Contemporary Educational Psychology 72 (2023) 102126

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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

Estimating the co-development of executive functions and math achievement throughout the elementary grades using a cross-lagged panel model with fixed effects

Haobai Zhang [a], Dana Miller-Cotto [a,b,1], Nancy C. Jordan [a,*]

[a] School of Education, University of Delaware, Newark, DE 19716, United States
[b] Kent State University, United States

ARTICLE INFO

Keywords:
Cross-lagged panel models with fixed effects
Executive functions
Mathematics
Longitudinal models

ABSTRACT

Executive functioning (EF) is associated with children's math skill development, both concurrently and longitudinally. However, it is not known how components of EF might be related to mathematics skills and vice versa over the course of elementary school. The present study addresses this issue by investigating relations between math achievement and two key components of EF – working memory (WM) and cognitive flexibility (CF) – from kindergarten to 5th grade, using the large-scale nationally representative dataset (N = 18,174) from the Early Childhood Longitudinal Study-Kindergarten (ECLS-K: 2011). Results from cross-lagged panel models with fixed effects support a transactional theoretical model, demonstrating a long-term reciprocal relationship between WM and math achievement from kindergarten to 5th grade and between CF and math achievement from 2nd grade to 5th grade. However, we found that reciprocal relations decrease as children grow older, suggesting that their math achievement relies less on EF and more on prior math knowledge over time.

1. Introduction

Early development of math skills is associated later math achievement and career trajectories (Jordan et al., 2009; Wang et al., 2017; Wrulich et al., 2014). To help explain this association, a rich body of literature has explored the role of executive function in young children's math skill development, both concurrently and longitudinally (for reviews, see Jacob & Parkinson, 2015; Peng & Kievit, 2020; Raghubar et al., 2010). Executive function (EF) involves three main subcomponents: working memory (holding and manipulating information in short-term memory), inhibitory control (controlling attention to override impulses), and cognitive flexibility (being flexible to changing demands or strategies) (Diamond, 2013; Dowsett & Livesey, 2000; Miyake et al., 2000). Although there are strong positive associations between EF subcomponents and math achievement (Finch, 2019; Nguyen et al., 2019), the direction and magnitude of these associations across the school years have not been fully investigated within a longitudinal context. Although longitudinal studies suggest that EF and math skills support each other in the early grades (Ellis et al., 2021; Miller-Cotto & Byrnes, 2020), research has not examined the direction and magnitude of these relationships across the elementary school years beyond third grade.

The goal of the present study was to examine the long-term development of two components of EF (working memory and cognitive flexibility) and math achievement using a large-scale longitudinal dataset. We expected the findings to shed a brighter light on the influence of fundamental cognitive processes (i.e., two components of EF) on math learning and, conversely, the influence of math learning on these processes over an extended period. We tested the data within the context of three theoretical explanations of the relation between EF and math skills: positive manifold (non-directional), cognitive filter (unidirectional), and transactional (bidirectional) models (Miller-Cotto & Byrnes, 2020).

1.1. *Theoretical explanations for the relationship between EF and math achievement*

Studies have shown that working memory (WM), cognitive flexibility (CF), and inhibitory control support academic performance in different ways (Clark et al., 2010; Friso-van den Bos et al., 2013; Magalhães et al.,

* Corresponding author.
E-mail addresses: haobaiz@udel.edu (H. Zhang), dmillerc@kent.edu (D. Miller-Cotto), njordan@udel.edu (N.C. Jordan).
[1] Dana Miller-Cotto is now at Kent State University, United States.

https://doi.org/10.1016/j.cedpsych.2022.102126

Available online 4 November 2022

2020; Yeniad et al., 2013). According to a recent meta-analysis (Cortés Pascual et al., 2019), of all three subcomponents, WM is the most dominant and thus widely studied factor regarding its relationship with math achievement.

Three theories have been identified that may potentially explain the directional relationship between WM and math. In Table 1, we summarize these three theoretical accounts, contextual explanation, and how the theories can be tested in empirical research (Miller-Cotto & Byrnes, 2020).

The *positive manifold* suggests that WM and mathematics skills share an underlying skill. Relatedly, one might assume that WM and mathematics skills would be highly correlated because they might share an underlying construct. Thus, no directionality is expected for the relationship between WM and math. An empirical test of this account would have paths of WM and math regressed onto prior WM and math respectively, with no directional paths between WM and math after controlling for between-person trait-like differences. The *cognitive filter* account asserts that learners have processing limitations that prevent them from adequately solving mathematics problems if too much information must be maintained for later use (e.g., Hitch, 1978). Thus, this theory would predict that reducing processing strains on the learner will promote better learning. An empirical test of this account would suggest math skills passing through WM to predict later performance. Therefore, there would be paths (arrows) extending from both prior WM (e.g., WM1) and prior math (e.g., Math 1) to later math (Math 2), but not from prior math to later WM.

Finally, the *transactional* account suggests that WM and math skills are related because stronger WM may equip individuals better in pulling information from long-term memory. One reason for this relationship could be that stronger WM helps to develop strategies to first offload excess details from a given problem to long term memory for later fact retrieval. Prior work shows that people with stronger WM show better fact retrieval (Baddeley, 2010) and that children with math learning disabilities demonstrate poor WM skills and retrieval deficits concurrently (Geary et al., 2012). As the accumulating math facts must be present to give children the opportunity to practice retrieving this information, WM would also be affected by math knowledge. Thus, WM tested at Time 1 would predict math performance at Time 2, given the cascading effects of knowledge on WM (Ericsson & Kintsch, 1995). Further, it would be expected that math achievement at Time 2 might also predict WM performance at Time 3. According to the transactional theory, there would be both concurrent and cross-lagged relationships over time.

Research provides some support for a transactional relationship between WM and math achievement in the early grades, using cross-lagged panel approaches to analyze the data. Miller-Cotto and Byrnes (2020) found statistically significant bidirectional paths between WM (as measured by a reversed digit span task) and math from kindergarten to second grade, also shown by McKinnon and Blair (2019) using a composite EF score. The findings contrast with the cognitive filter model, which posits that WM and math achievement share a common underlying factor that explains the relationship between the two domains. However, Miller-Cotto and Byrnes employed a cross-lagged panel model analysis with data only from time points between kindergarten and second grade. Moreover, the study did not fully examine how the predictive power from WM to math or from math to WM changed between kindergarten and second grade because only two time periods (from kindergarten to first grade and from first to second grade) were involved in the analysis.

A bidirectional relationship between EF and math from preschool to kindergarten is further supported by Schmitt et al. (2017) where WM (measured by a reversed digit span task) was used as an indicator of the EF latent construct. However, different conclusions were drawn in a more recent replication study with two different longitudinal datasets, showing that WM significantly predicted later math during this age period but not vice versa (Ellis et al., 2021). The way in which EF is

Table 1
Three Theories Characterizing Relationship Between WM and Math.

Hypothesis	Directionality	Contextual Explanation	Empirical Test of Theory
Positive manifold	No directional paths between WM and math	WM and math skills share underlying skills that are common to cognitive tests. WM may not promote or facilitate the acquisition or learning of math skills over time. Rather, the correlation between math and WM may simply reflect the factors such as general intelligence that are common to WM and math tasks.	After accounting for the stable between-person trait-like differences, models with no directional paths between WM and math fit the data better than adding any directional path between the two constructs over time.
Cognitive filter	Unidirectional (Earlier EF predicts subsequent math)	WM acts as a bottleneck filter that lets in only a subset of the information that is presented to children (Alloway & Alloway, 2010; Robison et al., 2018). Learners have processing limitations that prevent them from adequately solving mathematics problems if too much information must be maintained for later use (e.g., Hitch, 1978). This model does not assume prior math would predict later WM.	On within-person level, WM predicts changes in later math skills when prior math is controlled. But math does not predict later WM.
Transactional model	Bidirectional relationship between EF and math	WM and math are mutually supportive. When the number of items to be maintained in WM exceeds some level, the excess items are off-loaded to long-term memory for later retrieval. Stronger WM may equip individuals better in pulling information from long-term memory. Engaging in math practices at school which largely requires the uses of WM can be considered as a long-term training of WM and thus strengthen children's WM.	On within-person level, WM predicts the change in subsequent math when prior math is controlled. Math skills also predict later WM when prior WM is controlled.

modeled in the analysis (e.g., modeled as a reflective latent variable, as a composite variable, or as subcomponents specifically) may help explain the differing findings (Bull & Lee, 2014). Specifically, Camerota et al. (2020) showed that cross-lagged stability estimates between preschool and first grade were much larger when modeled as a reflective latent variable (RLV) than as a composite variable. In RLV models, the shared variance across different EF tasks is thought to measure the underlying ability and the remaining variance is measurement error. Researchers should be cautious when interpreting the results from different specification methods since they may link to different underlying theories. The present study considered WM and CF as separate constructs for a clearer test of the hypothesized theories for WM and CF, respectively. Given the lack of theoretical foundations to explicate the directionality of CF and mathematics skill relationship, the analyses for these two variables herein are more exploratory.

