Contemporary Educational Psychology 58 (2019) 276–287

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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

Theoretical and methodological implications of associations between executive function and mathematics in early childhood

Tutrang Nguyen[a,*], Robert J. Duncan[b], Drew H. Bailey[c]

[a] Curry School of Education & Human Development, University of Virginia, 405 Emmet Street South, Charlottesville, VA 22903, United States
[b] Department of Human Development and Family Studies and Public Health Graduate Program, Purdue University, 247 Hanley Hall, West Lafayette, IN 47907, United States
[c] School of Education, University of California, Irvine, 2072 Education, Irvine, CA 92697, United States

ARTICLE INFO

Keywords:
Executive function
Mathematics achievement
Development
Measurement models
Latent variables

ABSTRACT

Despite agreement about the importance of executive function (EF) for children's early math achievement, its treatment in correlational studies reflects a lack of agreement about the theoretical connection between the two. It remains unclear whether the association between EF and math operates through a latent EF construct or specific EF components. Specifying the correct measurement model has important theoretical implications for the predicted effects of EF interventions on children's math achievement. In the current study, we tested whether associations between EF and math operate via a latent EF factor, or via specific EF components using data from a large, nationally representative sample. We then replicated these same analyses with a meta-analytic database drawn from ten studies that collected measures of children's EF and math achievement. Our results lend support to explanations that a single EF factor accounts for most of the EF component-specific associations with math achievement. We discuss theoretical and methodological implications of these findings for future work.

1. Introduction

Executive function (EF) has been shown to be robustly associated with children's math achievement in early childhood (Best, Miller, & Naglieri, 2011; Blair, Ursache, Greenberg, & Vernon-Feagans, 2015; Clark, Pritchard, & Woodward, 2010; Espy et al., 2004; McClelland et al., 2007, 2014). Despite this consensus on the importance of EF to children's learning in academic contexts, its treatment in correlational studies of cognitive development reflects a lack of agreement about its relation to math achievement through (1) a single underlying EF construct and/or (2) the specific components of EF. Consequently, these theoretical distinctions have implications for the predicted effects of EF interventions on children's achievement.

Although EF is often conceptualized as a set of cognitive processes, including working memory, inhibitory control, and cognitive flexibility, individual differences in such processes appear to be largely psychometrically undifferentiated in correlational studies of young children's cognitive development (Wiebe, Espy, & Charak, 2008; Willoughby, Wirth, & Blair, 2012); although EF is usually found to be multidimensional in older children and adults (Lee, Bull, & Ho, 2013; Engelhardt, Briley, Mann, Harden, & Tucker-Drob, 2015; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000). Perhaps in part because of the theoretical distinctions but empirical overlap among EF measures in young children, sometimes non-experimental research in cognitive development includes models with academic outcomes regressed on specific components of EF (Bull, Espy, & Wiebe, 2008; Monette, Bigras, & Guay, 2011; Simanowski & Krajewski, 2017), and other times, achievement is regressed on a general EF construct (Bull, Espy, Wiebe, Sheffield, & Nelson, 2011; Hassinger-Das, Jordan, Glutting, Irwin, & Dyson, 2014; Schmitt, Geldhof, Purpura, Duncan, & McClelland, 2017). These differences in model specification reflects a lack of clarity about how the EF construct and its components relate to math achievement (for a recent discussion, see Rhemtulla, van Bork, & Borsboom, in press). Although these two model specifications are likely to produce some similar results – for example, they may explain similar amounts of total variance in children's academic outcomes – they imply very different types of causal relations among components of EF and academic achievement.

The current study addresses an unanswered question regarding the relation between EF and math achievement. Specifically, we investigate whether correlations among EF tasks and math achievement are consistent with the hypothesis that the association between EF and math achievement operates through specific components of EF, through a single latent EF factor, or both. Even under the assumption that EF

* Corresponding author.
E-mail addresses: tcn3rt@virginia.edu (T. Nguyen), duncan99@purdue.edu (R.J. Duncan), dhbailey@uci.edu (D.H. Bailey).

https://doi.org/10.1016/j.cedpsych.2019.04.002

Available online 11 April 2019

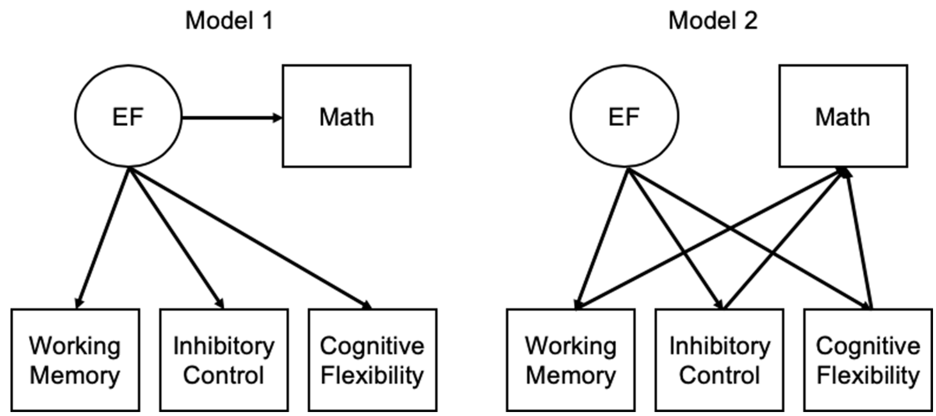

Fig. 1. Conceptual models for how the EF construct(s) affect/s math performance. Model 1 depicts latent EF as the primary influence on math and Model 2 depicts the specific components of EF as the primary influences on math.

consists of a single factor with multiple components, different modeling decisions reflect different theories about the causal relation between EF components and math, and yield substantively different results.

1.1. Specific components or underlying construct?

The current study aims to increase clarity about the robust associations between individual differences in young children's EF and math achievement. We consider two possibilities, displayed in Fig. 1: Model 1, which illustrates latent EF as the primary influence on math (represented in the model by a single common factor), and Model 2, which illustrates specific components of EF as the primary influences on math. Model 1, depicted in the left panel of Fig. 1, implies that the relations between children's EF components and their math achievement reflects the extent to which they measure the *same* construct. Model 2, depicted in the right panel of Fig. 1, implies that differences in the correlations between EF components and children's math achievement reflect the extent to which the EF components *differentially influence* children's math achievement. Is one of these models a more useful way of conceptualizing the influence of children's EF on math? We argue that previous research in this area has not yielded a clear answer and review the evidence for both possibilities below. We specifically focus on math achievement given the well-established relations between the two.

There are theoretical reasons to believe that EF components uniquely influence children's math achievement. For example, it may be that working memory helps children to temporarily store and manipulate relevant information, which allows them to learn complex arithmetic procedures. Cognitive flexibility may be uniquely predictive of children's math achievement because it helps children switch between different types of procedures during arithmetic practice. For example, adding fractions with the same denominator can be done without manipulating the denominator, whereas multiplying fractions with the same denominator requires squaring the denominator (Braithwaite, Pyke, & Siegler, 2017). Similarly, inhibitory control may help children suppress a dominant but incorrect solution strategy, such as completing operations by the order in which they are presented in the problem, rather than by the order of operations. This corresponds with the right panel of Fig. 1 (Model 2), which assumes that there are component-specific associations between EF and math that are unique from the shared variance among EF tasks. If mathematics achievement is influenced by specific EF components, then interventions that influence working memory, cognitive flexibility, or inhibitory control are predicted to have transfer effects to mathematics regardless of whether latent EF is affected. Consistent with this theory, some interventions that are designed to be EF component-specific, in addition to targeting other skills, have been found influence math achievement (Blair & Raver, 2014). Yet there are also some EF component-specific interventions that have produced null findings of transfer effects to math (Roberts et al., 2016).

