Educational Research Review 10 (2013) 29–44

Contents lists available at SciVerse ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/EDUREV

Review

Working memory and mathematics in primary school children: A meta-analysis

Ilona Friso-van den Bos [a,*], Sanne H.G. van der Ven [b], Evelyn H. Kroesbergen [a], Johannes E.H. van Luit [a]

[a] Department of Educational and Learning Sciences, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands
[b] Department of Psychological Research Methods, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 3 December 2012
Revised 6 May 2013
Accepted 6 May 2013
Available online 16 May 2013

Keywords:
Mathematics
Working memory
Executive functions
Children
Meta-analysis

ABSTRACT

Working memory, including central executive functions (inhibition, shifting and updating) are factors thought to play a central role in mathematical skill development. However, results reported with regard to the associations between mathematics and working memory components are inconsistent. The aim of this meta-analysis is twofold: to investigate the strength of this relation, and to establish whether the variation in the association is caused by tests, sample characteristics and study and other methodological characteristics. Results indicate that all working memory components are associated with mathematical performance, with the highest correlation between mathematics and verbal updating. Variation in the strength of the associations can consistently be explained by the type of mathematics measure used: general tests yield stronger correlations than more specific tests. Furthermore, characteristics of working memory measures, age and sample explain variance in correlations in some analyses. Interpretations of the contribution of moderator variables to various models are discussed.

Contents

* Corresponding author. Address: Department of Pedagogical and Educational Sciences, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands. Tel.: +31 30 253 3850.
E-mail address: i.vandenbos@uu.nl (I. Friso-van den Bos).

1747-938X/$ - see front matter
http://dx.doi.org/10.1016/j.edurev.2013.05.003

1. Introduction

Solving mathematical problems is an important activity in the lives of children. From an early age onwards, learning to count, acquiring number skills, and performing mathematical operations become part of children's daily activities. These activities remain important throughout a person's life. Research concerning the underpinnings of success and deficiencies in mathematical skills has expanded during the past two decades. A number of cognitive mechanisms underlying these mathematical skills have been proposed and their contribution to the development of mathematical skill has been investigated. A factor that is thought to play a central role in mathematic skill development is the capacity and the efficiency of working memory and the executive functions: inhibition, shifting and updating (e.g., Bull and Scerif, 2001; Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Passolunghi, Mammarella, & Altoè, 2008; St Clair-Thompson and Gathercole, 2006). Working memory capacity is frequently used as a predictor of skills in mathematics at a later point in time (see: LeFevre, DeStefano, Coleman, & Shanahan, 2005). The number of studies in which the predictive value of working memory and executive functions for mathematical performance is investigated has increased sharply, but the pattern of results is inconsistent: mathematics performance is not consistently predicted by one or all working memory components. Therefore, the present study serves as a meta-analysis of studies in which this relationship was investigated, to investigate whether each working memory component is related to mathematics performance. Moreover, to find an explanation for conflicting results, we investigated the influence of various moderator variables: the type of mathematics measurement used, characteristics of the working memory tasks, children's ages, the population (typical or atypical), the inclusion of control variables and the country and year of origin of the study.

1.1. Working memory

The most widely used model of working memory is the multi-component model originally proposed by Baddeley and Hitch (1974). This model comprises different subcomponents, each with its own function and capacity. Two slave systems, the visuospatial sketchpad and the phonological loop, are responsible for temporary storage of visual and spatial information, and phonological and auditory information, respectively. Capacity of the slave systems is usually measured through simple span tasks, in which increasingly longer strings of information must be replicated (e.g., a dot appearing in different consecutive locations for the visuospatial sketchpad and word lists for the phonological loop). A third component, the central executive, coordinates information stored within the slave systems. Capacity of the central executive is traditionally measured with complex span tasks. In these tasks, a series of items must also be replicated, but this information must first be manipulated: e.g., the items must be recalled backwards, or must be counted first. In other words, the slave systems require only storage of information, while the central executive also requires coordination of information (Oberauer, Süß, Wilhelm, & Wittman, 2003).

Since the formulation of the working memory model, it has been used extensively to inform research, but also educational practise, and it has been applied widely to a wide range of aspects of human thought (see: Baddeley, 2007).The three-factor model including the central executive, the visuospatial sketchpad, and the phonological loop (Baddeley and Hitch, 1974) provided good fit to working memory data of children of various ages (Gathercole, Pickering, Ambridge, & Wearing, 2004).

Years after the introduction of the working memory model, the notion arose that the coordinating role of the central executive may be differentiated further into different subprocesses (Baddeley, 1996). A framework based on both Baddeley (1996) and the executive function literature is now often used: the central executive remains an important part of the working memory model, but is subdivided into inhibition, shifting and updating (Miyake et al., 2000). In this framework, inhibition refers to the ability to suppress a dominant response in favour of another response or no response at all, shifting to the ability to switch between response sets, and updating to the ability to monitor and revise the information that is active in working memory. The terms working memory and updating are not used consistently in the literature: some studies use the term working memory to refer only to the central executive, or even more specifically to the executive function of updating, whereas others refer to the entire model of working memory. In the current study, the term working memory refers to the entire processing and storage unit as described by Baddeley and Hitch (1974), including the expansion of the central executive with the three executive functions. In the current study, all executive functions (inhibition, shifting and updating) and the two slave systems (visuospatial sketchpad and phonological loop) are referred to as working memory components. Measures of executive functioning (inhibition, shifting and updating) are seen as functionally detailed components of the central executive.

Using factor analysis, Miyake et al. (2000) confirmed the independence as well as showed the interrelatedness of the three executive functions in adults. The distinction between these three executive functions has also been confirmed in some factor analytical studies with children (e.g., Hughes, 1998; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Rose, Feldman, & Jankowski, 2011) but not in all (e.g., Van der Sluis, De Jong, & Van der Leij, 2007; Van der Ven, Kroesbergen, Boom, & Leseman, 2012b). Nevertheless, many studies have employed this distinction in their research on the relations between working memory, executive functioning and mathematical performance. Therefore, in this meta-analysis, the relation between mathematic skills and each specific component of the working memory system is analysed.

1.2. Relations between working memory and mathematical performance

The relations between working memory and various measures of mathematical skill have been studied extensively. Studies have shown that working memory is a strong predictor of mathematical skills across time (Gathercole, Tiffany, Briscoe, Thorn, & The ALSPAC team, 2005; Mazzocco and Kover, 2007; Toll, Van der Ven, Kroesbergen, & Van Luit, 2011), even when controlling for IQ (Alloway and Alloway, 2010), and that children with difficulties in mathematics score lower on measures of working memory (Swanson and Jerman, 2006). Yet, findings are inconsistent with regard to which components of working memory can predict mathematical performance, and which cannot. This is possibly a result of the complexity and variety of the tasks employed to use working memory and the range of aspects that comprise the working memory model (Baddeley, 2007).

Associations between inhibition and mathematical performance have been found in several studies that include either typically achieving children or children with (mathematical) difficulties and disorders. Rotzer et al. (2009) suggested that children with mathematical disorders have specific inhibitory deficits. Inhibition may be involved in the suppression of inappropriate strategies, such as addition when multiplication is required, or suppression of irrelevant information such as information from a context problem that is irrelevant to the problem itself. Inhibitory skills have also been found to reliably predict mathematics scores in typically developing young children (Bull and Scerif, 2001; Espy et al., 2004; St Clair-Thompson and Gathercole, 2006), and performance on inhibition tasks has been related to growth in mathematics scores across a time span of several months (Panaoura and Philippou, 2007). Moreover, inhibition has been found to already predict number sense, as an early form of mathematical proficiency (Kolkman, Hoijtink, Kroesbergen, & Leseman, 2013; Kroesbergen, Van Luit, Van Lieshout, Van Loosbroek, & Van de Rijt, 2009). However, differences between mathematics-disordered children and normal controls with regard to inhibition have not been found consistently (Censabella and Noël, 2008) and the involvement of inhibition in mathematical performance cannot be confirmed in every study (Andersson, 2008; Monette, Bigras, & Guay, 2011). Subsequently, it has been hypothesised that inhibition is domain-specific and only number or quantity-related inhibition is relevant for mathematical performance (Bull and Scerif, 2001).