It should also be noted that studies that apply different modeling techniques in examining reciprocal relationships have produced inconsistent results. For example, the cross-lagged model applied in Miller-Cotto and Byrnes (2020) study has been critiqued for not adequately separating the within-person (WP) and between-person (BP) variation in the estimates of cross-lagged effects. In other words, the estimated bidirectional effects reflected a mix of WP level and BP level effects which lead to the challenges in mapping the results onto the theories that were supposed to reflect within-person variation (as shown in Table 1). Using modeling methods to separate the WP and BP level differences, Willoughby et al. (2019) examined longitudinal relations between both WM and CF and math from kindergarten through second grade using the latent growth curve models with structured residuals (LCM-SR). Their findings demonstrated unstable cross-lagged effects for math both in terms of magnitudes and signals of the cross-lagged parameters on a within-person level in early grades. Using the same model, Feldon and Litson (2021) found a significant but weak bidirectional relationship between WM and math from fall to spring of kindergarten, and from first to second grade. Differently, Swanson et al. (2021) found that the influence of WM on math among bilingual students in early elementary school years was unidirectional using the random-intercept cross-lagged panel model (RI-CLPM). Although the modeling techniques in the above studies effectively control for time-invariant unobserved heterogeneity compared to traditional cross-lagged models, it has been recently argued that such "residual-level" models may produce biased estimates when the process (i.e., executive function development) under investigation has not reached equilibrium status (Andersen, 2021).

As discussion on the modeling techniques that examine longitudinal development are ongoing in recent methodology research (e.g., Curran & Hancock, 2021; Kearney, 2017; Usami et al., 2019), in the present study we chose the approach that maps closely with the EF-math theories to be tested. More specifically, we avoid interpreting biased estimates that combine WP and BP level effects and instead focus on WP effects, where children's prior performance serves as their own controls.

1.2. Do EF and math support each other across the elementary school age span?

The aforementioned longitudinal studies do not examine the longitudinal relation between EF and math beyond the early grades. As such, the developmental progression of this relation is not well understood. The present study thus extends findings from Miller-Cotto & Byrnes (2020) and Willoughby and colleagues (2019) by examining the development of EF and math through fifth grade. Cross-sectionally, Holmes and Adams (2006) found WM processes supported children's math differently among first graders and third graders where younger children relied more on visuospatial WM than older students in math problem-solving. It is possible that the predictive strength of EF fades with age because early arithmetic skills become more automatic as children gain more knowledge. These math processes may depend more on WM, for example, before they are automatic. It may also be the case that weak EF in older students make them tend to use more external scaffolds, such as calculators or worked examples, to achieve success in math.

In the other direction – where math predicts later WM – it is intriguing to conjecture that this directional relationship holds into later years of elementary school as math tasks (e.g., memorizing multiplication tables) practiced at school may serve as long-term training for cognitive abilities (Peng & Kievit, 2020; Finch, 2019). In a meta-analysis, Jacob and Parkinson (2015) report that the associations between EF and math are somewhat stronger for older (6–11, 12–18-year-olds) than younger children (3–5-year-olds) although the variation is not statistically significant. Thus, it is still uncertain how the accumulation of math knowledge and long-term math practice would consistently contribute to or facilitate WM as children mature. Overall, it is reasonable to predict that the transactional EF model may hold in later years as children continue schooling, but more exploration into the role of age in a mutual relationship is needed in longitudinal analyses.

Though not as widely studied as WM, meta-analyses reveal CF is significantly correlated with math achievement and has unique developmental pattern over time (e.g., Yeniad et al., 2013; de Santana et al., 2022). CF starts to develop in early childhood and increases rapidly during the middle childhood year. Recent longitudinal work found CF to be a unique predictor of math achievement when controlling for intelligence, attention, inhibitory control, working memory, and planning in the intermediate grades but not in second grade (Magalhães et al., 2020). In contrast, correlational data showed that age negatively impacts the magnitude of the overall correlation between CF and math, indicating that among younger children mathematics performance is more strongly impacted by CF (de Santana et al., 2022). Willoughby et al. (2019) found that CF predicted later math significantly with small effect sizes from kindergarten to first grade, but earlier math achievement did not predict CF in this age group. As many studies include CF as part of a composite score for general EF and test the relationship between overall EF and math (McKinnon & Blair, 2019; Nguyen et al., 2019; Schmitt et al., 2017), it is of interest to examine the role of age in the direction and magnitude of the relationship between CF and math achievement over time.

1.3. The present study

As discussed, emerging evidence reveals bidirectional relations between components of EF and math achievement for children in the primary grades (e.g., Ellis et al., 2021; Miller-Cotto & Byrnes, 2020; McKinnon & Blair, 2019; Schmitt et al., 2017; Willoughby et al., 2019; Wolf & McCoy, 2019). However, studies have not examined bidirectional relations involving time points from kindergarten through fifth grade to document how EF and math relate to each other from early to middle childhood. To extend findings from previous research, the present study examined relations between EF components (WM and CF) and math achievement throughout elementary school (i.e., kindergarten through fifth grade) to determine the directionality and magnitude of the relations.

We applied a relatively new statistical technique in psychology, namely a cross-lagged panel model with fixed effects, which allows for examining within-person effects (Allison et al., 2017). Previous studies have used modeling methods that could not distinguish between within- and between-person effects (e.g., McKinnon & Blair, 2019; Miller-Cotto & Byrnes, 2020), which combine effects and obscure estimates. The present study examined within-person differences, rather than between-person differences, in the development of EF and math. That is, we wanted to know whether children who score higher on a prior math test relative to their own score would score higher in EF later and vice versa, as opposed to whether children who score higher relative to other children would have higher EF in later grades and vice versa. This within-person approach to comparing the three theoretical models presented herein can offer more valid insights into how prior EF or math affect individual children as they progress through elementary grades.

In sum, the study addressed (1) whether the transactional model can still be strongly supported and be the better fitting model compared to the cognitive filter and positive manifold models when using the fixed effects model to control for unobserved time-invariant confounders by using SEM fit statistics and examining path coefficients; (2) whether a bidirectional relationship holds into later years in elementary school or, as hypothesized, fades away as children grow older for both WM and CF. Overall, we expected the findings to explicate the potentially intertwined relations between cognitive and educational processes.

2. Method

2.1. Participants

The present investigation used the ECLS-K: 2011 dataset, which collected longitudinal data of child development, early learning, and school progress from a nationally representative U.S. sample of 18,174 children from kindergarten (the 2010–11 school year) through fifth grade (the 2015–16 school year) (Tourangeau et al., 2019). Child-level data were collected with the three-stage sampling design in all 50 states and the District of Columbia across the United States to create a nationally representative sample. The ECLS-K:2011 EF and math data were collected over nine waves from kindergarten through the fifth grade. The data collection rounds of used measures are presented in Table 2. Table 3 shows descriptive statistics of included variables. To obtain unbiased estimates for the nationally representative population of children who entered kindergarten in 2010–11, we chose the ECLS-K:2011 sampling weight which is a child-based weight adjusted for the nonresponse pattern associated with child direct assessment at multiple waves covered in the present study (Tourangeau et al., 2019). The ECLS-K:2011 study only tested children in 1/3 of primary sampling units (PSUs) as the subsample to represent the full sample in the fall of first and second grades.

2.2. Measures

2.2.1. Math achievement

Children were given a set of 18 routing items in the first stage and entered the second stage with either high, middle, or low difficulty level based on their scores in the first stage. During the assessment, children saw the key information of the problems on the easel pages while the problems were read to them by an assessor. The math assessment covers conceptual and procedural knowledge in multiple areas which include number sense, properties, operations, measurement, geometry, data analysis, patterns, algebra, and functions.

Various scoring approaches were applied in the ECLS-K:2011 math assessment. We chose the IRT-based scaled scores for analysis since they are appropriate for longitudinal analyses and comparable across all time points. Different from raw scores, the IRT-based scoring estimated every child's math achievement on the same continuous scale by assessing the probability that a child produces a correct response for a given item based on their response pattern. Thus, the chosen scoring method estimates the number of items a child would have answered correctly if given all items. Internal reliability of IRT-based scores were high through all rounds of data collection from kindergarten to fifth grade (alpha ranged from 0.91 to 0.94).

Table 3
Descriptive Statistics of Related Variables.