In contrast, the left panel illustrated in Fig. 1 (Model 1) posits that factors common to EF components fully account for the relation between EF and children's math achievement. A latent EF factor can represent all the influences (e.g., shared developmental and/or cognitive processes) that are common to the components. For example, influences common to the three EF components may reflect general attention or goal maintenance skills (Blair, 2006; Duncan, Emslie, Williams, Johnson, & Freer, 1996; Garon, Bryson, & Smith, 2008). This may strike some as unlikely, given strong theoretical reasons to suggest the specific usefulness of working memory, inhibition, and switching for children's math learning. It could also be that there is a largely overlapping set of developmental factors (e.g., genes, home and schooling environments) that influence specific EF component development and children's math learning. However, the relations between EF components and children's math achievement are often found to be positive after statistically controlling for measures of such factors (Fuhs, Nesbitt, Farran, & Dong, 2014; Hughes, Ensor, Wilson, & Graham, 2010; Schmitt et al., 2017; Welsh, Nix, Blair, Bierman, & Nelson, 2010). Alternatively, the overlap between EF tasks labeled as measures of the same component may be due to factors other than shared cognitive processes (Oberauer, 2016). For example, the terms inhibition and task switching, while useful labels for many of the cognitive processes involved in children's math learning, may not accurately describe cognitive processes employed in all such tasks. It may be that the variation shared across inhibitory control tasks does not truly constitute a unitary cognitive process of "inhibition" but rather some more general set of developmental constraints on learning and performance common across a wide variety of tasks.

According to this model, children's math achievement is caused by influences common to the three EF components, and math achievement is not assumed to be any more distally related to EF components than they are to each other. Math achievement is an indicator, just as the three conceptualized EF components are, of the latent factor(s) general to EF. Of course, a major limitation of this model is that we do not know what the common factor in the model *is.* Several studies report low correlations between tasks designed to measure different EF components, as well as between tasks designed to measure the same EF component (Clark et al., 2010; Monette et al., 2011; Willoughby, Wirth, et al., 2012). Low loadings of EF tasks on latent factors are often observed (Espy et al., 2004; Lee et al., 2012; Miller, Müller, Giesbrecht, Carpendale, & Kerns, 2013; St. Clair-Thompson & Gathercole, 2006; Willoughby, Blair, Wirth, Greenberg, & The Family Life Project Investigators, 2012; Willoughby, Wirth, et al., 2012), and children's math achievement is sometimes found to correlate just as highly, or sometimes even higher, with EF tasks as EF tasks correlate with each other (Clark et al., 2010; Monette et al., 2011; Schmitt et al., 2017; Willoughby, Wirth, et al., 2012). The correlational literature on EF components generally finds similar average correlations between EF components and math achievement, with working memory as the strongest correlate of children's math achievement (see meta-analysis by Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013). However, along with the hypothesis that working memory is most causally important for math achievement, this finding is consistent with the hypothesis that working memory is simply the best indicator of common causes of both EF and math.

The implication of this version of the latent factor model (in which there are no paths from EF component residuals to math achievement) is that EF interventions will affect math achievement only if they affect the influences general to all EF tasks. To reiterate, these influences may be some combination of overlapping cognitive and developmental factors, or even the mutualistic co-development of EF components during earlier development (van Der Maas et al., 2006). If the influences general to EF tasks do not include individual cognitive processes themselves, but other characteristics of the child and context, then even training the EF components together may not be sufficient to facilitate broad transfer.

1.2. Challenges of EF construct conceptualization and measurement

It is difficult to disentangle whether the associations between children's EF and math achievement operate through specific components or through a single latent factor given the different conceptualizations and measures. Previous research has typically found that a single EF factor accounts for component-specific variations in early childhood (Allan & Lonigan, 2011, 2014; Bull et al., 2011; Clark, Sheffield, Wiebe, & Espy, 2013; Fuhs et al., 2014; Wiebe et al., 2008; Willoughby, Wirth, et al., 2012). Variation observed in the associations between EF components and math may be due differences in the strength to which they reflect an underlying EF factor. It could be that more valid and reliable measures of EF, or more complex tasks, load more highly onto the EF latent factor than less complex measures. Alternatively, certain EF measures may be related to math because the components that they tap (e.g., working memory) are causally related to math achievement. Despite the theoretical importance of this distinction, prior studies have not explicitly tested such models against each other. Across the wealth of studies examining the relation between EF and math achievement, there is also the issue of a lack of standardization of EF components and their measures and variation in the magnitude of associations between tasks within and across these constructs. The tasks used to measure a single component of EF are vastly different (Baggetta & Alexander, 2016; Morrison & Grammer, 2016). Some tasks have been used to assess multiple EF components such as Head-Toes-Knees-Shoulders (HTKS) that measure children's EF skills through gross motor responses (McClelland et al., 2014; Schmitt et al., 2017). Studies of EF and math achievement sometimes use a number of different measures for a particular EF component (see Carlson, 2005; Garon et al., 2008 for comprehensive lists). For example, to measure children's working memory, tasks such as Backwards Digit Span, Listening Recall, and the Auditory Working Memory subtest from the Woodcock-Johnson are used. Examples of measures of children's cognitive flexibility include card sorting, Shape School (switching), and Something is the Same. Examples of measures of children's inhibitory control include Stroop, Peg Tapping, Go/No-Go, Flanker, and Shape School (inhibition). It is difficult to administer all of these tasks to a large sample of preschoolers during a small number of testing sessions, and the tasks can vary widely across studies. Given the myriad of measures available, it is likely that children's performance can differ across a wide variety of tasks that purport to measure the same thing.

In addition to the number of tasks available to measure specific components of children's EF, the identification and labeling of many different individual components or processes that comprise EF has plagued its conceptualization. One example of this is evidenced by the differing conceptualizations of working memory. Major theories and psychometric studies have suggested that working memory is a multi-component model that can be further dissociated into several related but distinct processes (Baddeley & Hitch, 1974; Baddeley, 1986; Oberauer, Schulze, Wilhelm, & Süß, 2005; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Oberauer, Süß, Wilhelm, & Wittmann, 2003). Yet some studies conceptualize working memory to be just one of the components of EF and others use the term updating interchangeably with working memory (see Baggetta & Alexander, 2016 for a review; Lee, Ng, Bull, Pe, & Ho, 2011). One of the direct implications of this "conceptual clutter" (Morrison & Grammer, 2016) is the added complexity it presents for building models to rigorously test the relation between EF and math achievement, which consequently has implications for the findings from which we draw our conclusions. For the purposes of our study, we consider working memory and updating as interchangeable and as a component of EF along with cognitive flexibility and inhibitory control.

Further complicating the measurement and conceptualization of EF, the variation in performance on common tasks differs across age ranges (Carlson, 2005), and tasks that are intended to measure the same EF component often show weaker correlations than tasks intended to measure different components (Espy et al., 2004; Willoughby, 2016). Taken together, these issues make it difficult to triangulate across the literature to clarify the theoretical connections between EF, its subcomponents, and math achievement. As a way to summarize findings across the relevant published studies more systematically, the current study compares correlations across a set of studies to examine the consistency of associations between EF and children's math achievement with two different conceptual models: a model that implies the association is due to an underlying EF factor (i.e., Model 1) and a model that implies the association is due to EF component-specific performance (i.e., Model 2).

1.3. Current study

The current study asks a key question about the relations between EF and math in early childhood: does the association between EF and math achievement operate through specific components of EF, through a single latent EF factor, or both? We use data from a nationally representative sample of typically developing children across the U.S. along with a meta-analytic database of ten peer-reviewed studies on EF and math achievement in early childhood. We conduct a coordinated series of analyses using multiple datasets to determine which model best represents the observed data patterns. We first draw on a nationally representative, nonexperimental dataset, and then replicate these same analyses in a meta-analytic database of ten studies that collected measures of children's EF and math achievement. First, we present correlations between each EF component and math. Second, we use confirmatory factor analysis to estimate the loadings of each EF component onto the latent EF factor. Third, we correlate the factor loadings with the EF task-specific association with math to examine whether the EF components that load more strongly onto the latent EF factor are also more strongly related with math. Fourth, we compare the statistical fit of a model that allows for independent effects of EF components on math achievement (corresponding to Model 2 in Fig. 1), compared with a model that allows for an effect of an underlying EF factor on math achievement (corresponding to Model 1 in Fig. 1). Finally, for Model 1, which included a path from a latent EF factor to math achievement, we tested whether components of EF had significant residual correlations with math achievement. Because measures of EF administered during early childhood have widely been found to reflect a single factor (e.g., Fuhs et al., 2014; Wiebe et al., 2008), we hypothesize that the latent EF factor will largely account for the component-specific associations with math achievement.