Significant unique associations between shifting, a second component of working memory, and mathematical performance have been found in some studies (e.g., Andersson, 2007; Bull and Scerif, 2001). This relation may be explained by the task demands of many arithmetical achievement tests: children need to switch between operations, strategies and quantity ranges in order to successfully obtain an answer (Andersson, 2008; Bull and Scerif, 2001). Also, shifts between different steps of a multi-step problem need to be made to achieve a correct answer in these types of mathematical problems.

Other studies, however, have not found a relation between shifting and mathematical skills when controlling for other executive functions (e.g., Monette et al., 2011; Van der Ven, Boom, Kroesbergen, and Leseman, 2012b). Blair and Razza (2007) suggested that the lack of explained variance in mathematical performance attributed to shifting in some studies was due to variance already being accounted for by inhibitory control. Differences between arithmetic-disabled children and typically developing children in shifting tasks have been attributed to the need to use shifting, updating and inhibition simultaneously in these tasks (Van der Sluis, De Jong, & Van der Leij, 2004), suggesting that there is no unique association between shifting and mathematical performance. Yet, a recent meta-analysis showed that there was a significant association between shifting and mathematical performance (Yeniad et al., 2013). The current study adds to this report by exploring the association between shifting and mathematical skills in the much broader framework of the entire working memory model, and by using a more extensive set of moderator variables to explain variation in effect sizes, because no significant moderator variables were found in the cited study.

Of all executive functions, updating is most strongly related to mathematical performance (De Smedt et al., 2009; Passolunghi et al., 2008). It has been proposed that updating is involved in the storage and retrieval of partial results in mathematical problems, and remembering important information of the presented problem during the process of problem-solving (Dehaene, 1997). A child with insufficient updating skills may forget intermediate results, make procedural errors, or forget one part of the mathematical problem while working on a different part of the problem. Indeed, updating has been found to be correlated strongly with concurrent mathematical performance (Bull and Scerif, 2001; Van der Ven et al., 2012b) and to longitudinal growth in mathematical performance (Van der Ven et al., 2012b). Similar associations have been found on studies on a microgenetic level, a design in which many measurements within a short time span attempt to cover an important developmental stage (Van der Ven et al., 2012b).

Updating can be divided into verbal and visuospatial updating. Evidence of a distinction between these two factors has been found using factor analysis (Oberauer et al., 2003), which is why this distinction is employed in this study. Visuospatial updating has consistently been found to be associated with mathematical performance (St Clair-Thompson and Gathercole, 2006), whereas verbal updating has been found to have a unique association with mathematical skills in some studies (Jarvis and Gathercole, 2003; Navarro et al., 2011), but not in others (St Clair-Thompson and Gathercole, 2006).

Findings regarding the relations between functioning of the slave systems, which are the visuospatial sketchpad and phonological loop, and mathematic skills are also diverse. Whereas functioning of the visuospatial sketchpad has been found to be significantly associated with mathematics and number sense tasks in some studies (e.g., Andersson and Lyxell, 2007; Krajewski and Schneider, 2009a), others have reported only weak or no relations at all (e.g., Andersson, 2007; Andersson, 2008; Rasmussen and Bisanz, 2005). Visual and spatial encoding may make different contributions to mathematical performance (see Baddeley, 2003). The relation between visuospatial sketchpad and mathematics has also been found to be partially confounded by age: after controlling for age, correlations drop considerably (Kyttälä, Aunio, & Hautamäki, 2010), which suggests that if studies use a more narrow age range, correlations may not be found. Similarly, functioning of the phonological loop has been found to add unique explained variance to models predicting mathematical achievement (Panaoura and Philippou, 2007; Passolunghi et al., 2008). Again, findings are inconsistent (Jarvis and Gathercole, 2003; Passolunghi, Vercelloni, & Schadee, 2007).

In sum: all components of working memory, including all three executive functions, have been found to be related to mathematical performance, but there is no consistency with regard to the findings. Inconsistencies are often explained using characteristics of tasks and constructs specific to the working memory component (for an overview, see: Raghubar, Barnes, & Hecht, 2010). Two general hypotheses regarding the inconsistent relations between executive functioning and mathematics performance were brought forward by Best, Miller, and Naglieri (2011). In a large cross-sectional study, they found that different types of mathematical problems were related differently to executive functioning. They proposed that problem solving is strongly dependent on strategy formulation and implementation, and self-monitoring. Calculation, on the other hand, was suggested to be more related to fact retrieval, which requires less executive control (Best et al., 2011). Findings reported by Fuchs et al. (2005) suggested the same (see also: Raghubar et al., 2010). A second hypothesis proposed by Best et al. (2011) was that performance may be related to executive functioning to a different extent in various developmental stages. Age has been reported to moderate the relationship in other studies (Imbo and Vandierendonck, 2007; McKenzie, Bull, & Gray, 2003; Rasmussen and Bisanz, 2005), yielding the primary conclusion that older children rely less on working memory in mathematical problem solving than younger children do, suggesting that part of the inconsistencies between studies may be explained by the age of the participants.

1.3. The current study

1.3.1. Research questions

Given the proposal that working memory is especially important for acquiring mathematical operations and insights (Holmes, Gathercole, & Dunning, 2009), the current study explores the relations between working memory and (preparatory) mathematical performance at kindergarten and primary school age. The aim of the current study is twofold. The first aim is to further investigate the strength of the associations in children at primary school age, between 4 and 12 years, by means of a meta-analysis. We investigate the relation between mathematical skills and all working memory components: inhibition, shifting, visuospatial updating and verbal updating, visuospatial sketchpad and phonological loop. The accompanying research question is whether there is a significant association between each working memory component and mathematical performance. The second aim is to investigate whether variance in effect sizes of the different studies can be explained by characteristics of the tests, samples, methodological decisions, or studies. Moderator variables are outlined in Section 1.3.2.

1.3.2. Moderator variables

Throughout the literature, a variety of moderator variables has been proposed to explain the presence or absence of significant and relevant associations between working memory components and mathematical skills. Based on these hypotheses, a set of moderator variables was selected to explain variance in reported correlations. First, it has been proposed that problem solving in mathematics is more dependent on executive functioning than calculation (Best et al., 2011; Fuchs et al., 2005; Raghubar et al., 2010). Therefore, the current study makes a distinction between general mathematics measures, arithmetic, and a number of other categories of mathematics measures that cannot be directly classified as problem solving or calculation. The relations between working memory and these miscellaneous categories of mathematics are also explored (see Section 2.4.1).

Furthermore, several studies have proposed that associations between working memory and mathematical performance depend on the type of working memory measure used. For example, it has been proposed that only number-specific working memory tasks are relevant to mathematic skills (Bull and Scerif, 2001). Moreover, it has been suggested that visual and spatial processing are distinct skills of the visuospatial sketchpad and make different contributions to mathematical performance (Baddeley, 2003). Complex span tasks, in which information is processed before recalling information, and classical updating tasks, in which the processing factor contains selective replacement of information, are often seen as inherently different tasks, although both have been found to rely on the same processes (Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009; St Clair-Thompson and Gathercole, 2006). Therefore, to cover differences between tasks that measure working memory components, various typologies of working memory measures were included in the current study.

Sample characteristics have also been proposed to explain variance in associations between working memory and mathematical achievement. It has been suggested that age partly determines the correlation between the two (Best et al., 2011; Imbo and Vandierendonck, 2007; McKenzie et al., 2003; Rasmussen and Bisanz, 2005). Hence, age of the sample was included as a moderator variable. Moreover, it has been suggested that associations differ between typically achieving children and children with mathematical difficulties, the latter showing stronger dependencies on passive storage in contrast to active manipulation of working memory content (Kyttälä et al., 2010). Working memory has also been suggested to underlie mathematical difficulties in children with physical disabilities (Raches and Mazzocco, 2012), suggesting that they should be treated as a separate group in explaining variations in correlations. Therefore, sample type was divided into several categories of special needs groups, which were used as moderator variables (see Section 2.4.2).

Moreover, correlations reported in various studies may be influenced by decisions made with regard to study design. For instance, correlations may be controlled for other constructs, such as age, which could reduce the reported association between working memory and mathematical skills (Kyttälä et al., 2010), or correlations may concern longitudinal associations rather than concurrent associations, which also causes a reduction in the associations (Bull, Espy, & Wiebe, 2008). Inclusion of control variables and an indicator of time lag between working memory and mathematics measurements (being 0 in the case of studies that did not report on longitudinal data) are hence included as moderators.