	N	Mean	Minimum	Maximum
1. Math K-Fall	15,598	36.37	11.96	132.66
2. Math K-Spring	17,143	50.63	11.75	112.54
3. Math 1st-Fall	5,222	58.46	14.46	140.01
4. Math 1st-Spring	15,103	73.03	12.27	138.93
5. Math 2nd-Fall	4,729	77.51	13.63	139.96
6. Math 2nd-Spring	13,830	90.43	18.24	139.10
7. Math 3rd-Spring	12,866	104.28	43.41	147.89
8. Math 4th-spring	12,080	112.81	25.73	147.90
9. Math 5th-Spring	11,426	119.81	26.76	148.04
10. CF 2nd-Fall	4,708	6.37	1.25	9.67
11. CF 2nd-Spring	13,774	6.73	1.63	10.00
12. CF 3rd-Spring	12,744	7.22	1.75	10.00
13. CF 4th-spring	12,021	7.63	1.88	10.00
14. CF 5th-Spring	11,386	7.98	2.00	10.00
15. WM K-Fall	15,598	434.39	393.00	581.00
16. WM K-Spring	17,147	450.86	393.00	572.00
17. WM 1st-Fall	5,222	458.06	393.00	596.00
18. WM 1st-Spring	15,107	470.52	393.00	596.00
19. WM 2nd-Fall	4,727	474.71	403.00	567.00
20. WM 2nd-Spring	13,832	481.21	403.00	581.00
21. WM 3rd-Spring	12,877	490.11	403.00	603.00
22. WM 4th-spring	12,085	497.48	403.00	588.00
23. WM 5th-Spring	11,430	503.33	403.00	588.00

Note. WM = working memory; CF = cognitive flexibility.

2.2.2. Executive function

The ECLS-K:2011 measured working memory, cognitive flexibility, and inhibitory control in executive functions using the Number Reversed (Woodcock et al., 2001), the Dimensional Change Card Sort (DCCS) (Zelazo, 2006; Zelazo et al., 2013), and the Flanker Inhibitory Control and Attention Task (Zelazo et al., 2013), respectively. Inhibitory control was added into the direct assessment in fourth grade. Thus, the current analysis did not include this measure. While WM (Numbers Reversed) was measured with a consistent instrument across nine waves, the different scoring of CF measures made the scores comparable only from second grade to fifth grade. Therefore, the present study examined the mechanisms of WM and CF separately.

2.2.3. Number reversed task

The Number Reversed Task adapted from the Woodcock-Johnson III Test of Cognitive abilities (Woodcock et al., 2001) was conducted to measure children's WM over nine waves. Children were asked to repeat a sequence of numbers read to them by the assessor in the reverse order. The test stopped when children got three consecutive number sequences incorrect. There was a maximum of 30 items, starting with 2-digit number items and ending with 8-digit items.

We used the W-ability score computed with the WJ-III publisher's scoring norms. The W score provides an equal-interval scale. Growth in W scores was expected in the longitudinal data of ECLS-K:2011 because WJ-III standard mean was set to the average performance of ten-year-old children and in ECLS-K most children were younger than ten when they took the number reversed task. Reliability of the number reversed task range is 0.9 or greater (Mather & Woodcock, 2001; Tourangeau et al., 2019).

Table 2
Data Collection Rounds in ECLS-K:2011.

Measures	Subarea	Fall Kindergarten	Spring Kindergarten	Fall First grade	Spring First grade	Fall Second grade	Spring Second grade	Spring Third grade	Spring Fourth grade	Spring Fifth grade
Math Achievement	–	P	P	P	P	P	P	P	P	P
Executive Functions	Cognitive flexibility	P	P	P	P	C	C	C	C	C
	Working memory	P	P	P	P	P	P	P	P	P

Note. The table reports whether the measure was given at the certain wave. P represents paper version was administered; C represents that computerized version was administered.

2.2.4. Dimensional change Card Sort

The Dimensional Change Card Sort (DCCS) was used to measure children's cognitive flexibility. In kindergarten and first grade, the physical, table-top version of DCCS was administered and accuracy data were collected. Starting in second grade, a computerized version developed as a test in the NIH Toolbox for the Assessment of Neurological and Behavioral Function was conducted to assess performances on both accuracy and reaction time (Zelazo et al., 2013). Although the same construct was measured with both versions, the scoring was different, so the scores from two versions were not easily comparable. As such, we used data from the computerized version in second to fifth grade (5 waves), which took both accuracy and reaction time into account to avoid ceiling effects. Children were given 30 mixed-block trials on the computerized DCCS tasks, where they were asked to sort the stimulus picture based either on color or shape. Then two pictures, either of the same shape or same color as the stimulus picture, appeared at the bottom for children to choose based on the sorting rule. In each block, the sorting that appeared more often than the other was considered as a dominant rule. Reaction time on trials with non-dominant rules was collected to detect children's performance in inhibiting the dominant response and shift to the non-dominant rules.

The overall DCCS computed scores, which gave equal weights to accuracy (5 units) and reaction time (5 units) were used in the present study. The computed scores ranged from 0 to 10, which were comparable across all grades. The test-retest reliability for DCCS ranged between 0.90 and 0.94 in childhood (Beck et al., 2011; Zelazo et al., 2014).

2.3. Analytic approach

The cross-lagged panel model (CLPM) has been widely used to examine the reciprocal relationship between variables in developmental studies. It should be noted, though, that some criticism has been raised about the traditional CLPM model for producing biased results, which does not adequately separate the between-person (BP) and within person (WP) differences in estimating the reciprocal relationship (Hamaker et al., 2015; Keijsers, 2016; Paul et al., 2019). As the present study focuses on investigating the developmental change between EF and math within individuals over time, we use cross-lagged models with fixed effects, a technique that allows for controlling unobserved time-invariant confounders and a clearer focus on within-person variation (Allison et al., 2017). The goal is achieved by accommodating the traditional cross-lagged models into the fixed effects framework (Allison et al., 2017; Zyphur et al., 2020).

The equations to model the longitudinal relations of math and EF are given as followed:

$$Math_{it} = \mu_t + \beta_{1,t-1}EF_{i,t-1} + \beta_{2,t-1}Math_{i,t-1} + \alpha_i + \varepsilon_{it} \quad (1)$$

$$EF_{it} = \tau_t + \beta_{3,t-1}Math_{i,t-1} + \beta_{4,t-1}EF_{i,t-1} + \eta_i + \nu_{it} \quad (2)$$

where t denotes each observed time point ($t = 2 \ldots T$), i denotes each observed individual in the sample ($i = 1 \ldots .N$), μ_i and τ_i are the time-varying intercepts, βs are the autoregressive and cross-lagged path coefficients, ε and υ are the random disturbance associated with time-variant errors. The path diagram of the model is shown in Fig. 1. Latent variables α_i and η_i are added into the model as the fixed effects to achieve the goal of accounting for the unobserved confounders in the study. When each i is an individual, these fixed effects terms can be understood as the unit-specific effects that can capture time-invariant factors that make the individual similar to itself over time (Zyphur et al., 2020). In other words, they accounted for how individuals differ from each other (i.e., between-person variation). Such terms are commonly called as "fixed effects" in econometrics whereas the term "random intercept" is more widely used in multi-level models in psychology (e.g., random-intercept cross-lagged panel model, Hamaker et al., 2015). Both models are achieving the same goal to disentangle the WP and BP differences in longitudinal data (for thorough reviews and comparison of these models, see Curran & Hancock, 2021; Usami et al., 2019; Zyphur et al., 2020).

In Fig. 1, path a and path b represent the autoregressive paths within each variable respectively which reflect the stability of EF and math achievement development over time. Path c and path d represent the correlation between EF and math achievement at each time point. Path e represents the effect of math achievement at a later point on EF at an earlier time point while controlling for earlier math achievement. Alpha and Eta are the latent variables for the fixed effects.