2. Methods

2.1. Data

The current study uses data from multiple sources to address the research question. The first part of our study uses data from the federally-funded Early Childhood Longitudinal Study 2010-11 (ECLS-K: 2010). The ECLS-K: 2010 is an ongoing study focused on children's schooling experiences and it is the largest nationally representative sample to follow children from school entry into fifth grade. Unique to this dataset are three measures of EF that were not included in its precursor, the ECLS-K: 1998. Data collection for the ECLS-K: 2010 included parent interviews, surveys of teachers and school administrators, and direct child cognitive assessments. Data were collected for the full sample of children during the fall of kindergarten (wave 1), spring of kindergarten (wave 2), and spring of first grade (wave 4), and a subsample of children had their data collected for the fall of first grade (wave 3). Therefore, in the current study we only use waves 1, 2, and 4. The dataset includes 18,170 kindergartners (average age of 5.62 years) sampled from 1330 schools in 90 counties across the U.S (for more information on the ECLS-K: 2010, see Tourangeau, Nord, Lê, Sorongon, Hagedorn, Daly, & Najarian, 2014). See Table 1 for descriptive statistics of the sample included in our analyses.

The second part of the current study draws on correlation matrices from ten studies we identified that met the following criteria: (1) assessed each of the three components of EF – working memory, cognitive flexibility, and inhibitory control; (2) at least two of the EF measures were direct assessments; (3) assessed children using a standardized assessment of math achievement; (4) study samples must include children at or near school entry (i.e., children 3–6 years old); (5) did not select children on the basis of having a medical disorder or learning disability; (6) reported on original research (e.g., no commentaries or reviews); (7) published in an academic journal in a 10 year period spanning 2008–2017. The time period of our search was chosen because of the influx of literature examining the latent structure of EF (e.g. Garon et al., 2008; Wiebe et al., 2008).

A broad search of the literature by the first two authors was conducted using PsycINFO, PsycARTICLES, PubMed, and ERIC for articles that included the following search terms and their variants using the Boolean operator "AND": *executive function, executive control, mathematics skills, mathematics achievement, preschool, early childhood, elementary school*. To identify potential literature outside of indexed databases, we also used the Google Scholar search engine. The first author screened all titles and abstracts to ensure that the criteria listed above were met. All studies that appeared to meet the inclusion criteria were collected for full-text screening by the first and second authors for potential inclusion. Disagreements about study inclusion were resolved by consensus among all three authors. The majority of articles from this search included studies that examined one or two EF components, and were ruled out from inclusion. In all, our search process resulted in ten studies that examined the three EF components and met all other inclusion criteria.

Each study was independently coded by the first and second authors. A codebook was created specifically for this project in an Excel database. The codebook included relevant information about the studies, including all EF and mathematics measures used, sample size, age of the sample, and country in which the study took place. Any coding conflicts were discussed and resolved among all the authors. If there were multiple articles published using the same dataset, we selected the most recently published study using those data. Studies had to include correlations between the EF and math measures. If correlation matrices for the measures of interest were not presented in the original article, the first author contacted one of the authors of the included study to provide correlations for the measures at the earliest measurement time point. All contacted authors provided this information. See Table 2 for information on the ten studies included in the analyses. The measures of EF and math achievement associated with each data source are described in the sections below. All decisions in creating this meta-analytic database were made a priori and before conducting any analyses.

Table 1
Descriptive statistics for demographic variables, EF measures, and math by wave.

	N	% of sample/ Mean	SD	Range
Child characteristics collected at fall of kindergarten				
Female	18,132	48.80		
Black	2397	13.23		
Hispanic	4585	25.30		
Asian	1546	8.53		
Other	1107	6.11		
Non-English home language	2941	18.33		
Key measures collected at the fall of kindergarten				
Working memory	14,445	93.40	16.60	45–175
Inhibitory control	14,556	4.92	1.29	1–7
Cognitive flexibility	15,604	14.20	3.33	0–18
Math achievement	15,595	30.35	10.98	6.26–95.23
Key measures collected at the spring of kindergarten				
Working memory	17,124	94.98	17.13	40–175
Inhibitory control	15,925	5.08	1.30	1–7
Cognitive flexibility	17,149	15.14	2.80	0–18
Math achievement	17,143	43.40	11.51	6.26–81.12
Key measures collected at the spring of first grade				
Working memory	15,102	95.85	17.13	35–197
Inhibitory control	13,399	5.07	1.29	1–7
Cognitive flexibility	15,109	16.05	2.31	0–18
Math achievement	15,103	62.79	13.40	15.52–93.99

2.2. ECLS-K: 2010

2.2.1. EF measures

Children were assessed on three measures of EF at each wave with a combination of direct cognitive assessments and teacher reports: the Woodcock-Johnson Numbers Reversed subtest, the Dimensional Change Card Sort (DCCS), and the self-administered Children's Behavior Questionnaire (CBQ) – Short Form Inhibitory Control Subscale. Standard scores were used for Numbers Reversed and raw scores were used for the DCCS and CBQ.

Children's working memory was assessed with the Numbers Reversed subtest of the Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001). In this task, the assessor reads increasingly longer series of numbers to the child, up to a maximum of eight numbers, who must repeat the numbers in reverse order. For example, if presented with the sequence "4, 6, 8," the child would be expected to say "8, 6, 4." The assessment continued until the child gave three consecutive incorrect responses or completed all the number sequences. A maximum of 30 items could have been administered in all data collection rounds (5 two-digit number items; 5 three-digit number items; 4 four-digit number items; 4 five-digit number items; 4 six- digit number items; 4 seven-digit number items; and 4 eight-digit number items). Each item is scored as "correct" (i.e., the child correctly repeated the number sequence in reversed order), "incorrect" (i.e., the child did not correctly repeat the number sequence in reversed order), or "not administered" (i.e., the child was not administered the item because he or she did not answer enough items correctly to advance to the next item). This task involved 30 trials and raw scores ranged from 0 to 30. Raw scores can then be converted to standard scores. For the child who failed previous trials, his or her score for the later trials was imputed as 0. Children's standard score for this measure was used, which is normed to their age and created by the publisher ($M = 100$, $SD = 15$). The validity of the Numbers Reversed subtest has been extensively studied and established (LaForte, McGrew, & Schrank, 2014; McGrew & Woodcock, 2001). The technical manual

Table 2
Study descriptions from the meta-analytic database.

Study	N	Age (years)	Math measure	EF Component Measures		
				Working memory	Inhibitory control	Cognitive flexibility
Blair and Raver (2014) Tools of the Mind	759	5.72	WJ Applied Problems	Backward Digit Span	Flanker with Reverse Flanker	Dimensional Change Card Sort
Bull et al. (2008)	124	4.50	Performance Indicators in Primary School	Backward Digit Span	Shape School - Inhibition	Shape school - Switching
Clark et al. (2010)	104	4.00	WJ Math Fluency	BRIEF-P Working Memory[a]	Shape School - Inhibition	Shape school - Switching
Fuhs et al. (2014)	572	4.50	WJ Applied Problems	Backwards Digit Span	Peg Tapping	Dimensional Change Card Sort
Monette et al. (2011)	85	5.83	Wechsler Individual Achievement Test	Backward Word Span	Fruit Stroop	Card Sort
Schmitt et al. (2017)	424	4.69	WJ Applied Problems	WJ Auditory Working Memory	Simon Says	Card Sort Task
van der Ven et al. (2012)	211	6.42	Standardized Dutch national test	Digit Span Backwards	Animal Stroop	Sorting Task
Welsh et al. (2010) Head Start REDI	164	4.49	WJ Applied Problems	Backward Word Span	Peg Tapping	Dimensional Change Card Sort
Weiland and Yoshikawa (2013) Boston Pre-K Evaluation	2018	4.50	WJ Applied Problems	Backward Digit Span	Pencil Tapping	Dimensional Change Card Sort
Willoughby, Wirth, et al. (2012) Family Life Project	1036	5.80	WJ Applied Problems	Working Memory Span	Silly Sounds Stroop	Something's the Same

Note. WJ = Woodcock-Johnson.
[a] Indirect measure of an EF component.

reports correlations between Numbers Reversed and all other subtests in the Woodcock-Johnson Tests of Achievement, Cognitive Abilities, and Oral Language batteries. Correlations ranged from 0.23 to 0.40 in a sample of 3–5 year old children and 0.20 to 0.51 in a sample of 6–8 year old children. Concurrent validity evidence is also reported in the technical manuals, indicating moderate to high correlations between the subtest and other established measures of cognitive ability, such as the Wechsler Intelligence Scale for Children III (Wechsler, 1997) and Stanford-Binet: Fourth Edition Short Term Memory (Thorndike, Hagen, & Sattler, 1986). This measure has a reported reliability of 0.87 (Schrank, McGrew, & Woodcock, 2001; Woodcock et al., 2001).