Finally, it will be explored whether the strength of the association is influenced by the country of origin of the study. Children from different countries have been found to use different strategies in mathematical problem solving (Imbo and LeFevre, 2009), which may lead to children from various origins to show different degrees of working memory involvement in problem solving. Moreover, children from different countries use different number encoding strategies, with children from some cultures forming verbal representations and children from other cultures forming visuospatial representations (Tang et al., 2006), which may affect working memory involvement as well. Country of origin was therefore added to the models as an exploratory moderator variable. Publication year was also investigated as an exploratory moderator variable, because variations in the strength of association between mathematical performance and working memory may be the result of changes in the quality or format of measures due to increased digitalisation, or changes in the method of teaching, leading to similar strategy differences as found between countries (Imbo and LeFevre, 2009).

2. Method

The review process was divided into four stages: (1) listing of inclusion and exclusion criteria to identify relevant research; (2) searching for studies using preset search terms, and screening against inclusion and exclusion criteria; (3) extracting data and critically evaluating them; and (4) synthesising of the results reported in the included studies.

2.1. Inclusion criteria for the meta-study

Studies were considered eligible if they contained behavioural empirical data and outcomes. They had to report on a variable of mathematics or number sense performance, and one or more working memory components: inhibition, shifting, visuospatial and verbal updating, the visuospatial sketchpad and the phonological loop. Associations had to have been statistically tested, and children included in the studies had to be between 4 and 12 years of age. All mean ages had to be between 4;0 and 12;11 years and no children could be included younger than 4;0 years or older than 13:11 years (this somewhat broader range has been set to be able to include studies containing children in primary school who repeated a grade).

2.2. Identification and screening process of the studies

Searches were carried out between February and June 2012. A number of databases was searched: Scopus, PsycInfo, the Education Resources Information Center (ERIC) and the Social Sciences Citation Index (SSCI). Search terms included *math** or *arithmetic*; and *working memory* or *executive funct**; and *child**. Search terms produced a total of 1,326 hits, containing 630 unique results. Titles and abstracts of these studies were screened. Upon first scan, 236 studies were eligible for inclusion.

Final decisions regarding inclusion were made using the full text reports of these articles. Articles in which no correlation measures between mathematical performance and working memory measures were reported were excluded, leading to the removal of 114 articles. Eleven more studies were excluded because correlations were only reported for composite measures of working memory that could not be classified as a single component, such as a composite measure of all Automated Working Memory Assessment subtests (Alloway, Gathercole, Kirkwood, & Elliott, 2009). Thus, the final dataset contained 111 studies, reporting on a total of 16,921 participants. Appendix 1 presents an overview of these 111 studies, with relevant characteristics of each study.

2.3. Data extraction

After eligible studies had been selected, relevant information from the studies was coded. This includes the following variables of descriptive data: (1) number of participants, (2) mean age of the participants; range: 4.0–12.0 years, (3) time lag between first and last time point (0 in the case of studies that did not report on longitudinal data), (4) sample type: typical or atypical, (5) the type of measure used for each construct: name of the measure and the use of accuracy or reaction time data, (6) nationality of the participants, being one out of 20 nationalities listed in Section 2.4.4, and (7) year of publication; range: 1993–2012.

In the published studies, some measures were coded such that a negative correlation indicated a positive relation between executive functioning and mathematical performance. This was mainly the case when reaction time data were used to measure executive functioning. In these cases, the measures of effect size were recoded, such that a positive correlation always indicated a positive relationship.

When multiple instruments were used to measure one construct, all eligible effect sizes were included in the dataset. Multilevel modelling (see Section 2.5) was used to account for dependencies between effect sizes from a single study.

2.4. Data appraisal

After data were extracted from all eligible articles, several variables were created to quantify task, sample or study characteristics.

2.4.1. Characteristics of tasks

To investigate the effect of different tasks on the size of the correlation coefficient between working memory components and mathematics performance, various categories and characteristics were coded. For mathematics measures and visuospatial and verbal updating measures, categorical variables were created to distinguish between task categories. For inhibition, shifting, visuospatial updating and verbal updating measures, a number of task characteristics could be defined, such as the inclusion of numerical stimuli and whether a verbal of behavioural response was registered. Characteristics and categories can be found in Table 1. Note that whereas categories are mutually exclusive, e.g., a phonological loop measure can be classified as either word span or nonword span, characteristics are not necessarily mutually exclusive, e.g., a task can be classified as both an *N*-back task and as including spatial information.

For mathematics, inhibition and shifting measures, an additional variable was created, indicating whether accuracy or reaction time (RT) data were used for the effect size. To preserve power within analyses, questionnaires targeting working memory and mathematics, and measures recording frequency of retrieval strategy use (the number of times retrieval was used to solve a mathematical problem) were also coded as measurements of accuracy (please note that accuracy, frequency of retrieval, and parent or teacher ratings in questionnaires all refer to frequency data: frequency of a correct answer, frequency of retrieval strategy, and in rating scales, usually frequency of occurrence of behaviour, but also other types of questions). Excluding questionnaire data did not lead to different outcomes. When a combination of accuracy and RT was used, the variable was coded as RT. For updating and slave system measures, an insufficient number of RT measures was available to create such a variable.

2.4.2. Sample characteristics

Age of the children was coded in months. Various sample types were investigated in the body of studies. From these types, four categories of samples were created:

1. Typically developing children.
2. Children with physical disabilities.
3. Children with disorders or cognitive delays not specifically related to mathematics, but specified in any version of the *Diagnostic and Statistical Manual of Mental Disorders* (e.g., American Psychiatric Association, 2000), such as ADHD.
4. Children with delays or disorders in mathematics (e.g., Mathematics Disorder; APA, 2000). (Due to various criteria for mathematics disorder and inconsistency in reported norms, delays and disorders cannot be separated within the current sample of studies.).

If a study contained a sample of children that was a mixture of the types mentioned above, or if a sample had comorbid disorders, the study was categorised within the last category on the list that could apply. For the inhibition and visuospatial

Table 1
Characteristics of tasks: characteristics and categories of working memory and mathematics tasks.

Categorical variables of tasks			Characteristics across tasks			
Mathematics[a]	VSSP	2.2 PL	Inhibition	Shifting	VS updating	VE updating
1 National curriculum tests, composite measures or teacher ratings	1 Composite or other task (reference category)	1 Composite or other task (reference category)	1 Measuring verbal vs. behavioural responses of the child	1 Measuring verbal vs. behavioural responses of the child	1 *N*-back task vs. recall of all items	1 *N*-back task vs. recall of all items
2 Simple arithmetic (one step, terms below 10)	2 Dot matrix, knox blocks	2 Word recall	2 Task including numerical information	2 Task including numerical information	2 Visual vs. spatial information as input	2 Random generation vs. recall
3 Advanced arithmetic (written arithmetic, multiple step arithmetic)	3 Corsi blocks or variant	3 Digit recall	3 Distinction between random generation or limited response tasks			3 Including numerical information in the task
4 Word problems	4 Mazes recall	4 Nonword recall				
5 Counting and basic understanding of numerical concepts	5 Patterns recall, shape recall					
6 Geometry, shapes and algorithms						

Note: VSSP, visuospatial sketchpad; PL, phonological loop; VS, visuospatial; VE, verbal.
[a] For the inhibition and shifting analyses, mathematics category 6 (geometry) was merged with category 1 because of a low number of effect sizes concerning measures in category 6. For the visuospatial updating measure, both mathematics categories 4 (word problems) and 6 (geometry) were merged with category 1, and categories 2 (simple arithmetic) and 3 (advanced arithmetic) were merged to create one measure for arithmetic for the same reason.

updating analyses, categories 3 and 4 were merged, creating one category of "other disorders" because of the small number of effect sizes concerning samples in category 3. For shifting, category 3 was deleted because no effect sizes in the dataset concerned these samples.

2.4.3. Design characteristics: Inclusion of control variables

A dichotomous variable was created indicating whether control variables were included in the correlation analyses or not. Control variables that were encountered were: age, IQ, sex, prematurity of birth, year of schooling, measures of the phonological loop (controlling for the correlation between verbal updating and mathematical performance), and reading skills, or various combinations of variables. Further distinctions between control variables included in the studies could not be included because of the low frequency of occurrence of each (set of) control variable(s). Another variable was computed indicating the number of months between the first measurement moment and the last being reported on.