As CLPM with fixed effects only assumes sequential exogeneity, the error terms of math can correlate with all future EF in equation (1) as error terms can also correlate with later math in equation (2). To meet the assumption of no serial correlation, no correlation was specified between error terms in the models. With latent variables α_i and η_i, the

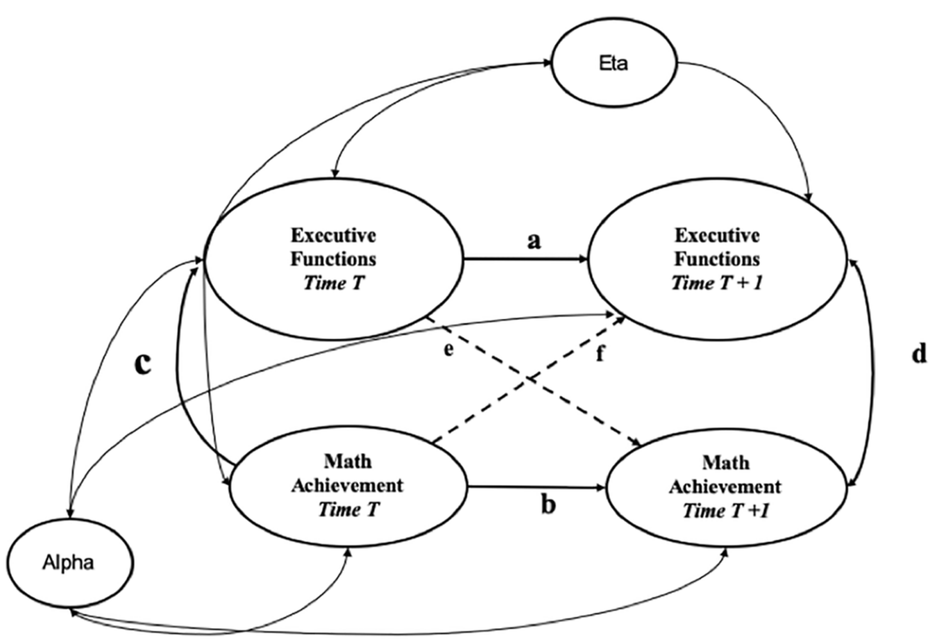

Fig. 1. Cross-lagged panel model with fixed effects. This figure demonstrates the mechanism of the cross-lagged panel model with fixed effect.

between-person effects are removed from the model and placed into these latent constructs. Therefore, with the inclusion of fixed effects, the cross-lagged effects estimated by this model only reflect the WP variation that we are interested in examining. In other words, when the effect is fixed within persons, each child would act as their own control as the prior measures and all unobserved time-invariant confounders were controlled. The ability to control unobserved confounders of this AR-CL model with fixed effects allows us to get closer to causal claim of the bidirectional between EF and math.

To examine potential bidirectional relationships between math achievement and executive functions (WM and CF) over time, separate sets of cross-lagged analyses were conducted for WM and CF separately in Mplus8 (Muthén & Muthén, 2017). To allow for estimating lagged, reciprocal causation while controlling for unobserved confounders, we fit the cross-lagged panel models with fixed effects using maximum likelihood estimation in structural equation modeling (ML-SEM) (Allison, 2009; Allison et al., 2017). We included the latent variables alpha and eta to represent the fixed effects which were allowed to covary with observed exogenous variables.

An additional benefit of using the maximum likelihood-SEM estimation is that missing data in ECLS-K can be easily handled by the full information maximum likelihood (FIML). Rather than piecewise deletion, FIML produced unbiased estimation with all observed data by using sample log-likelihood to do iterative computation (Enders & Bandalos, 2001; Williams et al., 2018).

We started with estimating the baseline model (model 1 in Fig. 2), which only included autoregressive paths within each variable through development and the effect of unobserved confounders accounted by two latent variables alpha and eta. The first-order autoregressive paths suggested that the score at each time point was regressed only on the immediately preceding score on the same measure. The autoregressive path coefficients reflected the stability of each measure over time. After fitting the baseline model, cross-lagged paths were added in subsequent models. In model 2 (model 2 in Fig. 2), we added unidirectional paths from each of the two components of EF (WM/CF) at the time point *t-1* to math achievement at the subsequent time point *t*. The unidirectional path coefficients reflected how EF at the prior time point predicted student math achievement at the later time point while controlling for the prior year of math achievement. In model 3 (model 3 in Fig. 2), the unidirectional paths from math achievement at time *t-1* to executive functions at the subsequent time point *t* were added onto the determined baseline model. Similarly, the path coefficients reflected how prior math achievement predicted later EF while controlling for prior EF. The final model (model 4 in Fig. 2) examined the cross-lagged relationship between EF and math achievement over time by adding the cross-lagged paths from both EF and math achievement at time *t-1* to math achievement and EF at the subsequent time point *t* respectively.

Two sets of models were run following the same procedure described above for each component in EF, namely, WM and CF. The only difference was that the analysis with CF included data at five waves as nine waves were included for WM. All models were evaluated with the model fit indices following the standard: the model fit would be considered as satisfactory when the comparative fit index (CFI) and Tucker–Lewis index (TLI) are greater than 0.9, Root Mean Square Error of Approximation (RMSEA) is less than 0.05, Standardized Root Mean Residual (SRMR) is less than 0.08 (Hu & Bentler, 1999). In the comparison of all nested models, besides the above fit indices, we chose the model with the smallest values in Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), and a significantly better fit than simpler models with the likelihood ratio (LR) tests.

2.3.1. *Transparency and openness*

We report the model selection process, all manipulations, and all measures in the study, and we followed JARS (Kazak, 2018). The data that support the findings of this study are openly available at https://nces.ed.gov/ecls/dataproducts.asp. All analysis code and related materials are available by emailing the corresponding author and will be available on OSF. Data were analyzed using R, version 4.1.0 and Mplus 8. This study's design and its analysis were not pre-registered.

3. Results

Bivariate correlations between each pair of variables are presented in Table S1 in the supplementary material. Both WM and CF positively correlated with math achievement measured at all waves.

3.1. *Relations between working memory and math achievement.*

A set of models examining WM and math achievement at nine waves was tested. The baseline model with first-order autoregressive paths fit the data poorly with model fit indices of $\chi^2(132) = 9515.84$, $p < .001$, CFI $= 0.881$, TLI $= 0.863$, RMSEA [90% CI] $= 0.063$ [0.062 0.064], SRMR $= 0.249$. The autoregressive effects within both math achievement (ranged from 0.83 to 0.93) and WM (ranged from 0.49 to 0.67)

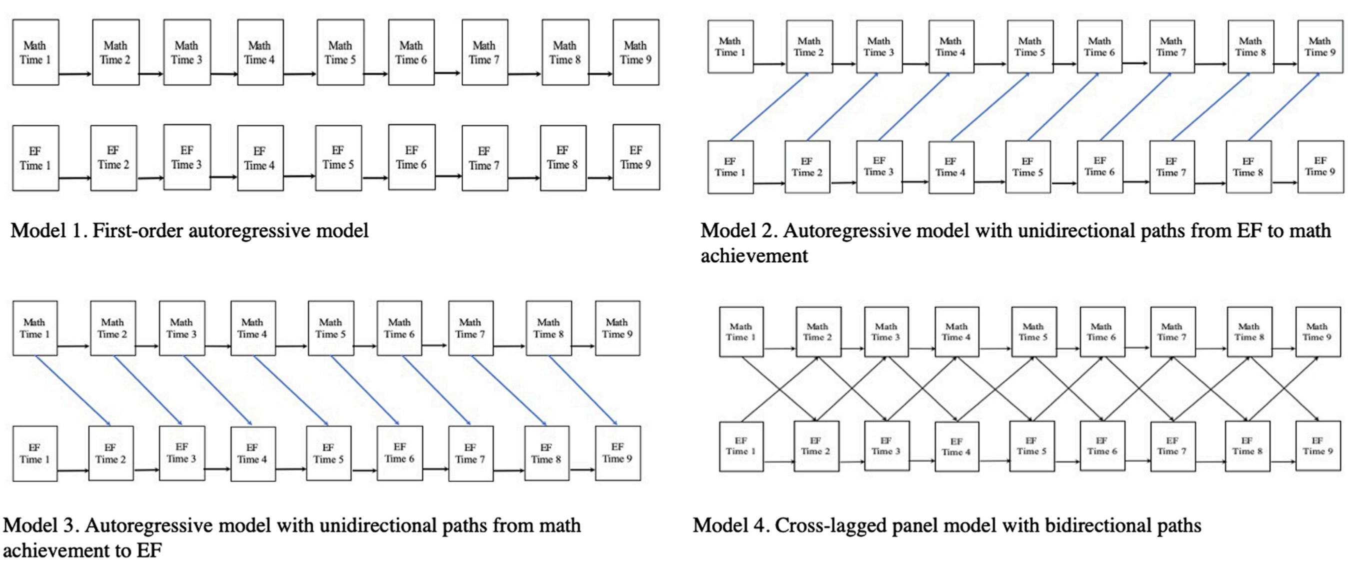

Fig. 2. Procedure for Model Developing and Testing. Only paths representing autoregressive effects and lagged effects are shown in above graphs. The latent variables alpha and eta not pictured in the graphs are included in all four models in the same way.

suggested strong stability of individual differences measured in these two variables from kindergarten to fifth grade.

Extending from the baseline model, we tested three additional models: a model with unidirectional paths from WM to later math achievement in the next time point (Model 2); a model with unidirectional paths from math achievement to later WM (Model 3); a cross-lagged model with bidirectional paths (Model 4). Table 4 shows all the model fit indices of this set of models. All three models fit the data better than the baseline model given the higher CFI and TLI (Table 4). The cross-lagged panel model with bidirectional paths had the best fit compared to models with unidirectional paths. We computed the chi-square difference tests for Model 4 and its nested models. The significant results show Model 4 best captures the data than other tested models. We noted that chi-square tests are sensitive to the large sample size and a relatively small improvement in model fit was found when comparing Model 3 and Model 4. The possible unidirectional relationship between WM and math will also be discussed.