Children were administered the DCCS task as an assessment of their cognitive flexibility (Frye, Zelazo, & Palfai, 1995). In this task, children sort cards into trays based on rules that change in the middle of the task. Children are presented with a series of picture cards that vary along three different dimensions – color, shape, and border – and are asked to sort each card into one of two trays depending on the sorting rule. The trays had a picture of a red boat and a blue rabbit. The first set of items were part of the Color Game, where the rule was to sort the cards by color (i.e., red or blue). For example, a blue boat card would be sorted into the blue rabbit tray. In the second game, the rules then switched and the child was asked to sort the cards by shape (i.e., rabbit or boat). In this scenario, a red rabbit card would be sorted into the blue rabbit tray. If the child correctly sorted four of the six cards in the Shape Game, then they moved on to the third and final game. In the Border Game, the sorting rule (color or shape) depended on whether the card had a black border around the edges. If the card had a black border, the child had to sort by color; if there was no border on the card, the child had to sort by shape. Items are scored as being "correct" (i.e., the card was sorted into the correct tray according to the sorting rule), "incorrect" (i.e., the card was sorted into the incorrect tray), or "not administered" because the child did not answer enough items correctly to advance to the next set of items. Children's scores are computed by combining scores on all three tasks, with a maximum score of 18 correct. This measure is reported to have shown high test-retest reliability of 0.90–0.94 (Beck, Schaefer, Pang, & Carlson, 2011; Weintraub et al., 2013; Zelazo, 2006). Weintraub et al. (2013) report a correlation of $r = -0.51$ for its convergent validity with the Delis-Kaplan Executive Function Inhibition measure (D-KEFS; Delis, Kaplan, & Kramer, 2001) and a correlation of $r = 0.14$ for its discriminant validity with the Peabody Picture Vocabulary Test – 4th edition (PPVT-4; Dunn & Dunn, 2012). This suggests that the DCCS has a relatively weak relationship with measures that tap different constructs. This measure of cognitive flexibility is widely used in studies predicting achievement in young children and is now a standardized measure in the NIH Toolbox (Weintraub et al., 2013; Zelazo, Anderson, Richler, Wallner-Allen, Beaumont, & Weintraub, 2013) and has also been used in studies also examining children's EF and achievement (Welsh et al., 2010).

To assess inhibitory control, teachers were asked to use the CBQ – Short Form Inhibitory Control Sub-Scale (Putnam & Rothbart, 2006). Teachers responded to how true or not a particular behavior is of the child on a 7-point Likert scale ranging from "extremely true" to "extremely untrue," with higher scores indicating that teachers rated individual children as demonstrating that particular behavior more frequently. This measure is comprised of six items. Sample items for this scale include whether the child "can wait before entering into new activities if s/he is asked to," "has trouble sitting still when s/he is told to," and "can easily stop an activity when s/he is told 'no.'" The item-level scores were computed as the mean of the items comprising the total score. Exploratory and confirmatory factor analyses have been used to substantiate the validity of the CBQ subscales (Putnam & Rothbart, 2006). Convergent validity was established from both reports of parent agreement and prediction of behavior patterns (Rothbart, Ahadi, Hershey, & Fisher, 2001). Although direct child assessments are preferred for measuring components of EF, the CBQ has been used as a reliable measure in prior studies examining EF and school readiness (e.g., Blair & Razza, 2007). Additionally, Allan, Hume, Allan, Farrington, and Lonigan (2014) showed in a meta-analysis that teacher ratings are actually the preferred type of measure for investigating relations between inhibitory control and academic achievement. The Cronbach's alpha for this assessment across all three waves was 0.87.

2.2.2. *Math achievement*

Math achievement was assessed with the ECLS-K math battery. This assessment was designed to measure skills in conceptual knowledge, procedural knowledge, and problem solving. The test consisted of questions on number sense, properties, and operations; measurement, geometry, and spatial sense; data analysis, statistics, and probability (measured with a set of simple questions assessing children's ability to read a graph); and pre-algebra skills such as identification of patterns.

Table 3
Correlations among EF components and Math in the ECLS-K Study.

Construct	1	2	3	4	5	6	7	8	9	10	11
1. Math T1											
2. WM T1	**0.55**										
3. IC T1	**0.31**	**0.23**									
4. CF T1	**0.34**	**0.27**	**0.18**								
5. Math T2	0.82	0.53	0.33	0.35							
6. WM T2	0.51	0.60	0.24	0.26	**0.57**						
7. IC T2	0.29	0.21	0.71	0.16	**0.34**	**0.24**					
8. CF T2	0.31	0.25	0.15	0.31	**0.36**	**0.27**	**0.17**				
9. Math T4	0.75	0.52	0.32	0.35	0.82	0.55	0.33	0.35			
10. WM T4	0.43	0.47	0.25	0.24	0.50	0.53	0.25	0.25	**0.55**		
11. IC T4	0.27	0.23	0.50	0.14	0.30	0.23	0.55	0.12	**0.30**	**0.24**	
12. CF T4	0.30	0.23	0.17	0.26	0.34	0.26	0.16	0.27	**0.38**	**0.29**	**0.16**

Note. All correlations significant at $p < .001$. Bolded are the within-wave correlations included in the analyses. T1 is the fall of kindergarten. T2 is the spring of kindergarten. T4 is the spring of first grade. WM = working memory. IC = inhibitory control. CF = cognitive flexibility.

The Cronbach's alphas for this assessment across the three waves were large (0.92, 0.94, and 0.93, respectively).

2.3. Meta-analytic database

2.3.1. EF measures

Each of the studies included in the present analyses had at least one measure for each of the three EF factors. For the studies that had multiple measures for each of the three components of EF, we compiled a list of measures to prioritize in the current study prior to requesting the correlations of those measures from the authors of the original studies. As an example, for measures of working memory we prioritized the Backwards Digit Span measure over the Listening Recall measure because the former measure was more closely related to the Numbers Reversed measure we used in the ECLS-K: 2010. We also elected to use a teacher-reported working memory measure from the Behavior Rating Inventory of Executive Function for Preschool (BRIEF-P) since that was the sole measure the authors used for that particular EF component in the study (Clark et al., 2010).

2.3.2. Math achievement

The ten studies selected for inclusion in our analyses each measured children's math achievement through direct assessment. Similar to the measures of EF in the selected studies, some studies had multiple measures of math achievement so we followed the same procedure for selecting a single measure from each study to include in our analyses. Collectively, the math achievement measures assessed children's number knowledge (counting, comparing and ordering of objects, beginning arithmetic), geometry and spatial thinking, and problem solving. Examples of measures of children's math achievement included the ECLS-K math battery, the Woodcock-Johnson Applied Problems subtest, the Woodcock-Johnson Math Fluency subtest, the Woodcock-Johnson Quantitative Concepts subtest, and the math composite from the Wechsler Individual Achievement Test. One study used a standardized national Dutch math assessment.

2.4. Analytic strategy

All analyses were conducted using either Stata 14 (StataCorp, 2015) or M*plus* 7 (Muthén & Muthén, 1998–2012). A series of analyses were conducted with the ECLS-K dataset, the meta-analytic dataset, and the study-specific correlations intended to inform our central research question on whether the association between children's EF and math achievement operates through specific components of EF, a single latent EF factor, or both. All results for the meta-analytic datasets are shown for each study separately.

First, correlation matrices are reported for each component of EF and math by dataset. We present the results for each wave of the ECLS-K dataset and for the meta-analytic study-specific correlations. Second, factor analyses were conducted to estimate the loadings of each EF component onto the latent EF factor for each study. Third, the factor loadings were then correlated with the task-specific association with math achievement to estimate to what extent the EF components that load more strongly onto latent EF are also more strongly associated with math. Of primary interest, we estimated models in which EF-math associations were via latent EF (Fig. 1, Model 1) or via components of EF (Fig. 1, Model 2) and compared their levels of statistical fit and theoretical interpretations. The BIC was compared across model specifications because one of the models (i.e., EF component explanation) was saturated and had no degrees of freedom for other model fit comparison tests. For the ECLS-K dataset, these models were estimated on the raw data at each time point. These models included a maximum likelihood robust estimator, which does not assume normal distributions of outcome variables (Muthén & Muthén, 1998–2012), and clustered data based on school assignment at the given time point. For the meta-analytic dataset, these models were estimated using a maximum likelihood estimator and were based on the correlation matrices only (i.e., the MLR estimator was not possible without the actual data).

