Psychological Bulletin

ISSN: 0033-2909

2022, Vol. 148, Nos. 5–6, 337–369
https://doi.org/10.1037/bul0000369

The Relation Between Executive Functions and Math Intelligence in Preschool Children: A Systematic Review and Meta-Analysis

Valentin Emslander[1] and Ronny Scherer[2]
[1] Faculty of Humanities, Education and Social Sciences, Luxembourg Centre for Educational Testing, University of Luxembourg
[2] Faculty of Educational Sciences, Centre for Educational Measurement at the University of Oslo, University of Oslo

Executive functions (EFs) are key skills underlying other cognitive skills that are relevant to learning and everyday life. Although a plethora of evidence suggests a positive relation between the three EF subdimensions, inhibition, shifting, and updating, and math skills for schoolchildren and adults, the findings on the magnitude of and possible variations in this relation are inconclusive for preschool children and several narrow math skills (i.e., math intelligence). Therefore, the present meta-analysis aimed to (a) synthesize the relation between EFs and math intelligence (an aggregate of math skills) in preschool children; (b) examine which study, sample, and measurement characteristics moderate this relation; and (c) test the joint effects of EFs on math intelligence. Utilizing data extracted from 47 studies (363 effect sizes, 30,481 participants) from 2000 to 2021, we found that, overall, EFs are significantly related to math intelligence ($\bar{r}$ = .34, 95% CI [.31, .37]), as are inhibition ($\bar{r}$ = .30, 95% CI [.25, .35]), shifting ($\bar{r}$ = .32, 95% CI [.25, .38]), and updating ($\bar{r}$ = .36, 95% CI [.31, .40]). Key measurement characteristics of EFs, but neither children's age nor gender, moderated this relation. These findings suggest a positive link between EFs and math intelligence in preschool children and emphasize the importance of measurement characteristics. We further examined the joint relations between EFs and math intelligence via meta-analytic structural equation modeling. Evaluating different models and representations of EFs, we did not find support for the expectation that the three EF subdimensions are differentially related to math intelligence.

Public Significance Statement
Executive functions (EFs) are key to learning, and children who score higher on EFs also show better scores in mathematics. This meta-analysis confirms that children who can better avoid (or *inhibit*) being distracted, *shift* easily between different tasks, or *update* the information they have just learned also score high on math intelligence tests. However, this link between EFs and math intelligence should always be interpreted in light of the measurement characteristics of EFs, as suggested by the moderator analyses. We found evidence to support the idea that the three EFs (inhibition, shifting, and updating) are equally important for math intelligence.

Keywords: cognitive skills, executive functions, mathematics, meta-analysis, preschool children

Supplemental materials: https://doi.org/10.1037/bul0000369.supp

Executive functions (EFs)—a set of mental processes that regulate human cognition and behavior (Miyake et al., 2000; Miyake & Friedman, 2012)—and their three arguably most investigated subdimensions (response inhibition, mental set shifting, and updating of working memory; aka inhibition, shifting, and updating) are considered prerequisites for many cognitive skills, such as reading, and correlate with fluid intelligence (Cassidy et al., 2016; Diamond, 2013; Follmer, 2018). In addition to providing empirical evidence of

Valentin Emslander https://orcid.org/0000-0003-3690-5883
Ronny Scherer https://orcid.org/0000-0003-3630-0710

The authors wish to express their sincere thanks to Marie Prunkl-Haberkorn and Tony C. A. Tan for their invaluable support during the screening and coding process, Caroline Bermes for her graphical support, senior librarian Toril M. Hestnes at the Library of Medicine and Science, University of Oslo, for performing the literature search as well as to Michael Schneider, University of Trier, for his valuable feedback on an earlier version of this article.

This research was supported in part by a grant from the Research Council of Norway to Ronny Scherer (NFR-FINNUT Grant 254744 "Adapt21") as well as the Emerging Researcher Award 2021 from the Division of Educational Psychology of the German Psychological Society (DGPs) and a grant from the Doctoral School in Humanities and Social Sciences of the University of Luxembourg to Valentin Emslander.

All supplemental material can be found under this link: https://osf.io/t67qa/?view_only=5ee953125f584687ae61af98b292464b.

Correspondence concerning this article should be addressed to Valentin Emslander, Faculty of Humanities, Education and Social Sciences, Luxembourg Centre for Educational Testing, University of Luxembourg, Campus Belval, Maison des Sciences Humaines, 11 Porte des Sciences, L-4366 Esch-sur-Alzette, Luxembourg. Email: valentin.emslander@uni.lu

the relation between EFs and these cognitive skills, researchers have repeatedly shown that EFs are linked to math skills, such as basic number knowledge, calculation, spatial skills, and mathematical reasoning, in schoolchildren and adults (see e.g., Best et al., 2011; Cragg et al., 2017; Friso-van den Bos et al., 2013; Peng et al., 2016; Yeniad et al., 2013) and play a crucial role in the development of math skills (van der Ven, 2011; Van der Ven et al., 2012). Therefore, it is critical to examine the preschool years and comprehensively investigate the constructs that have been found to contribute to the development of mathematical skills. For instance, development of an understanding of numbers begins before school entry (Passolunghi & Lanfranchi, 2012) and is related to EFs (Geary et al., 2019), suggesting that the preschool years are a formative period with rapid development of mathematical skills and EFs (Zelazo & Carlson, 2012). These developmental patterns might even differ between EF subdimensions in this age group (Diamond, 2013; Garon et al., 2008), as some EFs develop earlier than others. As preschool children are on the brink of structured schooling, their EFs constitute an important part of their school readiness and predict their academic success throughout their school careers (Blair, 2002; Diamond, 2013; G. J. Duncan et al., 2007). EFs and early mathematical skills are not only core areas of school readiness (the very goal of the preschool years) but also predict academic success later in the school career. Thus, both constructs are crucial to investigate in preschool children due to their age and preschool status.

Following the literature on the differentiation of cognitive skills over time, some researchers have argued that EFs and basic math skills actually measure the same underlying construct in children, as both are strongly linked to intelligence and can be integrated into the Cattel–Horn–Carroll (CHC) theory of intelligence (Ackerman et al., 2005; N. P. Allan et al., 2014; Fleming & Malone, 1983; Jewsbury et al., 2016; Roth et al., 2015). Specifically, Jewsbury et al. (2016) suggested subsuming inhibition and shifting under a general speed factor (*Gs*), and updating can be captured by a general memory factor (*Gsm*). Math skills have been categorized as quantitative knowledge (*Gq*), fluid intelligence (*Gf*), and visual processing (*Gv*; Schneider & McGrew, 2018; Uttal et al., 2013). However, other scholars have argued that EFs and math skills measure correlated but distinct constructs (see e.g., Best et al., 2011; Cragg et al., 2017; Friedman et al., 2006; Peng et al., 2016; Yeniad et al., 2013). These perspectives differ in whether math skills are conceptualized as broader (e.g., students' grades or performance on teacher-constructed math tests) or narrower (e.g., only intelligence tests with math components). Typically, EFs and math skills in preschool children are not measured with standardized questionnaires that require the children to read and write (cf. the Behavior Rating Inventory of Executive Function; Gioia et al., 2000), and researchers have to resort to innovative ways of assessing EFs and math skills. Thus, EFs might vary in their relation to math skills in terms of the measurement properties (e.g., N. P. Allan et al., 2014; Cortés Pascual et al., 2019) or study and sample characteristics (David, 2012; Friso-van den Bos et al., 2013; Peng et al., 2016). Investigating the influence of diverse measurement, sample, and study characteristics on the relation between EFs and math skills can provide valuable information to practitioners who want to streamline the assessment of these constructs and researchers who aim to better understand the nature of the relation.

The extant body of literature also reveals another issue. There is some disagreement about the magnitude of the joint relations of inhibition, shifting, and updating with math skills in preschool children. While some researchers found that all three are key predictors of math skills (e.g., R. Duncan et al., 2016; Purpura et al., 2017), others found that updating (Friso-van den Bos et al., 2013) or shifting (Jacob & Parkinson, 2015) is superior to the other EF subdimensions. With this equivocal discussion at hand, researchers should avoid jumping to conclusions about the differential effects of specific EFs, as strong claims call for even stronger evidence, and the field of EF research is no stranger to critical discussion (see e.g., the discussion on the relation between working memory and intelligence; Ackerman et al., 2005; Beier & Ackerman, 2005; Kane et al., 2005; Oberauer et al., 2005). Concerning the joint and individual contributions of inhibition, shifting, and updating to math skills, some of these divergent findings may depend on the ways in which EFs are represented in measurement models, for instance, as single or multiple latent variables or composite factors (Camerota et al., 2020). The common practice of utilizing single, reflective latent variables has been questioned, especially when describing the relations between EFs and math skills. For instance, Nguyen et al. (2019) examined an alternative model to the single latent EF variable model and showed that relations with math skills could also operate via unique components of EFs. To integrate these different approaches, we performed meta-analytic structural equation modeling (MASEM) and examined the effects of EFs on math intelligence jointly rather than separately.

To examine the relations between the three EF subdimensions and narrow math skills, in the present study, we synthesized 363 effect sizes from 47 studies from 2000 to 2021. Most meta-analyses have investigated the relation between a specific EF and math skills (e.g., N. P. Allan et al., 2014; David, 2012; Peng et al., 2016) or the relation between EFs and general cognitive abilities in preschool children (Ackerman et al., 2005; Brydges et al., 2012; Jewsbury et al., 2016). However, the present meta-analysis combines these two approaches by focusing on math skills related to general cognitive abilities, that is, math intelligence and the three EF subdimensions. Furthermore, we identified which study, sample, and measurement characteristics may moderate these relations and examined their multivariate nature. To the best of our knowledge, this study is the first to investigate the relation between EFs and math intelligence using a multilevel and multivariate meta-analysis of preschool children.

Theoretical Framework

Executive Functions and Their Measurement

EFs are a set of general-purpose control mechanisms that regulate human cognition and behavior and are important for self-control and self-regulation (Miyake et al., 2000; Miyake & Friedman, 2012). Miyake and Friedman (2012) characterized these control mechanisms in their EF framework based on four conclusions drawn from the literature on EFs. First, the three EF subdimensions in their framework—inhibition, shifting, and updating—tend to be distinct in adults, showing a diversity factor structure with three distinct factors. Yet, the three EFs are correlated, especially in preschool children, yielding a unitary factor for all three EFs (Wiebe et al., 2008). Second, twin studies have suggested that a large proportion of EFs are genetically inherited (Friedman et al., 2008). Third, measures of executive functions can be used to differentiate between clinical and nonclinical behaviors (Young et al., 2009). Fourth, longitudinal studies have shown that EFs are expressed in a stable way over the course of one's life (Mischel et al., 2011), although the structure of EFs develops over time and therefore differs between age groups. Other researchers have investigated the differences between emotionally arousing ("hot"; e.g., due to a punishment or reward) and emotionally neutral ("cool") EFs (Brock et al., 2009; Zelazo & Carlson, 2012). Their findings suggested that, when considered jointly, "hot" EFs are uniquely related to disruptive behavior in preschool children, and "cool" EFs are uniquely related to academic achievement (Brock et al., 2009; Willoughby et al., 2011). This distinction was further corroborated for inhibitory control by N. P. Allan et al.'s (2014) meta-analysis, hinting at a variety of executive function dimensions that can be distinguished beyond Miyake et al.'s (2000) work. Additionally, higher level EFs have been examined, such as "planning," tapping into multiple more basic EF processes (Miyake & Friedman, 2012). For the purpose of this meta-analysis, we further describe the EF framework proposed by Miyake et al. (2000) and its three subdimensions (inhibition, shifting, and updating), the ways EFs are measured in children, and their implications for math skills.

Inhibition is the ability to deliberately override a dominant, automatic, or prepotent response when needed (Miyake et al., 2000; Miyake & Friedman, 2012). Inhibition tasks generally include a conflict between the child's automatic response and the correct response. For instance, in the head–toes–knees–shoulders task (McClelland et al., 2014), a child is instructed to touch their head when prompted to touch their toes and vice versa, as well as to touch their knees when prompted to touch their shoulders and vice versa. One of the dependent variables is the number of correct trials, and the prepotent response that needs to be inhibited is touching the stated body part (see Garon et al., 2008, for a comprehensive review of EF tasks). In math learning, this ability comes into play whenever the obvious answer to a question is not the correct one, and the task demands a more thorough approach (Graziano et al., 2016). When a child is trading cards, stickers, or marbles with a friend, for instance, the friend might ask how many items (i.e., marbles) the child has. A specific answer requires the child to avoid shouting out the answer that comes to mind first (e.g., "many, many marbles") and instead count the marbles one by one and then tell their friend ("I have four marbles"). In addition, for children working on a math question, inhibition is important to stay focused and avoid the urge to give up or be distracted by more attractive alternatives (Clements et al., 2016).

Shifting—also referred to as attention shifting or cognitive flexibility (Diamond, 2013)—is the ability to switch back and forth between mental sets or tasks (Miyake et al., 2000; Miyake & Friedman, 2012). Tasks that measure shifting generally have an element of novelty to them, be it changing rules or adjusting priorities. For example, in the dimensional change card sorting task (Zelazo, 2006), the child is presented with cards displaying various shapes in different colors. The child is asked to sort the cards by color, and after some time, they are asked to sort the cards by shape. The number of correct responses after the sorting rule is changed serves as the dependent variable. This change in sorting rules represents the novelty inherent in shifting tasks. In a math skill-testing situation, shifting is represented by the act of switching to a new question and then applying the appropriate rule to the new question rather than sticking with one rule, although the question demands otherwise. Shifting also comes into play when the solutions to math questions require children to change their perspectives. To be a successful marble trader, for instance, the child in the previous example must constantly switch from receiving marbles and adding them to their count to giving marbles to their friend and subtracting them.

Updating is the ability to constantly monitor and rapidly manipulate the content of working memory (Miyake et al., 2000; Miyake & Friedman, 2012). Although this definition is only slightly different from working memory itself (viz., the ability to store and process information at the same time; Baddeley, 1992), the two constructs can be distinguished regarding several points. In Baddeley's (1992) memory model, working memory is an overarching brain system with three components: the central executive, the phonological loop, and the visual sketch. Both updating and working memory tasks go beyond merely holding content in working memory (Baddeley, 1992) and add the aspect of manipulation (Lehto, 1996), while updating tasks introduce an aspect of monitoring and replacing old pieces of information with new ones. Consequently, working memory can be assessed with a backward digit span. The child is asked to repeat a list of digits (or words) backward, with the number of correct responses in a row representing the dependent variable. This task requires the child to listen to and remember the list of digits (holding them in their working memory) while simultaneously reversing the order of the digits (processing the content of working memory). To make this an updating task, the child can be asked to use only the third last digit, which adds the aspect of monitoring and manipulating, rather than repeating all digits backward. In the present study, we refer to working memory and updating as *updating* due to the similarities in their assessment and considerable ambiguity in primary studies. In the mathematics context, updating is crucial for solving multistep problems. Using the marble trading example, the child needs to remember the preceding steps or preliminary results (e.g., "I have two marbles") to decide on the next step and then update their previous results mentally (Harvey & Miller, 2017). The latter step could be to add one marble to the count (e.g., "My friend gave me one marble. So, I have three now") without having to count the marbles again.

Despite the distinction between multiple EF subdimensions, this distinction may not be clear-cut for complex EF tasks. In fact, such tasks may require multiple EF processes or processes other than executive functioning. Friedman et al. (2008) described this issue of "task impurity" using an example from the Wisconsin Card Sorting Test (WCST): While some WCST tasks require shifting due to changing sorting rules, the tests may also require perceptual and motor skills. Moreover, EF and math tasks may share design and assessment features, such as relying on numbers or operations (i.e., "method overlap"). Therefore, the relations between EFs and other constructs could be biased. To circumvent such bias, latent variable models have been proposed to account for common and unique variations among EF tasks and processes (Camerota et al., 2020).

Several models that describe the structure of EF assessments have been developed and reported in primary studies. These models include, but are not limited to, latent variable models with a single latent EF variable, multiple-correlated latent EF variables, or multiple latent EF variables with a general EF factor and some specific factors (Camerota et al., 2020; Friedman & Miyake, 2017).

The question of the unity or diversity of executive functions is a key question in EF research and has initiated many empirical studies that have explored the fit of appropriate measurement models (Karr et al., 2018). However, such measurement models may vary among age groups, and most evidence supports the representation of EFs by a single latent variable for young and preschool-age children (Wiebe et al., 2008; Willoughby et al., 2012). At the same time, some evidence also points to the representation of EFs by a two-dimensional model (Karr et al., 2018; Lerner & Lonigan, 2014). Nonetheless, over time, the various EF processes may become more differentiated in later childhood (Brydges et al., 2014; Lerner & Lonigan, 2014).

Math Intelligence and Its Measurement

In the present study, we conceptualized math skills as part of intelligence due to their inherent fluid and crystallized aspects (Woodcock et al., 2001; Wechsler, 2003). Some theoretical frameworks subsume math skills into subfacets of intelligence. For instance, the CHC theory of intelligence categorizes math skills mainly under the narrow stratum I of fluid intelligence (*Gf*) as quantitative reasoning—that is, the inductive and deductive reasoning abilities involving mathematical relations and properties—and under stratum I quantitative knowledge (*Gq*), which includes mathematical knowledge and achievement (Schneider & McGrew, 2018). Mathematics skills related to geometry and space may also involve visual processing (*Gv*) generally and spatial reasoning specifically (Uttal et al., 2013). This conceptualization of math skills as part of general intelligence, which we refer to as "math intelligence" (e.g., Dweck, 2014; Martens et al., 2006; Rattan et al., 2012), is based on two elements. First, the present sample was comprised of preschool children who had not yet been exposed to formal schooling; thus, school achievement scores or teacher grades were not included. Second, we reasoned that intelligence and EFs are related (Friedman et al., 2006). Although much research attention has been paid to examining the relations between EFs and domain-general intelligence components (Ackerman et al., 2005), little is known about the relation between EFs and math-specific intelligence components.

Further, we differentiate between three subdimensions of math intelligence tasks: calculation and reasoning, basic number knowledge, and spatial tasks. Calculation and reasoning, which encompass numerical operations and applied mathematical problems, can be measured by, for instance, the Applied Problems subscale of the Woodcock–Johnson Test of Cognitive Ability (Version III; Woodcock et al., 2001). This is a widely used intelligence test validated for individuals ages 2 to older than 90. The Applied Problem scale includes subitizing, simple subtraction, and calculation tasks, among others. Basic number knowledge, which comprises counting and cardinality, can be measured with simple tasks in which children are asked to count for as long as they can. The highest number to which the children can correctly count serves as the dependent variable. Spatial tasks involve aspects of geometry, shapes, and algorithms. Examples include shape recognition or shape-matching tasks in which children must match a probe with a target shape (e.g., Child Math Assessment; Starkey et al., 2004). In the present study, spatial tasks were later merged with the "other" category due to the small number of available effect sizes.

The three subdimensions of math intelligence were derived from previous theory and research. From a theoretical perspective, they reflect the three aspects of the CHC theory of intelligence adapted to math intelligence in preschool-age children: Calculation and reasoning and basic number knowledge represent fluid intelligence (*Gf*) and quantitative knowledge (*Gq*), whereas spatial tasks represent math-related visual processing (*Gv*) in preschool children (Schneider & McGrew, 2018). We merged the calculation and reasoning categories into one subdimension following previous factor-analytic research, which indicated that mathematical operations (i.e., calculation) and reasoning load on quantitative knowledge (*Gq*; Parkin & Beaujean, 2012). Friso-van den Bos et al. (2013) specified similar categories (simple arithmetic; counting and basic understanding of numerical concepts; geometry, shapes, and algorithms) but included additional categories that did not apply to preschool children (e.g., advanced arithmetic or teacher rating). Peng et al. (2016) drew on similar but more differentiated categories of math tasks (i.e., basic number knowledge, whole-number calculations, fractions, geometry, algebra, and word-problem solving). Thus, we derived the math intelligence construct and its three subdimensions in the present study from previous theory and research.

Relations Between Executive Functions and Math Skills

EFs contribute to the process of learning and performing math skills. Not only are there applied settings where EFs and math skills are related, as described previously, but there is also strong evidence that preschool children's EFs can predict their math skills later (Bull et al., 2008; Clements et al., 2016; Cragg & Gilmore, 2014; McClelland et al., 2014). For researchers and educators, it is crucial to better understand this relationship in preschool children in order to pave the way for success in their (academic) careers (Ancker & Kaufman, 2007; G. J. Duncan et al., 2007). In addition, measures of EFs can be deployed to identify children who need more help learning mathematical skills (C. A. C. Clark et al., 2010).