To test the stability of the cross-lagged and autoregressive effects over time, the bidirectional model was also tested against the same model with fixed autoregressive path coefficients and fixed cross-lagged coefficients (Model 4a in Table 4). The constrained models with invariant coefficients showed a worse fit than the corresponding freely estimated models based on the results from the chi-square difference test using the Satorra-Bentler scaled chi-square values ($\chi^2(28) = 636.33$, $p < .01$). The results from the comparison suggested that imposing constraints was not tenable in the context, especially when time lags were unequal. Model fit indicators of the constrained models of Model 4 along with the constrained models of other possible bidirectional models with slight differences in latent variable specification are reported in Table S2 in the supplemental material.

These results confirm the cross-lagged association existing between WM and math achievement from kindergarten to fifth grade. The final model, which best described the reciprocal longitudinal relationships between WM and math achievement, is shown in Fig. 3. Significant path coefficients from WM to math achievement suggest that WM at the prior time point consistently predicts later math achievement when the prior math achievement was controlled. The predictive power from WM to math achievement was statistically significant throughout but in an overall descending trend with fluctuations from kindergarten ($\beta = 0.103$, $p < .01$) to fifth grade ($\beta = 0.056$, $p < .01$). The results suggest that the influence of early WM on later math is observed to be decreasing as children mature.

A similar pattern was also found in the other direction where math achievement at the prior time point significantly predicted WM at the later time point while controlling for the previous WM. The finding held at all nine waves. The transfer from math achievement to later WM was stable but also in an observed descending trend from kindergarten ($\beta = 0.347$, $p < .01$) through fifth grade ($\beta = 0.206$, $p < .01$).

Following the effect size guideline for cross-lagged effects established by Orth et al. (2022), we used the benchmark values of 0.03 as a small effect, 0.07 as a medium effect, and 0.12 as a large effect when interpreting the effect size of cross-lagged effects. These benchmarks were established for both traditional CLPM and CLPM with a focus on the within-person difference in the context of various content disciplines. The benchmark values were smaller in magnitude compared to the effect size conventions for correlations coefficients because (1) some outcome variance has been accounted by the autoregressive effects; (2) the developmental changes in traits make longitudinal associations systematically smaller in size than concurrent associations (Orth et al., 2022). Thus, these newly developed benchmarks are more appropriate to evaluate the effect sizes in the present study. The standardized coefficients on the paths from math to later WM indicate large effect across all waves whereas the effect of WM on math was medium to large in kindergarten and first grade and then falls to small-to-medium effects in later years of elementary school. The coefficients on paths from WM to math achievement and from math to WM were visualized as the dashed line and solid line respectively in Fig. 5 (a). The 95% confidence intervals of the cross-lagged coefficients were shown in the shaded grey area to better visualize the differences among the effects. It can be clearly seen that the effect of prior math on later WM (shown as the solid line) is smaller than the influence of prior WM on later math (shown as the dashed line).

3.2. Relationship between cognitive flexibility and math achievement

A series of models were fit following the procedures described above. The model fit indices of all models are shown in Table 5. We started with testing the baseline model (Model 1), which only contained first-order autoregressive paths within each measure. The fit for the first-order autoregressive model was good indicating that the baseline autoregressive model fit the data well ($\chi^2(31) = 1375.39$, $p < .001$, CFI = 0.957, TLI = 0.939, RMSEA [90% CI] = 0.056 [0.053, 0.058], SRMR = 0.076). The autoregressive path coefficients for math achievement ranged from 0.77 to 0.84 suggesting stable autoregressive longitudinal effects within math achievement.

After fitting the baseline model, we tested Model 2, which included the unidirectional paths from CF to math achievement at the subsequent time point. Both CFI and TLI of Model 2 increased and were greater than 0.95. The model fit improved comparing to Model 1 (see model fit in Table 5) suggesting CF at the prior time point significantly predicted later math achievement while controlling for the math score at the preceding time point. According to the path coefficients, the predictive power from CF to math was small but consistent longitudinally across five waves from second to fifth grade ($r = 0.039$- 0.115, *ps* less than 0.01). The path coefficients can be interpreted as the percent change in the math achievement at the subsequent time point for every percent change in the CF at the preceding period.

Similarly, Model 3, which contains the unidirectional paths from prior math achievement to the subsequent CF, was tested. The model fit the data better than the baseline model. The significant path coefficients ranged from 0.152 to 0.309 (*ps less than 0.01*) suggesting that math achievement at the prior wave consistently predicted later CF after controlling for the prior CF. The path coefficients in Model 3 can be interpreted similarly to Model 2.

Lastly, the autoregressive cross-lagged model (Model 4) with fixed effects was tested and produced the best model fit indices among nested models. The chi-square difference tests also showed significant results when compared to both Model 2 and Model 3, suggesting that model 4 with cross-lagged effects best captured the data. Prior math achievement predicted later CF at the subsequent time point and, conversely, prior CF

Table 4
Model Fit Statistics for Models of Working Memory and Math Achievement.

Model	Model description	Model fit						
		χ2	df	CFI	TLI	RMSEA	90% CI	SRMR
Model 1	First-order autoregressive model	9515.84	132	0.881	0.863	0.063	0.062 0.064	0.249
Model 2	Autoregressive model with unidirectional path *from WM to math*	8629.85	124	0.892	0.868	0.062	0.061 0.063	0.224
Model 3	Autoregressive model with unidirectional path *from Math to WM*	6305.59	124	0.922	0.904	0.053	0.052 0.054	0.138
Model 4	Autoregressive c ross-lagged model with *bidirectional paths*	5773.72	116	0.928	0.906	0.052	0.051 0.053	0.107
Model 4a	Cross-lagged model with equality constraints on path coefficients	6430.05	144	0.920	0.916	0.049	0.048 0.050	0.143

Note. WM = working memory; math = math achievement.

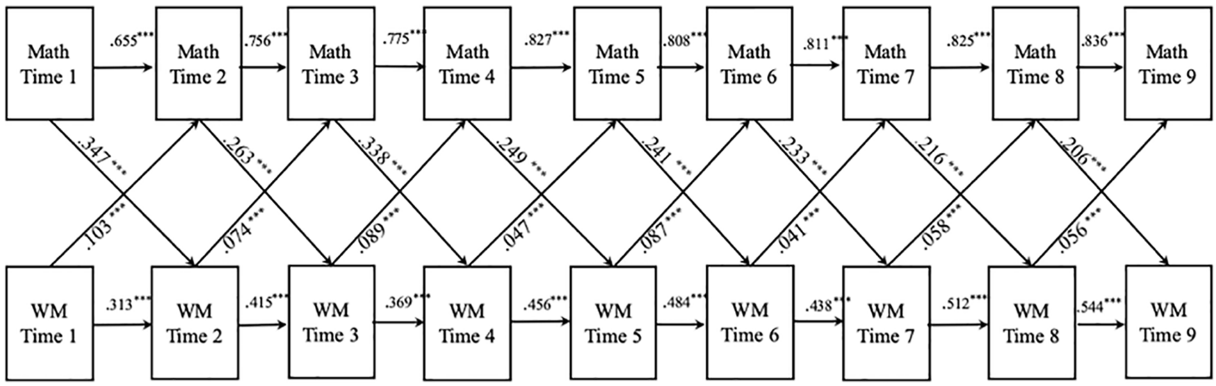

Fig. 3. Cross-Lagged Panel Model with Bidirectional Paths between WM and Math Achievement *Note*. Time 1 = Fall of kindergarten; time 2 = spring in kindergarten; time 3 = fall of 1st grade; time 4 = spring of 1st grade; time 5 = fall of 2nd grade; time 6 = spring of 2nd grade; time 7 = spring of 3rd grade; time 8 = spring of 4th grade; time 9 = spring of 5th grade. WM = working memory; math = math achievement. **p* <.05. ***p* <.01. ****p* <.001.

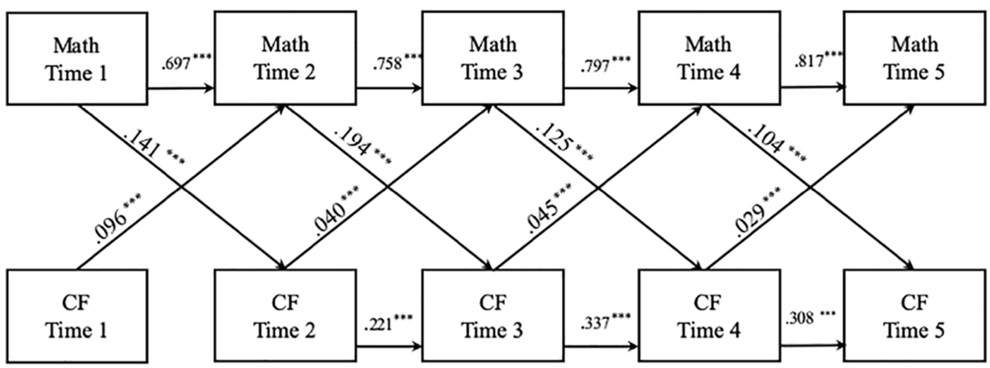

Fig. 4. Cross-Lagged Panel Model with Bidirectional Paths between CF and Math Achievement. *Note.* Time 1 = fall of 2nd grade; time 2 = spring of 2nd grade; time 3 = spring of 3rd grade; time 4 = spring of 4th grade; time 5 = spring of 5th grade. **p* <.05. ***p* <.01. ****p* <.001.