3. Results

3.1. Correlations among EF components and math

Correlations among EF components and math performance at waves 1, 2, and 4 in the ECLS-K dataset are reported in Table 3. All correlations were significant at $p < .001$. Within each specific wave, math was more closely associated with the specific components of EF (Time 1: $rs = 0.31$–0.55; Time 2: $rs = 0.34$–0.57; Time 4: $rs = 0.30$–0.55) than the components of EF were with each other (Time 1: $rs = 0.18$–0.27; Time 2: $rs = 0.17$–0.27; Time 4: $rs = 0.16$–0.29). All study-specific correlations are available in Appendix Tables 1–10 in the online supplementary materials.

3.2. EF factor loadings

All results for factor loadings onto a latent EF are presented in Table 4. No clear pattern of results emerged for the factor loadings across datasets. In other words, knowing the specific EF component is not informative about the magnitude of the loading. Excluding one study that had convergence issues, working memory had the strongest loading 4 times, inhibitory control 5 times, and cognitive flexibility 4 times (one was a tie between working memory and inhibitory control).

Under the assumption that the associations of specific tasks are related to math due to the latent EF factor, the expectation would be that the size of the loadings onto latent EF would be associated with the tasks correlation with math. The factor loadings and the EF task-math

Table 4
Loadings for each EF component on latent EF.

Study	EF Factor Loadings		
	WM λ	IC λ	CF λ
ECLS-K T1	0.60	0.39	0.65
ECLS-K T2	0.64	0.39	0.37
ECLS-K T4	0.65	0.43	0.44
Blair and Raver (2014)	0.37	0.41	0.77
Bull et al. (2008)	0.46	0.53	0.84
Clark et al. (2010)	0.29[a]	1.04[a]	0.17[a]
Fuhs et al. (2014)	0.36	0.83	0.50
Monette et al. (2011)	0.62	0.55	0.42
Schmitt et al. (2017)	0.52	0.63	0.48
van der Ven et al. (2012)	0.35	0.59	0.66
Weiland and Yoshikawa (2013)	0.57	0.71	0.64
Welsh et al. (2010)	0.57	0.62	0.41
Willoughby, Wirth, et al. (2012)	0.42	0.42	0.33

[a] Convergence issues (IC λ > 1).

correlation are correlated at $r = 0.47$ (see Fig. 2). In other words, the tasks that load more strongly onto latent EF were more highly correlated with math achievement. This trend holds when restricted only to tasks intended to measure the same EF component (see Fig. 3). For the study specific patterns of these relations, see Appendix Fig. 1 in the online supplementary materials.

3.3. Latent EF construct or components of EF accounting for the EF-Math Link?

The primary analyses used to address the research question are comparing model fit statistics of two hypothetical model specifications for EF associations with math. In Model 1, latent EF is assumed to be the primary influence of math and not the specific components. In Model 2, the specific components are assumed to fully mediate the effects of the latent EF factor on math. Model 2 is completely saturated (i.e., no degrees of freedom); therefore, the BIC is used for the statistical comparison. The results for all models and studies are included in Table 5. For the three time points in the ECLS-K dataset, Model 1 was preferred each time and fit the data exceptionally well: CFI/TLI = 1.00, RMSEA = 0–0.02, and SRMR = 0–0.01. In other words, for the ECLS-K dataset, the associations between EF and math are consistent with a model in which a single EF factor influences math achievement across all waves. The study-specific results from our meta-analysis were less conclusive. Seven of the ten samples had smaller BIC values for Model 1, two of the ten had smaller BIC values for Model 2, and one was equal.

The path estimates from Model 1, Model 2, and Model 1 with residuals included between each specific component and math are shown in Table 6. For the three time points in the ECLS-K dataset, Model 1 showed that latent EF had estimated βs = 0.85–0.88. Model 2 showed that working memory was the most closely associated component of EF, βs = 0.44–0.46, although only time point 1 found working memory to have a positive residual correlation beyond the latent EF factor. Across all of the specific meta-analytic studies, latent EF was a very robust predictor of math, with no clear pattern in terms of residual correlations. Furthermore, in Model 2 across the specific meta-analytic studies, there were no clear patterns for which component was most predictive of math. To test the robustness of our findings to the inclusion of demographic covariates, we ran two additional analyses including age, gender, race/ethnicity, and socioeconomic status as covariates in the ECLS-K Model 1 and Model 2 across all three time points. The results were not substantively different from our main analysis models. These estimates are presented in the online supplementary materials in Appendix Table 11.

4. Discussion

The current study examined whether correlations among EF tasks and math achievement are consistent with the hypothesis that the association between EF and math achievement operates through specific components of EF, through a single latent EF factor, or both. We use a large-scale database as well as meta-analytic techniques to pursue this methodological and conceptual question. The correlations between EF tasks and the close associations between EF and math achievement largely replicate previous research findings (e.g., Bull et al., 2011; Clark et al., 2010, 2013; Fuhs et al., 2014; Willoughby, Blair, et al., 2012; Willoughby, Wirth, et al., 2012; Wiebe et al., 2008). Most of our models and analyses found support for Model 1, namely, that the underlying latent EF factor largely accounts for component specific associations with math. The standardized estimates of the latent EF effect on math in Model 1 can statistically be interpreted as factor loadings. Notably, when considering the data from this perspective, the factor loading for math on latent EF would be the largest loading on the EF factor in 9 of

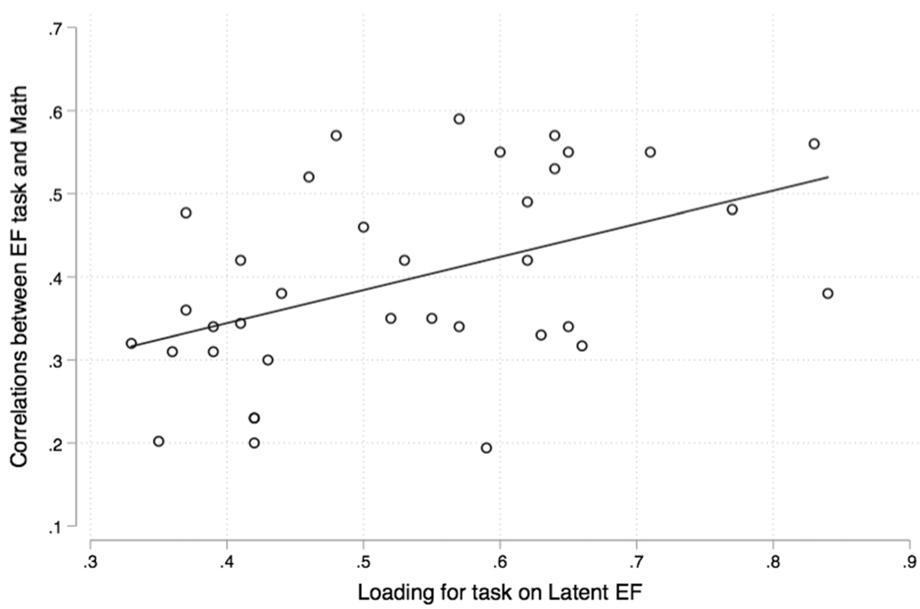

Fig. 2. Correlation between task-specific EF loading and task-specific math correlation (r = 0.47). Clark et al. (2010) excluded due to factor loading issue (IC λ > 1).