2.4.4. Study characteristics

Countries of origin of the studies were divided into the following categories, based on geographical location:

1. North America (reference category).
2. United Kingdom.
3. Europe (Austria, Belgium, Estonia, Finland, France, Germany, Greece, Italy, the Netherlands, Spain, Sweden).
4. Other (Australia, Brazil, China, New Zealand, Singapore, Taiwan).

This subdivision was composed based on a sufficient number of studies to be placed in the same category, so as to preserve power. For inhibition, categories 2 and 3 were merged because of the small number of effect sizes within category 2. For shifting, categories 1 and 4, and categories 2 and 3 were merged because of the small number of effect sizes within categories 1 and 2. For visuospatial updating, category 1 was merged with category 4 for the same reason.

2.5. Statistical analyses on the selected studies

To answer the research questions, Hierarchical Linear Modelling (HLM) was applied using the software package HLM version 6.06. Effect sizes (correlation coefficients) were nested within studies to account for dependencies between multiple effect sizes within a study. Two HLM models were tested to answer the research questions regarding the associations between mathematical performance and each working memory component. First, an unconditional model was tested in which an overall effect size was estimated and the variance around the overall average was computed. Second, all moderator variables were entered simultaneously into a conditional model, and moderators that made no significant contribution to the model were removed one by one, until the most parsimonious model was found. Moderator variables were: (1) mathematics measure categories, (2) working memory measure characteristics, (3) sample characteristics (mean age and categories of typical and atypical samples), (4) other methodological decisions: time between first and last measurement (coded as 0 when mathematics and working memory had been measured concurrently) and whether control variables had been included, and (5) study characteristics (publication year, country of origin).

Funnel plot analyses showed reasonable symmetry in the reported correlations, suggesting that there was little influence of publication bias (see: Egger, Smith, Schneider, & Minder, 1997), and trim and fill procedures did not lead to relevant changes in the weighted mean correlation coefficients (See: Duval and Tweedie, 2004). Therefore, the original dataset was used in all reported analyses.

3. Results

3.1. Inhibition

The unconditional model, based on 131 correlations drawn from 29 studies, indicated a medium-sized significant correlation between inhibition and mathematics measures, $r(28) = .27$, $p < .001$. Homogeneity analyses revealed a significant variance around the mean effect size, χ^2 (28, $N = 29$) = 73,821.63, $p < .001$.

The final regression model including all significant predictors of the correlation between mathematical performance and inhibition measures is presented in Table 2. The model indicates that type of mathematics measures was important: correlations were higher when national curriculum tests or composite measures were included than when more specific types of tests, such as arithmetic tests, were included. Differences between categories other than reference category, general mathematics, were not significant, all *p*s > .05. There was a positive effect of random generation tasks as inhibition measure, indicating that they yielded higher effect sizes than other measures, while the inclusion of control variables was associated with lower effect sizes.

3.2. Shifting

The unconditional model, based on 94 correlations drawn from 18 studies, indicated a medium-sized significant positive correlation between shifting and mathematics measures, $r(17) = .28$, $p < .001$. Homogeneity analyses revealed a significant variance around the mean effect size, χ^2 (17, $N = 18$) = 70,109.23, $p < .001$.

The final regression model, including all significant predictors of the correlation between mathematical performance and shifting measures, is presented in Table 3. The model indicates an effect of mathematics measure: correlations were lower when arithmetic measures rather than other measures were included in the effect size. There was no difference between simple and advanced arithmetic, $p = .37$. Effect sizes were higher for younger children, and in samples of children with mathematical difficulties. More time between measurements was predictive of lower effect sizes, and European studies (including the UK) yielded higher correlations than studies from elsewhere.

3.3. Visuospatial updating

The unconditional model, based on 63 correlations drawn from 21 studies, indicated a medium-size significant correlation between visuospatial updating and mathematics measures, $r(20) = .34$, $p < .001$. Homogeneity analyses revealed a significant variance around the mean effect size, χ^2 (20, $N = 21$) = 7,044.75, $p < .001$.

The final regression model is presented in Table 4. Both arithmetic and counting and conceptual skills showed lower correlations with visuospatial updating than all other measures. Regression weights between these categories of mathematical skills differed significantly in all cases (all *p*s < .001). Visuospatial updating measures that were based on processing (rather

Table 2
Variables significantly predicting the size of the correlation between mathematics and inhibition.

Type	Variable	*B*	*β*	SE	*t*	*df*	*p*
	Intercept	0.29	–	0.03	10.57	28	<.001
Maths	Simple arithmetic	−0.13	−0.27	0.06	−2.05	123	<.001
Maths	Advanced arithmetic	−0.16	−0.25	0.06	−2.73	123	<.01
Maths	Word problems	−0.16	−0.26	0.07	−2.46	123	.02
Maths	Counting and concepts	−0.12	−0.27	0.06	−2.16	123	.03
WM	Random generation	0.18	0.34	0.04	4.17	123	<.001
Design	Control variables	−0.27	−0.50	0.01	−33.42	123	<.001

Note: Type refers to the moderator type, as listed in Section 2.5 and Table 1; WM, working memory. For mathematics measures, the reference category was formed by national curriculum tests, composite measures, and teacher ratings. The degrees of freedom at the intercept are determined by the number of cases at the second level.

Table 3
Variables predicting the size of the correlation between mathematics and shifting.

Type	Variable	B	β	SE	t	df	p
	Intercept	0.32	–	0.05	6.11	16	<.001
Maths	Simple arithmetic	−0.06	−0.13	0.02	−2.29	87	.02
Maths	Advanced arithmetic	−0.06	−0.11	0.02	−2.64	87	.01
Sample	Age	−0.01	−0.78	<0.01	−2.15	87	.03
Sample	Maths difficulties	0.17	0.24	0.02	7.73	87	<.001
Design	Time lag	−0.01	−0.82	0.01	−2.12	87	.04
Study	Origin	0.27	0.72	0.10	2.75	16	.02

Note: Type refers to the moderator type, as listed in Section 2.5 and Table 1. For mathematics measures, the reference category was formed by national curriculum tests, composite measures, and teacher ratings. The degrees of freedom at the intercept are determined by the number of cases at the second level. The variable *Origin* was added as a second level predictor (study-level), leading to a lower number of degrees of freedom.

Table 4
Variables predicting the size of the correlation between mathematics and visuospatial updating.

Type	Variable	B	β	SE	t	df	p
	Intercept	0.39	–	0.07	5.89	20	<.001
Maths	Arithmetic	−0.19	−0.39	0.02	−7.59	55	<.001
Maths	Counting & concepts	−0.32	−0.81	0.04	−7.55	55	<.001
WM	Replace vs. recall all	−0.26	−0.46	<0.01	−151.01	55	<.001
WM	Spatial vs. visual	−0.15	−0.41	<0.01	−84.03	55	<.001
Sample	Age	0.01	1.47	<0.001	54.21	55	<.001
Sample	Maths difficulties	0.04	0.09	<0.01	36.47	55	<.001
Design	Time lag	−0.01	0.04	<0.01	−4.93	55	<.001

Note: Type refers to the moderator type, as listed in Section 2.5; WM, working memory. For mathematics measures, the reference category was formed by national curriculum tests, composite measures, and teacher ratings. The degrees of freedom at the intercept are determined by the number of cases at the second level.

than replacing) information and measures that addressed visual rather than spatial skills also yielded higher correlations with mathematics measures. There was a positive effect of age, indicating that effect sizes were higher for older children, and effect sizes in samples with children with mathematical difficulties were higher. Finally, less time between measures was predictive of higher correlation coefficients.

3.4. Verbal updating

The unconditional model, based on 411 correlations drawn from 85 studies, indicated a medium-sized significant correlation between verbal updating and mathematics measures, r (84) = .38, $p < .001$. Homogeneity analyses revealed a significant variance around the mean effect size, χ^2 (84, $N = 85$) = 86,502.69, $p < .001$.