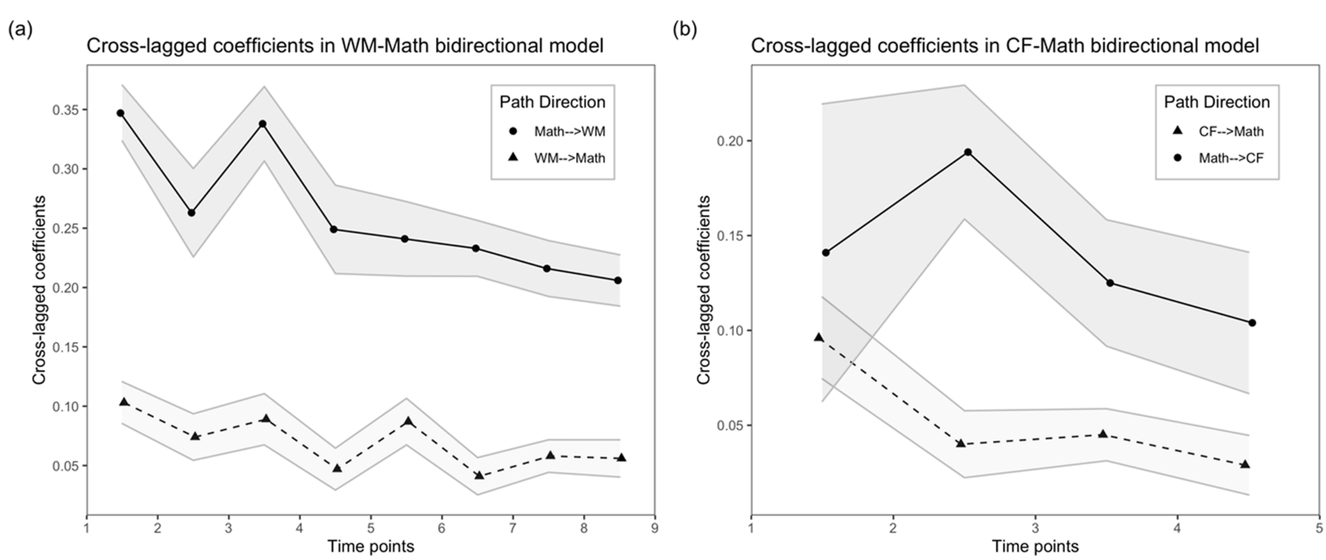

Fig. 5. Cross-lagged Path Coefficients in Full Models. Figure on the left side: (a) the coefficients on the cross-lagged paths in the full model of WM and math achievement over nine waves; figure on the right side: (b) the coefficients on the cross-lagged paths in the full model of CF and math achievement over five waves. The numbers on the x-axis represent the order of the period between each two neighboring time points among all observed time points. The numbers on the y-axis represent the magnitude of the path coefficients. The solid line represents the effect of prior math on later WM/CF. The dashed line represents the effect of prior WM/CF on later math. The shaded grey areas show the 95% confidence intervals of the cross-lagged coefficients.

predict later math achievement after controlling for prior CF and math achievement, respectively, and unobserved time-invariant confounders. The path diagram of the final model is shown in Fig. 4.

This final model with cross-lagged and autoregressive coefficients freely estimated was also compared with the bidirectional model with equality constraints on cross-lagged coefficients and autoregressive

Table 5
Model Fit Statistics for Models of Cognitive Flexibility and Math Achievement.

Model	Model description	Model fit						
		$\chi2$	df	CFI	TLI	RMSEA	RMSEA 90% CI	SRMR
Model 1	First-order autoregressive model	1375.39	31	0.957	0.939	0.056	0.053 0.058	0.076
Model 2	Autoregressive model with unidirectional path from CF to math	1120.99	27	0.965	0.943	0.054	0.051 0.056	0.049
Model 3	Autoregressive model with unidirectional path from math to CF	1078.06	27	0.967	0.946	0.053	0.050 0.055	0.040
Model 4	Cross-lagged model with bidirectional paths	975.06	23	0.970	0.942	0.054	0.051 0.057	0.027
Model 4a	Cross-lagged model with equality constraints on path coefficients	1229.30	36	0.962	0.954	0.049	0.046 0.051	0.151

Note. The fit indices of each fitted model are reported in the sequence of fitting order.

coefficients respectively (model 4a in Table 5). The aim of this comparison is to test the stability of the carry-over effects and the bidirectional effects between math and CF over time. The results from the chi-square difference test showed that model with equal path coefficients over time yielded worse fit than the freely estimated model ($\chi^2(13) = 306.99$, $p < .01$). Both the autoregressive effects and cross-lagged effects tended to be changing as children grow older. Model comparisons between restricted models (i.e., fixing cross-lagged effects and autoregressive effects to be invariant over time) and the current less restricted models were reported in Table S3.

The coefficients on paths from CF to math achievement and from math to CF in the final model were visualized as the dashed line and solid line respectively in Fig. 5 (b). Although the predictive power in both directions was statistically significant from second grade to fifth grade, the coefficients on the paths from CF to later math is smaller than the ones from math to later CF, especially from the spring of the second grade (Fig. 5 (b)). This finding suggests that increase in prior math has a greater influence in later CF than vice versa. For example, if a second grader's math score increased one percent in the fall semester, then the student's CF performance will increase 0.141 percent in the next spring semester; if the second grader increased one percent in CF test in Fall, then the student's math performance will increase 0.096 percent in the later spring semester. Following the effect size guideline by Orth et al. (2022), we found the effect of math on CF were large on average from second to fifth grade and the effect size of CF on math was medium to large beginning in second grade but falls to small in 5th grade.

4. Discussion

The goal of the present study was to extend our knowledge of concurrent and longitudinal relations between EF and mathematics skills. Extending prior work using the same national data set (Miller-Cotto & Byrnes, 2020; Willoughby et al., 2019), we employed causal inference methods that separate the within and between-person effects of these relations, examined relations between WM and CF, respectively, and math, and examined the extent to which the magnitude in these relations changes as children progress through school. Our analyses allow interpretation within the context of theoretical accounts of the directional relation between EF and math learning.

Results from our cross-lagged panel models with fixed effects demonstrated a reciprocal relationship between WM and math achievement from kindergarten to fifth grade, and between CF and math achievement from second grade to fifth grade. Most notably, to our knowledge, this is the first study to document an observed overall decreasing trend in the size of the reciprocal relationships over the course of the nine data points between WM and math achievement.

The data support a transactional model of development across a wide age range and across both components of EF, even under a stricter control (which used fixed-effect models to eliminate the between-person variation and control for unobserved time-invariant confounders) than was used in previous work. Unlike prior work, the present findings provide estimates closer to causal methods regarding the relationship between EF and math, such that EF skills may cause an increase in math performance and math skills may cause an increase in EF performance. Although this transactional relation does not automatically prove causality, employing the cross-lagged panel model with a fixed-effects approach brings us closer to making a causal inference by controlling for unobserved time-invariant confounders.

4.1. Working memory and math achievement

The overall pattern of the co-development of WM and math achievement demonstrated a significant bidirectional association over all nine assessment waves, from kindergarten through fifth grade. Better WM is shown to be associated with better performance in later math while controlling for their prior math achievement. However, better math achievement in the prior year also led to stronger WM in the subsequent year. This finding strengthens evidence of a mutually supportive relation over a longer developmental period than previous studies (e.g., Clements et al., 2016; Miller-Cotto & Byrnes, 2020) because we observed that the bidirectional pattern remains on the within-person level data with more time points– although the magnitude of support decreases over time.

We compared our results with previous studies that explored the bidirectional relationship using the ECLS-K:2011data set, although different developmental ranges and subsamples were covered and different modeling strategies were applied. In the present study, we found stable and meaningful carry-over effects for math performance and WM respectively across elementary school years and stronger effects for math (ranging from 0.66 to 0.84) than WM (ranging between 0.31 and 0.54). This is aligned with the significant carry-over effects found in both Miller-Cotto and Byrnes (2020; 0.75 and 0.64 for math; 0.33 and 0.25 for WM) and Willoughby et al. (2019; 0.23 and 0.34 for math, 0.08 and 0.10 for WM), although both of those studies studied only up to first and second grade. We extended the previous findings and found carry-over effects continued to last into later grades in elementary school, which suggests the math knowledge students gained in later years still largely builds on students' own prior knowledge.