Because EFs are crucial for learning and performing math skills, a lack of EFs can predict difficulties in math skills. This has been suggested by experimental dual-task studies, such as those focused on updating (see Raghubar et al., 2010, for a review). To adequately assist children who are struggling with math skills early on, it is important to clarify the structure of the EF construct and the potential differential contribution of the three EF subdimensions. In contrast to most of the evidence suggesting a unitary EF construct in preschool children (Wiebe et al., 2008), several researchers have shown results that support the notion of distinct EF subdimensions in that age group (Carlson, 2005; Espy et al., 2001). Thus, it is unclear whether EFs are best represented as one construct or as several subdimensions. However, this knowledge could help educators and parents assist struggling children to thrive with tailored instruction later in school (e.g., by reducing updating demands, as in Case et al., 1996) or could help parents provide support and strengthen EFs in a playful way (see e.g., Hutchison & Phillips, 2018). However, there seems to be no strong evidence of far transfer from, for instance, updating to reading comprehension and arithmetic (Melby-Lervåg & Hulme, 2013) or between EFs (Kassai et al., 2019). Furthermore, although it might be possible to train EFs, the benefits of such training seem to vanish quickly (Takacs & Kassai, 2019). From an economic perspective, knowledge of the EF subdimensions is relevant for the construction of measures for EFs and math skills. If the constructs cannot be differentiated, they could be measured with one instrument, saving time and financial resources as well as reducing the effort required of children. Examining the relation between EFs and math skills will inform not only theorization but also the current assessment practices for EFs in preschool educational contexts.

The unity and diversity of EFs have implications for describing the EF–math intelligence relationship. Different representations of EFs can result in different inferences drawn from this relationship. Nguyen et al. (2019) noted that there is a lack of agreement about the theoretical assumptions underlying the link between EFs and math skills. Specifically, the authors observed the common practice of representing EFs as a single latent variable or a composite score (see also Camerota et al., 2020). With this representation, the relation between EFs and math skills operates through a variable that captures what is common among the manifest EF indicators (Borsboom et al., 2003). Such a model ultimately provides information about the extent to which a unitary EF construct explains variations in the measures of skills in mathematics and sheds light on the joint contribution to math skills. However, if researchers are interested in the unique effects of EFs, the EF–math relations could be described via the residuals of the manifest EF indicators or multiple latent EF variables (Gignac & Kretzschmar, 2017; Nguyen et al., 2019). Finally, models describing the direct relations between multiple manifest or latent EF variables and math skills have the potential to unravel whether differential or uniform effects exist (Arán Filippetti & Richaud, 2017). Overall, the joint relations between multiple EFs and math intelligence can be modeled and interpreted in several ways.

Previous Meta-Analyses

We identified eight meta-analyses that investigated the relation between math skills and EFs as a whole, as well as the specific subdimensions (inhibition, shifting, and updating). This body of research revealed the need to bring together multiple EF subdimensions, a focus on age groups, and conceptualizations of math skills. Table 1 presents an overview of the previous meta-analyses.

In general, studies have found moderate correlations between all three subdimensions of EF and math skills. N. P. Allan et al. (2014) found a moderate average correlation ($\bar{r}$ = .34) between inhibition and the development of academic skills, including math skills, in their meta-analysis of preschool children. For shifting, Yeniad et al.'s (2013) meta-analysis yielded a correlation of $\bar{r}$ = .26 between shifting and math skills (operationalized as math performance) in children ages 4–14 years. Five meta-analyses (Cortés Pascual et al., 2019; David, 2012; Friso-van den Bos et al., 2013; Peng et al., 2016; Swanson & Jerman, 2006; see Table 1) focused on the relation between updating and math skills and found slightly higher correlations than those reported previously for inhibition and shifting. Two meta-analyses (David, 2012; Swanson & Jerman, 2006) found medium-to-large group differences in updating between children and young adults with and without math difficulties in favor of the group without math difficulties. Extending the sample to participants between 3 and 52 years of age, Peng et al. (2016) found a correlation of $\bar{r}$ = .35 between updating (operationalized as working memory) and math skills. Recently, Cortés Pascual et al. (2019) examined the links between composite EF and updating scores and math performance in children ages 6–12 years, finding similar average correlations ($\bar{r}$ = .37) for both links (Table 1).

Two meta-analyses examined the connections between math skills and all three EF subdimensions separately. Friso-van den Bos et al. (2013) examined the relation between math skills and inhibition ($\bar{r}$ = .27), shifting ($\bar{r}$ = .28), and updating (visuospatial: $\bar{r}$ = .34; verbal: $\bar{r}$ = .38) in samples of 4–12-year-olds. Jacob and Parkinson (2015) investigated the link between the three EF subdimensions and math skills (indicated by math achievement) in preschool children and schoolchildren, as well as in adolescents (ages 3–5 years, 6–11 years, and 12–18 years, respectively). Their results showed moderate correlations across the age groups and EF subdimensions, with an overall correlation of $\bar{r}$ = .31. However, Jacob and Parkinson (2015) did not examine inhibition, shifting, and updating separately across the three age groups. Thus far, only this meta-analysis and one other have focused on preschool children (N. P. Allan et al., 2014; Jacob & Parkinson, 2015). All the meta-analyses broadly conceptualized math skills, including a wide range of subdimensions, measures, and tasks. Two meta-analyses showed substantial heterogeneity in correlations due to variations in math skills and tasks (Friso-van den Bos et al., 2013; Peng et al., 2016). In addition, key aspects of the unique and joint effects of EFs on math skills have not yet been investigated meta-analytically.

Possible Moderators

Previous meta-analyses of the relations between EFs and math skills have suggested a wide array of possible moderators, focusing on differences in age and measurement (see Table 1). In the present meta-analysis, we focused on study, sample, and measurement characteristics as three key sources of heterogeneity (for an extensive list of all coded moderators, see the Coding of Studies section and the codebook in Supplemental Material S1). As in any meta-analysis, study characteristics, such as variables indicating the context of the primary study or whether a primary study was published, can reveal research trends and design issues (see Ferguson & Heene, 2012). Sample characteristics, such as gender composition, country of origin, average age, and whether the children were in kindergarten, preschool, or still at home, have been found to explain substantial heterogeneity (see Friso-van den Bos et al., 2013; Yeniad et al., 2013). In particular, differences in age may lead to divergent findings because the development of EFs and math skills evolves rapidly before children enter school (N. P. Allan et al., 2014; Garon et al., 2008). David (2012) and Swanson and Jerman (2006) found that the links between updating and math skills are stronger in younger children than in older ones, and Friso-van den Bos et al. (2013) reported similar findings for shifting. We coded the authors' descriptions of the preschool status of the respective samples. As for gender composition, some studies found no effect on the relation between EFs and math skills (Bull et al., 2008; C. A. C. Clark et al., 2010). In a meta-analysis of children ages 6–12 years, however, Cortés Pascual et al. (2019) found that gender composition significantly moderates the relationship of updating and an EF composite with math performance. At the same time, most studies in the present sample did not explicitly examine the possible influence of gender, leaving a research gap in the literature on preschool children. Additionally, there is some evidence that children from different countries and continents vary in terms of performance speed and problem-solving strategies, which might

Table 1

Prior Meta-Analytic Findings on the Relation Between Executive Functions and Math Skills in Children

Reference	Description and method	Significant moderators (selected)	Age range	Examined relation	k_S	$\bar{r}$	95% CI
Allan et al. (2014)	Meta-analysis of the link between inhibition and academic skills (encompassing literacy and math)	• EF task type (hot vs. cold) • Test mode (behavioral vs. rating)	2.5–6.5	Inhibition and the development of math as a school subject	75	.34	[.29, .39]
Cortés Pascual et al. (2019)	Meta-analysis of the link of WM and an EF composite score with math performance	• EF subdimensions	6–12	EF composite and math performance	18	.37	[.30, .42]
		• Gender composition	6–12	WM and math performance	11	.37	[.29, .45]
David (2012)	Meta-analysis comparing groups with and without math learning difficulties on three aspects of WM	• WM task type (numerical vs. nonnumerical)	9–21	Phonological loop $d = -0.36$	12	—	[−0.58, −0.14]
			9–21	Visuospatial sketchpad $d = -0.59$	9	—	[−0.87, −0.31]
		• Age	9–21	Central executive $d = -0.93$	17	—	[−1.53, −0.33]
Friso-van den Bos et al. (2013)	Multilevel meta-analysis of the link of inhibition, shifting, and updating with math skills	• Math skills task type	4–12	Inhibition and math skills	29	.27	—
		• WM task type	4–12	Shifting and math skills	18	.28	—
		• Age	4–12	Visuospatial updating and math skills	21	.34	—
			4–12	Verbal updating and math skills	85	.38	
Jacob and Parkinson (2015)	Meta-analysis looking at cross-sectional and longitudinal data on the EF subdimensions, EF composite, and three age groups accounting for dependencies by clustering standard errors by lead authors	• EF subdimension	3–5	EF composite and math achievement	26	.29	[.23, .36]
			6–11	EF composite and math achievement	34	.35	[.28, .41]
			12–18	EF composite and math achievement	8	.33	[.25, .42]
			3–18	EF composite and math achievement	60	.31	[.26, .37]
			3–18	Inhibition and math achievement	33	.31	[.25, .38]
			3–18	Shifting and math achievement	17	.34	[.24, .44]
			3–18	Updating and math achievement	40	.31	[.22, .39]
Peng et al. (2016)	Meta-analysis of the WM–math skills link, accounting for dependencies of effect sizes with robust standard error estimation	• WM type (e.g., verbal, numerical) • Types of math skills	3–52	WM and math skills	110	.35	[.32, .37]
Swanson and Jerman (2006)	Multilevel meta-analysis comparing memory performance of students with and without math disabilities	• Type of cognitive measure	6–13	Verbal WM $d = -0.70$	43[a]	—	[−0.79, −0.61]
		• Age	6–13	Visuospatial WM $d = -0.63$	13[a]	—	[−0.77, −0.48]
Yeniad et al. (2013)	Meta-analysis of the links between shifting, math performance, and intelligence	• None, due to the small number of included studies	4–14	Shifting and math performance	18	.26	[.15–.35]
			4–14	Shifting and intelligence	11	.30	[.18–.41]

Note. Larger positive effect sizes indicate a closer relationship between EFs and math intelligence. Common meta-analysis refers to a meta-analysis that takes only one effect size per study into account. k_S = number of included studies; age range = mean age of the youngest and the oldest sample within the meta-analysis in years; EF = executive function; $\bar{r}$ = weighted mean correlation; d = Cohen's d; WM = working memory; CI = confidence interval.
[a] Number of included effect sizes (*not* studies).

hint at differences in instruction and number language (Imbo & LeFevre, 2009). Furthermore, studies have reported that socioeconomic status is related to EFs (Blair, 2010) and influences their relation to math skills (Noble et al., 2007; Riggs et al., 2006). Finally, measurement characteristics are especially important in the context of preschool children because children at that age generally are not yet able to read or write. Therefore, commonly used measures based on reading and writing (e.g., paper-and-pencil word problem-solving tasks) cannot be applied. As a result, a wide array of alternative measures of inhibition, shifting, and updating have been used, introducing heterogeneity to the pool of measurements for preschool children. Some previous meta-analyses have supported the possible differential relation between different types of EF tasks and math skills (see N. P. Allan et al., 2014; Jacob & Parkinson, 2015). Table S6.1 presents a summary and examples of all types of EF tasks examined in the present meta-analysis. Regarding the measurement characteristics of math skills, different task types (e.g., basic number knowledge tasks vs. numerical reasoning tasks) have been proposed (see N. P. Allan et al., 2014; Peng et al., 2016) and have been found to moderate the relation between updating and math skills (Friso-van den Bos et al., 2013). Despite this evidence, several key characteristics have been tested to only a limited extent, including whether using a verbal, paper-and-pencil, behavioral, or computer-based test of EFs or math skills makes a difference (see N. P. Allan et al., 2014); whether the constructs were tested in a group setting or individually; and whether the test was a performance test or a third-person rating. In the present meta-analysis, we investigated the possible moderating effects of such characteristics on the relation between measures of EFs and math intelligence within and between studies. In the Coding of Studies section, we provide a list of all moderators.

The Present Meta-Analysis

The present meta-analysis was aimed at elucidating the relation between EFs and math skills—the latter of which is conceptualized as a facet of intelligence—in preschool children. The meta-analysis also aims to quantify variations within and between studies through multilevel meta-analysis, explaining these variations based on the study, sample, and measurement characteristics, and to test the amount of variation in math skills that can be jointly explained by inhibition, shifting, and updating. To answer the research questions, we followed the steps of the Meta-Analysis Reporting Standards (APA Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008), and we utilized advanced meta-analytic approaches, including multilevel meta-analysis and MASEM. Specifically, the present meta-analysis addresses three research questions:

1. To what extent are EFs (represented by a composite and by the three subdimensions of inhibition, shifting, and updating) and math skills (conceptualized as math intelligence) related in preschool children? (Overall correlations)

2. To what extent do these relations vary within and between studies, and which sample, study, and measurement characteristics explain this variation? (Heterogeneity and moderators)

3. To what extent do the three subdimensions of EFs (i.e., inhibition, shifting, and updating) differ in their ability to explain variations in math intelligence, and how much variation do they jointly explain? (Model testing)

Method

Literature Search

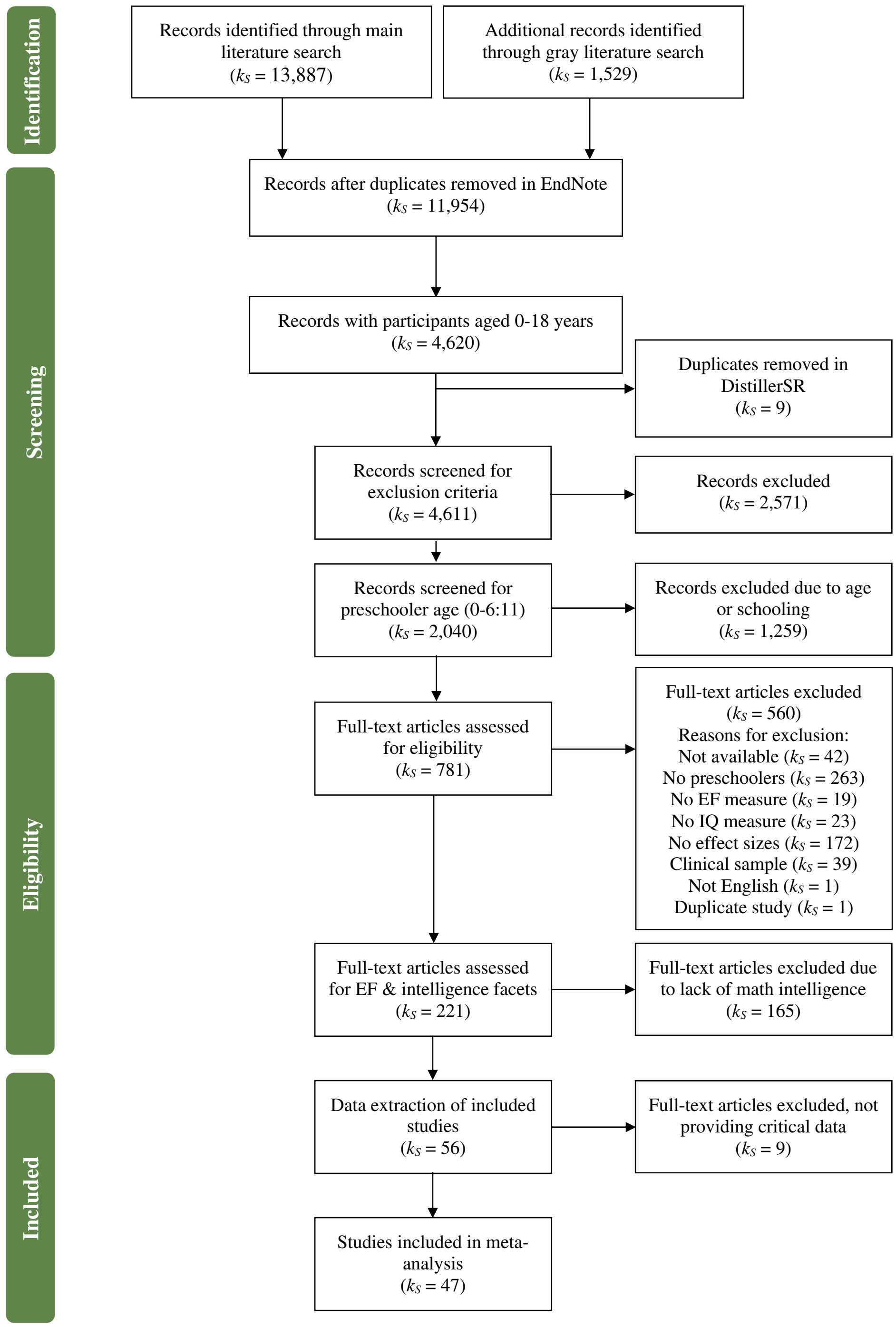
We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. The literature search and parts of the study selection were part of a broader project conducted by the authors that focused on the interrelations between EFs and different facets of intelligence in the general population. For this reason, the first steps of the literature search, screening, and coding procedure included the broad construct of intelligence. The literature search focused on published peer-reviewed articles and dissertations listed in Medline, Embase, ERIC, PsycINFO, ISI Web of Science (Core Collection), and ProQuest Dissertations and Theses. To mitigate potential publication bias, we also searched for gray literature in Google Scholar, OpenGrey, PsyArXiv, and ResearchGate. We performed an initial search of these databases in November 2018 and updated the search in March 2019 and in February–March 2021. The two updates yielded about 1,800 additional publications. We adjusted the search terms to fulfill the requirements of the different databases. The details of these searches, the search terms, and the full search strategy are provided in Supplemental Material S2. All searches were limited to results from between January 1, 2000, and the dates above; English-language publications; and studies with human subjects (Boyle et al., 2018). The start date was chosen to ensure that the studies utilized the core EF dimensions defined in Miyake et al.'s (2000) seminal paper. More practical reasons for this start date were to minimize cohort effects and keep the number of studies manageable. Given that math skills are considered an element of intelligence and cognate reasoning skills, we first based the search terms on EFs, intelligence, and reasoning:

> ((executive (function* or dysfunction* or control* or abilit*)) OR (dysexecutive (function* or dysfunction* or control* or abilit*)) OR (cognitive control)) OR (short term memor*) OR (working memor*) OR (updating) OR (Multitasking Behavior) OR (Inhibition) OR (Shifting) OR (Switching) OR ((cognitive or mental) flexibilit*)) AND ((Intelligence) OR (IQ) OR ((Intelligence or IQ) (test* or measure* or examination* or tool* or score* or scoring or scale* or instrument* or assessment* or rating* or evaluat* or questionnaire*)) OR (reasoning) OR (Problem Solving) OR (Decision Making)).

The first step of the search (the main literature search) yielded 13,887 publications, and the second step (gray literature search) yielded 1,529 publications. We removed duplicates and all publications with participants older than 18 years to ensure a focus on children and youths (Bramer et al., 2016), resulting in 4,620 publications. We then transferred these publications to DistillerSR (Evidence Partners, 2021) and deduplicated them once more, resulting in 4,611 publications for further screening. The literature search, deduplication, and screening procedure for all references are summarized in Figure 1. As part of the screening, we extracted relevant publications that focused on math intelligence.

Figure 1

Flowchart Summarizing the Literature Search and Study Selection



Note. k_S = number of studies; EF = executive function. See the online article for the color version of this figure.

Study Selection

The screening procedure comprised three steps—initial screening of titles and abstracts, screening for preschool children, and full text screening (see Table 1). Table S6.2 in the Supplemental Material presents detailed descriptions of all three screening steps. After following the three steps, we included all studies that fulfilled the following five criteria: (a) The study had to be empirical and report original research findings. (b) The mean age of the sample had to be at or less than 6 years and 11 months, with more than half of the sample not in primary school yet. We chose this high cutoff age to possibly include children from countries such as Germany, where children tend to start first grade in the year they turn 7 years old. In combination with including only samples of which the majority were not yet in school, we tried to create an age-inclusive but predominantly preschool overall sample. (c) The study had to contain at least one sample in which the majority of participants were healthy and not diagnosed with a disorder or medical condition. This criterion maximizes the generalizability of the findings to the general public, as recent meta-analyses have found impaired EFs in children with conditions such as Type 1 diabetes mellitus, fetal alcohol spectrum disorder, or high-functioning autism spectrum disorder (Broadley et al., 2017; Kingdon et al., 2016; Lai et al., 2017). (d) The facet of intelligence and at least one EF had to be measured. (e) The zero-order correlations of this relation had to be reported, or sufficient information to compute such an effect size had to be given. We did not include partial correlations or the results of multivariate analyses, as they do not purely represent the relation between EFs and facets of intelligence and, therefore, are not equivalent to zero-order correlations (Lipsey & Wilson, 2001).

We excluded studies during the three screenings if they fulfilled one or more of the following seven exclusion criteria: (a) The study was published before January 1, 2000. (b) The language of reporting was not English. (c) The subjects were nonhuman. (d) The results of the same sample were included in another study. (e) The abstract, full text, and secondary sources reporting the results of the study were unavailable. (f) Intelligence was operationalized as anything other than math intelligence, numerical intelligence, figural intelligence, verbal intelligence, or literacy skills (e.g., emotional intelligence, theory of mind). (g) Results were reported only for samples with a medical condition or a disorder or for samples with a mean age older than 6 years and 11 months.

A total of 221 studies were eligible for the next screening step after we applied the inclusion and exclusion criteria. The screening was conducted by one of the authors together with another graduate coder who had undergone training with a screening manual in 2020. After the update in 2021, the coding author worked with the DistillerSR (Evidence Partners, 2021) artificial intelligence module (see Baguss, 2020, for an introduction). Between 20% and 30% of the studies in each screening step were double screened. Any disagreement was settled through discussion among the authors, and the interrater reliability of the coding (κ) ranged from 93% to 98%.