The final regression model is presented in Table 5. There were effects of both mathematics task type and updating task type. Specific mathematics measures, including arithmetic or counting and concepts, yielded significantly lower effect sizes than general measures. Regression weights of these categories did not differ amongst each other (all *ps* > .05). Also, verbal updating measures that required the replacement of information in working memory yielded higher effect sizes, and measures that did not require random generation yielded higher effect sizes than random generation tasks. Studies containing

Table 5
Variables predicting the size of the correlation between mathematics and verbal updating.

Type	Variable	B	β	SE	t	df	p
	Intercept	0.39	–	0.02	25.29	84	<.001
Maths	Simple arithmetic	−0.05	−0.11	0.02	−2.17	402	.03
Maths	Advanced arithmetic	−0.06	−0.10	0.03	−2.07	402	.04
Maths	Counting & concepts	−0.06	−0.10	0.01	−5.70	402	<.001
WM	Replace vs. recall all	0.08	0.13	0.01	11.66	402	<.001
WM	Random generation	−0.04	−0.05	0.001	−54.49	402	<.001
Sample	DSM diagnosis	0.12	0.15	0.02	5.99	402	<.001
Sample	Physical disabilities	0.15	0.17	0.02	7.96	402	<.001
Design	Control variable	−0.08	−0.14	<0.001	−458.92	402	<.001

Note: Type refers to the moderator type, as listed in Section 2.5 and Table 1; WM, working memory. For mathematics measures, the reference category was formed by national curriculum tests, composite measures, and teacher ratings. The degrees of freedom at the intercept are determined by the number of cases at the second level.

children with DSM disorders or physical disabilities also showed higher correlations between verbal updating and mathematical performance than other studies. The inclusion of control variables was associated with lower correlations.

3.5. Visuospatial sketchpad

The unconditional model, based on 239 correlations drawn from 55 studies, indicated a medium-sized significant correlation between measures of the visuospatial sketchpad and mathematics measures, $r\,(54) = .34$, $p < .001$. Homogeneity analyses revealed a significant variance around the mean effect size, $\chi^2\,(54, N = 55) = 23{,}747.67$, $p < .001$.

The final regression model is presented in Table 6. The model shows higher effect sizes for general mathematical skills than for advanced arithmetic, word problems and counting and concepts. Furthermore, arithmetic measures yielded higher correlation coefficients than word problems, $p < .001$, but did not differ from counting and concepts, $p = .20$, nor did counting and concepts differ from word problems, $p = .41$. Use of Corsi blocks rather than other measures of visuospatial sketchpad was associated with higher effect sizes. Correlations between mathematical performance and visuospatial sketchpad were slightly lower for children with mathematical difficulties, and slightly higher for children with physical disabilities. Finally, age of the sample was negatively related to effect size.

3.6. Phonological loop

The unconditional model, based on 295 correlations drawn from 65 studies, indicated a medium-sized significant correlation between verbal updating and mathematics measures, $r\,(64) = .31$, $p < .001$. Homogeneity analyses revealed a significant variance around the mean effect size, $\chi^2\,(64, N = 65) = 58{,}171.86$, $p < .001$.

The final regression model is presented in Table 7. The model shows higher effect sizes for general mathematical skills than for specific mathematics measures. Differences between categories of specific mathematics measures, however, are not significant, *ps* > .05, except for the difference between advanced arithmetic and word problems, $p < .001$. Moreover, samples including children with mathematical difficulties, other DSM disorders or physical disabilities showed higher effect sizes than typical samples. The inclusion of control variables was associated with lower effect sizes.

3.7. Comparison between models

To investigate which working memory component shows the highest correlation with mathematical performance, weighted mean correlation coefficients were compared. Comparisons showed the following pattern:

$$r_{VerbalUpdating} > r_{VisuospatialSketchpad} = r_{VisuospatialUpdating} > r_{PhonologicalLoop} > r_{Inhibition} = r_{Shifting}$$

The strongest correlation was between mathematical performance and verbal updating. This relationship was significantly stronger than the relationship with other components, all *ps* < .001. The second strongest relation was between mathematical performance and both the visuospatial sketchpad and visuospatial updating: these two relations did not significantly differ from each other, $z = 0.03$, $p = .98$, but were significantly stronger than the relations with all remaining components, all *ps* < .001. Phonological loop measures correlated more strongly with mathematic skills than inhibition and shifting. The strength of the correlations between inhibition and shifting and mathematical performance did not differ, $z = 0.53$, $p = .59$; both showed lower correlations with mathematical performance than all other components, all *ps* < .001.

Table 8 summarises the sets of significant predictors for the association between mathematic skills and each working memory component. The type of mathematics measure was a significant moderator in all models, and therefore the most consistent of moderator variables. Moreover, sample type contributed to all models except the model concerning the relation between inhibition and mathematical performance. All other classes of moderator variables contributed to half of the models

Table 6
Variables predicting the size of the correlation between mathematics and visuospatial sketchpad.

Type	Variable	*B*	*β*	SE	*t*	*df*	*p*
	Intercept	0.36	–	0.03	11.72	54	<.001
Maths	Advanced arithmetic	−0.13	−0.18	0.04	3.58	231	.001
Maths	Word problems	−0.07	−0.11	0.03	−2.36	231	.02
Maths	Counting & concepts	−0.07	−0.19	0.02	−3.24	231	<.01
WM	Corsi blocks	0.11	0.43	0.03	3.25	231	<.01
Sample	Maths difficulties	−0.05	−0.10	0.02	−2.56	231	.01
Sample	Physical disabilities	0.12	0.22	0.02	8.24	231	<.001
Sample	Age	−0.01	−1.19	<0.01	−3.20	231	<.01

Note: Type refers to the moderator type, as listed in Section 2.5 and Table 1; WM, working memory. For mathematics measures, the reference category was formed by national curriculum tests, composite measures, and teacher ratings. The degrees of freedom at the intercept are determined by the number of cases at the second level.

Table 7
Variables predicting the size of the correlation between mathematics and phonological loop.

Type	Variable	*B*	*β*	SE	*t*	*df*	*p*
	Intercept	0.32	–	0.02	13.62	64	<.001
Maths	Simple arithmetic	−0.08	−0.18	0.03	−2.43	285	.02
Maths	Advanced arithmetic	−0.14	−0.18	0.03	−4.58	285	<.001
Maths	Word problems	−0.07	−0.11	0.03	−2.40	285	.02
Maths	Counting & concepts	−0.07	−0.15	0.03	−2.73	285	<.01
Maths	Geometry	−0.10	−0.09	0.03	−2.94	285	<.01
Sample	Maths difficulties	0.12	0.22	0.04	3.26	285	<.01
Sample	DSM-disorders	0.10	0.16	0.04	2.66	285	<.01
Sample	Physical disabilities	0.21	0.29	0.02	10.93	285	<.001
Design	Control variables	−0.17	−0.25	<0.01	−519.26	285	<.001

Note: Type refers to the moderator type, as listed in Section 2.5 and Table 1. For mathematics measures, the reference category was formed by national curriculum tests, composite measures, and teacher ratings. The degrees of freedom at the intercept are determined by the number of cases at the second level.

Table 8
Summary of variables significantly predicting the relation between mathematics performance and each working memory component.

Component	Moderator Mathematics measure	Working memory measure	Age	Sample type	Control variable included	Time lag	Origin of study
Inhibition	X	X			X		
Shifting	X		X	X		X	X
Visuospatial updating	X	X	X	X		X	
Verbal updating	X	X		X	X	X	
Visuospatial sketchpad	X	X	X	X			
Phonological loop	X			X	X		

or less. Distinctions between accuracy and reaction time measures did not contribute to any of the models, nor did publication year.

After inclusion of these moderator variables into the models, unexplained heterogeneity of effect sizes between studies dropped significantly, all *ps* < .001, but in each model, significant heterogeneity remained unexplained, all *ps* < .001.

4. Discussion

A large body of research has reported on the relations between working memory components and mathematical performance, yielding inconsistent effects. The current meta-analysis had two aims: to investigate the significance and relevance of these relations by means of a research synthesis, and to explain variation in effect sizes using various predictor variables in a multilevel meta-analysis. Another recent meta-analysis (Yeniad et al., 2013) has reported on similar associations between shifting and mathematics. The current study adds to this work by including the other components of working memory and using a more extensive set of moderator variables.