The present study also resolved some issues related to comparisons of cross-lagged effects. Although Miller-Cotto and Byrnes (2020) found positive and statistically significant cross-lagged coefficients between WM and math from kindergarten to second grade, due to the modeling strategy of their study, the positive coefficients could reflect a mix of two types of variations and interpretations: (a) between-person level: students who have higher WM than others also have better math, and (b) within-person level: when each student is acting as his/her own control, higher WM will lead to better math performance. Thus, their results did not necessarily reflect a true bidirectional relationship as hypothesized in the transactional model in which component (b) is of the main interest.

In contrast, we applied a different modeling strategy, which allowed us to focus on this component (b). In other words, the examination of cross-lagged coefficients obtained in the present analysis is a more direct test of the transactional model than the earlier study. Extending previous research, we found that the cross-lagged effects between WM and math still exist even after we brought the focus back to the within-person variation, and importantly such pattern lasted into later elementary school grades. Supporting the rationale that we discussed, after partitioning out the stable between-person trait differences, we found smaller cross-lagged coefficients in both directions compared to those in Miller-

Cotto & Byrnes (2020).

Using the same modeling strategy (i.e., LCM-SR) to separate between- and within-person variation on the same dataset across the early grades, Willoughby et al. (2019) and Feldon and Litson (2021) yielded different bidirectional findings. Willoughby et al. found unstable and typically non-significant cross-lagged effects, both negative and positive, between WM and math from kindergarten to second grade whereas Feldon and Litson found very small, but significant, positive relationships from fall to spring of kindergarten and from first grade to second. Our investigation included the same early period and showed consistent stable cross-lagged effects between WM and math, which is partially aligned with Feldon and Litson. The difference between our findings and Willoughby et al. might come from the different sample selection standards since the latter only included children who had complete data on all six time points, whereas all available samples were included (missingness handled with FIML) in the present study. Taken together, findings favor a transactional model over unidirectional models, such as the cognitive filter account.

Although it may be the case that WM influences math skills by helping children pull information from long-term memory to solve math problems, the influence of mathematics on WM remains unclear. It could be the case that mathematics instruction offers children opportunities to hold information in short-term memory for later use, thereby practicing domain-specific WM strategies – especially since the WM task in the present study involves remembering digits. Alternatively, each time a student gains new knowledge and skill in a domain, this increased knowledge seems to form the basis of increased WM ability. Increased WM ability, in turn, may help them process more information as they interact with new math problems or develop fluency in mathematics. The characterization of WM that is most consistent with our results is those of Ericsson and Kintsch (1995), Jones and Macken (2015), and Unsworth et al. (2012). These authors argue that the factor that contributes most to performance on WM is the ability to retrieve information from long-term memory. Thus, as previously mentioned, it may be the case that these reciprocal relations exist because opportunities to hold information in short-term memory for later use are opportunities to practice domain-specific WM strategies.

The current work extends findings from prior work (Miller-Cotto & Byrnes, 2020; Willoughby et al., 2019), demonstrating that the predictive power from prior WM ability to later math achievement showed an observed fluctuating but still overall decreasing trend in the full model; the largest effect found during kindergarten (Time 1 - Time 2) was almost twice as big as the smallest effect in fourth to fifth grades (Time 8 – Time 9, see Fig. 5 (a)). Although some fluctuations were observed during the school year which needs to be interpreted with caution, we still observed an overall decreasing in the cross-lagged effects when we qualitatively inspected the changes in both the coefficients and the confidence intervals. Peng and Kievit (2020) argued that performing math tasks at the early stage of learning recruits more WM for problem-solving than in later learning stages, which may explain the decrease in the relationship between WM and math achievement as children get older. Relatedly, Evans and Stanovich (2013) proposed dual-processing theories of higher cognition, suggesting that individuals prefer rapid autonomous processes and yield default responses over the distinctive higher-order reasoning processes which causes a heavy load on WM. That is, when solving a math problem with familiar information, older children with more exposure to math knowledge and instruction may use the strategy that uses fewer processing demands (e.g., such as stored math facts or procedures that can be retrieved directly). One example might be the use of addition and subtraction facts in a math test. Younger children with less schooling experience and less accumulating knowledge may rely on counting processes (e.g., counting on from the larger addend), which require more WM resources (Geary et al., 2004). Moreover, previous studies examining the cross-domain associations showed increasing differentiation of domains after formal schooling and found that domain-specific skills become more closely related to domain-specific achievement than domain-general predictors as children grow older (Kwok et al., 2021). Working memory, as an underlying domain-general cognitive factor, thus exhibits a weaker relationship with math which is consistent with prior findings.

When examining the path coefficients in the other direction from math achievement to WM (the solid line in Fig. 5 (a)), we found that the predictive power from math to WM fluctuated around a higher level from kindergarten to second grade and then decreased when children entered third grade. For older children (i.e., children in fourth and fifth grades), the math-related practices at school may rely more on information retrieval from long-term memory rather than recruiting WM, although WM is needed for solving multi-step problems.

4.2. *Cognitive flexibility and math achievement*

Less studied than WM, previous research has produced equivocal results regarding the association between CF and math achievement (Clark et al., 2010; Nguyen & Duncan, 2019; van der Sluis et al., 2004; Magalhães et al., 2020; Yeniad et al., 2013). In the present study, we found CF and math achievement to be reciprocally predictive at later time points while controlling for previous math scores, prior CF, and all time-invariant unobserved confounders. The effect of math on CF was large to medium from second to fifth grade (Orth et al., 2022). On the other hand, the effect sizes on the paths from CF to math achievement were generally small, ranging from 0.029 to 0.096, showing that children's prior CF was associated with positive but small increases in math achievement in the subsequent year. The magnitude of the effect sizes in the present study aligns with results from Nguyen and Duncan (2019), which examined the effect of CF at school entry on math achievement in the third grade. In their final model, the range of effect sizes on math achievement was also small, although the authors speculated that larger effects might emerge at later grades. Willoughby et al. (2019) found small and significant cross-lagged effects between CF and math from kindergarten to first grade (using the physical version of the DCCS). When the data from second grade to fifth grade were included in the present analysis (using the computerized version of the DCCS), we still found weak predictive power from CF to math achievement.

While many theories of WM and academic skills have been asserted, it remains unclear how CF, at least as measured by the Dimensional Change Card Sorting Task, is related to mathematics skills. It may be the case that CF allows students to refine their attention to parts of a problem that are important, flexibly shift between strategies when they observe that they are no longer working and define goals appropriately. This process is also embodied in previous findings that students sorted the problems and exhibited different reasoning patterns based on the problem characteristics in mental arithmetic (Rathgeb-Schnierer & Green, 2017; Green & Rathgeb-Schnierer, 2020). However, what about the reverse? In other words, what does having strong mathematics skills mean for improved CF? It may be the case that opportunities to practice mathematics works to improve CF. When students become better at refining their attention to parts of a problem that are important, they learn to shift strategies when they observe that they are no longer working, and define goals appropriately, which may in turn transfer to their CF. This finding is also aligned with results from some qualitative studies on small samples suggesting that CF can be promoted by teaching activities in math classrooms (e.g., Serrazina & Rodrigues, 2021; Heinze et al., 2009).

4.3. *Limitations and future directions*

The present study is constrained by the measures available in ECLS-K:2011 data set, as there was only one indicator for WM and CF, respectively. The reversed digit span task, which was used to measure WM at all time points, is a relatively narrow (but reliable and valid) measure that involves attention span in a numerical context. Using measures of WM in a non-numerical context might amplify the context of the relation between math achievement and more general WM. In recent work, Wilkey et al. (2020) found that number-specific EF was a predictor of math achievement when controlling for magnitude perception, but non-numeric EF was not. It is worth noting, however, that Miller-Cotto and Byrnes (2020) tested their models with both math and reading, demonstrating similar reciprocal relationship patterns. This finding suggests that children may not be interpreting the numbers in the digit span task as numerical symbols but rather as lexical representations. Nevertheless, future research should consider multiple measures of WM to get a fuller picture of the underlying mechanisms. Likewise, ECLS-K:2011 only uses a broad measure of math achievement. EF may be differentially related to various components of math, such as computation and applied problem-solving, prompting the need for future work to examine the causal relationship between EF and math domains individually.