After this screening, we conducted a finer-grained screening to identify which studies examined EFs and math intelligence in preschool children. Table 2 presents an evidence gap map of the frequency of intercorrelations between EF subdimensions and facets of intelligence. As we aimed to sort the studies by EF and intelligence facets, we differentiated among four categories of EFs: inhibition, shifting, updating, and a composite of multiple EFs. Intelligence measures were divided into four categories (subcategories are listed in parentheses): standardized intelligence tests (figurative, verbal, or numerical), math intelligence tests (standardized math test or math-related test), literacy skills tests (reading comprehension, reading fluency, receptive vocabulary, productive vocabulary, language fluency, letter identification, or language comprehension), and working memory tests. Of the 221 studies included in the facet screening, 56 studies reported results for

Table 2
Evidence Gap Map of the Number of Studies Measuring EF Subdimensions and Intelligence Facets in Preschool Children

Intelligence	Inhibition	Shifting	Updating	EF composite	**EF sum**
Standardized IQ test	58	37	48	39	102
Figurative intelligence	41	28	36	29	77
Verbal intelligence	32	17	23	22	57
Numerical intelligence	8	3	6	1	9
Math intelligence	39	21	34	18	56
Standardized math test	32	16	28	18	48
Math-related test	11	9	10	2	14
Literacy skills	86	53	60	40	124
Reading comprehension	8	5	7	5	15
Reading fluency	6	4	4	1	8
Receptive vocabulary	60	37	38	31	89
Productive vocabulary	24	15	15	10	31
Language fluency	11	9	10	4	16
Letter identification	22	10	15	11	32
Language comprehension	15	9	12	3	20
Working memory	90	63	95	21	100
Intelligence sum	160	104	129	65	

Note. Darker shaded cells identify sufficient amounts of research to conduct a meta-analysis. Lighter cells indicate a need for more primary studies. Numbers are based on $k_S = 221$ studies with preschool children identified in the screening process. EF = executive function. See the online article for the color version of this table.

at least one EF and math intelligence and therefore met the inclusion criterion for the present meta-analysis.

Coding of Studies

Three independent coders extracted data from the full texts and coded data on the study, sample, and measurement characteristics for moderator analyses beyond the statistical relation between EFs and math intelligence. The following study characteristics were coded: publication year (2000–2021), publication status (published study vs. gray literature), type of publication (journal article, conference paper, dissertation, or preprint), study background (whether the research question of the study was educational vs. clinical), and study design (whether the data were longitudinal vs. cross-sectional). The coded sample characteristics were children's mean age (in years), gender composition (percentage of girls in the sample), percentage of school students in the sample (samples with 50% or more school students were excluded), country of sample origin (e.g., the United States, Spain, and China), continent of sample origin (North America vs. Australia, Asia, and Europe), sample size, preschool status of the majority of the children in the sample (prekindergarten, kindergarten, preschool), and socioeconomic status of the sample (low, low to medium, medium, medium to high, high). We coded socioeconomic status based on either an explicit indication (e.g., "families reported a medium-high socioeconomic level"; Cueli et al., 2020, p. 239) or, if that was not available, a combination of factors (e.g., parents' education, eligibility for free meals, household income in comparison to the national average) rated by two coders.

For the measurement characteristics of math intelligence, we coded the number of items used to assess math intelligence, math intelligence subdimension (basic number knowledge, calculation and reasoning, spatial, composite), mode of math intelligence testing (paper-and-pencil, verbal, behavioral, verbal and paper-and-pencil, computer-based), whether math intelligence was tested with a performance-based test or a third-party rating, whether math intelligence was tested in a group setting or individually, and the internal consistency of the math intelligence measure (Cronbach's α). Measurement characteristics of EFs included the number of items used to assess EFs, EF subdimensions (inhibition, shifting, updating, or composite of at least two of the other subdimensions), type of EF task (e.g., Stroop task, Simon task, dimensional change task, span task; see Table S6.1, for descriptions and examples of all 15 task types), mode of EF testing (verbal, behavioral, apparatus-based, computer-based), whether EFs were tested with a performance-based test or a third-party rating, whether EFs were tested in a group setting or in an individual setting, and the Cronbach's α value of the EF measure. We considered the study characteristics and sample characteristics moderators at the study level and the measurement characteristics moderators at the effect size level. The three coders, one of the authors and two graduate coders, had undergone training with a coding manual. The two graduate coders showed excellent interrater reliability with the coding author (κ = 91% and 94%) for the 23 studies they had double coded. Disagreement was resolved through discussion after the interrater reliability estimation. The codebook, with a description of all coded variables and their possible values as well as all data, can be found in Supplemental Material S1.

Twenty-five studies did not report crucial information for the meta-analysis, such as the correlation coefficients between EFs and math intelligence. In such cases, we contacted the authors and asked them to provide the missing information. After 2–6 weeks and a reminder, 16 of 25 authors provided the missing information. Therefore, it was possible to include 47 studies. From these studies, we extracted 363 effect sizes for a total sample of 30,481 preschool children. The studies included three gray literature publications.

Data-Analytic Approaches

After conducting preliminary analyses for publication bias and influential cases, we applied multilevel meta-analysis, which has two major advantages over conventional meta-analysis, to address Research Questions (RQ) 1 and 2. First, this approach enabled us to include multiple effect sizes per study, in contrast to conventional meta-analyses, and enhanced statistical power (see Soveri et al., 2017). Second, it is possible to examine heterogeneity and moderator effects at the effect size and study levels. All the data sets as well as the general analytic code for RQ1 and RQ2 can be found in Supplemental Materials S1 and S3, respectively. To address RQ3, we performed MASEM. Supplemental Materials S1 and S4 represent the corresponding data set and the analytic code for the MASEM approach, respectively.

Publication Bias and Influential Cases

In psychological research, nonsignificant results are less likely to be published—a phenomenon called publication bias (Egger et al., 1997; Ferguson & Heene, 2012). To avoid replicating selectively published findings in the meta-analysis, we assessed the publication bias in the present data. We graphically inspected the 363 effect sizes of this meta-analysis via contour-enhanced funnel plots.

A contour-enhanced funnel plot depicts all effect sizes in comparison to their respective standard errors, and different shades of color indicate their level of statistical significance (Peters et al., 2008). In cases of publication bias, the plot should be skewed to one side at the base of the funnel. Otherwise, the plot should be symmetrical, and studies with larger standard errors should be evenly spread around the base of the funnel (Egger et al., 1997). We performed Begg's rank correlation test to check for a significant association between the effect size estimates and their sampling variances (Begg & Mazumdar, 1994). Moreover, Fernández-Castilla et al. (2021) recently extended trim-and-fill analyses to multilevel meta-analysis and developed a procedure for obtaining estimates of the number of effect sizes that may have been suppressed due to selection bias (L_0^+ and R_0^+). Estimates above 2 (L_0^+) or 3 (R_0^+) respectively, could indicate the presence of selection bias.

Extending the baseline (correlated and hierarchical effects [CHE]) model, we performed the precision effect test by estimating the moderator effects of the sampling standard errors, the funnel plot test by estimating the moderator effects of the study sample sizes, and the precision effect estimate with standard error (PEESE) test by estimating the moderator effects of the sampling variances.

Multilevel Meta-Analyses

For the meta-analysis, we used the R packages metafor (Viechtbauer, 2010), metaSEM (Cheung, 2015b), dmetar (Harrer et al., 2019), robumeta (Fisher et al., 2017), and clubSandwich (Pustejovsky, 2021). To address all the research questions, we conducted the analyses with five data sets: the complete data set (encompassing the entirety of k_{ES} = 363 effect sizes), the inhibition data set (with $k_{ES} = 137$ effect sizes for inhibition as the EF), the shifting data set (with $k_{ES} = 96$ effect sizes for shifting as the EF), the updating data set (with $k_{ES} = 107$ effect sizes for updating as the EF), and the pooled data set with study-level data (i.e., one average effect size per study, $k_S = k_{ES} = 47$). To obtain the average effect size of the pooled data set, we aggregated the correlations between EFs and math intelligence to weighted-average, study-level correlations utilizing the aggregate()-function in the R package metafor. This aggregation was based on inverse-variance weighting and accounted for the dependencies between the multiple correlations within the studies (Viechtbauer, 2021). Specifically, we assumed a within-study correlation of $\rho = 0.3$ and specified a compound symmetric structure to represent this dependency.

As noted, most of the 47 studies provided more than one correlation and reported a relation between multiple EF tasks and multiple measures of math intelligence. Consequently, the effect sizes were dependent (Borenstein et al., 2011; Cheung, 2014). Specifically, due to the inclusion of multiple correlations within studies and the inclusion of multiple measures of EFs or math intelligence in the same study, the dependency structure may be correlational and hierarchical (Pustejovsky & Tipton, 2021). To account for these two forms of dependencies, we specified a three-level random-effects model with a constant sampling correlation (ρ) to obtain an estimate of the average correlation between EFs and math intelligence in preschool children and the respective variance components. Pustejovsky and Tipton (2021) referred to this model as the CHE model and emphasized that it is suitable in situations in which the data structure may be correlational and hierarchical, and where an informed estimate of the within-study correlation between effect sizes (ρ) can be obtained. Compared to multilevel meta-analysis (Van den Noortgate et al., 2013), the Level 1 sampling errors in the CHE model are assumed to be correlated. This assumption provides a suitable representation of the nature of the present meta-analytic data. Compared to robust variance estimation (Fernández-Castilla et al., 2020), the CHE model quantifies the variance components at the different levels of analysis and thus could shed light on the level at which heterogeneity exists or is explained by moderators.

In a review of the existing measurement models describing the structure of executive functions, Camerota et al. (2020) found support for average correlations between EF tasks of $\rho = 0.20$–0.40. Drawing from existing reviews and influential studies on the relations between EF tasks (Ackerman et al., 2005; Friedman & Miyake, 2017; Jewsbury et al., 2016) and the existing meta-analyses on the relations between mathematics skills and EFs (see Table 1), we assumed a constant sampling correlation of $\rho = 0.30$ for the complete, composite, and pooled data sets (as they include a broad range of EF and math tasks), and $\rho = 0.40$ for the inhibiton, shifting, and updating data sets (as they include EF tasks assessing only one EF subdimension). However, we examined the sensitivity of choosing these values, varying ρ between 0.20 and 0.60 and examining the effects on model parameters.

The CHE model without moderators represents a three-level random-effects model, incorporating ρ as follows (Cheung, 2015a; Pustejovsky & Tipton, 2021):

$$\text{Level 1 (Sampling variation)}: \quad \hat{\theta}_{ij} = \theta_{ij} + \varepsilon_{ij}, \tag{1}$$

$$\text{Level 2 (Within-study variation)}: \quad \theta_{ij} = \kappa_j + \zeta_{(2)ij}, \tag{2}$$

$$\text{Level 3 (Between-study variation)}: \quad \kappa_j = \beta_0 + \zeta_{(3)j}, \tag{3}$$

where, the estimator of the ith effect size, $\hat{\theta}_{ij}$, in the jth study is decomposed into the true effect size, θ_{ij}, and the residuals, ε_{ij}, at Level 1 with $\varepsilon_{ij} \sim N(0, s_j^2)$ and $\text{Cov}(\varepsilon_{hj}, \varepsilon_{ij}) = \rho s_j^2$ for $h \neq i$ and s_j^2, representing the average known sampling variance of study j. At Level 2, the true effect size, θ_{ij}, is then decomposed into the average study-level effect size, κ_j, and some deviation of the effect size, θ_{ij}, from κ_j. The variance $\tau^2_{(2)}$ indicates the within-study heterogeneity, $\zeta_{(2)ij} \sim N(0, \tau^2_{(2)})$. Similarly, the average effect size, κ_j, is decomposed into the average population effect, β_0, and the study-specific deviation to quantify the between-study heterogeneity, $\zeta_{(3)j} \sim N(0, \tau^2_{(3)})$. This three-level approach allowed us to quantify three components of variance via restricted maximum likelihood estimation: variance between studies (Level 3), variance between the effect sizes within a study (Level 2), and sampling variance (Level 1; see Van den Noortgate et al., 2013).

To support our choice of the CHE model as the baseline model, we specified alternative models with random effects and a fixed-effects model and compared them to the CHE model. For these comparisons, we examined the following information criteria: Akaike's information criterion (AIC), the Bayesian information criterion (BIC), and the corrected AIC (AICc; Cavanaugh, 1997). Smaller values indicate a preference for a model.

To assess whether variations other than the sampling error existed, we calculated Cochran's Q (Cochran, 1950). A statistically insignificant Q value indicates homogeneity, whereas a statistically significant Q value indicates heterogeneity within the distribution of effect sizes (Ellis, 2010). To complement this overall heterogeneity test, we calculated the level-specific I^2 indices. The I^2 index quantifies the proportion of variance caused by heterogeneity, and it is categorized as low (25%), moderate (50%), and high (75%) heterogeneity (Higgins et al., 2003). The I^2 indices for Levels 2 and 3 in a three-level random-effects meta-analysis are defined as follows (Cheung, 2014):

$$I^2_{(2)} = \frac{\hat{\tau}^2_{(2)}}{\hat{\tau}^2_{(2)} + \hat{\tau}^2_{(3)} + \tilde{\upsilon}} \quad \text{and} \quad I^2_{(3)} = \frac{\hat{\tau}^2_{(3)}}{\hat{\tau}^2_{(2)} + \hat{\tau}^2_{(3)} + \tilde{\upsilon}}, \tag{4}$$

where, $\hat{\tau}^2_{(2)}$ represents the Level 2 variance, $\hat{\tau}^2_{(3)}$ represents the Level 3 variance, and $\tilde{\upsilon}$ represents the within-study sampling variance (Cheung, 2015a). We supplemented these indices at the respective levels.

Moderator Analyses

To address the second research question, we examined possible moderator effects on the relation between EFs and math intelligence by entering the study, sample, and measurement characteristics into the CHE model. To facilitate interpretation, we mean-centered some of the continuous moderators, z-transformed the children's average age, arcsine-transformed proportional measures (e.g., gender composition coded as the percentage of girls in the sample; see Schwarzer et al., 2019), and dummy-coded categorical and ordinal moderators with more than two categories. The respective mixed-effects metaregression models with the constant sampling correlation ρ and moderator variables $\mathbf{x}_{ij}$ represented a direct extension of the CHE model (Pustejovsky & Tipton, 2021): $\hat{\theta}_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta} + \zeta_{(3)j} + \zeta_{(2)ij} + \varepsilon_{ij}$, with $\varepsilon_{ij} \sim N(0, s_j^2)$, $\mathrm{Cov}(\varepsilon_{hj}, \varepsilon_{ij}) = \rho s_j^2$ for $h \neq i$, $\zeta_{(2)ij} \sim N(0, \tau_{(2)}^2)$, and $\zeta_{(3)j} \sim N(0, \tau_{(3)}^2)$. For moderators with many categories, we extended the CHE model by an additional level of analysis (CHE + model; Pustejovsky & Tipton, 2021). For instance, to assess the moderator effects of the EF task types, we specified a cross-classified random-effects model with the level of task type next to the effect size and study levels (Figure 2). This model allowed us to explicitly quantify the variance between EF task types.

Meta-Analytic Structural Equation Modeling

To examine the joint effects of EFs on math intelligence and test for possible differential effects (RQ3), we performed correlation-based MASEM. To distinguish clearly between the three EF subdimensions (i.e., inhibiting, shifting, and updating), we excluded correlation matrices that contained only composite EF measures. Specifically, we tested four structural equation models (see Figure 3) and evaluated their goodness-of-fit to the meta-analytic data at the study level. Model 1 represented a regression model in which math intelligence was predicted by the three EFs. This model quantified the overall variance explanation, allowing for differential effects of the EF subdimensions. Model 2 further constrained the regression coefficients in Model 1 to equality across the three executive functions. This model assumed equal effects on math intelligence (Marsh, Dowson, et al., 2004). Model 3 represented the three EF subdimensions with a latent variable (i.e., a general factor underlying the EFs; Miyake & Friedman, 2012) that predicted math intelligence. This model captured the covariation among the three EFs in a latent variable and quantified the explanation of joint variance (e.g., Scherer et al., 2018). As in Model 3, Model 4 assumed a latent variable of EFs but used the residuals of the EF subdimensions as predictors of math intelligence. Model 3 focused on a common latent EF variable as the primary source of variance in math intelligence ("common factor model"); Model 4 focused on the residuals of EFs as the sources of variation ("residual factors model"). In other words, these two models targeted common or unique effects on math intelligence (see also Nguyen et al., 2019). Notably, Model 4 estimated the unique effects after we controlled for the common latent EF variable but not after we controlled for the effect of the common latent EF variable. In fact, adding a direct path between the latent EF variable and math intelligence resulted in a model that was not identified ($df_M = -1$). Following Nguyen et al.'s (2019) procedure, we specified a hybrid model that estimated the direct effect of the latent EF variable *and* the direct effect of one EF residual at a time on math intelligence. This model allowed us to include the latent variable and the residual effect and resulted in an identified model. We refer to these models (one for each EF residual) as *hybrid models*.[1] Supplemental Materials S1 and S5 represent the corresponding data set and the analytic code for the hybrid models, respectively.

Correlation-based MASEM allows researchers to synthesize not only single correlations but also entire correlation matrices across primary studies to test multivariate hypotheses (Cheung, 2015a). Several procedures have been developed to perform correlation-based MASEM, and they benefit from these advantages (Sheng et al., 2016). In a review of these procedures, Scherer and Teo (2020) highlighted that they comprise two key elements: pooling of correlation matrices and structural equation modeling based on the pooled correlation matrix. Two-stage MASEM pools the correlation matrices utilizing maximum likelihood estimation and allows researchers to quantify the between-study heterogeneity of the correlation coefficients (i.e., random-effects modeling) in the first stage (Cheung & Cheung, 2016). In the second stage, the structural equation model is fitted to the pooled correlation matrix utilizing weighted least squares estimation (Cheung, 2015a) but without quantifying the heterogeneity of the model parameters. One-stage MASEM performs these two stages in one step and utilizes maximum likelihood estimation (Jak & Cheung, 2020). In the present meta-analysis, we performed two-stage MASEM to address the third research question and compared the resultant model parameters with those obtained by performing one-stage MASEM as a sensitivity test. At the time of writing, these two MASEM approaches had not been extended to accommodate hierarchical or correlation dependencies among correlation matrices (Jak & Cheung, 2020). Therefore, we estimated Models 1–4 based on the study-level correlation matrices, combining multiple correlations for the same pairs of variables within studies via Fisher's z transformation (Borenstein et al., 2011). Thus, each primary study contributed only one correlation matrix. This data set and the corresponding analytic code can be found in Supplemental Materials S1 and S4, respectively.

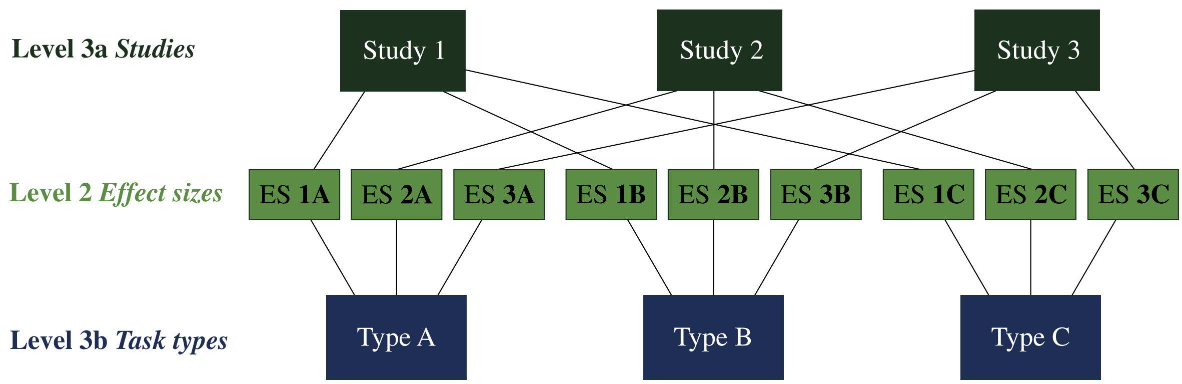
To evaluate the fit of the four models, we relied on common guidelines for the comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR; acceptable fit was defined as CFI $\geq$ .95, RMSEA $\leq$ .08, SRMR $\leq$ .10; e.g., Hu & Bentler, 1999; Marsh et al., 2005). However, these guidelines have been validated only on a limited set of specific structural equation models; therefore, we did not apply them as "golden rules" (Marsh, Hau, et al., 2004). Model comparisons were performed via chi-square difference testing and differences in information criteria (Kline, 2015). We performed correlation-based MASEM in the R package metaSEM (Cheung, 2015b).

Sensitivity Analyses

To evaluate the sensitivity of the results, we examined the effects of (a) including versus excluding the study with the largest sample (Nguyen & Duncan, 2019; more than 17,000 participants); (b) different choices of the constant sampling correlation ρ between 0.20 and 0.60, as previously noted (Fisher et al., 2017); (c) one-stage versus two-stage MASEM (Jak & Cheung, 2020); and (d) unattenuated versus attenuated correlations with the reliabilities of EF (α_{EF}) and math intelligence (α_{Math}) measures (Hunter & Schmidt, 2004). For (d), we considered only Cronbach's α values as reliability coefficients. If Cronbach's α was not provided in a primary study, we used the average reliabilities of all available EF

[1] We also examined alternative modeling approaches to identify Model 4 with an additional direct path, such as constraining factor loadings (and ultimately residual variances) to equality or fixing the parameters of the EF measurement model to values that had been estimated from MASEM of a confirmatory factor analysis model with only the EF subdimensions (the two-step estimation approach). However, none of these approaches resulted in robust and trustworthy standard errors and, thus, they were not considered for further analyses.

