (BAPS), Liège, Belgium.

Bellon, E., Fias, W., De Smedt, B. (2019). Online metacognition in children's arithmetic. Presented at the Groupe de contact annual meeting, Blankenberge, Belgium, 09 May 2019-10 May 2019.

Bellon, E., Fias, W., De Smedt, B. (2019). Neural correlates of metacognitive monitoring in arithmetic. Presented at the Leuven Brain Institute meeting, Leuven, Belgium.

Bellon, E. (2018). Numerical magnitude processing, executive functions and metacognition in children’s arithmetic. Presented at Research meeting Numerical cognition - Brain and Mind institute (BMI), Western University, London, Ontario, Canada, 18 Oct 2018.

Bellon, E., Fias, W., De Smedt, B. (2018). The additional role of cognitive control and metacognition in arithmetic. Presented at the International Mind, Brain and Education Society Conference, Los Angeles, California, 27 Sep 2018-29 Sep 2018.

Bellon, E., Fias, W., De Smedt, B. (2018). More than number sense: The additional role of cognitive control and metacognition in arithmetic. Presented at the Junior Researchers of Earli (JURE) annual conference, Antwerp, Belgium.

Bellon, E., Fias, W., De Smedt, B. (2018). Who is ‘typically developing’? - The impact of different categorizations of children’s math performance on research conclusions within a longitudinal study on arithmetic. Presented at the 42nd annual International Academy for Research in Learning Disabilities (IARLD) Conference, Ghent, Belgium, 02 Jul 2018-03 Jul 2018.

Bellon, E., Fias, W., De Smedt, B. (2018). More than number sense: the additional role of cognitive control and metacognition in arithmetic. Presented at the Mathematical Cognition and Learning Society Conference (MCLS), Oxford, United Kingdom, 08 Apr 2018-09 Apr 2018.

Bellon, E., Fias, W., De Smedt, B. (2017). Learning from your mistakes: Associations between cognitive control, metacognition and arithmetic. Presented at the Earli 2017 - Education in the crossroads of economy and politics - Role of research in the advancement of public good, Tampere, Finland, 29 Aug 2017-02 Sep 2017.

Bellon, E., Fias, W., De Smedt, B. (2017). Learning from your mistakes in arithmetic: Associations between cognitive control, metacognition and arithmetic ability. Presented at the Summer School:

Cognitive control and consciousness: behavioural and neural mechanisms, Weggis, Switzerland, 19 Jun 2017-22 Jun 2017.

Bellon, E., Fias, W., De Smedt, B. (2017). Learning from your mistakes in arithmetic: Associations between cognitive control, metacognition and arithmetic ability. Presented at the Improving mathematical cognition and learning: formal and informal instructional influences & interventions, Nashville (TN), USA, 15 May 2017-16 May 2017.

Bellon, E., Fias, W., De Smedt, B. (2017). More than number sense: cognitive control & metacognition in arithmetic. Presented at the Advances in Numerical Cognition, Tournai, Belgium, 04 May 2017-05 May 2017.

Bellon, E., Fias, W., De Smedt, B. (2017). Learning from your mistakes during fact retrieval: the association between cognitive control, calibration of confidence, number processing and single-digit addition and multiplication. Presented at the SRCD biennial meeting, Austin (TX), USA, 06 Apr 2017-08 Apr 2017.

Bellon, E., Fias, W., De Smedt, B. (2016). Are individual differences in arithmetic fact retrieval related to inhibition? Presented at the Domain-general and domain-specific foundations of numerical and arithmetic processing, Tübingen, Germany, 28 Sep 2016-30 Sep 2016.

Bellon, E., Fias, W., De Smedt, B. (2016). Are individual differences in arithmetic fact retrieval related to inhibition? Presented at the EARLI SIG 15, Leuven, Belgium, 29 Aug 2016-30 Aug 2016.

Bellon, E., Fias, W., De Smedt, B. (2016). Are individual differences in arithmetic fact retrieval related to inhibition? Presented at the SIG 22 "Neuroscience and Education", Amsterdam, The Netherlands, 23 Jun 2016-25 Jun 2016.

Bellon, E., Fias, W., De Smedt, B. (2016). Are individual differences in arithmetic fact retrieval related to inhibition? Presented at the Heterogeneous Contributions to Numerical Cognition, Ghent, Belgium, 01 Jun 2016-03 Jun 2016.

Science Outreach

Bellon, E. (2019). Is zelfkennis het begin van alle rekenknobbels? (https://www.eoswetenschap.eu/psyche-brein/zelfkennis-het-begin-van-alle-rekenknobbels)

Bellon, E. (2019). Is self-knowledge the beginning of all math nodes? (https://www.sciencefiguredout.be/self-knowledge-beginning-all-math-nodes)