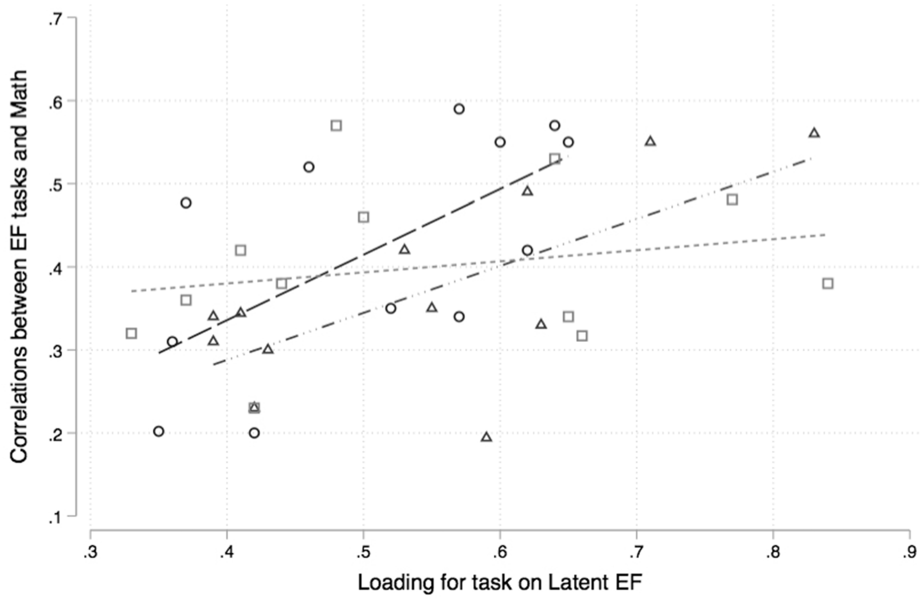

Fig. 3. *Component specific correlations between task-specific EF loading and task-specific math correlation.* Working memory, $r = 0.64$, is the triangles with dash-dot line. Inhibitory control $r = 0.69$, is the circles with long dashes. Cognitive flexibility, $r = 0.23$, is the squares with short dashes. Clark et al. (2010) excluded due to factor loading issue (IC $\lambda > 1$).

the 12 models (see Tables 4 and 6; excluding the one study that had a factor loading larger than one). Additionally, math correlations with EF tasks were consistently higher than those among EF tasks. This likely partially reflects the modest reliability of EF tasks (Willoughby, Kuhn, Blair, Samek, & List, 2017); however, the relatively high factor loadings of EF component tasks compared with their residual correlations with math achievement require a more substantive theoretical explanation. And finally, EF loadings were positively correlated with task-specific correlations to math. Thus, the ECLS-K and the meta-analytic sample results are largely consistent with the predictions of Model 1, in which a single EF factor accounts for EF component-specific associations with math achievement.

Although examining the factor structure of latent EF was not the focus of the current study, our results are consistent with previous work finding a single underlying factor underlying individual differences in EF (e.g., Bull et al., 2011; Clark et al., 2013; Wiebe et al., 2008; Willoughby, Blair, et al., 2012; Willoughby, Wirth, et al., 2012), suggesting that variation in the associations between EF tasks and math could be due to differences in their associations with an EF factor. However, results from Model 1 cannot rule out the possibility of small but positive EF component effects. Our findings, although cross-sectional, are consistent with a recent longitudinal investigation of the relation between EF and math achievement by Nesbitt, Fuhs, and Farran (2019) and Willoughby, Wylie, and Little (2018) who both found a very high correlation between the latent factors influencing math achievement and EF across early childhood (see Table 6, column 1). We conducted cross-sectional analyses to address conceptual and theoretical questions of how EF and math relate, whether that is due to an underlying factor or specific components. It is important to note here that we are not addressing how one or the other predicts growth in the other construct. Some previous longitudinal studies regress later math on earlier EF, some with (e.g., Blair & Raver, 2014; Bull et al., 2008; Fuhs et al., 2014; Welsh et al., 2010) and others without (e.g., Clark et al., 2010; van der Ven, Kroesbergen, Boom, & Leseman, 2012; Weiland & Yoshikawa, 2013) statistically controlling for prior math achievement. Both methods do some to reduce threats to internal validity. Using later achievement as a criterion precludes the possibility that the outcome fully causes the predictors, although this makes little difference in the absence of some event that directly influences EF and not math (Foster, 2010). Controlling for the autoregressor controls for some but not all possible confounds influencing both EF and math across time (Bailey, Duncan, Watts, Clements, & Sarama, 2018; Hamaker, Kuiper, & Grasman, 2015). All three approaches – cross-sectional analysis, longitudinal regression, and longitudinal regression

Table 5
Comparison of model fit indices for Latent EF predicting Math (Model 1) and specific components of EF predicting Math (Model 2).

		Model 1						Model 2		
Study	N	Chi-square	BIC	CFI/TLI	RMSEA	SRMR	R^2	BIC	R^2	Preferred Model
ECLS-K T1	15,831	(2) 11.30**	362,822	1.00/1.00	0.02	0.01	0.72	362,829	0.62	1
ECLS-K T2	17,312	(2) 1.47	404,615	1.00/1.00	0.00	0.00	0.77	404,633	0.40	1
ECLS-K T4	15,198	(2) 1.43	353,746	1.00/1.00	0.00	0.00	0.73	353,764	0.39	1
Blair and Raver (2014)	289	(2) 6.44*	3153	0.98/0.93	0.09	0.03	0.76	3158	0.39	1
Bull et al. (2008)	104	(2) 8.76*	1154	0.92/0.75	0.18	0.05	0.56	1154	0.37	Equal
Clark et al. (2010)	104	(2) 1.66	1182	1.00/1.02	0.00	0.03	0.66	1189	0.31	1
Fuhs et al. (2014)	562	(2) 2.67	6013	1.00/1.00	0.02	0.01	0.63	6023	0.40	1
Monette et al. (2011)	85	(2) 0.16	973	1.00/1.16	0.00	0.01	0.41	982	0.23	1
Schmitt et al. (2017)	409	(2) 15.48***	4435	0.95/0.86	0.13	0.04	0.64	4432	0.38	2
van der Ven et al. (2012)	211	(2) 1.71	2381	1.00/1.01	0.00	0.02	0.19	2390	0.12	1
Weiland and Yoshikawa (2013)	2018	(2) 35.17***	20,641	0.99/0.96	0.09	0.02	0.77	20,621	0.48	2
Welsh et al. (2010)	164	(2) 3.37	1815	0.99/0.96	0.07	0.03	0.63	1822	0.35	1
Willoughby, Wirth, et al. (2012)	1058	(2) 7.61*	11,850	0.98/0.93	0.05	0.02	0.43	11,856	0.15	1

Note. Meta-analysis results did not include data from the ECLS-K Study. Model 2 was completely saturated, thus only the BIC is reported. Preferred model is based on the smaller BIC from the two models.

Table 6
Comparison of path estimates for Latent EF predicting Math (Model 1) and specific components of EF predicting Math (Model 2).

	Model 1	Model 1 with residuals added			Model 2		
Study	Latent EF	WM residual	IC residual	CF residual	WM	IC	CF
ECLS-K T1	0.85***	0.13***	−0.00	−0.04*	0.46***	0.18***	0.19***
ECLS-K T2	0.88***	0.03	0.01	−0.02	0.46***	0.19***	0.21***
ECLS-K T4	0.85***	−0.01	−0.01	0.02	0.44***	0.17***	0.23***
Blair and Raver (2014)	0.87***	0.23**	−0.07	−0.67[a]	0.36***	0.19***	0.32***
Bull et al. (2008)	0.75***	0.30**	0.04	−1.36[a]	0.46***	0.28**	0.10
Clark et al. (2010)	0.81***	0.26	−5.86[a]	0.02	0.38***	0.28**[b]	0.11
Fuhs et al. (2014)	0.79***	0.01	−0.53[a]	0.09	0.14***	0.41***	0.27***
Monette et al. (2011)	0.64***	0.08	0.03	−0.05	0.32**	0.22*	0.10
Schmitt et al. (2017)	0.80***	−0.02	−0.30[a]	0.37***	0.19***	0.12**	0.48***
van der Ven et al. (2012)	0.44***	0.08	−0.18	0.09	0.13	0.07	0.26***
Weiland and Yoshikawa (2013)	0.88***	0.20***	−0.21**[a]	−0.08	0.38***	0.28***	0.27***
Welsh et al. (2010)	0.79***	−0.29[a]	−0.00	0.18*	0.14*	0.37***	0.30***
Willoughby, Wirth, et al. (2012)	0.66***	−0.12	−0.06	0.17***	0.13***	0.17***	0.28***

Note. Latent EF column is from model without any residuals. Residuals for Model 1 were added one at a time and estimated independently of any other residual.
[a] The standardized effect of latent EF on math was greater than 1.
[b] The loading on latent EF was greater than 1.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

with an autoregressor – are susceptible to residual confounding, and estimates from all three kinds of models may be prone to bias from factors common to EF if they are not modeled explicitly. Notably, two longitudinal studies that attempt to model factors influencing math and EF scores similarly over time estimated effects of EF on subsequent math achievement in the range of 0–0.10 (Nesbitt et al., 2019; Willoughby et al., 2018), estimates not far from the median residual correlations between two of the components and math achievement in Model 1 for our study (see Table 6).