Figure 2
Exemplary Structure of a Cross-Classified Model



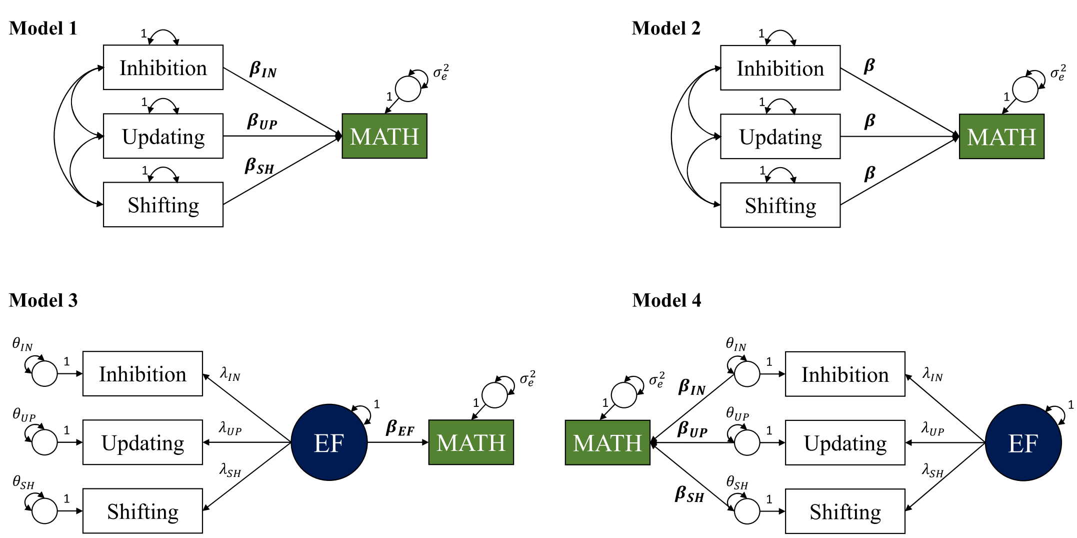
Note. See the online article for the color version of this figure.

or math intelligence measures in the data set. These average reliabilities were obtained with a random-effects meta-analysis of all available reliabilities for the respective constructs (see Supplemental Material S3).

Evaluation of Primary Study Quality

Evaluating the quality of the primary-level studies is an important step toward crafting an argument for the evidential value of the meta-analytic data (Johnson, 2021). In this meta-analysis, we extracted the following quality indicators from the primary studies: Publication status (0 = *gray*, 1 = *published*), overall sample size N, the reporting of the reliability coefficients for the study sample (0 = *not reported*, 1 = *reported*), the reliability coefficients of the executive functions and math intelligence assessments (Cronbach's α), and the reporting of p values associated with the EF–math intelligence correlations (0 = *not reported*, 1 = *reported*). To synthesize these indicators, we created a study quality index as an emergent composite variable via structural equation modeling. This model contained the EF–math intelligence correlation as an outcome variable that is predicted by the quality index (see Supplemental Material S7). To identify the model, the residual variance of the composite variable was fixed to zero, and the regression coefficient connecting study quality and the correlations was fixed to 1 (Henseler, 2021). Creating a composite quality index via SEM has several advantages: First, it allows meta-analysts to combine a diverse set of quality indicators, such as binary, ordinal, count, and continuous variables, into a single index. Second, it handles missing data in the quality indicators efficiently via multiple imputation or model-based procedures. Third, the quality indicators receive individual rather than

Figure 3
Meta-Analytic Structural Equation Models (Models 1–4) Representing the Relations Between Executive Functions and Math Intelligence



Note. MATH = math intelligence; IN = inhibition; UP = updating; SH = shifting; EF = latent variable representing general executive functions. See the online article for the color version of this figure.

equal weights. Fourth, it accounts for the hierarchical structure of the meta-analytic data (i.e., multiple samples or effect sizes per study). After estimating the SEM parameters in the R package lavaan (Rosseel, 2012), we extracted the resultant composite index and explored the extent to which it explained heterogeneity in the EF–math intelligence correlations.

Transparency and Openness

This review's protocol was not preregistered. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidelines for our systematic review and meta-analysis. We share all supplemental material, analysis syntaxes, data, the codebook, and the documentation of the systematic search at https://osf.io/vn5pe/?view_only=d4a5f0670ca5469599d13828918957ba (Emslander & Scherer, 2022).

Results

Description of the Included Studies

The study, sample, and measurement characteristics of all (k_S = 47) studies included in the present meta-analysis are described in Table S6.3, and the frequencies and missing values of the categorical moderators are shown in Table S6.4 in the Supplemental Material. Thirty-eight of the 47 studies reported multiple effect sizes, ranging from 1 to 36 per study, with an average of 8. Overall, the present meta-analysis included 65 samples and k_{ES} = 363 effect sizes. These effect sizes stem from 44 published journal articles and three gray literature items: one dissertation (Ahmed, 2019), one preprint (Costa et al., 2021), and one conference proceeding (R. Duncan et al., 2016; see Table S6.3). Figure S6.1 in the Supplemental Material displays the number of effect sizes and studies by the country of sample origin. Most studies (66%) drew on samples from the United States (representing North America; none of the studies were based on Canadian samples), whereas 17% of the studies used samples from Europe, 13% drew from Asia, and 4% drew from Australia. All studies were conducted in an educational context. Most EF measures (98%) and math intelligence measures (99%) were administered in an individual setting, although one study measured EFs in a group setting (Ahmed, 2019), and another measured both constructs in group settings (Träff et al., 2020; see Table S6.4).

Figure S6.2 in the Supplemental Material presents the number of effect sizes by EF subdimension, math intelligence subdimension, EF task type, EF test mode, and math intelligence test mode. The largest number of effect sizes (38%) indicated the relation between inhibition and math intelligence (k_{ES} = 137), followed by the relation between updating and math intelligence (k_{ES} = 107) and the relation between shifting and math intelligence (k_{ES} = 96). Thus, the literature slightly emphasizes the correlation between inhibition and math intelligence. Regarding math intelligence, most effect sizes were associated with the correlations of EFs and basic number knowledge (k_{ES} = 157), calculation and reasoning (k_{ES} = 105), or a math intelligence composite (k_{ES} = 93). The tasks that were most frequently used to measure EFs were Stroop-like tasks (k_{ES} = 79) to measure inhibition, dimensional change tasks (k_{ES} = 83) to measure shifting, and difficult span tasks (k_{ES} = 50) to measure updating. These EF measures were mostly administered verbally (k_{ES} = 120) followed by apparatus-based (k_{ES} = 93), computer-based (k_{ES} = 71), and behaviorally (k_{ES} = 62). Math measures were administered verbally even more often than EF measures (k_{ES} = 273; 77% of effect sizes).

Table 3 shows the descriptive statistics of the continuous moderators. The total sample of the present meta-analysis included 30,481 preschool children, ranging in age from 3.0 to 6.6 years. Half of the children were girls (*M* = 50.4%, *SD* = 4.2%). Although 46 studies drew on samples of children who had not yet entered first grade, 15% of Mills et al.'s (2019) sample were first graders. All studies were published between 2007 and 2021 (see Table 3).

Preliminary Analyses

Identifying a Baseline Model

First, we specified and estimated the baseline model (i.e., the CHE model) with the respective constant sampling correlations. To support this model, we identified alternative meta-analytic models with different variance components. Table 4 displays the fit statistics for the complete data set. Based on these results, we found that the three-level random-effects model with a constant sampling correlation best represented the data (CHE in Table 4). Consequently, this model served as the baseline model for additional moderator analyses and for reporting an overall effect size. The CHE model also served as a baseline model for the inhibition, updating, and shifting data sets. Given the small sample sizes of the pooled and composite data sets, we chose the random-effects model with robust variance estimation and a constant sampling correlation of $\rho = 0.30$.

Detecting Influential Effect Sizes

Outlier analysis identified two of the 363 effect sizes (Ahmed et al., 2019; Mills et al., 2019) as influential. The confidence interval

Table 3
Descriptive Results of Continuous Moderators

Continuous moderator	k_S	k_{ES}	*M*	*SD*	*Mdn*	Min	Max
Publication year	47	363	2016.21	1.13	2017	2007	2021
Gender composition	47	363	50.36	4.23	50	37.59	63.90
Age (in years)	46	355	4.93	0.84	5	3	6.58
Reliability of math intelligence measures	22	169	0.85	0.10	0.88	0.46	0.97
Number of math intelligence items	42	268	32.46	22.51	33	3	120
Reliability of EF measures	24	63	0.87	0.08	0.87	0.56	0.95
Number of EF items	39	237	35.52	40.30	20	3	240

Note. k_S = number of included studies; k_{ES} = number of effect sizes; gender composition = percentage of girls in the sample; EF = executive function.

Table 4
Model Fit Indices for the Baseline and Cross-Classified Models

Variable	*df*	LogLik	AIC	BIC	AICc
CHE model, $\tau^2_{(2)}$ and $\tau^2_{(3)}$ freely estimated, hierarchical nesting, $r = \rho = 0.3$	3	210.51	−415.01	−403.34	−414.94
HE model, $\tau^2_{(2)}$ and $\tau^2_{(3)}$ freely estimated, hierarchical nesting	3	207.78	−409.57	−397.89	−409.50
RE model, $\tau^2_{(2)} = 0$ and $\tau^2_{(3)}$ freely estimated	2	−625.45	1254.91	1262.69	1254.94
RE model, $\tau^2_{(2)}$ freely estimated and $\tau^2_{(3)} = 0$	2	162.39	−320.79	−313.00	−320.75
FE model, $\tau^2_{(2)} = 0$ and $\tau^2_{(3)} = 0$	1	−1364.99	2731.98	2735.88	2732.00
CHE+ model, $\tau^2_{(2)}$ and $\tau^2_{(3)}$ freely estimated, hierarchical nesting, $r = \rho = 0.3$	4	257.67	−507.35	−491.78	−507.23
HE+ model, $\tau^2_{(2)}$ and $\tau^2_{(3)}$ freely estimated, hierarchical nesting	4	251.49	−494.98	−479.41	−494.87

Note. LogLik = log-likelihood; AIC = Akaike information criterion; BIC = Bayesian information criterion; AICc = corrected Akaike information criterion; CHE = correlated and hierarchical effects; CHE model = three-level random-effects model with hierarchical nesting and constant sampling correlation; HE model = three-level random-effects model with hierarchical nesting; RE model = common random-effects model; FE model = fixed-effects model; CHE+ model = cross-classified four-level random-effects model with hierarchical nesting, constant sampling correlation, and executive function task type as an extra level; HE+ model = cross-classified four-level random-effects model with hierarchical nesting and executive function task type as an extra level.

of these effect sizes exceeded the average correlation confidence interval and displayed considerable difference in fits values (DFFITS > 0.4 *SD*s), Cook's (1977) distance values (>0.15), and covariance ratio (<1), suggesting that these effect sizes should be removed for greater precision (Viechtbauer & Cheung, 2010). However, removing these two influential cases yielded a similar average correlation as the meta-analysis that included all effect sizes ($\bar{r}$ = .347 vs. $\bar{r}$ = .341), showed almost equal population variances, $\tau_{(2)}$ = .101 and $\tau_{(3)}$ = .084 versus $\tau_{(2)}$ = .106 and $\tau_{(3)}$ = .085, and displayed only marginally different confidence intervals (95% CI [.32, .38] vs. [.31, .37]). As a result, we assumed that the meta-analysis was robust against influential cases. Furthermore, both effect sizes stemmed from studies drawing on samples larger than the median sample size of all 47 studies, exhibited good psychometric quality, and reported core study characteristics. Therefore, we refrained from excluding the two effect sizes.

Analyses of Publication Bias

The contour-enhanced funnel plots for the five data sets are shown in Figure 4. For the complete data, the precision estimate test indicated the presence of publication bias, because the sampling errors moderated the overall effect, B = –13.6, SE = 0.6, p < .001. Similarly, the PEESE test resulted in a statistically significant moderator effect of the sampling variances, B = –69.5, SE = 3.7, p < .001. The funnel plot test did not suggest any dependency between the effect sizes and the sample sizes, B = 0.0, SE = 0.0, p = .58. The two trim-and-fill estimates were zero; thus, no extra effect sizes were needed to counterbalance potential asymmetry in the funnel plot. However, these trim-and-fill results should be interpreted with caution, as the added value of this method is controversial (Duval & Tweedie, 2000; Peters et al., 2007). The funnel plots for the complete, inhibition, and shifting data sets yielded nonsignificant Kendall's τ values ranging from −.04 to −.10, providing no evidence of publication bias. The Kendall's τ value in the updated data set, −.15, was statistically significant and suggested some evidence of publication bias (see Figure 4). We did not adjust for publication bias in the subsequent models.

Main Analysis: Overall Correlations (RQ1)

Table 5 reports the results of the meta-analyses for the five data sets: the complete, pooled, inhibition, shifting, and updating data sets. To address RQ1, we calculated a three-level random-effects meta-analysis with the complete data set, which yielded a moderately positive average correlation ($\bar{r}$ = .34) between EFs and math intelligence in preschool children. Similarly, the average correlation was $\bar{r}$ = .40 for the pooled data set (see Table 5) and $\bar{r}$ = .47 for all composite EF measures, testing at least two EF subdimensions. Figure S6.3 in the Supplemental Material displays a forest plot for the pooled data set with effect sizes and their confidence intervals (see Supplemental Material S3).

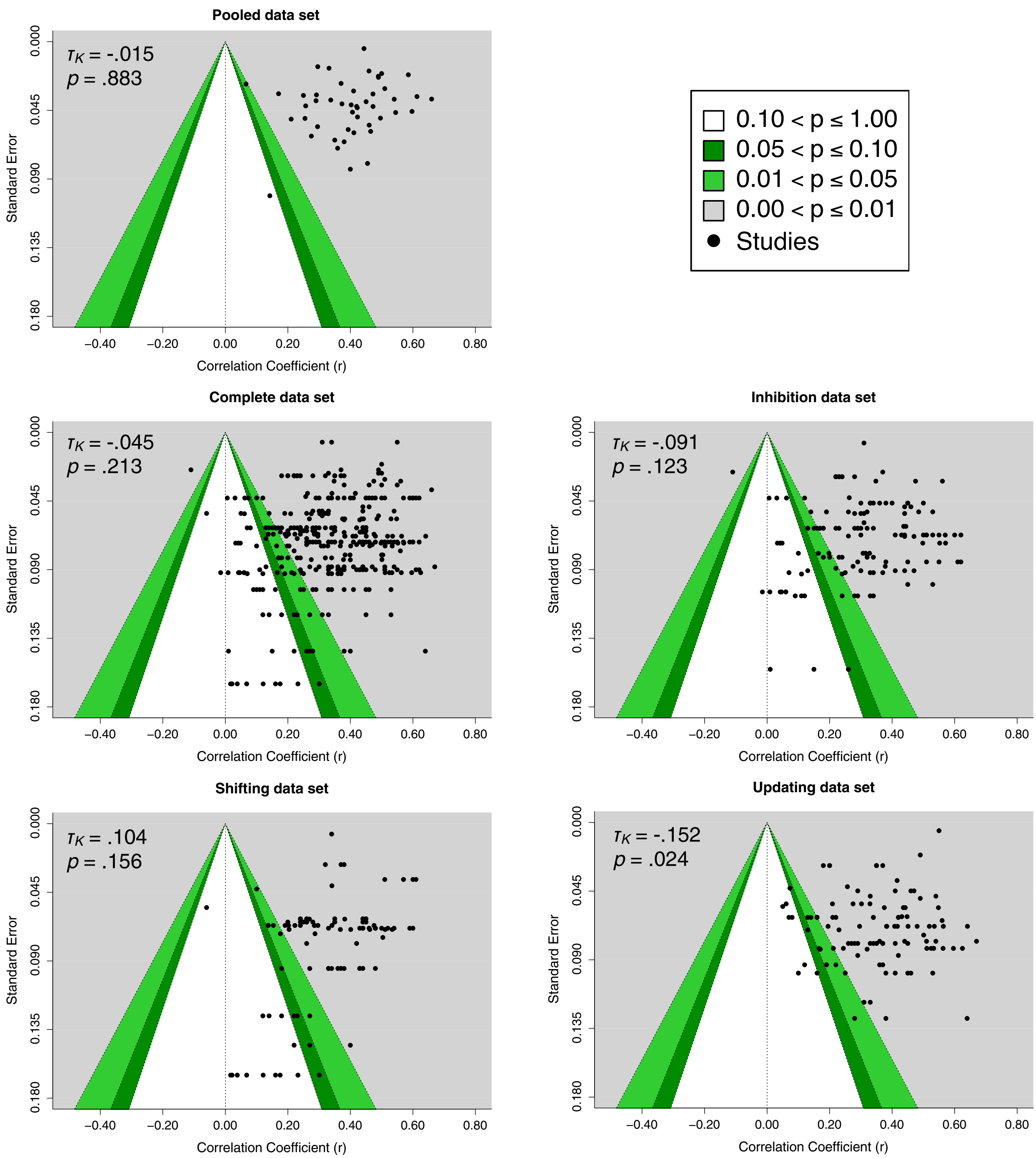
Investigating the three EF subdimensions separately, we found a substantial average correlation between inhibition and math intelligence ($\bar{r}$ = .30), shifting and math intelligence ($\bar{r}$ = .32), and updating and math intelligence ($\bar{r}$ = .36). However, given the overlapping confidence intervals, there was no evidence of statistically significant differences between the three average correlations (see Table 5). The complete and pooled data sets and the corresponding analytic codes are presented in Supplemental Materials S1 and S3, respectively.

Heterogeneity and Moderator Analyses (RQ2)

Heterogeneity Indices

The effect sizes of the five data sets were highly heterogeneous (p_Q < .001; see Table 5, for the exact Q values). For the three-level random-effects meta-analysis with constant sampling correlation of the complete data set, the heterogeneity at the effect size level and the study level was 55% and 35% of the total variance, respectively, which was not due to sampling error. The univariate meta-analysis

Figure 4
Contour-Enhanced Funnel Plots of All Data Sets



Note. Larger positive effect sizes indicate a closer relationship between EFs and math intelligence. Correlation coefficients on the *x*-axis are plotted against the standard errors on the *y*-axis for every effect size. τ_K = Kendall's τ value; EF = executive function. See the online article for the color version of this figure.

Table 5
Results of the Meta-Analysis of the Relation of EFs and Their Subdimensions With Math Intelligence

Relation	Data set	$\bar{r}$	95% CI	*SE*	k_S	k_{ES}	*t*	$\tau_{(2)}$	$\tau_{(3)}$	*Q*	$I^2_{(2)}$	$I^2_{(3)}$
EF and math intelligence	Complete	.34*	[.31, .37]	.02	47	363	21.28	.106	.085	3690.13*	.55	.35
EF and math intelligence	Pooled	.40*	[.36, .44]	.02	47	47	21.70	—	.117	658.69*	—	.93
Inhibition and math intelligence	Inhibition	.30*	[.25, .35]	.02	30	137	12.35	.089	.110	955.49*	.36	.54
Shifting and math intelligence	Shifting	.32*	[.25, .38]	.03	20	96	9.94	.063	.124	362.17*	.18	.69
Updating and math intelligence	Updating	.36*	[.31, .40]	.02	27	107	14.89	.105	.088	1032.50*	.52	.37

Note. Larger positive effect sizes indicate a closer relationship between EFs and math intelligence. EF = executive function; $\bar{r}$ = weighted-average correlation, pooled from all effect sizes within the respective study; CI = confidence interval; *SE* = standard error; k_S = number of included studies; k_{ES} = number of effect sizes; $\tau_{(2;3)}$ = heterogeneity at effect size (2) and study level (3), respectively; *Q* = sum of squared deviations of each effect size's estimate from overall meta-analytic estimate; $I^2_{(2;3)}$ = heterogeneity indices at effect size (2) and study level (3), respectively.
* $p < .001$.

of the pooled data set showed high total heterogeneity (90%), as did the data set with only composite EF measures (81%). The effect sizes of the inhibition, shifting, and updating data sets also varied substantially within and between studies (see Table 5). The large heterogeneity throughout the data sets motivated subsequent moderator analyses, in accordance with RQ2.

Continuous Moderators

To explain the heterogeneity (RQ2), we performed a moderator analysis of the complete data set, as well as separate analyses for the inhibition, shifting, and updating data sets. In the complete data set, we examined a total of seven continuous and 14 categorical moderators. Table 6 reports the results of the continuous moderators in the complete data set, and Table S6.5 reports the results for inhibition, shifting, and updating data sets separately. Publication year, gender composition, age, number of math items, EF reliability, and number of EF items did not moderate the relation between EFs and math intelligence. The only continuous characteristic that marginally statistically significantly moderated this relation was the reliability of the math intelligence measures ($B = 0.30$, $SE = 0.15$, 95% CI [−0.01, 0.60], $p = .055$). The higher the reliability of the math intelligence test, the higher the correlation with EFs tended to be (see Table 6). This moderating effect was exclusive to the effect size level, $R^2_{(2)} = 3\%$; $R^2_{(3)} = 0\%$.