4.1. Unconditional models

The analyses confirmed a positive and significant relation between mathematical performance and each working memory component under investigation. This indicates that better working memory performance on each component is associated with better mathematical performance, as suggested in several studies (e.g., Geary et al., 2004; Kroesbergen et al., 2009; Passolunghi et al., 2008). The working memory component that showed the strongest relation with mathematical skills was verbal updating, whereas inhibition and shifting showed the weakest relations, confirming findings reported in previously published studies (Kolkman et al., 2013). This may imply that verbal updating has the highest relevance for mathematics performance, and inhibition and shifting are less relevant. Alternatively, this may be due to measurement characteristics, such as the increasing complexity of updating tasks (e.g., Alloway and Alloway, 2010), in comparison to mostly static difficulty levels of inhibition and shifting tasks (A. R. A. Conway, personal communication, March 18, 2013).

Furthermore, each of the relations between working memory components and mathematics performance showed significant variance around the mean effect size, and was therefore entered into a model with a set of predictor variables. Several moderator variables were identified with these models, the most consistent of which the type of mathematics tests, which moderated relations between all working memory components and mathematical performance. A number of other moderator variables contributed to some models, but not all.

4.2. Moderators of correlations

4.2.1. Effect of mathematics measures

The size of the relation between working memory and mathematical performance was found to be dependent on the type of mathematics test that was used in a study (Best et al., 2011; Fuchs et al., 2005; see: Raghubar et al., 2010). National curriculum tests and composite measures consistently showed higher correlations with working memory components than any of the other categories of mathematics tests. This may indicate that these tests, or composites, appealed to working memory capacity to a greater extent than measures that specifically measured one type of mathematic skill. In national curriculum tests or other broad measures of mathematics, this may be due to the necessity to switch between various operations and mental models and the need to update sets of information. Moreover, in both general mathematics tests and composite measures, higher correlations may be found because effects of specific knowledge or practising attenuated when other domains of mathematics were included in the same correlation. These findings confirm that calculation depends less on executive functioning than general mathematical problem solving (Best et al., 2011; Fuchs et al., 2005; see: Raghubar et al., 2010). All working memory components except verbal updating correlated more strongly with general mathematics tests than with purely arithmetical measures.

An additional explanation for the higher correlations between working memory components and general mathematics tests than other mathematics measures is that general measures of mathematics are more often validated measures with high reliability (e.g., Schrank, McGrew, & Woodcock, 2001). In other words, the elevated correlation in the case of general mathematics measures may be a measurement issue, and the estimate of the correlation between general mathematics measures and working memory components may be more accurate.

4.2.2. Effect of working memory measures

Characteristics and categories of the working memory measure also moderated the correlations between mathematical performance and executive functioning. For both visuospatial and verbal updating, correlations differed depending on whether the updating task required the replacement of information (traditional updating) or the recall of all presented stimuli (complex span). This suggests that these two types of updating play different roles in mathematic skills, contrary to previously reported unity of the constructs (Schmiedek et al., 2009; St Clair-Thompson and Gathercole, 2006).

It is notable, however, that both effects were in different directions: for visuospatial updating, replacement was associated with lower correlations than recall of all stimuli, while for verbal updating, replacement was associated with higher correlations with mathematical performance. This may be due to a difference in strategy selection between visuospatial and verbal tasks. For example, visuospatial information may be remembered using a constant mental map, whereas verbal information may be remembered through cyclic rehearsal (Baddeley, 2007). The updating of verbal information may also have greater relevance for mathematical performance because of its frequent involvement in task demands, for example, when a child solves an arithmetic problem and needs to maintain and replace intermediate answers and problem-solving steps in working memory (Dehaene, 1997), while a parallel form of visuospatial updating may be less relevant for mathematics. Replacement of visuospatial information, however, may carry less relevance to mathematical performance than maintaining all visuospatial information, because visuospatial updating is not associated with approximate nonsymbolic mathematical skills (Xenidou-Dervou, Van Lieshout, & Van der Schoot, in press), but it is with the shape of the mental number line (Friso-van den Bos, Kolkman, Kroesbergen, & Leseman, in press). The latter may serve as a reference point in judging the likelihood of an obtained answer to a problem, and must include all information of a problem to remain accurate. Finally, it has been argued that the primary distinction between traditional updating and complex span tasks resides in the involvement of inhibition when no longer relevant information needs to be disregarded (Schmiedek et al., 2009). Inhibition of visuospatial information may require less effort than inhibition of verbal information due to differences in encoding and rehearsal strategies, accounting for the dissociation.

Moreover, the inclusion of spatial information was associated with lower correlations between mathematical performance and visuospatial updating than the inclusion of visual information. It has been postulated that spatial representation of number, in the form of a mental number line, is causally related to mathematical performance (Booth and Siegler, 2008), which suggests that spatial rather than visual updating is involved in mathematical performance. However, these data may suggest that it is the mapping between symbolic and nonsymbolic representations, which depends more on visual updating, that predicts success in mathematics (Mundy and Gilmore, 2009).

Random generation tasks have been presented as both tests of inhibitory skills (e.g., Baddeley, 2007; Swanson, 2006b) because of the need to inhibit overlearned sequences, such as counting strings, as well as tests of verbal updating (e.g., Andersson, 2007) because of the need to update the set of generated items. Random generation tasks have been classified according to the classification of the authors of each article including these tasks, but analyses suggest that the strength of the relation is somewhere in between. They yield higher correlations in the case of inhibition analyses, but lower correlations in the case of verbal updating analyses.

Distinctions between word, digit and nonword recall did not moderate the associations between phonological loop and mathematical performance. This may indicate that type of information included is not of any relevance to the size of the correlation with mathematical skills in the case of phonological loop measures, which only require the recall of information as presented (Gathercole, Pickering, Knight, & Stegmann, 2004).

Finally, the inclusion of numerical information did not contribute to any of the models, suggesting that the relation between working memory capacity or efficiency and mathematical performance does not differ depending on the inclusion of numerical information in tasks. This might indicate that working memory processing is not modular, as proposed in special needs research (Noël and Rousselle, 2011), but domain-general.

4.2.3. Age of the samples included

A frequently posed hypothesis (Best et al., 2011; Imbo and Vandierendonck, 2007; McKenzie et al., 2003; Rasmussen and Bisanz, 2005) is that the association between working memory and mathematical performance depends on age, with younger children relying more on visuospatial processing. This could be confirmed for visuospatial sketchpad and visuospatial updating. In older children the correlation between mathematic skills and visuospatial updating is higher. In contrast, the correlation with visuospatial sketchpad measures is lower than in younger children. This suggests that mathematical performance becomes more and more dependent on visuospatial updating capacities rather than storage capacities, probably because children learn to solve increasingly advanced problems. Alternatively, the associations between mathematical performance and visuospatial sketchpad may drop because of increased reliance on the phonological loop (McKenzie et al., 2003; Raghubar et al., 2010). These claims could not be confirmed in the current analyses, because correlation analyses, used in this study, do not allow for partitioning of variance. Moreover, relations between shifting and mathematical performance were found to be lower for older children, which may be a result of children relying more on cognitive strategies when they get older (cognitive strategies are strategies in which no manipulatives are used, such as mental addition and retrieval), and less on the use of manipulatives, which could lead to lower demands on shifting ability. Alternatively, older children may find shifting tasks in general less cognitively demanding than younger children, which could also explain the decrease in association.

For analyses concerning other working memory components, age did not make a significant contribution to the models. This suggests that the involvement of inhibition, verbal updating and phonological loop in mathematical performance does not undergo quantitative changes.

4.2.4. Sample types and study origins

It is worth noting that associations between mathematic skills and working memory differed between various sample types in all analyses, except those concerning the relation with shifting. Children with mathematical difficulties, psychological disorders or physical disabilities all show higher associations between mathematical performance and working memory functioning than typically developing children. This might imply that typically developing children use different strategies in mathematical problem solving than children with difficulties or disorders. Strategies employed by typically developing children may be more working memory efficient, which results in the performance of typically developing children being less dependent on working memory capacity. For example, children with disorders may make less effective use of long-term memory retrieval and therefore depend on working memory functioning more than typically developing children. Second, the differences in associations may be a result of their use of manipulatives: children with lower confidence in their mathematic skills (predominantly girls) have a greater tendency to persevere in their use of manipulatives (Carr, Steiner, Kyser, & Biddlecomb, 2008), which requires them to actively compare various modes of representation when solving a mathematical problem. Finally, inflated correlations between working memory components and mathematical performance in children with various types of disorders may be a consequence of greater variance in both working memory and mathematical skills within groups with disabilities, leading to stronger statistical associations. This could also be seen as a restriction in range problem in typically developing children (Aron and Aron, 2002). In other words: the relative uniformity of the group may cause correlation coefficients to drop.