4.1. Implications

Selecting among the models considered in the current study has important implications for thinking about the design of early childhood EF interventions focused on promoting transfer to academic domains in addition to improving children's EF, math achievement, or both. Our results suggest that interventions will be effective to the degree that they can improve the mechanisms that influence factors general to all components as opposed to specific components. Interventions targeting specific EF components (e.g., card-sort task performance or working memory task performance alone), and not factors common to EF, may not reliably transfer to mathematics. Currently, the intervention literature does not provide us with clear answers as to whether there is reliable transfer from EF to math achievement. Jacob and Parkinson (2015) demonstrate in their meta-analysis that there is little evidence that EF interventions alone can boost children's math achievement. Non-experimental studies with the strongest statistical controls show weak predictive power of EF to math achievement, and EF training has shown inconsistent evidence of transfer to math achievement. Processes underlying the positive effects of EF interventions on math achievement likely vary across interventions and samples, and may be attributable to changes in EF, changes to children's learning environments independent of EF, or some combination of both (Jacob & Parkinson, 2015).

The argument that EF interventions should target factors general to EF components rather than the specific EF components in order to observe transfer is not new (Diamond, 2012), but it is unclear what approaches are best for boosting the influences general to performance across EF tasks. Diamond (2012) suggests that early childhood EF interventions will be most effective when they are increasingly challenging and sustained over time. One-time intervention boosts to improve children's EF will not be enough to have substantial or lasting effects on children's math achievement. A potentially promising approach for interventions will be those that are broadly beneficial for EF components that are also likely to be beneficial for children's math skills. This could include multifaceted interventions that directly target both domain general cognitive skills and early math knowledge, or those that broadly target aspects that are likely common to both domains.

To the extent that factors general to EF tasks can be influenced by early intervention, such interventions are predicted to have the largest effects on children's mathematics achievement. We propose that both intervention and correlational research on EF will progress more efficiently if predictions from the correlational literature are incorporated into future intervention designs. Further, we must also carefully consider the mechanisms through which EF interventions influence children's achievement (via a higher-order factor or through specific components). We recommend that evaluations of EF-targeted interventions should model the effects of the treatment on EF tasks in addition to a latent EF factor simultaneously to determine which set of EF skills the intervention is affecting (i.e., specific EF components, underlying EF construct, or both; for discussion of these issues, see Protzko, 2017).

4.2. Limitations and future directions

Several issues make the meaning of the latent EF factor unclear. First is task impurity of EF assessments in early childhood (e.g., Miyake & Friedman, 2012), and the types of EF tasks used in the study. The EF tasks chosen for this study likely tap other EF components in addition to the EF component specifically targeted by the task. We were unable to include all of the EF tasks that were available in our meta-analytic database in our models, and the range of EF tasks selected for each component should be another focus for future research. Tasks that measure multiple components of EF but are represented by a single indicator may produce results that appear to be a single component, but in fact consist of multiple components. If so, perhaps latent EF is not unitary but a mix of the components that tasks intend to measure. Some have proposed that this mix of components could be some combination of causes that affect a variety of cognitive skills through a single causal pathway (Rhemtulla et al., in press; Tucker-Drob, 2013; Willoughby, Blair, & The Family Life Project Investigators, 2015), an overlap of cognitive processes (Kovacs & Conway, 2016), or other causes apart from shared cognitive processes (Oberauer, 2016).

Second, it might just be the case that the two models tested in this study are overly simplified and incorrect. A better understanding of EF components, their co-development, and their causal structure would improve the potential for developing useful theories about the co-development of EF and academic skills. For example, if EF components are related in a formative, rather than reflective way, it is possible that neither of the types of models considered in the current study yields causally informative estimates (Borsboom, Mellenbergh, & Van Heerden, 2003; Kline, 2006; Rhemtulla et al., in press; van Der Maas et al., 2006; Willoughby et al., 2015). Both of these possibilities demonstrate the need for more experimental psychometric validation of EF components, in which subskills are manipulated and transfer to other components and "upward" to EF are evaluated (Protzko, 2017). However, existing evidence of small transfer effects from EF training across tasks (Diamond, 2012) suggests that latent variable models may be a useful starting place. Our findings most directly point to the importance of specifying a measurement model for EF in future experimental research. If a common factor influences math achievement via pathways that do not go through EF components, then non-experimental studies will need to account for this possibility. Statistically controlling for two other EF components in order to isolate the effect of a third component does not mean that latent EF is also controlled for (see Schmidt, 2017 for a discussion).

Further, our analyses were not well designed to rule out small additional effects of the EF components on math achievement. Even if Model 1 is the correct model, our results do not convincingly rule out the hypothesis that the influences of specific EF components on math achievement are small and positive. We found some support for Model 2, implying relatively large direct effects of EF components on math achievement. Interventions and non-experimental longitudinal studies that differentiate between *changes* in general and specific cognitive skills would help to test this hypothesis more directly.

Finally, the study is limited by the quantity and quality of measures available in these datasets. Task impurity is a potential problem with EF assessments in early childhood (Miyake & Friedman, 2012). Most of the EF tasks used in the study likely capture some aspects of other EF components (i.e., inhibitory control, working memory, or cognitive flexibility) in addition to the EF component targeted by the task. Additionally, we analyzed just one task intended to tap each of the EF components. Future studies should attempt to test whether EF components show more discriminant validity during this age range when more measures are used to assess each component. However, in young children, discriminant validity is questionable even when there are many measures used per construct. Further, among the studies that do have multiple measures per construct, there has not been strong evidence for EF component-specific factors influencing math in early childhood (e.g., Lee et al., 2012; Monette et al., 2011; van der Ven et al., 2012; Willoughby, Wirth, et al., 2012).

Including some indirect assessments in our study along with direct tasks may exacerbate the task impurity problem. Our meta-analytic sample includes nine studies that have direct measures of each EF component. Seven of the nine studies provide support for Model 1 of latent EF as the primary influence on math, similar to our results from the ECLS-K. Of course, we cannot completely rule out this measurement issue given that two of the studies in our meta-analysis support Model 2 of specific EF components as the primary influence on math (Schmitt et al., 2017; Willoughby, Wirth, et al., 2012), and one study (Clark et al., 2010) used an indirect measure, which presented convergence issues. To the extent measurement error contributes to these results, this would also be true for all studies that combine EF indicators into latent variables, a relatively common practice observed in our meta-analytic articles. The major contribution of this paper is not to show a large correlation between latent EF and math achievement, but to show that a common factor model fits as well or better than models commonly used to estimate effects of EF components on math achievement, which has very different theoretical implications.

Acknowledgements

We are grateful to the Institute of Education Sciences (IES) for supporting this work through grant R305B120013 awarded to the University of California, Irvine and R305B170002 awarded to the University of Virginia. D.H. Bailey is supported by a Jacobs Fellowship. Research reported in this publication was also supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (NIH) under Award Number P01-HD065704 to the University of California, Irvine. The content is solely the responsibility of the authors and does not necessarily represent the official views of IES, the U.S. Department of Education, NIH, or the Jacobs Foundation. We thank the authors whose studies are included in our meta-analytic dataset for sharing their correlation matrices with us. We also thank Tyler Watts and two anonymous reviewers for their helpful comments on previous versions of this manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cedpsych.2019.04.002.

References

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.

Allan, N. P., & Lonigan, C. J. (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. *Developmental Psychology, 47*(4), 905–915.

Allan, N. P., & Lonigan, C. J. (2014). Exploring dimensionality of effortful control using hot and cool tasks in a sample of preschool children. *Journal of Experimental Child Psychology, 122*, 33–47.

Baddeley, A. D. (1986). *Working memory.* Oxford, UK: Oxford University Press.

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

Bailey, D. H., Duncan, G. J., Watts, T., Clements, D., & Sarama, J. (2018). Risky business: Correlation and causation in longitudinal studies of skill development. *American Psychologist, 73*, 81–94. https://doi.org/10.1037/amp0000146.