Categorical Moderators

Table 7 displays the results for categorical moderators in the complete data set, and Table S6.6 shows the results for the inhibition, shifting, and updating data sets separately. To ensure meaningful interpretation of the results, moderators with fewer than four effect sizes per category were excluded from the testing (see Bakermans-Kranenburg et al., 2003). This criterion was applied to publication type, study background, and math intelligence performance test versus third-party rating. To facilitate interpretability, we reduced the categories of the moderators' mode of EF testing (to verbal, behavioral, apparatus-based, and computer-based) and socioeconomic status (to low, medium-low, medium-high, and high).

Most categorical characteristics (publication status, preschool status, socioeconomic status, study design, math intelligence subdimension, mode of math intelligence testing, group vs. individual math intelligence test, EF performance test vs. third-party rating, and group vs. individual EF test) did not show a statistically significant moderating effect on the relation between EFs and math intelligence (see Table 7). However, several categorical characteristics—country, continent, EF subdimension, mode of EF testing, and EF task type—statistically significantly moderated this relation for effect size and study level. The country where the study was conducted was identified as a statistically significant moderator at the study level, $R^2_{(2)} = 0\%$; $R^2_{(3)} = 30\%$. We found nonsignificant correlations in the two studies that drew on samples from Turkey (Söğüt et al., 2021) and Pakistan (Armstrong-Carter et al., 2020). All other countries yielded significant correlations, which were especially high in studies with samples from Japan (Fujisawa et al., 2019) and the United Kingdom (Blakey & Carroll, 2015; Costa et al., 2021). Closely related, the continent on which the study was conducted significantly moderated the relation between EFs and math intelligence almost exclusively at the study level, $R^2_{(2)} = 0\%$; $R^2_{(3)} = 7\%$. Specifically, studies conducted in the United States on

Table 6
Results of the Continuous Moderators

Continuous moderator	k_S	k_{ES}	*B*	*SE*	95% CI	*p*
Publication year	47	363	0.00	0.01	[−0.03, 0.02]	.744
Gender composition	47	363	0.05	0.31	[−0.55, 0.65]	.882
Age (in years)	46	355	0.00	0.01	[−0.02, 0.03]	.758
Reliability of math intelligence measures	23	169	0.30	0.15	[−0.01, 0.60]	.055
Number of math intelligence items	42	268	0	0	[−0.00, 0.00]	.373
Reliability of EF measures	13	63	0.11	0.30	[−0.50, 0.71]	.730
Number of EF items	39	237	0	0	[−0.00, 0.00]	.840

Note. A significant *p* value indicates that there is a statistical difference between the levels of the moderator. k_S = number of included studies; k_{ES} = number of effect sizes; gender composition = percentage of girls in the sample; *SE* = standard error; CI = confidence interval; EF = executive function.

Table 7
Results of the Categorical Moderators

Moderator	k_S	k_{ES}	$\bar{r}$ [95% CI]	*SE*	$R^2_{(2)}$	$R^2_{(3)}$	*p*
Baseline ($I^2_{(2;3)}$: .55; .35)	47	363	.34 [.31, .37]	0.02	—	—	—
Country	47	363	—	—	.00	.30	.021*
United States	31	247	.36 [.33, .40]	0.02	—	—	—
Australia	2	20	.32 [.15, .49]	0.09	—	—	—
Spain	2	21	.28 [.11, .45]	0.09	—	—	—
Netherlands	2	11	.34 [.13, .54]	0.10	—	—	—
United Kingdom	2	22	.37 [.20, .55]	0.09	—	—	—
Sweden	2	9	.25 [.07, .42]	0.09	—	—	—
Japan	1	4	.47 [.25, .69]	0.11	—	—	—
China	3	14	.30 [.14, .46]	0.08	—	—	—
Turkey	1	9	.13 [−.16, .42]	0.15	—	—	—
Pakistan	1	6	.07 [−.14, .28]	0.11	—	—	—
Continent	47	363	—	—	.00	.07	.041*
North America	31	247	.36 [.33, .40]	0.02	—	—	—
Asia, Australia, and Europe	16	116	.30 [.19, .40]	0.05	—	—	—
Publication status	47	363	—	—	.00	.09	.09
Preschool status	47	363	—	—	.00	.02	.31
Socioeconomic status	38	304	—	—	.00	.00	.77
Study design	47	363	—	—	.00	.00	.58
Math intelligence subdimension	47	355	—	—	.00	.08	.06
Mode of math intelligence testing	46	353	—	—	.00	.00	.48
Math performance testing	46	360	—	—	.00	.00	.46
Math group testing	46	360	—	—	.00	.00	.31
EF subdimension	47	363	—	—	.06	.24	<.001**
Inhibition	32	137	.30 [.26, .33]	0.02	—	—	—
Shifting	19	96	.32 [.25, .40]	0.04	—	—	—
Updating	27	107	.35 [.28, .41]	0.03	—	—	—
Composite	9	23	.45 [.35, .55]	0.05	—	—	—
EF task type	46	356	—	—	.24	.17	<.001**
Composite	6	11	.47 [.38, .56]	0.05	—	—	—
Stroop (inhibition)	19	79	.24 [.05, .43]	0.10	—	—	—
Simon (inhibition)	10	21	.40 [.20, .60]	0.10	—	—	—
Slow (inhibition)	2	5	.26 [.03, .49]	0.12	—	—	—
Shape school (inhibition)	1	4	.38 [.14, .62]	0.12	—	—	—
Delay (inhibition)	5	9	.30 [.09, .51]	0.11	—	—	—
Flanker (inhibition)	4	12	.31 [.10, .52]	0.11	—	—	—
Tap (inhibition)	6	14	.44 [.23, .66]	0.11	—	—	—
Dimensional change (shifting)	17	83	.31 [.12, .50]	0.10	—	—	—
Flexible selection (shifting)	4	14	.37 [.16, .58]	0.11	—	—	—
Random generation (updating)	2	8	.45 [.23, .67]	0.11	—	—	—
Easy span (updating)	9	18	.30 [.10, .50]	0.10	—	—	—
Difficult span (updating)	15	50	.39 [.20, .59]	0.10	—	—	—
Last word (updating)	1	12	.14 [−.08, .35]	0.11	—	—	—
Search (updating)	2	16	.26 [.04, .47]	0.11	—	—	—
Mode of EF testing	45	346	—	—	.03	.00	.03*
Verbal	23	120	.31 [.27, .35]	.02	—	—	—
Behavioral	22	62	.37 [.29, .46]	.04	—	—	—
Apparatus-based	20	93	.35 [.26, .43]	.04	—	—	—
Computer-based	13	71	.36 [.27, .46]	.05	—	—	—
EF performance testing	46	349	—	—	.00	.00	.79
EF group testing	45	347	—	—	.00	.00	.63

Note. Larger positive effect sizes indicate a closer relationship between EFs and math intelligence. A significant *p* value indicates that there is a statistical difference between the levels of the moderator. $I^2_{(2;3)}$ = heterogeneity indices at Levels 2 and 3, respectively; k_S = number of included studies; k_{ES} = number of effect sizes; $\bar{r}$ = weighted-average correlation; CI = confidence interval; *SE* = standard error; $R^2_{(2;3)}$ = variance explanation at Levels 2 and 3, respectively; EF = executive function.
* $p < .05$. ** $p < .001$.

the North American continent yielded higher correlations than studies conducted in countries on the Asian, Australian, and European continents (see Table 7). The EF subdimension was a significant moderator as well, explaining most variance at the study level, $R^2_{(2)} = 6\%$; $R^2_{(3)} = 24\%$. Inhibition, shifting, and updating correlated significantly with math intelligence. However, the EF composite showed the greatest relation. Examination of the confidence intervals revealed that the EF composite had a significantly stronger correlation with math intelligence than inhibition (see Table 7). The mode of EF testing significantly moderated the relation between EFs and math intelligence, with descriptively higher correlations for behavioral, apparatus-based, and computer-based assessments than for verbal assessments. This moderator explained only a small amount of variance between effect sizes and none between studies, $R^2_{(2)} = 3\%$; $R^2_{(3)} = 0\%$. The EF task type also showed a statistically significant influence on the relation between EFs and math intelligence and had a moderating effect at the effect size and study levels, $R^2_{(2)} = 24\%$; $R^2_{(3)} = 17\%$. All EF task types exhibited statistically significant correlations with math intelligence. The Simon and tap tasks revealed the largest correlations for inhibition, the flexible selection tasks for shifting, and the random generation and difficult span tasks for updating (see Table 7). For the data set with composite EF measures, there was a statistically significant difference in the effects between prekindergarten and kindergarten with higher correlations for pre-K children ($B = -0.07$, $SE = 0.02$, $p < .05$). For separate results of the categorical moderators of inhibition, shifting, and updating, please see Table S6.6. Overall, mostly the location of the sample origin and the characteristics of the EF measurement moderated the relation between EFs and math intelligence.

To further test the significant effects of the EF task types, we specified this moderator as another level of analysis in addition to the study level. The resultant cross-classified model, in which the EF task types represented a fourth analytic level, was favored over the three-level model with constant sampling correlation, $\Delta\chi^2(1) = 94.3$, $p < .001$, and resulted in a between-task variance of $\tau^2_{(4)} = .01$, 95% CI [.01, .02]. These results further indicated that the EF–math intelligence correlations varied across EF task types (see Table 4 and Supplemental Material S3).

Table 8

Pooled Correlation Matrix Under the Stage-1 Random-Effects Model ($k_S = 23$, $N = 5{,}402$)

Constructs	1	2	3	4
1. Math intelligence				
$\bar{r}$	1.000			
2. Inhibition				
$\bar{r}$	0.312	1.000		
95% CI	[0.245, 0.379]			
τ^2	0.020			
I^2	85.0%			
3. Shifting				
$\bar{r}$	0.381	0.294	1.000	
95% CI	[0.311, 0.451]	[0.217, 0.371]		
τ^2	0.007	0.007		
I^2	63.1%	60.2%		
4. Updating				
$\bar{r}$	0.379	0.279	0.245	1.000
95% CI	[0.323, 0.435]	[0.190, 0.367]	[0.178, 0.312]	
τ^2	0.009	0.021	0.001	
I^2	71.7%	84.4%	16.5%	

Note. Larger positive effect sizes indicate a closer relationship between executive functions (EFs) and math intelligence. The pooled correlation coefficients ($\bar{r}$) were based on the Stage 1 random-effects model of the two-stage meta-analytic structural equation modeling (MASEM) procedure. k_S = number of included studies; N = number of included students; 95% CI = Wald 95% confidence intervals; τ^2 = between-study variation of the correlation coefficients; I^2 = heterogeneity index.

MASEM (RQ3)

To test the hypothesized models that shed light on the extent to which the three EFs explain variation in math intelligence jointly and uniquely (RQ3), we performed a two-stage MASEM utilizing study-level data. The analytic code for the MASEM approach and the hybrid models can be found in Supplemental Materials S4 and S5, respectively.

Stage 1: Pooling Correlation Matrices

Excluding the correlations between the composite scores of executive functions and math intelligence, we were able to include 38 studies and 120 correlation coefficients based on the data of 26,281 children in these analyses. Before pooling the correlation matrices across the primary studies, we tested them for positive definiteness, a key prerequisite for structural equation modeling (Cheung, 2015a; Kline, 2015). All matrices were positive definite and could be submitted to the pooling stage.

Next, we pooled the correlation matrices under a random-effects model via the tssem1()-function in the R package metaSEM. This model resulted in statistically significant between-study heterogeneity of the correlation matrices, $Q_E(114) = 1039.5$, $p < .001$. The corresponding pooled correlation matrix, between-study variances, and heterogeneity indices are shown in Table 8.

Stage 2: Structural Equation Modeling

Model 1 (Regression Model). Fitting Model 1 to the pooled correlation matrix via the tssem2()-function in metaSEM resulted in positive and statistically significant regression coefficients for inhibition ($\beta_{Inhibition} = 0.19$, 95% CI [0.14, 0.24]), shifting ($\beta_{Shifting} = 0.19$, 95% CI [0.12, 0.26]), and updating ($\beta_{Updating} = 0.26$, 95% CI [0.20, 0.33]). With a residual variance of $\sigma^2_e = 0.79$ (95% CI [0.75, 0.83]), about 21% of the variance in math intelligence can be explained. Dominance analyses indicated that updating contributed to explaining the overall variance with 10%, while inhibition and shifting contributed less (about 6% each). Although this result may indicate that inhibition and shifting have the lowest explanatory power, the differences in the average contribution among the three EFs were not substantial (see Supplemental Material S4). Model 1 was just identified and exhibited a perfect fit to the data.

Model 2 (Regression Model With Equal Regression Coefficients). Model 2 assumed equal regression coefficients for the EFs and exhibited a good fit to the data: $\chi^2(2) = 2.9$, $p = .23$, CFI = 0.999, RMSEA = 0.004, SRMR = 0.018. The overall regression coefficient was $\beta = 0.21$ (95% CI [0.19, 0.24]), and the residual variance was $\sigma^2_e = 0.80$ (95% CI [0.76, 0.83]). Overall, about 20% of the math intelligence variance was explained by this model. Given that Model 1 was exactly identified, the chi-square difference test statistic corresponded to the chi-square value of Model 2, $\chi^2(2) = 3.6$, $p = .16$, and indicated that the equality of the regression coefficients could be assumed. Thus, for the sample of primary studies, there was no evidence of differential effects of the three EFs on math intelligence (RQ3).

Model 3 (Structural Equation Model With a Latent EF Variable). Fitting Model 3 to the pooled correlation matrix, we obtained fit indices that indicated a close-to-perfect fit to the data: $\chi^2(2) = 0.6$, $p = .75$, CFI = 1.000, RMSEA = 0.000, SRMR = 0.008. Factor loadings were positive and statistically significant for inhibition ($\lambda_{Inhibition}$ = 0.47, 95% CI [0.41, 0.53]), shifting ($\lambda_{Shifting}$ = 0.47, 95% CI [0.42, 0.53]), and updating ($\lambda_{Updating}$ = 0.54, 95% CI [0.48, 0.60]). The overall regression coefficient was β = 0.65 (95% CI [0.58, 0.72]), and the residual variance was $\sigma_e^2 = 0.58$ (95% CI [0.48, 0.66]). Overall, about 42% of the math intelligence variance was explained in this model. Therefore, representing EFs with a latent variable and thus capturing the covariance among the three EFs explained substantially more variance in math intelligence than representing them as distinct but correlated variables (RQ3).

Model 4 (Structural Equation Model With Unique EF Effects). We further specified and estimated Model 4, a model proposed by Nguyen et al. (2019), which describes the unique effects of EFs on math intelligence after a common trait shared among the three subdimensions is controlled for. This model was just identified and exhibited a perfect fit to the meta-analytic data. As in Model 3, the factor loadings were positive and statistically significant for inhibition ($\lambda_{Inhibition}$ = 0.47, 95% CI [039, 0.57]), shifting ($\lambda_{Shifting}$ = 0.49, 95% CI [0.41, 0.57]), and updating ($\lambda_{Updating}$ = 0.51, 95% CI [0.42, 0.61]), and the three EF residuals explained 21% of the variance in math intelligence. Furthermore, the regression coefficients were positive and statistically significant for inhibition ($\beta_{Inhibition}$ = 0.19, 95% CI [0.14, 0.24]), shifting ($\beta_{Shifting}$ = 0.19, 95% CI [0.12, 0.26]), and updating ($\beta_{Updating}$ = 0.26, 95% CI [0.20, 0.33]), and matched those in Model 1. Similar to Model 2, constraining the regression coefficients to equality did not diminish the fit of Model 4 significantly, $\chi^2(2) = 2.9$, $p = .23$. After their shared variation was controlled for, the unique effects did not differ significantly among the three EF subdimensions. Finally, Models 3 and 4 exhibited a somewhat similar fit to the data, $\chi^2(2) = 0.6$, $p = .75$. Although both models may explain the joint relations between EF subdimensions and math intelligence, the unitary Model 3 fit slightly better and was more parsimonious, as it represented the data with a smaller number of parameters (Kline, 2015).[2]

Hybrid Models (Structural Equation Models With a Latent EF Variable and One Unique EF Effect). Finally, we estimated three hybrid models specifying the effects of the latent EF variable and one residual effect on math intelligence (Nguyen et al., 2019). These models exhibited a very good fit to the meta-analytic data (e.g., CFIs = 1.000, RMSEAs = 0.000, SRMRs ≤ 0.008; see Supplemental Material S5). In all models, the latent EF variable was positively and statistically significantly related to math intelligence (βs = 0.60–0.68, $p < .001$); however, each of the three residual effects was statistically insignificant ($\beta_{Inhibition}$ = –0.02, 95% CI [–0.15, 0.08]; $\beta_{Shifting}$ = –0.03, 95% CI [–0.24, 0.10]; $\beta_{Updating}$ = 0.05, 95% CI [–0.06, 0.16]). Thus, in these models, the variation in math intelligence was primarily explained by the latent EF variable but not the EF residuals (between 39% and 45% in total).

Subgroup Analysis

As the final step, we examined the extent to which the four models applied to two key groups of primary studies in the samples: studies with prekindergarten and kindergarten children (k_S = 19, N = 21,401) and studies with preschool children (k_S = 19, N = 4,880). Figure 5 shows the meta-analytic structural equation models with model parameters for these two subgroups. These models were based on two separate Stage 1 random-effects models. Although the variance explanations in Model 1 were similar (20% vs. 21%), the dominance of the EF subdimension updating was more pronounced in the studies that included preschool children. However, given the few studies in this subgroup, the differences in the regression coefficients were not statistically significant (Model 1 vs. Model 2; for the MASEM syntaxes, see Supplemental Material S4). Model 3 revealed a slightly higher variance explanation by the latent EF variable for the prekindergarten and kindergarten samples (45%) than for the preschool samples (39%). Model 4 supported the dominance of shifting in the preschool samples after we controlled for the latent EF variable. This model did not exhibit a fit superior to that of Model 3 for both study subgroups. Overall, the conclusions drawn for the total meta-analytic samples held for the two subgroups of primary studies and revealed a tendency toward more pronounced unique effects of shifting for preschool samples.

Sensitivity Analyses

Exclusion of a Large Primary Study

Table S6.7 shows the results of all sensitivity analyses. Comparing the results of the complete data set with and without the largest study (both rounded to $\bar{r}$ = .34; Nguyen & Duncan, 2019), we found very similar correlations. Additionally, the confidence intervals of the two results were identical (95% CI [.31, .39]). After Nguyen and Duncan's (2019) study was excluded, the heterogeneity between effect sizes remained similar, $I^2_{(2)}$ = 55% versus 53%; $I^2_{(3)}$ = 35% vs. 33%, as did the variance in the effect size and study levels, $\tau^2_{(2)}$ = 0.01 versus 0.01; $\tau^2_{(3)}$ = 0.01 vs. 0.01; see Table S6.7 in the Supplemental Material S6.

Within-Study Correlation Between Effect Sizes

We further examined the extent to which the parameters of the baseline model were sensitive to the choice of the within-study correlation between effect size (ρ), varying its values between 0.2 and 0.6 (see Supplemental Material S3). Across all data sets, the estimates of the weighted-average effect sizes and the Levels 2 and 3 variance components were not substantially affected by the choice of ρ. For instance, the effect sizes of the baseline model for the complete data were 0.3405 (ρ = 0.2), 0.3410 (ρ = 0.3), 0.3415 (ρ = 0.4), 0.3419 (ρ = 0.5), and 0.3422 (ρ = 0.6). The heterogeneity indices were similar, $I^2_{(2)}$ = 37.0% and $I^2_{(3)}$ = 53.4% (ρ = 0.2), $I^2_{(2)}$ = 35.0% and $I^2_{(3)}$ = 55.4% (ρ = 0.3), $I^2_{(2)}$ = 33.0% and $I^2_{(3)}$ = 57.4% (ρ = 0.4), $I^2_{(2)}$ = 31.2% and $I^2_{(3)}$ = 59.4% (ρ = 0.5), and $I^2_{(2)}$ = 29.3% and $I^2_{(3)}$ = 61.4% (ρ = 0.6). Overall, the sensitivity to the choice of ρ was marginal.

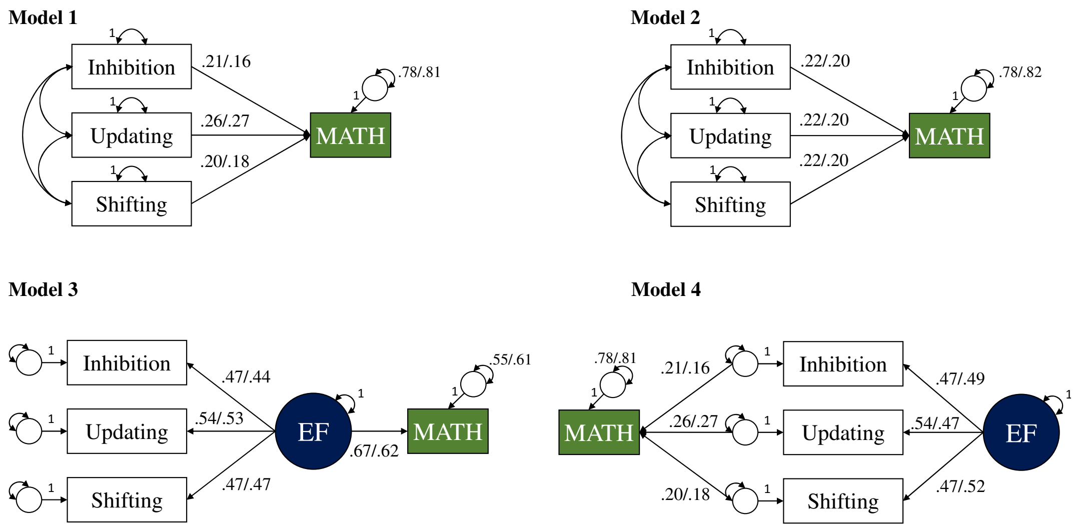
MASEM Estimation

Estimating Models 1–4 via one-stage MASEM resulted in the same directions and sizes of effects as those resulting from

[2] To further substantiate the unique effects of the EFs, we performed a Cholesky decomposition (Dang et al., 2015) and found support for the positive relation between inhibition and math intelligence (b = 0.30, 95% CI [0.25, 0.34]), the positive residual effect of shifting on math intelligence after partialling out inhibition (b = 0.23, 95% CI [0.17, 0.29]), and the unique residual effect of updating after partialling out inhibition and shifting (b = 0.24, 95% CI [0.18, 0.30]). This decomposition explained about 20% in variance and represented an approach for extracting unique relations while controlling for what is shared by the three EF subdimensions. The equality of the regression parameters remained, $\chi^2(2) = 3.4$, $p = .18$.