Country of origin only contributed to the prediction of the correlation between mathematical performance and shifting, but not any other working memory component. Studies from Europe yielded higher correlation coefficients than studies from other countries. This might be caused by differences in teaching methodology or design of mathematics tests, or by different strategies used by children from different countries. However, further research is needed to investigate the cause of these differences in study outcomes.

4.3. Limitations and future research

Publication bias is a potential threat to the validity of the results in any meta-analysis. With a bias towards publishing studies that report significant effects between constructs, this might be a problem for the current study as well. Yet, correlation analysis, as the dependent measure in this study, is hardly ever a primary result of any article, as most publications focus on more advanced analyses to answer their research questions. We assume that this reduces the threat of publication bias to this study. Funnel plot analyses and trim and fill procedures confirmed that publication bias was probably not a large problem in this study.

Furthermore, the current study was limited to studies in which correlation coefficients were reported, which limits the number of studies included, amongst which the studies that do not report on correlations between constructs. Moreover, studies with various alternative designs may be informative about the relations between working memory and mathematical performance. One of these types is the experimental design, in which one of the constructs is trained and training gains in both constructs are investigated (e.g., Holmes et al., 2009). A second design that was not included in the current study was the dual-task study, in which performance on a primary task is limited by resources consumed by a secondary task (e.g., Hecht, 2002). Finally, studies reporting exclusively on differences between children clustered in various samples of special needs and typical samples (such as comparisons of working memory span between children performing below-average or average in mathematics) were excluded (e.g., Geary, Hoard, & Bailey, 2012).

Investigating working memory, including executive functioning, is complicated because the impurity problem poses a challenge that cannot be resolved. The impurity problem refers to the fact that executive functions by definition operate on other processes, such as language skills, and therefore cannot be measured purely (Miyake et al., 2000). This problem can be partially addressed by the inclusion of measures that rely on different processes in the conditional models, such as in the current study. Inclusion of moderator variables that distinguish between tasks addressing different processes may be indicative of variation in relevance between the processing of different types of information, but could also be influenced by variation in impurity between categories of tasks, and greater relevance of certain processes for mathematical performance than others. For example, the distinction between visual and spatial information significantly contributes to the model explaining variation in the correlation between mathematical skills and visuospatial updating. This may indicate that visual information processing is indeed more relevant to mathematical performance than spatial information processing, contradicting previous findings concerning the involvement of spatial encoding in mathematics performance (Knops, Thirion, Hubbard, Michel, & Dehaene, 2009), but may also indicate that visual tasks are contaminated more by processes relevant to mathematic skills (or less by processes not relevant to mathematical performance). Disentangling variations in executive processes and influences of related processes poses a challenge to research in executive functions that has not been resolved so far, but may benefit from latent variable analysis of behavioural data or neuroimaging studies (Baddeley, 2007).

Furthermore, not only can measures of executive functions be contaminated by the processes they directly operate on, but also by other executive functions. It has been argued that different executive functions partially rely on the same processes. For example, inhibition and shifting both rely on a process of conflict resolution. In inhibition tasks, a predominant response must be inhibited in favour of another, and in shifting, there is competition between two response sets of which one must be chosen (see: Garon, Bryson, & Smith, 2008). The notion of cross-contamination between executive functions has been confirmed by the fact that factor analyses concerning the structure of executive functions do not consistently find the structure as proposed by Miyake et al., 2000; (e.g., Van der Sluis et al., 2007; Van der Ven, Kroesbergen, Boom, and Leseman, 2012b). Moreover, inhibition and shifting tasks also require continuous refreshing of the rules of the task at hand in working memory while the task is being performed. If cross-contamination of different executive functions is indeed of influence on the measures of the constructs, it is possible that an association between one component of working memory and mathematical performance is in fact a result of performance being dependent on another executive function. Relations between inhibition and shifting and mathematical performance have previously been found to be the result of shared explained variance with updating measures (Van der Ven et al., 2012b). Future studies should attempt to fractionate different aspects of executive functioning, and map the overlap in skill required to successfully perform various working memory tasks in order to disentangle their differential relationship to mathematical performance.

5. Conclusion

The current review contributes to the extant body of literature by drawing together the findings from various studies and investigating the relationship between working memory components and mathematics performance. Past research has extensively investigated the relations between working memory functioning and mathematical performance, but has yielded inconsistent results. The current meta-study confirmed the relationship between mathematical performance and all components of working memory. The working memory component that correlated most strongly with mathematical performance is verbal updating. It also clearly showed that correlations are influenced consistently by characteristics of the mathematics measure and sample type, and less consistently by characteristics of the working memory measure and age. Unresolved issues, primarily in the measurement of executive functions, pose challenges to this type of research that require methodological attention in future studies.

Acknowledgements

This research was supported by a project grant from the Netherlands Organisation for Scientific Research (NWO), grant 411-07-113. We would like to thank Joop J.C.M. Hox, Leoniek Wijngaards-de Meij and Marjan Bakker for their advice concerning the analytical models, Elise de Bree for her advice with regard to presentation of the theories, and Anna van Nunen for her assistance in searching literature and coding data. In addition, we would like to thank the three anonymous reviewers for their constructive feedback to the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.edurev.2013.05.003.

References

References with ∗ were included in the meta-analysis. A full list of studies included in the meta-analysis can be found in the Appendix.

∗Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*, 20–29. http://dx.doi.org/10.1016/j.jecp.2009.11.003.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioural characteristics of children with low working memory. *Child Development, 80*, 606–621. http://dx.doi.org/10.1111/j.1467-8624.2009.01282.x.

American Psychiatric Association, Diagnostic and statistical manual of mental disorders, 4th ed., 2000, American Psychiatric Association, 10.1176/appi.books.9780890423349.

∗Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive Psychology, 21*, 1201–1216. http://dx.doi.org/10.1002/acp.1317.

∗Andersson, U. (2008). Working memory as a predictor of written arithmetic skills in children: The importance of executive functions. *British Journal of Educational Psychology, 78*, 181–203. http://dx.doi.org/10.1348/000709907X209854.

∗Andersson, U., & Lyxell, B. (2007). Working memory deficit in children with mathematical problems: A general or specific deficit? *Journal of Experimental Child Psychology, 96*, 197–228. http://dx.doi.org/10.1016/j.jecp.2006.10.001.

Aron, A., & Aron, E. N. (2002). *Statistics for Psychology, Third Edition*. Upper Saddle River,NJ: Prentice-Hall.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology, 49A*, 5–28.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839. http://dx.doi.org/10.1038/nrn1201.

Baddeley, A., Working memory, thought and action, 2007, Oxford University Press; Oxford.

Baddeley, A. D. and Hitch, G. J., Working memory. In G.H. Bower (Ed.), The psychology of learning and motivation, Vol. 8 (47–89), 1974, Academic Press; New York.

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*, 327–336. http://dx.doi.org/10.1016/j.lindif.2011.01.007.

∗Blair, C., & Razza, R. P. (2007). Relating effortful control, executive functioning, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 87*, 647–663. http://dx.doi.org/10.1111/j.1467-8624.2007.01019.x.

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

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

Bull, R., & Scerif, G. (2001). Executive function as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*, 273–293. http://dx.doi.org/10.1207/S15326942DN1903_3.

Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*, 61–75. http://dx.doi.org/10.1016/j.lindif.2007.04.005.

Censabella, S., & Noël, M. P. (2008). The inhibition capacities of children with mathematical disabilities. *Child Neuropsychology, 14*, 1–20. http://dx.doi.org/10.1080/09297040601052318.

Dehaene, S., The number sense: How the mind creates mathematics, 1997, Oxford University Press; New York.

∗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 to second grade. *Journal of Experimental Child Psychology, 103*, 186–201. http://dx.doi.org/10.1016/j.jecp.2009.01.004.

Duval, S., & Tweedie, R. (2004). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2004), 455–463. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(1997), 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629.

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*, 465–486. http://dx.doi.org/10.1207/s15326942dn2601_6.