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education, 10*(1), 10–33.

Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test–retest reliability. *Journal of Cognition and Development, 12*(2), 169–193.

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.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences, 29*(2), 109–125.

Blair, C., & Raver, C. C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PLoS ONE, 9*(11), https://doi.org/10.1371/journal.pone.0112393.

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.

Blair, C., Ursache, A., Greenberg, M., & Vernon-Feagans, L. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter–word knowledge in the early elementary grades. *Developmental Psychology, 51*(4), 459–472.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219.

Braithwaite, D. W., Pyke, A. A., & Siegler, R. S. (2017). A computational model of fraction arithmetic. *Psychological Review, 124*(5), 603–625.

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.

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.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616.

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.

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*, 662–677. https://doi.org/10.1111/j.1467-8624.2012.01854.x.
Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system.* Odessa, FL: Psychological Assessment Resources.
Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science, 21*(5), 335–341.
Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30*(3), 257–303.
Dunn, L. M., & Dunn, D. M. (2012). *Peabody Picture Vocabulary Test, (PPVT-4).* Johannesburg: Pearson Education Inc.
Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P., & Tucker-Drob, E. M. (2015). Genes unite executive functions in childhood. *Psychological Science, 26*(8), 1151–1163.
Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*(1), 465–486.
Foster, E. M. (2010). Causal inference and developmental psychology. *Developmental Psychology, 46*(6), 1454–1480.
Friso-van den Bos, I., van der Ven, S. H., Kroesbergen, E. H., & van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29–44.
Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*(4), 483–527.
Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology, 50*(6), 1698–1709.
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.
Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116.
Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology, 118*, 78–92.
Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology, 35*, 20–36.
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*, 512–552. https://doi.org/10.3102/0034654314561338.
Kline, R. B. (2006). Reverse arrow dynamics: Formative measurement and feedback loops. In G. R. Hancock, & R. O. Mueller (Eds.). *Structural equation modeling: A second course* (pp. 43–68). Greenwich, CT: Information Age Publishing.
Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*(3), 151–177.
LaForte, E. M., McGrew, K. S., & Schrank, F. A. (2014). WJ IV technical abstract (Woodcock-Johnson IV assessment service bulletin no. 2). Rolling Meadows, IL: Riverside.
Lee, K., Bull, R., & Ho, R. M. (2013). Developmental changes in executive functioning. *Child Development, 84*(6), 1933–1953.
Lee, K., Ng, S. F., Bull, R., Pe, M. L., & Ho, R. H. M. (2011). Are patterns important? An investigation of the relationships between proficiencies in patterns, computation, executive functioning, and algebraic word problems. *Journal of Educational Psychology, 103*(2), 269–281.
Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology, 82*(1), 82–99.
Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*(1), 59–80.
McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947–959. https://doi.org/10.1037/0012-1649.43.4.947.
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*. https://doi.org/10.3389/fpsyg.2014.00599.
McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III Technical Manual.* Itasca, IL: Riverside.
Miller, M. R., Müller, U., Giesbrecht, G. F., Carpendale, J. I., & Kerns, K. A. (2013). The contribution of executive function and social understanding to preschoolers' letter and math skills. *Cognitive Development, 28*(4), 331–349.
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.
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.
Monette, S., Bigras, M., & Guay, M. C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *Journal of Experimental Child Psychology, 109*(2), 158–173.
Morrison, F. J., & Grammer, J. K. (2016). Conceptual clutter and measurement mayhem: Proposals for cross-disciplinary integration in conceptualizing and measuring 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. 327–348). Washington, DC: American Psychological Association.
Muthén, L. K., & Muthén, B. O. (1998–2012). Mplus user's guide (7th ed.). Los Angeles, CA: Muthén & Muthén.
Nesbitt, K. T., Fuhs, M. W., & Farran, D. C. (2019). Stability and instability in the co-development of mathematics, executive function skills, and visual-motor integration from prekindergarten to first grade. *Early Childhood Research Quarterly, 46*(1), 262–274.
Oberauer, K. (2016). Parameters, not processes, explain general intelligence. *Psychological Inquiry, 27*(3), 231–235.
Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. M. (2005). Working memory and intelligence—Their correlation and their relation: A comment on Ackerman, Beier, and Boyle. *Psychological Bulletin, 131*(1), 61–65.
Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—Facets of a cognitive ability construct. *Personality and Individual Differences, 29*(6), 1017–1045.
Oberauer, K., Süß, H. M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*(2), 167–193.
Protzko, J. (2017). Effects of cognitive training on the structure of intelligence. *Psychonomic Bulletin & Review, 24*(4), 1022–1031.
Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*(1), 102–112.
Rhemtulla, M., van Bork, R., & Borsboom, D. (2019). Worse than measurement error: Consequences of inappropriate latent variable models. *Psychological Methods* (in press).
Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P. J., Gathercole, S., Gold, L., ... Wake, M. (2016). Academic outcomes 2 years after working memory training for children with low working memory. *JAMA Pediatrics, 170*(5), https://doi.org/10.1001/jamapediatrics.2015.4568 e154568-e154568.
Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development, 72*(5), 1394–1408.
Schmidt, F. L. (2017). Beyond questionable research methods: The role of omitted relevant research in the credibility of research. *Archives of Scientific Psychology, 5*(1), 32–41.
Schmitt, S. A., Geldhof, G. J., Purpura, D. J., Duncan, R. J., & McClelland, M. M. (2017). Examining the relations between executive function, math, and literacy during the transition to kindergarten: A multi-analytic approach. *Journal of Educational Psychology, 109*(8), 1120–1140. https://doi.org/10.1037/edu0000193.
Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). *Technical Abstract (Woodcock-Johnson III Assessment Service Bulletin No.2).* Itasca, IL: Riverside Publishing.
Simanowski, S., & Krajewski, K. (2017). Specific preschool executive functions predict unique aspects of mathematics development: A 3-year longitudinal study. *Child Development,* 1–18. https://doi.org/10.1111/cdev.12909.
StataCorp (2015). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.
St. Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*(4), 745–759.
Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet intelligence scale* (4th ed.). Itasca, IL: Riverside Publishing.
Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., Hagedorn, M. C., Daly, P., & Najarian, M. (2014). Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLSK: 2011), User's Manual for the ECLS-K:2011 Kindergarten Data Files and Electronic Codebook. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
Tucker-Drob, E. M. (2013). How many pathways underlie socioeconomic differences in the development of cognition and achievement? *Learning and Individual Differences, 25*, 12–20.
van Der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842.
van der Ven, S. H., Kroesbergen, E. H., Boom, J., & Leseman, P. P. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology, 82*(1), 100–119.
Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, TX: Psychological Corporation.
Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102*(1), 43–53.
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.
Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112–2130. https://doi.org/10.1111/cdev.12099.
Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., ... Gershon, R. C. (2013). Cognition assessment using the NIH toolbox. *Neurology, 80*(11 Suppl. 3), S54–S64.
Willoughby, M. T. (2016). Commentary on the changing nature of executive control in preschool. *Monographs of the Society for Research in Child Development, 81*(4), 151–165.
Willoughby, M. T., Blair, C. B., & The Family Life Project Investigators (2015). Measuring

executive function in early childhood: A case for formative measurement. *Psychological Assessment, 28*(3), 319–330.

Willoughby, M. T., Blair, C. B., Wirth, R. J., Greenberg, M., & The Family Life Project Investigators (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment, 24*(1), 226–239.

Willoughby, M. T., Kuhn, L. J., Blair, C. B., Samek, A., & List, J. A. (2017). The test–retest reliability of the latent construct of executive function depends on whether tasks are represented as formative or reflective indicators. *Child Neuropsychology, 23*(7), 822–837.

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment, 24*(2), 418–431.

Willoughby, M. T., Wylie, A. C., & Little, M. H. (2018). Testing longitudinal associations between executive function and academic achievement. *Developmental Psychology*. https://doi.org/10.1037/dev0000664 Advance online publication.

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

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., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). NIH Toolbox Cognition Battery (NIHTB-CB): Measuring executive function and attention. In P. D. Zelazo & P. J. Bauer (Eds.), National Institutes of Health Toolbox Cognition Battery (NIHTB-CB): Validation for children between 3 and 15 years. Monographs of the Society for Research in Child Development (Vol. 78, pp. 1–172). https://doi.org/10.1111/mono.12032.