Figure 5

Meta-Analytic Structural Equation Models for Prekindergarten/Kindergarten and Preschool Samples



Note. Larger positive effect sizes indicate a closer relationship between EFs and math intelligence. MATH = math intelligence; EF = latent variable representing general executive functions. The coefficients for the pre-K and K samples are shown first, and the coefficients for the preschool samples second. All coefficients were statistically significant ($p < .05$). See the online article for the color version of this figure.

the two-stage approach (see Supplemental Material S4). Marginal differences occurred in the estimated standard errors and goodness-of-fit indices. However, the conclusions drawn from the four models did not change. The model parameters were not sensitive to the MASEM estimation procedure.

Attenuation of Correlation Coefficients

Effect sizes were attenuated for their corresponding reliability or, if no reliability had been reported, for the average reliability of the EF composite ($\alpha_{composite}$ = .86), inhibition ($\alpha_{inhibition}$ = .84), shifting ($\alpha_{shifting}$ = .81), updating ($\alpha_{updating}$ = .90), and math intelligence (α_{math} = .84) measures. Comparing the average correlation between EFs and math intelligence with unattenuated and attenuated effect sizes, revealed a higher correlation with attenuated effect sizes ($\bar{r}$ = .34 vs. $\bar{r}$ = .40). The average correlation remained statistically significant, of moderate size, and positive, meaning that its overall interpretation did not change. In addition, the confidence intervals of the results with unattenuated and attenuated effect sizes overlapped (95% CI [.31, .37] vs. [.36, .44]). Descriptive comparisons with the other data sets yielded similar results. It should be noted that the difference between unattenuated and attenuated average correlations ($\bar{r}$ = .40 vs. $\bar{r}$ = .49) for the relation between EFs and math intelligence in the pooled data set was statistically significant (see Table S6.7). As we averaged all available Cronbach's α values for this attenuation, we refrained from interpreting this difference in average correlations to avoid overgeneralizing the unreliability of EF measures. Due to the very similar average correlations, identical confidence intervals, and comparable heterogeneity, we decided to use unattenuated effect sizes, including the three effect sizes reported by Nguyen et al. (2019). Overall, these findings indicate that the meta-analysis was only marginally sensitive to the specified conditions.

Influence of Primary Study Quality

The composite index of primary study quality ranged between −0.02 and 0.25 and deviated from a normal distribution at the two tails (skewness = 0.31, kurtosis = 0.87). On average, the quality index was 0.12 (SD = 0.05) with a median of 0.12. Using the quality index as a moderator, we found that it explained about 29.1% of the within-study heterogeneity, yet no between-study heterogeneity. Primary study quality moderated the EF-math intelligence correlation significantly, $F(1, 361) = 20.2$, $p < .001$, and the respective effect was positive (B = 0.05, SE = 0.01, 95% CI [0.02, 0.08]). Hence, primary studies with higher quality tended to report larger correlations between EFs and math intelligence. After controlling for study quality and considering the cluster-robust standard errors, the effects of all other moderators remained and were similar in size and direction (see Supplemental Material S7).

Discussion

Summary of Key Findings

To contribute to the debate on the relations among cognitive skills in preschool children, we addressed three research questions by performing a meta-analysis of a total of 363 effect sizes from 47 studies. First, we synthesized the relations between EFs and math intelligence in preschool children, differentiating among three EF subdimensions: inhibition, shifting, and updating (RQ1). Second, we identified moderators that explain heterogeneity within and between studies (RQ2). Table 9 displays the key results for RQ1 and RQ2. Third, we examined the differential contributions of inhibition, shifting, and updating to the explanation of variance in math intelligence (RQ3). Overall, the meta-analysis provided evidence of the relation between math intelligence and EFs as a

Table 9
Summary of the Main Findings of RQ1 and RQ2

Relation (data set)	RQ1: Correlations $\bar{r}$ [95% CI]	RQ2: Significant moderators
EF and math intelligence (complete)	.34 [.31, .37]	• Country: Samples from the United Kingdom and Japan showed large effects; samples from Turkey and Pakistan small, nonsignificant ones • Continent: Larger effect for American samples • EF subdimension: Order of effects, composite > updating > shifting > inhibition • EF task type: Largest effects for composite, tap (inhibition), Simon (inhibition), random generation (updating), and difficult span (updating) tasks • Mode of EF testing: Order of effects, verbal < behavioral ≈ apparatus-based ≈ computer-based testing
Inhibition and math intelligence (inhibition)	.30 [.25, .35]	• Reliability of math intelligence measures: Math intelligence measures with greater reliability showed closer link to inhibition • Continent: Larger effect for American samples • Inhibition task type: Largest effects for Simon, shape school, and tap tasks • Mode of inhibition testing: Order of effects, verbal < behavioral = apparatus-based = computer-based testing
Shifting and math intelligence (shifting)	.32 [.25, .38]	• Reliability of math intelligence measures: Math intelligence measures with greater reliability showed closer link to shifting • Publication status: Larger effect for published studies
Updating and math intelligence (updating)	.36 [.31, .40]	• Math intelligence subdimension: Order of effects, basic number knowledge < calculation and reasoning < composite • Updating task type: Largest effects for random generation and difficult span tasks

Note. Larger positive effect sizes indicate a closer relationship between EFs and math intelligence. $\bar{r}$ = weighted-average correlation; CI = confidence interval; EF = executive function; RQ = research question.

composite as well as separate subdimensions. Moreover, we found that several study, sample, and measurement characteristics moderated this relationship and explained variations within and between studies. Finally, there was no evidence of differential relations between math intelligence and inhibition, shifting, and updating. After the relation between a latent variable representing the EFs and math intelligence was controlled for, none of the EF residuals were related to math intelligence.

Overall Correlations (RQ1)

The results of the meta-analysis indicated a moderate relation between EFs and math intelligence ($\bar{r} = .34$, $p < .001$; see Table 9). Investigated separately, the three EF subdimensions—inhibition, shifting, and updating—each showed moderately statistically significant relations to math intelligence ($\bar{r} = .30$, $\bar{r} = .32$, and $\bar{r} = .36$, respectively; $p < .001$). Answering RQ1, these results indicate that EFs and their subdimensions are substantially related to math intelligence in preschool children similarly to other age groups (see Best et al., 2011; Cragg et al., 2017; Friso-van den Bos et al., 2013; Peng et al., 2016; Yeniad et al., 2013). From a conceptual perspective, this finding corroborates existing frameworks that integrate EFs and math skills (Diamond, 2013; Miyake et al., 2000). As they share some, but not all, variations, EFs and math intelligence can be differentiated but might have several processes in common (Kovacs & Conway, 2016) and be based on a similar collection of cognitive and metacognitive skills, which aid children in solving EF and math intelligence tasks (van Der Maas et al., 2006). Similar to the assumptions underlying the CHC theory, math intelligence and EFs may represent distinct but related constructs that share some common skills. Jewsbury et al. (2016) argued that EFs and facets of intelligence represent different facets of general cognition. In practice, this implies that assessing one of the two constructs does not make assessing the other redundant, although the performance on one may predict the performance on the other to some extent (van Aken et al., 2019). Thus, reducing the testing burden for young test-takers by focusing on only one construct may come at the cost of insufficient construct coverage. Additionally, this implies that educators should promote EFs and math intelligence in young children, as training in one may not necessarily transfer to the other (Webb et al., 2018). However, EFs might not be causally related to academic achievement, as suggested by a fixed-effect analysis of longitudinal data on the link between inhibition and several achievement measures (Willoughby et al., 2012). Additionally, Willoughby et al. (2019) only found a small within-person association between shifting and working memory and academic achievement from kindergarten to second grade. The transferability of EF training has also been questioned meta-analytically (Melby-Lervåg & Hulme, 2013). More research is needed to test the causality and, thus, the transferability of EF trainings to evaluate the economic sense of a more general use of EFs in teaching (see e.g., Vaughn et al., 2012).

Overall, previous meta-analytic research that included mostly schoolchildren and adolescents reported somewhat similar relations between EFs as a whole and math skills. For instance, Cortés Pascual et al. (2019) observed an effect ($\bar{r} = .37$) in primary school children (6–12 years old) that was comparable to the present results. This might imply that there is no drastic decrease in the strength of the relation between EFs and math intelligence from preschool to primary school. After integrating the results for the separate EF subdimensions, however, we observed a slight trend of decreasing relations between math intelligence and inhibition with older age. The relation between inhibition and math skills seemed to be stronger in meta-analyses with younger children and weaker with older children, with correlations ranging from $\bar{r}$ = .34 (N. P. Allan et al., 2014; 2.5–6.5 years) to the present result of $\bar{r}$ = .30 (95% CI [.24, .36]; 3–6.5 years) to $\bar{r}$ = .27 (Friso-van den Bos et al., 2013; 4–12 years). The relation between shifting and math skills, while yielding considerably lower average correlations, showed more variation over all the meta-analyses we reviewed. We found an average correlation of $\bar{r}$ = .32 for the preschool sample, which is descriptively larger than those reported by Yeniad et al. (2013; $\bar{r}$ = .26; 4–14 years) and Friso-van den Bos et al. (2013; $\bar{r}$ = .28; 4–12) and comparable to that reported by Jacob and Parkinson (2015; $\bar{r}$ = .34, 3–18 years). The link between updating and math intelligence did not vary substantially, and the correlation of $\bar{r}$ = .36 identified in the present study was similar to those reported in previous meta-analyses ($\bar{r}$ = .37, Cortés Pascual et al., 2019; $\bar{r}$ = .34 and .38, Fuhs et al., 2016; $\bar{r}$ = .35, Peng et al., 2016; see Table 1). Given these similarities, our narrower conceptualization of math skills as a facet of intelligence did not result in substantially different relations compared to other conceptualizations of these skills.

As noted previously, one key issue challenging the interpretation of the EF–math intelligence relation is task impurity (Nguyen et al., 2019). We addressed this issue in two ways. First, effect sizes based on complex EF tasks that required children to engage in multiple processes rather than a single EF process were meta-analyzed separately as part of the "composite" data set. In this way, possible bias due to task impurity may have been reduced in the meta-analyses of the data specific to the EF subdimensions. Second, as a common practice, we represented EFs with a latent variable (Model 3) to single out what was common across the tasks measuring the three EF subdimensions and what was considered a measurement error (Camerota et al., 2020). However, Nguyen et al. (2019) argued that this representation may not account for possible task impurity in relation to mathematics skills. Thus, we followed their recommendation and examined the EF–math intelligence relation via the residuals, that is, the latent variables describing what is unique to each EF subdimension (Model 4). In this way, common variation that might be due to overlapping EF processes required in the tasks may be controlled by the single latent EF variable. Although these approaches could counter possible task impurity bias to some extent, alternative latent variable models allowing for cross-loadings or task-specific factors may construct explicitly overlapping EF processes at the level of tasks. To specify and estimate such models, meta-analytic correlations among EF tasks are needed (Scherer & Teo, 2020). Moreover, accounting for the overlap between EF processes does not address the possible impurity due to the processes involved in EF tasks outside executive functioning (Friedman et al., 2008). In this sense, some processes involved in EFs and math intelligence may also be shared. A more detailed differentiation of math intelligence into its subdimensions and a data set providing a sufficiently large sample of correlations at the level of subdimensions are needed to control for this form of task impurity via latent variables.

Heterogeneity and Moderators (RQ2)

We found substantial heterogeneity in the data between and within studies. Five study and measurement characteristics could explain significant amounts of this heterogeneity in the relation between EFs and math intelligence: (a) the country and (b) the continent on which the study was conducted; (c) the EF subdimension; (d) the EF task type; and (e) the mode of EF testing. For the relation between inhibition and math intelligence, heterogeneity was explained by the reliability of the math intelligence and EF measures, the continent, the inhibition task type, and the mode of inhibition testing. Only the reliability of math intelligence measures and publication status moderated the correlation with shifting, and only the math intelligence subdimension and the updating task type explained the variability in the correlation with updating (Table 9). Regarding RQ2, these findings stress the importance of measurement characteristics (compared to study and sample characteristics) as moderators of the relation between EFs and math intelligence in preschool children (Table 9). At the same time, they contrast some of the moderator effects identified in previous meta-analyses with a different conceptualization of math skills.

Study and Sample Characteristics

In the present meta-analysis, the only sample characteristics moderating the relation between EFs and math intelligence were the country and the continent on which the primary studies had been conducted. These results were most likely driven by two studies, drawing on samples from Turkey (Söğüt et al., 2021) and Pakistan (Armstrong-Carter et al., 2020), which found nonsignificant correlations between EFs and math intelligence. This finding aligns somewhat with the notion that the involvement of EFs in math intelligence differs between children from different countries and perhaps cultures (Friso-van den Bos et al., 2013). However, we refrain from further interpreting cultural differences in these skills as the only reason for these effects. To the best of our knowledge, such cultural effects have yet to be examined in detail. Moreover, 31 of the 47 primary studies included in the present meta-analysis were conducted in the United States, limiting the possible inferences that could be drawn regarding country differences.

Several promising moderators did not yield statistically significant results. For example, children's age did not statistically significantly explain any heterogeneity in the present meta-analysis. Similarly, Cortés Pascual et al. (2019) found that age does not significantly moderate the EF–math relation in children between 6 and 12 years of age. One explanation for these results could be that the age range was restricted in the present meta-analysis as well as in previous ones (i.e., N. P. Allan et al., 2014; Cortés Pascual et al., 2019; Friso-van den Bos et al., 2013; Yeniad et al., 2013). This result contrasts with that of David's (2012) meta-analysis, in which age was identified as a significant moderator. However, we can observe the previous trend of younger children exhibiting stronger relations of inhibition with math intelligence than older children when reviewing previous meta-analyses. In line with previous studies (Bull et al., 2008; C. A. C. Clark et al., 2010), but contrary to one of eight prior meta-analyses (Cortés Pascual et al., 2019), the gender composition of the samples did not show any moderator effects, probably due to the small variation in this characteristic between studies. These results are very plausible, as even in the clearest case of gender composition in education—namely, single-sex education—girls and boys are unlikely to show large differences other than in their interests (Eliot, 2013).

Executive Function Measurement Characteristics

Three characteristics of the EF measurement moderated the relation between EFs and math intelligence in preschool children. First, this relation varied slightly between the EF subdimensions under investigation, such that the correlation was descriptively weaker for inhibition, moderate for shifting, and stronger for updating (see RQ2). Similar differences in the correlations with math skills between the EF subdimensions have been found in previous meta-analyses (Cortés Pascual et al., 2019; Friso-van den Bos et al., 2013; Jacob & Parkinson, 2015). In the moderator analysis, we also examined a composite EF score that summarizes all EF measures, encompassing at least two of the three EF subdimensions. With a high correlation with math intelligence, the composite EF is likely to be the driver of this moderation effect. The relation between this score and math intelligence was stronger than the relations for each subdimension individually. This might be because tests yielding a composite score for EFs tend to be similar in design to tests of general cognitive ability and, therefore, may overlap substantially with math intelligence tests (see e.g., Yeniad et al., 2013). Additionally, these findings are in line with the notion that the differentiation among the three EF subdimensions might be greater in the later years of childhood, explaining the small differences between subdimensions in the present preschool sample.

Second, the EF task type significantly explained heterogeneity within and between studies. The extent of the explained heterogeneity was so great that a cross-classified random-effect model in which EF task type was the fourth level suggested substantial variation across the EF task types. The present meta-analysis categorized more than 70 EF tasks into 15 distinguishable categories. The strongest relations to math intelligence were found for the composite task type (which measures at least two EF subdimensions), tap and Simon tasks (which measure inhibition), and random generation and difficult span tasks (which measure updating). These results are not surprising; the composite, random generation, and difficult span tasks are very close in structure to math intelligence tasks and require individuals to handle numbers, especially in updating tasks such as the backward number span task (Peng et al., 2016). This is another example of task impurity (Friedman et al., 2008), which describes the common phenomenon that one task requires multiple EF subdimensions, making a clear-cut interpretation more difficult. Overall, the results reflect the vast variety of EF tasks used in the field and simultaneously draw attention to inconsistencies related to measuring EFs at an age at which children cannot read.

Third, the mode of EF testing moderated the relationship between EFs and math intelligence at the effect size level, with very small descriptive differences among verbal, behavioral, apparatus-based, and computer-based testing. Looking more closely into the single EF subdimensions, we found that this moderation effect occurred only in inhibition tasks and not in shifting or updating tasks. Therefore, inhibition drives the overall moderator effect. The higher effects for behavioral, apparatus-based, and computer-based testing might also be due to their overlap with common actions involving math intelligence. The source for this speculation is the similarities between the behavioral tap task and counting on one's fingers, for instance, or the computer-based shape school task and preschool geometry puzzles. Although this finding might raise the question of how testing mode-general versus mode-specific inhibition is, we should not overinterpret it, as the mode of EF testing explained only a very small amount of variance between effect sizes and none between studies. Previous meta-analyses had somewhat similar findings for working memory domains (David, 2012; Friso-van den Bos et al., 2013; Peng et al., 2016). Overall, this indicates the disadvantage of verbal (inhibition) testing compared to the other modes of testing for preschool children.

Model Testing: Explaining Variations in Math Intelligence (RQ3)

Utilizing the analytic framework of correlation-based MASEM, we substantiated the evidence that the EF subdimensions statistically significantly and jointly explain variation in math intelligence via testing and comparing several structural equation models (Models 1–4). Examining the joint effects via multiple regression (Model 1), we found that all EFs exhibited statistically significant relations with math intelligence and these relations did not differ among the EF subdimensions (Model 2). Representing EFs as a single latent variable (Model 3), we obtained a strong and positive effect on math intelligence, an effect much larger than the EF factor loadings. An alternative interpretation of this finding might be that math intelligence tests are better single measures of general EF than any single EF measure because the common variance of inhibition, shifting, and update can be better explained by math intelligence than with any one of the EF subdimensions. Further testing with a broader set of measures representing the constructs more comprehensively is needed to corroborate this finding conceptually.

Finally, we examined the unique effects via the EF residuals (Model 4) and found positive and statistically significant path coefficients that did not differ between the three EF subdimensions. Overall, the variance explanations in math intelligence were substantial, ranging from 21% to 42%. Of the three EFs, updating descriptively, but not statistically significantly, explained more variance than inhibition and shifting. This trend may point to possible overlaps in the processes or assessment methods involved in updating and math intelligence tasks and supports previous findings (Bull & Lee, 2014). Despite the interpretation of these overlaps as method effects or commonalities among the processes involved in EFs, another interpretation lies in the domain specificity of EFs, bringing together these two elements. Specifically, Peng et al. (2018) sought evidence of the domain specificity of working memory dimensions (numeric, verbal, and visuospatial) and argued that working memory operates through domain knowledge, skills, and procedures in task-specific situations. This perspective aligns with the situational model by Doebel (2020), who considered the development of EFs to be a set of skills activating specific knowledge, beliefs, values, norms, and preferences. Thus, EFs are directed at meeting the demands and goals of specific tasks. In this sense, the shared EF process and specific task demands cannot be strictly separated.

From a substantive perspective, several findings are worth noting. First, the evidence of uniform relations with math intelligence if the three EF subdimensions are considered jointly may indicate the unitary (rather than multifaceted) nature of the EF construct. Gonzalez et al. (2021) argued that uniform relations to external criteria are key to establishing that constructs may represent the same underlying processes or traits. Although the present results allow for this conclusion, this criterion is not sufficient to establish that the three EFs are the same. Additional evidence of the factor structure of EF measures across samples, contexts, and time is needed to substantiate this conclusion (Marsh et al., 2019). However, to generate such meta-analytic evidence, item- or subscale-level correlation matrices are needed, and the diversity of EF assessments across studies limits the possibility of such research syntheses (Karr et al., 2018). The present meta-analysis revealed small-to-moderate correlations among the three EFs. At the same time, there is some existing evidence of the unitary nature of EFs for preschool children (Garon et al., 2008; Nelson et al., 2016). These observations highlight that relations to external criteria (i.e., math intelligence) may not represent the only source of information for or against the unity of the three EFs.