Friso-van den Bos, I., Kolkman, M. E., Kroesbergen, E. H. and Leseman, P. P. M. (in press). Explaining variability: Numerical representations in 4- to 8-year old children. *Journal of Cognition and Development.* DOI: 10.1080/15248372.2012.742900.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*, 493–513. http://dx.doi.org/10.1037/0022-0663.97.3.493.

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

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177–190. http://dx.doi.org/10.1037/0012-1649.40.2.177.

∗Gathercole, S. E., Tiffany, C., Briscoe, J., Thorn, A., & The ALSPAC team (2005). Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of Child Psychology and Psychiatry, 46*, 598–611. http://dx.doi.org/10.1111/j.1469-7610.2004.00379.x.

Geary, D. 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*, 291–307. http://dx.doi.org/10.1177/0022219410392046.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*(2004), 121–151. http://dx.doi.org/10.1016/j.jecp.2004.03.002.

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory & Cognition, 30*(2002), 447–455. http://dx.doi.org/10.3758/BF03194945.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*, F9–F15. http://dx.doi.org/10.1111/j.1467-7687.2009.00848.x.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233–253. http://dx.doi.org/10.1111/j.2044-835X.1998.tb00921.x.

Imbo, I., & LeFevre, J. –A. (2009). Cultural differences in complex addition: Efficient Chinese versus adaptive Belgians and Canadians. *Journal of Experimental Psychology, 35*, 1465–1476. http://dx.doi.org/10.1037/a0017022.

*Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96*, 284–309. http://dx.doi.org/10.1016/j.jecp.2006.09.001.

Jarvis, H. E., & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on National Curriculum tests at 11 and 14 years of age. *Educational and Child Psychology, 20*(3), 123–140.

Knops, A., Thirion, B., Hubbard, E. M., Michel, V., & Dehaene, S. (2009). Recruitment of an area involved in eye movements during mental arithmetic. *Science, 324*, 1583–1585. http://dx.doi.org/10.1126/science.1171599.

Kolkman, M. E., Hoijtink, H. J. A., Kroesbergen, E. H., & Leseman, P. P. M. (2013). The role of executive functions in numerical magnitude skills. *Learning and Individual Differences, 24*, 145–151. http://dx.doi.org/10.1016/j.lindif.2013.01.004.

*Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visio-spatial working memory, and pre-school quantity number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*, 516–531. http://dx.doi.org/10.1016/j.jecp.2009.03.009.

*Kroesbergen, E. H., Van Luit, J. E. H., Van Lieshout, E. C. D. M., Van Loosbroek, E., & Van de Rijt, B. M. (2009). Individual differences in early numeracy: The role of executive functions and subitizing. *Journal of Psychoeducational Assessment, 27*, 226–236. http://dx.doi.org/10.1177/0734282908330586.

*Kyttälä, M., Aunio, P., & Hautamäki, J. (2010). Working memory resources in young children with mathematical difficulties. *Scandinavian Journal of Psychology, 51*, 1–15. http://dx.doi.org/10.1111/j.1467-9450.2009.00736.x.

LeFevre J.A., DeStefano D., Coleman B. and Shanahan T., Mathematical cognition and working memory, In: Campbell J.I.D., (Ed), Handbook of mathematical cognition 2005, Psychology Press; Hove, England, 361–377.

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

*Mazzocco, M. M. M., & Kover, S. T. (2007). A longitudinal assessment of executive functioning skills and their association with math performance. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence, 13*(2007), 18–45. http://dx.doi.org/10.1080/09297040600611346.

McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visual spatial interference on children's arithmetical performance. *Educational and Child Psychology, 20*, 93–108.

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. http://dx.doi.org/10.1006/cogp.1999.0734.

*Monette, S., Bigras, M., & Guay, M. –C. (2011). The role of executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology, 109*, 158–173. http://dx.doi.org/10.1016/j.jecp.2011.01.008.

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology, 103*, 490–502. http://dx.doi.org/10.1016/j.jecp.2009.02.003.

*Navarro, J. I., Aguilar, M., Alcalde, C., Ruiz, G., Marchena, E., & Menacho, I. (2011). Inhibitory Processes, Working Memory, Phonological Awareness, Naming Speed, and Early Arithmetic Achievement. *Spanish Journal of Psychology, 14*, 580–588. http://dx.doi.org/10.5209/rev_SJOP.2011.v14.n2.6.

Noël, M. –P and Rousselle, L., Developmental changes in the profiles of dyscalculia: An explanation based on a double exact – and – approximate number representation model. Frontiers in Human Neuroscience 5, 2011, 165. DOI: 10.3389/fnhum.2011.00165.

Oberauer, K., Süß, H. –M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*, 167–193. http://dx.doi.org/10.1016/S0160-2896(02)00115-0.

Panaoura, A., & Philippou, G. (2007). The developmental change of young pupils' metacognitive ability in mathematics in relation to their cognitive abilities. *Cognitive Development, 22*, 149–164. http://dx.doi.org/10.1016/j.cogdev.2006.08.004.

*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*, 229–250. http://dx.doi.org/10.1080/87565640801982320.

*Passolunghi, M.-C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22*, 165–184. http://dx.doi.org/10.1016/j.cogdev.2006.09.001.

Raches, D., & Mazzocco, M. M. (2012). Emergence and nature of mathematical difficulties in young children with Barth Syndrome. *Journal of Developmental & Behavioral Pediatrics, 33*, 328–335. http://dx.doi.org/10.1097/DBP.0b013e31824c4090.

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. http://dx.doi.org/10.1016/j.lindif.2009.10.005.

*Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*, 137–157. http://dx.doi.org/10.1016/j.jecp.2005.01.004.

*Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2011). Modeling a cascade of effects: the role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science, 14*, 1161–1175. http://dx.doi.org/10.1111/j.1467-7687.2011.01068.x.

Rotzer, S., Loenneker, T., Kucian, K., Martin, E., Klaver, P., & Von Aster, M. (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. *Neuropsychologia, 47*, 2859–2865. http://dx.doi.org/10.1016/j.neuropsychologia.2009.06.009.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology, 35*, 1089–1096. http://dx.doi.org/10.1037/a0015730.

Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). *Technical Abstract (Woodcock-Johnson III Assessment Service Bulletin No. 2)*. Itasca, IL: Riverside Publishing.

*St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*, 745–759. http://dx.doi.org/10.1080/17470210500162854.

*Swanson, H. L. (2006). Cross-Sectional and Incremental Changes in Working Memory and Mathematical Problem Solving. *Journal of Educational Psychology, 98*, 265–281. http://dx.doi.org/10.1037/0022-0663.98.2.265.

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research, 76*, 249–274. http://dx.doi.org/10.3102/00346543076004429.

Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., & Liu, Y. (2006). Arithmetic processing in the brain shaped by cultures. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 10775–10780. http://dx.doi.org/10.1073/pnas.0604416103.

Toll, S. W. M., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2011). Executive functions as predictors of math learning disabilities. *Journal of Learning Disabilities, 44*, 521–532. http://dx.doi.org/10.1177/0022219410387302.

Van der Sluis, S., De Jong, P. F., & Van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *Journal of Experimental Child Psychology, 87*, 239–266. http://dx.doi.org/10.1016/j.jecp.2003.12.002.

*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*, 427–449. http://dx.doi.org/10.1016/j.intell.2006.09.001.

Van der Ven, S. H. G., Boom, J., Kroesbergen, E. H., & Leseman, P. P. M. (2012a). Microgenetic patterns of children's multiplication learning: Confirming the overlapping waves model by latent growth modeling. *Journal of Experimental Child Psychology, 113*, 1–19. http://dx.doi.org/10.1016/j.jecp.2012.02.001.

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012b). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82*, 100–119. http://dx.doi.org/10.1111/j.2044-8279.2011.02035.x.

Xenidou-Dervou, I., Van Lieshout, E. C. D. M. and Van der Schoot, M. (in press). Working memory in nonsymbolic approximate arithmetic processing: A dual-task study with preschoolers. *Cognitive Science.*

Yeniad, N., Malda, M., Mesman, J., Van IJzendoorn, M. H. and Pieper, S., Shifting ability predicts math and reading performance in children: A meta-analytical study. Learning and Individual Differences 23, 2013, 1–9. DOI: 10.1016/j.lindif.2012.10.004.