Second, the MASEM results did not support the differential relations to math intelligence identified in experimental and longitudinal studies of EFs (Cragg & Gilmore, 2014; Jacob & Parkinson, 2015). One possible reason for these divergent results may be the way in which math skills were operationalized in the present study. In contrast to measures of school performance and educational achievement, we conceptualized math skills within the intelligence framework, defining them as math intelligence. In this sense, the uniform relations between EFs and math intelligence align with mounting evidence of substantial relations between EFs and general intelligence (Jewsbury et al., 2016). The strength of these relations is comparable to that reported for EFs and other intelligence measures (Friedman & Miyake, 2017). Thus, to meaningfully interpret the EF–math link, researchers must carefully consider the conceptualization of math skills.

Third, Model 3 revealed a strong relation between the single latent EF variable and math intelligence, with a regression coefficient that was higher than the factor loadings of the EF subdimensions. Reviewing the data from a smaller set of primary studies, Nguyen et al. (2019) obtained the same results as we did in the present meta-analysis. This result may have several explanations and implications. From a substantive perspective, the results may highlight that the processes common to the three EFs and the processes captured by the math intelligence assessment tasks overlap substantially (Kovacs & Conway, 2016). In other words, the two constructs may be based on a common set of cognitive and metacognitive or mutually reinforcing skills that support children in performing the respective assessment tasks (van Der Maas et al., 2006). Similar to the notion of "*g*" in describing what is considered common across intelligence tests, these overlapping processes may also be considered indicators of a general factor underlying EFs and math intelligence (Webb et al., 2018). Identifying what is common across the EF subdimensions and math intelligence could aid in understanding possible transfer effects between the two constructs (Melby-Lervåg et al., 2016). From a measurement perspective, the high structural parameter connecting the single latent EF variable and math intelligence may also be due to a misrepresentation of the EF construct. Despite a very good fit to the meta-analytic data, the reflective measurement model in Model 3 may not represent the true EF structure. In fact, Rhemtulla et al. (2020) showed that common factor models can introduce severe bias to the structural parameters in structural equation models, for instance, if the true model is a formative rather than a reflective measurement model. Similarly, Camerota et al. (2020) argued for considering alternative representations of EFs, for instance, as composites.

Fourth, after what was shared among the EF subdimensions was controlled for, the residuals were statistically significantly and positively related to math intelligence in Model 4. This finding is in line with some of the results reported by Nguyen et al. (2019). Of the nine empirical studies the authors reviewed, two supported Model 4 and showed the same pattern we observed in the meta-analysis. The positive relations between the EF residuals and the measure of math intelligence indicate that unique effects might exist. However, these unique effects did not differ among the three EF residuals, possibly due to the limited number of studies included. This finding calls into question the extent to which the possible unique processes underlying the three EF subdimensions are substantively unique. Although there was a slight preference for Model 3 over Model 4 in the pooled meta-analytic data set (Model 3 was more parsimonious and showed a descriptively better fit than Model 4), both models described the data well and shed light on different aspects of the EF–math intelligence relation.

To further test for the uniqueness of the EF effects, we also adopted Nguyen et al.'s (2019) hybrid models, in which the relationships between the EF residuals and math intelligence were controlled for the effect of the latent EF variable—that is, what is common to all three EF subdimensions. These models supported the dominance of the latent EF variable, as none of the residual effects existed. Similar to Nguyen et al.'s (2019) observations, "latent EF was a very robust predictor of math" (p. 282), and this latent variable accounted largely for the EF–math intelligence relationship.

Fifth, we observed a tendency toward more differentiated relations among the EF subdimensions and math intelligence for the preschool samples than for the prekindergarten and kindergarten samples. Although age ranges in preschool, prekindergarten, and kindergarten differ between countries, this result may inform the discussion on the differentiation of EFs and other cognitive skills over time (Lerner & Lonigan, 2014). However, a much broader age range is needed in future meta-analyses to investigate the possible breaking points at which differentiation occurs.

From a methodological perspective, the MASEM approach offered a powerful procedure for jointly describing and examining the relations among the four constructs and exploring the fit of different models. This allowed us to obtain additional evidence of the relations between EFs and math intelligence beyond univariate relations (Cheung, 2015a). In addition, we were able to evaluate the fit of structural equation models representing different assumptions (Cheung & Cheung, 2016). Specifically, in the meta-analysis, we followed extant modeling approaches, assuming either correlated but distinct EFs or a unitary EF construct (Friedman & Miyake, 2017). However, as noted previously, correlations among EF and math intelligence indicators at the level of tasks or task paradigms (Lehtonen et al., 2018) could shed more light on the structure of EFs and the extent to which math intelligence may be incorporated into this structure. Based on such meta-analytic data, finer-grained models could be tested, such as models that assume a general factor underlies EFs and math intelligence and specific factors representing their unique components (Friedman et al., 2008). In this way, the relationship between EFs and math intelligence could be described factor analytically and may inform existing factor models that incorporate EFs and other intelligence measures (Jewsbury et al., 2016; Webb et al., 2018). However, the preference for one or another factor model depends on the assignment of the indicators to the respective latent variables (Ackerman et al., 2005). Therefore, establishing clear EF assessment frameworks and examining the sensitivity toward the choice of alternative frameworks are key (Oberauer et al., 2005).

Finally, although the choice of the four models and the hybrid models we tested via MASEM was based on different theoretical assumptions on the EF–math intelligence relationship, we could not support the preference for one model over the others. In this sense, all four models were useful for examining the joint relations between the EF subdimensions and math intelligence from different perspectives. Models 1 and 2 focused on direct joint relations and their equality across EF subdimensions; Models 3 and 4 focused on the effect of a single EF variable or, respectively, the unique effects of the EF subdimensions. Although these models are commonly specified in EF research and exhibited a very good fit to the meta-analytic data, they may oversimplify the complex relationship between EFs and math skills and be subject to task impurity bias (Camerota et al., 2020; Nguyen et al., 2019). Therefore, we encourage researchers in the field to evaluate multiple models rather than a single model for describing the EF–math relationship, consider alternative representations of the respective EF and mathematics constructs, such as composite scores or formative measurement models (Rhemtulla et al., 2020), and examine in greater detail the processes that are common or unique to EFs and math skills (Cragg & Gilmore, 2014; Kovacs & Conway, 2016).

Limitations and Future Directions

Several limitations of the meta-analysis are worth noting. First, there are several limitations inherent in the inclusion/exclusion criteria and the language restriction. Thus, the results are limited to preschool children without a diagnosed medical condition or disorder. Thus, the results do not generalize across all possible samples of preschoolers and do not allow one to test specific disadvantages in EFs and math intelligence due to medical conditions or disorders. Including clinical studies might have been especially interesting because EFs are impaired in children with specific disorders (Broadley et al., 2017; Kingdon et al., 2016; Lai et al., 2017). As we restricted our search to English records, a monolanguage bias might be present, which limits the generalizability of our results (Johnson, 2021). Future meta-analytic research should, therefore, test the robustness of the meta-analytic evidence across multiple languages. Further, we excluded studies published before the year 2000, potentially missing otherwise eligible studies and their contribution to evidence formation in the field.

Second, 66% of all included studies used American samples. Although 34% of the included studies examined samples from Asia, Australia, or Europe, the study pool did not include samples from South America or Africa. Therefore, further evidence of samples from a broader spectrum of countries is needed to be able to generalize the present findings further and test for cultural differences in cognitive processes (Imbo & LeFevre, 2009). Third, the study pool was too small to investigate all the moderators of interest. This was partially due to the strict inclusion and exclusion criteria applied during the study selection. For example, we had to exclude eligible clinical studies that did not report crucial results for the healthy control group. Therefore, we encourage researchers to extend and update this meta-analytic sample to examine additional moderators of the EF–math link. Fourth, the categorization of EF task types in the meta-analysis represented a compromise between precision and manageability due to the large variety of possible categorizations in the extant EF literature (Baggetta & Alexander, 2016; Garon et al., 2008). In line with Miyake et al. (2000), we did not explicitly distinguish between tasks measuring hot and cool EFs, which might have explained further variations between task types (Brock et al., 2009). In the same vein, we could not distinguish clearly between pure working memory tasks and updating tasks and thus subsumed them under the updating category. Future meta-analyses should try to examine these two constructs separately if the reporting precision of primary studies allows. To account for the methodological overlap between EF and math intelligence tasks, the content of a task (e.g., whether a span task uses numbers, letters, or pictures as stimuli) should be coded and examined as a methodological moderator in future research. To further enhance precision, n-back tasks and difficult span tasks could be investigated separately in accordance with meta-analytic findings (Redick & Lindsey, 2013) if enough effect sizes are available. Generally, variation between task definitions is a well-known phenomenon in EF research and has led to divergent findings in the relations between EFs and other cognitive abilities (Ackerman et al., 2005; Friso-van den Bos et al., 2013; Lehtonen et al., 2018). A clear and consistent framework of EF tasks for different age groups and comprehensive reporting in primary studies might be the much-needed solutions to this problem. Although it was not possible in this meta-analysis due to the small number of primary studies providing intercorrelations among math intelligence tasks or subdimensions, we argue that a further differentiation of the math intelligence construct, especially in Models 1–4, could provide more detailed evidence of the connections between specific EFs and specific math skills. As several meta-analyses have demonstrated (Hjetland et al., 2020; Scherer et al., 2020; Yang et al., 2022), hypotheses regarding the connections between multifaceted constructs could be tested with a range of factor models, such as single-, correlated-, or nested-factor models. To obtain such evidence, however, a more detailed reporting of primary studies, especially the correlation matrices containing not only the correlations among EF subdimensions and correlations between EF subdimensions and math skills but also the correlations among math skills, is needed. In the future, finer-grained reporting could further allow researchers to tease apart the multiple facets of EFs and math intelligence, as well as partial out the influence of general intelligence, which could not have been fully achieved with the present data.

A comprehensive framework of EFs and academic skills could be established in the future to streamline EF assessment and account for several trends in the literature. Namely, such a framework should aim to explain (a) the very similar correlations across EF subdimensions (see Table 1), (b) an age trend of the EF–academics relationship increasing/decreasing with age for different EF subdimensions, (c) very small training and transferability effects (Kassai et al., 2019; Melby-Lervåg et al., 2016; Melby-Lervåg & Hulme, 2013; Takacs & Kassai, 2019), and (d) the domain specificity of EFs in task-specific situations (Doebel, 2020; Peng et al., 2018). Such a framework would introduce exciting new hypotheses to test and take the field forward in leaps and bounds.

Similar to the present study, new research helps to create a solid basis for this comprehensive framework. Spiegel et al. (2021), for instance, recently examined the relationship between EFs and reading, language, and mathematics in primary school children. The authors' findings suggest that a simple age trend—that EF subdimensions diverge but their link to academic skills strengthens with age—is very unlikely. Instead, the predictive strength of specific EF subdimensions seems to depend on the developmental stage and the exact academic skill being measured. Combined with insights from Peng and Kievit's (2020) review of the mutual effects of cognitive functions and academic skills, these findings provide further grounds for a comprehensive framework of EFs and academic skills in the future.

Conclusions

The present meta-analysis provides a comprehensive overview of the literature on the relation between EFs and math intelligence in preschool children from 2000 to 2021. The present findings suggest that EFs, represented by a composite as well as three subdimensions, are positively and significantly related to math intelligence in preschool children. This relation testifies to the overlap in some skills and measures and, ultimately, the involvement of EFs in solving math intelligence tasks and vice versa. To some degree, the performance on one construct measure could be used to predict the performance on the other. Nevertheless, the evidence presented in this meta-analysis does not suggest that assessing one of the two constructs may make assessment of the other redundant. In addition to the positive correlations, there is substantial heterogeneity within and between studies, suggesting that these effect sizes are reproducible across studies. As measurement characteristics, rather than sample or study characteristics, primarily explain parts of this heterogeneity, we highlight the importance of considering the psychometric quality of EFs and math intelligence assessments when interpreting their correlation. When considered jointly, the relations between the EF subdimensions and math intelligence were similar for the study-level data. This finding does not provide evidence of the differential relations between a single EF and math intelligence after the other EFs are controlled for. Further research is needed to establish a comprehensive framework of EF task types to streamline the EF assessments, clarifying, for instance, the impact of age on the relation between EFs and math intelligence.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. https://doi.org/10.1037/0033-2909.131.1.30

*Ahmed, S. (2019). *Measuring executive function during early childhood: The utility of direct assessments, teacher ratings, and group-based tasks*. http://hdl.handle.net/2027.42/151502

*Ahmed, S., Tang, S., Waters, N. E., & Davis-Kean, P. (2019). Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. *Journal of Educational Psychology*, *111*(3), 446–458. https://doi.org/10.1037/edu0000296

*Allan, D. M., Allan, N. P., Lerner, M. D., Farrington, A. L., & Lonigan, C. J. (2015). Identifying unique components of preschool children's self-regulatory skills using executive function tasks and continuous performance tests. *Early Childhood Research Quarterly*, *32*(3), 40–50. https://doi.org/10.1016/j.ecresq.2015.02.001

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology*, *50*(10), 2368–2379. https://doi.org/10.1037/a0037493

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

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*(9), 839–851. https://doi.org/10.1037/0003-066X.63.9.839

Arán Filippetti, V., & Richaud, M. C. (2017). A structural equation modeling of executive functions, IQ and mathematical skills in primary students: Differential effects on number production, mental calculus and arithmetical problems. *Child Neuropsychology*, *23*(7), 864–888. https://doi.org/10.1080/09297049.2016.1199665

*Armstrong-Carter, E., Finch, J. E., Siyal, S., Yousafzai, A. K., & Obradović, J. (2020). Biological sensitivity to context in Pakistani preschoolers: Hair cortisol and family wealth are interactively associated with girls' cognitive skills. *Developmental Psychobiology*, *62*(8), 1046–1061. https://doi.org/10.1002/dev.21981

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

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

Baguss, J. (2020, June 20). Artificial intelligence solutions—One of these things is not like the other. *Evidence Partners Blog*. https://blog.evidencepartners.com/artificial-intelligence-solutions-one-of-these-things-is-not-like-the-other

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin*, *129*(2), 195–215. https://doi.org/10.1037/0033-2909.129.2.195

*Barnes, M. A., Clemens, N. H., Fall, A.-M., Roberts, G., Klein, A., Starkey, P., McCandliss, B., Zucker, T., & Flynn, K. (2020). Cognitive predictors of difficulties in math and reading in pre-kindergarten children at high risk for learning disabilities. *Journal of Educational Psychology*, *112*(4), 685–700. https://doi.org/10.1037/edu0000404

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. https://doi.org/10.2307/2533446

Beier, M. E., & Ackerman, P. L. (2005). Working memory and intelligence: Different constructs. Reply to Oberauer et al. (2005) and Kane et al. (2005). *Psychological Bulletin*, *131*(1), 72–75. https://doi.org/10.1037/0033-2909.131.1.72

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

Blair, C. (2002). School readiness. Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist*, *57*(2), 111–127. https://doi.org/10.1037/0003-066X.57.2.111

Blair, C. (2010). Stress and the development of self-regulation in context. *Child Development Perspectives*, *4*(3), 181–188. https://doi.org/10.1111/j.1750-8606.2010.00145.x

*Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

*Blakey, E., & Carroll, D. J. (2015). A short executive function training program improves preschoolers' working memory. *Frontiers in Psychology*, *6*(8), Article 1827. https://doi.org/10.3389/fpsyg.2015.01827

*Blankson, A. N., Gudmundson, J. A., & Kondeh, M. (2019). Cognitive predictors of kindergarten achievement in African American children. *Journal of Educational Psychology*, *111*(7), 1273–1283. https://doi.org/10.1037/edu0000346

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Wiley.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203

Boyle, K., Felling, R., Yiu, A., Battarjee, W., Schwartz, J. M., Salorio, C., & Bembea, M. M. (2018). Neurologic outcomes after extracorporeal membrane oxygenation: A systematic review. *Pediatric Critical Care Medicine*, *19*(8), 760–766. https://doi.org/10.1097/PCC.0000000000001612

Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association : JMLA*, *104*(3), 240–243. https://doi.org/10.3163/1536-5050.104.3.014

Broadley, M. M., White, M. J., & Andrew, B. (2017). A systematic review and meta-analysis of executive function performance in type 1 diabetes mellitus. *Psychosomatic Medicine*, *79*(6), 684–696. https://doi.org/10.1097/PSY.0000000000000460

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of ‘hot’ and ‘cool’ executive function to children’s academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*(3), 337–349. https://doi.org/10.1016/j.ecresq.2009.06.001

Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). The differentiation of executive functions in middle and late childhood: A longitudinal latent-variable analysis. *Intelligence*, *47*, 34–43. https://doi.org/10.1016/j.intell.2014.08.010

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence*, *40*(5), 458–469. https://doi.org/10.1016/j.intell.2012.05.006

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*(3), 205–228. https://doi.org/10.1080/87565640801982312

*Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T. D., & Nelson, J. M. (2011). Using confirmatory factor analysis to understand executive control in preschool children: Sources of variation in emergent mathematic achievement. *Developmental Science*, *14*(4), 679–692. https://doi.org/10.1111/j.1467-7687.2010.01012.x

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

*Cai, D., Zhang, L., Li, Y., Wei, W., & Georgiou, G. K. (2018). The role of approximate number system in different mathematics skills across grades. *Frontiers in Psychology*, *9*, Article 1733. https://doi.org/10.3389/fpsyg.2018.01733

*Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child Development*, *83*(4), 1229–1244. https://doi.org/10.1111/j.1467-8624.2012.01768.x

Camerota, M., Willoughby, M. T., & Blair, C. B. (2020). Measurement models for studying child executive functioning: Questioning the status quo. *Developmental Psychology*, *56*(12), 2236–2245. https://doi.org/10.1037/dev0001127

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

Case, R., Okamoto, Y., Griffin, S., McKeough, A., Bleiker, C., Henderson, B., Stephenson, K. M., Siegler, R. S., & Keating, D. P. (1996). The role of central conceptual structures in the development of children’s thought. *Monographs of the Society for Research in Child Development*, *61*(1/2), i–295. https://doi.org/10.2307/1166077

Cassidy, A. R., White, M. T., DeMaso, D. R., Newburger, J. W., & Bellinger, D. C. (2016). Processing speed, executive function, and academic achievement in children with dextro-transposition of the great arteries: Testing a longitudinal developmental cascade model. *Neuropsychology*, *30*(7), 874–885. https://doi.org/10.1037/neu0000289

Cavanaugh, J. E. (1997). Unifying the derivations for the Akaike and corrected Akaike information criteria. *Statistics & Probability Letters*, *33*(2), 201–208. https://doi.org/10.1016/S0167-7152(96)00128-9

Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211–229. https://doi.org/10.1037/a0032968

Cheung, M. W.-L. (2015a). *Meta-analysis: A structural equation modeling approach*. Wiley.

Cheung, M. W.-L. (2015b). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*, Article 1521. https://doi.org/10.3389/fpsyg.2014.01521

Cheung, M. W.-L., & Cheung, S. F. (2016). Random-effects models for meta-analytic structural equation modeling: Review, issues, and illustrations. *Research Synthesis Methods*, *7*(2), 140–155. https://doi.org/10.1002/jrsm.1166

*Chu, F. W., vanMarle, K., Hoard, M. K., Nugent, L., Scofield, J. E., & Geary, D. C. (2019). Preschool deficits in cardinal knowledge and executive function contribute to longer-term mathematical learning disability. *Journal of Experimental Child Psychology*, *188*, Article 104668. https://doi.org/10.1016/j.jecp.2019.104668

*Clark, C. A., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development*, *84*(2), 662–677. https://doi.org/10.1111/j.1467-8624.2012.01854.x

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology*, *46*(5), 1176–1191. https://doi.org/10.1037/a0019672

Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, *36*, 79–90. https://doi.org/10.1016/j.ecresq.2015.12.009

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*(3–4), 256–266. https://doi.org/10.1093/biomet/37.3-4.256

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, *19*(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493

Cortés Pascual, A., Moyano Muñoz, N., & Quílez Robres, A. (2019). The relationship between executive functions and academic performance in primary education: Review and meta-analysis. *Frontiers in Psychology*, *10*, Article 1582. https://doi.org/10.3389/fpsyg.2019.01582

*Costa, H. M., Partanen, P., & Van Herwegen, J. (2021). *The role of working memory, processing speed and approximate number system abilities in low maths achievement among preschoolers* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/b6p8d

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

*Cueli, M., Areces, D., García, T., Alves, R. A., & González-Castro, P. (2020). Attention, inhibitory control and early mathematical skills in preschool students. *Psicothema*, *32*(2), 237–244. https://doi.org/10.7334/psicothema2019.225

Dang, C.-P., Braeken, J., Colom, R., Ferrer, E., & Liu, C. (2015). Do processing speed and short-term storage exhaust the relation between working memory capacity and intelligence? *Personality and Individual Differences*, *74*, 241–247. https://doi.org/10.1016/j.paid.2014.10.012

David, C. V. (2012). Working memory deficits in math learning difficulties: A meta-analysis. *International Journal of Developmental Disabilities*, *58*(2), 67–84. https://doi.org/10.1179/2047387711Y.0000000007

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

Doebel, S. (2020). Rethinking executive function and its development. *Perspectives on Psychological Science*, *15*(4), 942–956. https://doi.org/10.1177/1745691620904771

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., 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

*Duncan, R., Nguyen, T., Miao, A., McClelland, M., & Bailey, D. (2016). *Executive function and mathematics achievement: Are effects construct- and time-general or specific?* Society for research on educational effectiveness. ERIC. https://eric.ed.gov/?id=ED567239

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Dweck, C. S. (2008). *Mindsets and math/science achievement*. Carnegie Corp. of New York-Institute for advanced Study Commission on Mathematics and Science Education. http://www.growthmindsetmaths.com/uploads/2/3/7/7/23776169/mindset_and_math_science_achievement_-_nov_2013.pdf

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *The BMJ*, *315*(7109), Article 629. https://doi.org/10.1136/bmj.315.7109.629

Eliot, L. (2013). Single-sex education and the brain. *Sex Roles*, *69*(7–8), 363–381. https://doi.org/10.1007/s11199-011-0037-y

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

Emslander, V., & Scherer, R. (2022, July 16). *Supplemental Material (the relation between executive functions and math intelligence in preschool children: A systematic review and meta-analysis)*. https://doi.org/10.17605/OSF.IO/VN5PE

Espy, K. A., Kaufmann, P. M., Glisky, M. L., & McDiarmid, M. D. (2001). New procedures to assess executive functions in preschool children. *The Clinical Neuropsychologist*, *15*(1), 46–58. https://doi.org/10.1076/clin.15.1.46.1908

Evidence Partners. (2021). *DistillerSR* (Version 2.35) [Computer Software]. https://www.evidencepartners.com/products/distillersr-systematic-review-software

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2021). Detecting selection bias in meta-analyses with multiple outcomes: A simulation study. *Journal of Experimental Education*, *89*(1), 125–144. https://doi.org/10.1080/00220973.2019.1582470

Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, *52*(5), 2031–2052. https://doi.org/10.3758/s13428-020-01373-9

Fisher, Z., Tipton, E., & Hou, Z. (2017). *robumeta: Robust variance meta-regression* (R package Version 2.0) [Computer software]. https://CRAN.R-project.org/package=robumeta

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

Fleming, M. L., & Malone, M. R. (1983). The relationship of student characteristics and student performance in science as viewed by meta-analysis research. *Journal of Research in Science Teaching*, *20*(5), 481–495. https://doi.org/10.1002/tea.3660200510

Follmer, D. J. (2018). Executive function and reading comprehension: A meta-analytic review. *Educational Psychologist*, *53*(1), 42–60. https://doi.org/10.1080/00461520.2017.1309295

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

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

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., Hornburg, C. B., & McNeil, N. M. (2016). Specific early number skills mediate the association between executive functioning skills and mathematics achievement. *Developmental Psychology*, *52*(8), 1217–1235. https://doi.org/10.1037/dev0000145

*Fujisawa, K. K., Todo, N., & Ando, J. (2019). Changes in genetic and environmental influences on cognitive ability, executive function, and preacademic skills in Japanese preschool age twins. *Developmental Psychology*, *55*(1), 38–52. https://doi.org/10.1037/dev0000627

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

*Geary, D. C., & vanMarle, K. (2016). Young children's core symbolic and nonsymbolic quantitative knowledge in the prediction of later mathematics achievement. *Developmental Psychology*, *52*(12), 2130–2144. https://doi.org/10.1037/dev0000214

*Geary, D. C., vanMarle, K., Chu, F. W., Hoard, M. K., & Nugent, L. (2019). Predicting age of becoming a cardinal principle knower. *Journal of Educational Psychology*, *111*(2), 256–267. https://doi.org/10.1037/edu0000277

Gignac, G. E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence*, *62*, 138–147. https://doi.org/10.1016/j.intell.2017.04.001

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). TEST REVIEW Behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235–238. https://doi.org/10.1076/chin.6.3.235.3152

Gonzalez, O., MacKinnon, D. P., & Muniz, F. B. (2021). Extrinsic convergent validity evidence to prevent jingle and jangle fallacies. *Multivariate Behavioral Research*, *56*(1), 3–19. https://doi.org/10.1080/00273171.2019.1707061

*Gray, S. A., & Reeve, R. A. (2014). Preschoolers' dot enumeration abilities are markers of their arithmetic competence. *PLOS ONE*, *9*(4), Article e94428. https://doi.org/10.1371/journal.pone.0094428

Graziano, P. A., Garb, L. R., Ros, R., Hart, K., & Garcia, A. (2016). Executive functioning and school readiness among preschoolers with externalizing problems: The moderating role of the student–teacher relationship. *Early Education and Development*, *27*(5), 573–589. https://doi.org/10.1080/10409289.2016.1102019

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. (2019). *dmetar: Companion R package for the guide "Doing Meta-Analysis in R"* (R package Version 0.0.9000) [Computer software]. https://dmetar.protectlab.org

*Harvey, H. A., & Miller, G. E. (2017). Executive function skills, early mathematics, and vocabulary in head start preschool children. *Early Education and Development*, *28*(3), 290–307. https://doi.org/10.1080/10409289.2016.1218728

Henseler, J. (2021). *Composite-Based structural equation modeling: Analyzing latent and emergent variables*. Guilford Press.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *The BMJ*, *327*(7414), Article 557. https://doi.org/10.1136/bmj.327.7414.557

Hjetland, H. N., Brinchmann, E. I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review*, *30*, Article 100323. https://doi.org/10.1016/j.edurev.2020.100323

*Howard, S. J., & Vasseleu, E. (2020). Self-regulation and executive function longitudinally predict advanced learning in preschool. *Frontiers in Psychology*, *11*, Article 49. https://doi.org/10.3389/fpsyg.2020.00049

*Hu, B. Y., Johnson, G. K., Teo, T., & Wu, Z. (2020). Relationship between screen time and chinese children's cognitive and social development. *Journal of Research in Childhood Education*, *34*(2), 183–207. https://doi.org/10.1080/02568543.2019.1702600

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

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. SAGE Publications.

Hutchison, J., & Phillips, D. (2018, March 27). *Executive functions: Supporting foundational skills for early math learning*. DREME. Development and Research in Early Math Education. https://dreme.stanford.edu/news/executive-functions-supporting-foundational-skills-early-math-learning

Imbo, I., & LeFevre, J.-A. (2009). Cultural differences in complex addition: Efficient Chinese versus adaptive Belgians and Canadians. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1465–1476. https://doi.org/10.1037/a0017022

Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research*, *85*(4), 512–552. https://doi.org/10.3102/0034654314561338

Jak, S., & Cheung, M. W.-L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods*, *25*(4), 430–455. https://doi.org/10.1037/met0000245

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell-Horn-Carroll model. *Journal of Experimental Psychology: General*, *145*(2), 220–245. https://doi.org/10.1037/xge0000119

Johnson, B. T. (2021). Toward a more transparent, rigorous, and generative psychology. *Psychological Bulletin*, *147*(1), 1–15. https://doi.org/10.1037/bul0000317

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Kassai, R., Futo, J., Demetrovics, Z., & Takacs, Z. K. (2019). A meta-analysis of the experimental evidence on the near- and far-transfer effects among children's executive function skills. *Psychological Bulletin*, *145*(2), 165–188. https://doi.org/10.1037/bul0000180

*Kidd, J. K., Lyu, H., Peterson, M. S., Hassan, M. Z., Gallington, D. A., Strauss, L. I., Patterson, A. B., & Pasnak, R. (2019). Patterns, mathematics, early literacy, and executive functions. *Creative Education*, *10*(13), 3444–3468. https://doi.org/10.4236/ce.2019.1013266

Kingdon, D., Cardoso, C., & McGrath, J. J. (2016). Research review: Executive function deficits in fetal alcohol spectrum disorders and attention-deficit/hyperactivity disorder—A meta-analysis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *57*(2), 116–131. https://doi.org/10.1111/jcpp.12451

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Press.

*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. https://doi.org/10.1016/j.lindif.2013.01.004

*Korucu, I., Litkowski, E., & Schmitt, S. A. (2020). Examining associations between the home literacy environment, executive function, and school readiness. *Early Education and Development*, *31*(3), 455–473. https://doi.org/10.1080/10409289.2020.1716287

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946

*Kroesbergen, E., Van Luit, J., Van Lieshout, E., Van Loosbroek, E., & Van de Rijt, B. (2009). Individual differences in early numeracy: The role of executive functions and subitizing. *Journal of Psychoeducational Assessment*, *27*(3), 226–236. https://doi.org/10.1177/0734282908330586

Lai, C. L. E., Lau, Z., Lui, S. S. Y., Lok, E., Tam, V., Chan, Q., Cheng, K. M., Lam, S. M., & Cheung, E. F. C. (2017). Meta-analysis of neuropsychological measures of executive functioning in children and adolescents with high-functioning autism spectrum disorder. *Autism Research*, *10*(5), 911–939. https://doi.org/10.1002/aur.1723

Lehto, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *49*(1), 29–50. https://doi.org/10.1080/713755616

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, *144*(4), 394–425. https://doi.org/10.1037/bul0000142

Lerner, M. D., & Lonigan, C. J. (2014). Executive function among preschool children: Unitary versus distinct abilities. *Journal of Psychopathology and Behavioral Assessment*, *36*(4), 626–639. https://doi.org/10.1007/s10862-014-9424-3

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE Publications.

*Mann, T. D., Hund, A. M., Hesson-McInnis, M. S., & Roman, Z. J. (2017). Pathways to school readiness: Executive functioning predicts academic and social–emotional aspects of school readiness. *Mind, Brain and Education*, *11*(1), 21–31. https://doi.org/10.1111/mbe.12134

Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why multicollinearity matters: A reexamination of relations between self-efficacy, self-concept, and achievement. *Journal of Educational Psychology*, *96*(3), 518–522. https://doi.org/10.1037/0022-0663.96.3.518

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). *Goodness of fit in structural equation models*. Lawrence Erlbaum.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in Overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of*

Educational Psychology, *111*(2), 331–353. https://doi.org/10.1037/edu0000281

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, *42*(2), 236–243. https://doi.org/10.1016/j.jesp.2005.04.010

*Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of Educational Psychology*, *101*(3), 689–704. https://doi.org/10.1037/a0014240

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology*, *5*, Article 599. https://doi.org/10.3389/fpsyg.2014.00599

*McCoy, D. C., Gonzalez, K., & Jones, S. (2019). Preschool self-regulation and preacademic skills as mediators of the long-term impacts of an early intervention. *Child Development*, *90*(5), 1544–1558. https://doi.org/10.1111/cdev.13289

*McKinnon, R. D., Blair, C., & the Family Life Project Investigators. (2018). Does early executive function predict teacher–child relationships from kindergarten to second grade? *Developmental Psychology*, *54*(11), 2053–2066. https://doi.org/10.1037/dev0000584

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612

*Mills, B., Dyer, N., Pacheco, D., Brinkley, D., Owen, M. T., & Caughy, M. O. (2019). Developmental transactions between self-regulation and academic achievement among low-income African American and Latino children. *Child Development*, *90*(5), 1614–1631. https://doi.org/10.1111/cdev.13091

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., Kross, E., Teslovich, T., Wilson, N. L., Zayas, V., & Shoda, Y. (2011). 'Willpower' over the life span: Decomposing self-regulation. *Social Cognitive and Affective Neuroscience*, *6*(2), 252–256. https://doi.org/10.1093/scan/nsq081

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

*Mou, Y., Berteletti, I., & Hyde, D. C. (2018). What counts in preschool number knowledge? A Bayes factor analytic approach toward theoretical model development. *Journal of Experimental Child Psychology*, *166*, 116–133. https://doi.org/10.1016/j.jecp.2017.07.016

*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. *The Spanish Journal of Psychology*, *14*(2), 580–588. https://doi.org/10.5209/rev_SJOP.2011.v14.n2.6

Nelson, J. M., James, T. D., Chevalier, N., Clark, C. A. C., & Espy, K. A. (2016). Structure, measurement, and development of preschool executive function. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 65–89). American Psychological Association. https://doi.org/10.1037/14797-004

*Nguyen, T., & Duncan, G. J. (2019). Kindergarten components of executive function and third grade achievement: A national study. *Early Childhood Research Quarterly*, *46*, 49–61. https://doi.org/10.1016/j.ecresq.2018.05.006

Nguyen, T., Duncan, R. J., & Bailey, D. H. (2019). Theoretical and methodological implications of associations between executive function and mathematics in early childhood. *Contemporary Educational Psychology*, *58*, 276–287. https://doi.org/10.1016/j.cedpsych.2019.04.002

Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, *10*(4), 464–480. https://doi.org/10.1111/j.1467-7687.2007.00600.x

Oberauer, K., Schulze, R., Wilhelm, O., & Süss, H.-M. (2005). Working memory and intelligence—Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 61–65. https://doi.org/10.1037/0033-2909.131.1.61

Parkin, J. R., & Beaujean, A. A. (2012). The effects of Wechsler Intelligence Scale for Children-Fourth Edition cognitive abilities on math achievement. *Journal of School Psychology*, *50*(1), 113–128. https://doi.org/10.1016/j.jsp.2011.08.003

Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *The British Journal of Educational Psychology*, *82*(Pt. 1), 42–63. https://doi.org/10.1111/j.2044-8279.2011.02039.x

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, *144*(1), 48–76. https://doi.org/10.1037/bul0000124

Peng, P., & Kievit, R. A. (2020). The development of academic achievement and cognitive abilities: A bidirectional perspective. *Child Development Perspectives*, *14*(1), 15–20. https://doi.org/10.1111/cdep.12352

Peng, P., Namkung, J., Barnes, M., & 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

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine*, *26*(25), 4544–4562. https://doi.org/10.1002/sim.2889

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996. https://doi.org/10.1016/j.jclinepi.2007.11.010

*Purpura, D. J., Schmitt, S. A., & Ganley, C. M. (2017). Foundations of mathematics and literacy: The role of executive functioning components. *Journal of Experimental Child Psychology*, *153*, 15–34. https://doi.org/10.1016/j.jecp.2016.08.010

Pustejovsky, J. (2021). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections*. (R package Version 0.5.3) [Computer software]. https://CRAN.R-project.org/package=clubSandwich

Pustejovsky, J. E., & Tipton, E. (2021). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*, 425–438. https://doi.org/10.1007/s11121-021-01246-3

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*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok—Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology*, *48*(3), 731–737. https://doi.org/10.1016/j.jesp.2011.12.012

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *20*(6), 1102–1113. https://doi.org/10.3758/s13423-013-0453-9

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30–45. https://doi.org/10.1037/met0000220

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social–emotional competence. *Journal of Applied Developmental Psychology*, *27*(4), 300–309. https://doi.org/10.1016/j.appdev.2006.04.002

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Scherer, R., Siddiq, F., & Tondeur, J. (2020). All the same or different? Revisiting measures of teachers' technology acceptance. *Computers & Education*, *143*, Article 103656. https://doi.org/10.1016/j.compedu.2019.103656

Scherer, R., & Teo, T. (2020). A tutorial on the meta-analytic structural equation modeling of reliability coefficients. *Psychological Methods*, *25*(6), 747–775. https://doi.org/10.1037/met0000261

Scherer, R., Tondeur, J., Siddiq, F., & Baran, E. (2018). The importance of attitudes toward technology for pre-service teachers' technological, pedagogical, and content knowledge: Comparing structural equation modeling approaches. *Computers in Human Behavior*, *80*, 67–80. https://doi.org/10.1016/j.chb.2017.11.003

*Schmitt, S. A., Duncan, R. J., Budrevich, A., & Korucu, I. (2020). Benefits of behavioral self-regulation in the context of high classroom quality for preschoolers' mathematics. *Early Education and Development*, *31*(3), 323–334. https://doi.org/10.1080/10409289.2019.1660555

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford Press.

Schwarzer, G., Chemaitelly, H., Abu-Raddad, L. J., & Rücker, G. (2019). Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Research Synthesis Methods*, *10*(3), 476–483. https://doi.org/10.1002/jrsm.1348

Sheng, Z., Kong, W., Cortina, J. M., & Hou, S. (2016). Analyzing matrices of meta-analytic correlations: Current practices and recommendations. *Research Synthesis Methods*, *7*(2), 187–208. https://doi.org/10.1002/jrsm.1206

*Söğüt, M., Göksun, T., & Altan-Atalay, A. (2021). The role of numeracy skills on the Wisconsin card sorting test (WCST) performances of 5- to 8-Year-old turkish children. *British Journal of Developmental Psychology*, *39*(1), 231–246. https://doi.org/10.1111/bjdp.12353

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin & Review*, *24*(4), 1077–1096. https://doi.org/10.3758/s13423-016-1217-0

Spiegel, J. A., Goodrich, J. M., Morris, B. M., Osborne, C. M., & Lonigan, C. J. (2021). Relations between executive functions and academic outcomes in elementary school children: A meta-analysis. *Psychological Bulletin*, *147*(4), 329–351. https://doi.org/10.1037/bul0000322

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *19*(1), 99–120. https://doi.org/10.1016/j.ecresq.2004.01.002

*Strauss, L. I., Peterson, M. S., Kidd, J. K., Choe, J., Lauritzen, H. C., Patterson, A. B., Holmberg, C. A., Gallington, D. A., & Pasnak, R. (2020). Evaluation of patterning instruction for kindergartners. *Journal of Educational Research*, *113*(4), 292–302. https://doi.org/10.1080/00220671.2020.1806195

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis of the literature. *Review of Educational Research*, *76*(2), 249–274. https://doi.org/10.3102/00346543076002249

Takacs, Z. K., & Kassai, R. (2019). The efficacy of different interventions to foster children's executive function skills: A series of meta-analyses. *Psychological Bulletin*, *145*(7), 653–697. https://doi.org/10.1037/bul0000195

*Thorell, L. B., & Nyberg, L. (2008). The childhood executive functioning inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology*, *33*(4), 536–552. https://doi.org/10.1080/87565640802101516

*Träff, U., Olsson, L., Östergren, R., & Skagerlund, K. (2020). Development of early domain-specific and domain-general cognitive precursors of high and low math achievers in grade 6. *Child Neuropsychology*, *26*(8), 1065–1090. https://doi.org/10.1080/09297049.2020.1739259

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

van Aken, L., van der Heijden, P. T., Oomens, W., Kessels, R. P. C., & Egger, J. I. M. (2019). Predictive value of traditional measures of executive function on broad abilities of the Cattell–Horn–Carroll theory of cognitive abilities. *Assessment*, *26*(7), 1375–1385. https://doi.org/10.1177/1073191117731814

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

van der Ven, S. H. G. (2011). *The structure of executive functions and relations with early math learning*. Utrecht University.

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. *The British Journal of Educational Psychology*, *82*(Pt. 1), 100–119. https://doi.org/10.1111/j.2044-8279.2011.02035.x

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive interventions for students struggling in reading and mathematics. A practice guide*. Center on Instruction. ERIC.

*Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). I. Spatial skills, their development, and their links to mathematics. *Monographs of the Society for Research in Child Development*, *82*(1), 7–30. https://doi.org/10.1111/mono.12280

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W. (2021). Aggregate multiple effect sizes or outcomes within studies. *Metafor*. https://wviechtb.github.io/metafor/reference/aggregate.escalc.html

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Webb, S. L., Loh, V., Lampit, A., Bateman, J. E., & Birney, D. P. (2018). Meta-analysis of the effects of computerized cognitive training on executive functions: A cross-disciplinary taxonomy for classifying outcome cognitive factors. *Neuropsychology Review*, *28*(2), 232–250. https://doi.org/10.1007/s11065-018-9374-8

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). The Psychological Corporation.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, *44*(2), 575–587. https://doi.org/10.1037/0012-1649.44.2.575

Willoughby, M., Kupersmidt, J., Voegler-Lee, M., & Bryant, D. (2011). Contributions of hot and cool self-regulation to preschool disruptive behavior and academic achievement. *Developmental Neuropsychology*, *36*(2), 162–180. https://doi.org/10.1080/87565641.2010.549980

*Willoughby, M. T., Kupersmidt, J. B., & Voegler-Lee, M. E. (2012). Is preschool executive function causally related to academic achievement? *Child Neuropsychology*, *18*(1), 79–91. https://doi.org/10.1080/09297049.2011.578572

*Willoughby, M. T., Magnus, B., Vernon-Feagans, L., Blair, C. B., Investigators, F. L. P., & the Family Life Project Investigators. (2017). Developmental delays in executive function from 3 to 5 years of age predict kindergarten academic readiness. *Journal of Learning Disabilities*, *50*(4), 359–372. https://doi.org/10.1177/0022219415619754

Willoughby, M. T., Wylie, A. C., & Little, M. H. (2019). Testing longitudinal associations between executive function and academic achievement. *Developmental Psychology*, *55*(4), 767–779. https://doi.org/10.1037/dev0000664

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III test of cognitive abilities*. Riverside Publishing.

*Yang, X., Peng, P., & Meng, X. (2019). How do metalinguistic awareness, working memory, reasoning, and inhibition contribute to Chinese character reading of kindergarten children? *Infant and Child Development*, *28*(3), Article e2122. https://doi.org/10.1002/icd.2122

Yang, X., Yan, M., Ruan, Y., Ku, S. Y. Y., Lo, J. C. M., Peng, P., & McBride, C. (2022). Relations among phonological processing skills and mathematics in children: A meta-analysis. *Journal of Educational Psychology*, *114*(2), 289–307. https://doi.org/10.1037/edu0000710

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

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology*, *118*(1), 117–130. https://doi.org/10.1037/a0014657

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*(1), 297–301. https://doi.org/10.1038/nprot.2006.46

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354–360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

Received October 23, 2020
Revision received July 1, 2022
Accepted July 15, 2022 ■