5. Conclusion

Perhaps the most important lesson from the present study is that EF (especially WM) and math achievement develop hand over hand throughout the elementary grade span, although the reciprocal relation decreases over time. The present study also suggests that math learning increases domain-general cognitive skills, perhaps even beyond EF. Although the methods used in the present longitudinal study come closer to supporting a causal model, it is necessary to employ alternative study designs to test the theories presented here. One way to do this might be to determine whether growth in EF also leads to growth in math skills and vice versa through randomized controlled studies. Although studies have sought to ascertain whether training EF in the context of mathematics leads to improved mathematics skills (see EF + Math Program, n.d.; Takacs & Kassai, 2019), there is still little evidence for the transfer of EF training to improved math skills or that math instruction improves EF. Future randomized studies should test competing models, like the cognitive filter vs. the transactional model. For example, across different grade levels, studies could examine whether explicit instruction in general WM (e.g., recalling digits) improves math performance (e.g., accuracy addition and subtraction combinations) and WM performance and, conversely, whether explicit instruction in math improves both WM and math performance.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are openly available at https://nces.ed.gov/ecls/dataproducts.asp. All analysis code and related materials are available by emailing the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cedpsych.2022.102126.

References

Allison, P. D. (2009). *Fixed effects regression models* (Vol. 160). London, UK: Sage.

Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum Likelihood for Cross-lagged Panel Models with Fixed Effects. *Socius*. https://doi.org/10.1177/2378023117710578

Andersen, H. K. (2021). Equivalent Approaches to Dealing with Unobserved Heterogeneity in Cross-Lagged Panel Models? Investigating the Benefits and Drawbacks of the Latent Curve Model with Structured Residuals and the Random Intercept Cross-Lagged Panel Model. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000285.

Baddeley, A. (2010). Working memory. *Current biology : CB, 20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive Function in Preschool Children: Test-Retest Reliability. *Journal of cognition and development: Official journal of the Cognitive Development Society, 12*(2), 169–193. https://doi.org/10.1080/15248372.2011.563485

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

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

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

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

Curran, P., & Hancock, G. (2021). The Challenge of Modeling Co-Developmental Processes over Time. *Child Development Perspectives., 15*. https://doi.org/10.1111/cdep.12401

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

Dowsett, S. M., & Livesey, D. J. (2000). The development of inhibitory control in preschool children: Effects of "executive skills" training. *Developmental Psychobiology, 36*(2), 161–174. https://doi.org/10.1002/(SICI)1098-2302(200003)36:2<161:AID-DEV7>3.0.CO;2-0

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Ellis, A., Ahmed, S. F., Zeytinoglu, S., Isbell, E., Calkins, S. D., Leerkes, E. M., Grammer, J. K., Gehring, W. J., Morrison, F. J., & Davis-Kean, P. E. (2021). Reciprocal Associations Between Executive Function and Academic Achievement: A Conceptual Replication of Schmitt et al. (2017). *Journal of Numerical Cognition*, *7*(3), 453-472. https://doi.org/10.5964/jnc.7047.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245. https://doi.org/10.1037/0033-295X.102.2.21

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*, 223–241. https://doi.org/10.1177/1745691612460685

Feldon, D. F., & Litson, K. (2021). Modeling Theories and Theorizing Models: An Attempted Replication of Miller-Cotto & Byrnes' (2019) Comparison of Working Memory Models Using ECLS-K Data. *Educational Psychology Review*, 1–28. https://doi.org/10.1007/s10648-021-09596-8

Finch, J. E. (2019). Do Schools Promote Executive Functions? Differential Working Memory Growth Across School-Year and Summer Months. *AERA Open*. https://doi.org/10.1177/2332858419848443.

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

Geary, D. C., Hoard, M. K., Byrd-Craven, J., & DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, *88*(2), 121–151. https://doi.org/10.1016/j.jecp.2004.03.002.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical Cognition Deficits in Children With Learning Disabilities and Persistent Low Achievement: A Five-Year Prospective Study. *Journal of Educational Psychology, 104*(1), 206–223. https://doi.org/10.1037/a0025398

Green, M., & Rathgeb-Schnierer, E. (2020). Summing Up: Cognitive Flexibility and Mental Arithmetic. *Journal of Mathematics Education, 13*(1), 1–17. https://doi.org/10.26711/007577152790051.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116. https://doi.org/10.1037/a0038889

Heinze, A., Star, J. R., & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education. *ZDM, 41*, 535–540. https://doi.org/10.1007/s11858-009-0214-4

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology, 10*(3), 302–323. https://doi.org/10.1016/0010-0285(78)90002-6

Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology, 26*(3), 339–366. https://doi.org/10.1080/01443410500341056

Hu, L.-T., & 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

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

Jones, G., & Macken, B. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition, 144*, 1–13. https://doi.org/10.1016/j.cognition.2015.07.009

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental psychology, 45*(3), 850–867. https://doi.org/10.1037/a0014939

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist, 73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kearney, M. W. (2017). Cross-lagged panel analysis. In M. R. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. 312–314). Thousand Oaks, CA: Sage.

Keijsers, L. (2016). Parental monitoring and adolescent problem behaviors: How much do we really know? *International Journal of Behavioral Development, 40*(3), 271–281. https://doi.org/10.1177/0165025415592515

Kwok, F. Y., Bull, R., & Muñez, D. (2021). Cross- and Within-Domain Associations of Early Reading and Mathematical Skills: Changes Across the Preschool Years. *Frontiers in psychology, 12*, Article 710470. https://doi.org/10.3389/fpsyg.2021.710470

Magalhães, S., Carneiro, L., Limpo, T., & Filipe, M. (2020). Executive functions predict literacy and mathematics achievements: The unique contribution of cognitive flexibility in grades 2, 4, and 6. *Child Neuropsychology*, 1–19 10.1080/09297049.2020.1740188.

Mather, N., & Woodcock, R. W. (2001). *Woodcock-Johnson® III tests of achievement: Examiner's manual standard and extended batteries*. Itasca, IL: Riverside.

McKinnon, R. D., & Blair, C. (2019). Bidirectional relations among executive function, teacher–child relationships, and early reading and math achievement: A cross-lagged panel analysis. *Early Childhood Research Quarterly, 46*, 152–165. https://doi.org/10.1016/j.ecresq.2018.03.011

Miller-Cotto, D., & Byrnes, J. P. (2020). What's the best way to characterize the relationship between working memory and achievement? An initial examination of competing theories. *Journal of Educational Psychology, 112*(5), 1074–1084. https://doi.org/10.1037/edu0000395

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

Muthén, L.K. and Muthén, B.O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

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

Orth, U., Meier, L. L., Bühler, J. L., Dapp, L. C., Krauss, S., Messerli, D., & Robins, R. W. (in press). Effect size guidelines for cross-lagged effects. *Psychological Methods.* https://doi.org/10.1037/met0000499.

Paul, J. M., Gray, S. A., Butterworth, B. L., & Reeve, R. A. (2019). Reading and math tests differentially predict number transcoding and number fact speed longitudinally: A random intercept cross-lagged panel approach. *Journal of Educational Psychology, 111*(2), 299–313. https://doi.org/10.1037/edu0000287

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

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

Rathgeb-Schnierer, E., & Green, M. (2017). Profiles of Cognitive Flexibility in Arithmetic Reasoning: A cross country comparison of German and American elementary students. *Journal of. Mathematics Education, 10*(1), 1–16. https://doi.org/10.21125/inted.2017.0252

de Santana, A. N., Roazzi, A., & Nobre, A. P. M. C. (2022). The relationship between cognitive flexibility and mathematical performance in children: A meta-analysis. *Trends in Neuroscience and Education, 28*, Article 100179. https://doi.org/10.1016/j.tine.2022.100179

Schmitt, S. A., Geldhof, G. J., Purpura, D. J., Duncan, R., & 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

Serrazina, L., & Rodrigues, M. (2021). Teacher's Actions to Promote Flexibility in Mental Arithmetic. *Journal of. Mathematics Education., 13*, 56–72. https://doi.org/10.26711/007577152790054

Swanson, H., Arizmendi, G., & Li, J.-T. (2021). What mediates the relationship between growth in math problem-solving and working memory in English language learners? *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000718

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

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., et al. (2019). Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011) User's Manual for the ECLS-K:2011 Kindergarten-Fifth Grade Data File and Electronic Codebook. *Public Version (NCES 2019–051). U.S. Department of Education.* Washington, DC: National Center for Education Statistics.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Working Memory Capacity and Retrieval Limitations from Long-Term Memory: An Examination of Differences in Accessibility. *The Quarterly Journal of Experimental Psychology, 65*, 2397–2410. https://doi.org/10.1080/17470218.2012.690438

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods, 24*, 637–657. https://doi.org/10.1037/met0000210

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

Wang, M., Ye, F., & Degol, J. L. (2017). Who chooses STEM careers? Using A relative cognitive strength and interest model to predict careers in science, technology, engineering, and mathematics. *Journal of Youth and Adolescence, 46*(8), 1805–1820. https://doi.org/10.1007/s10964-016-0618-8

Wilkey, E. D., Pollack, C., & Price, G. R. (2020). Dyscalculia and typical math achievement are associated with individual differences in number-specific executive function. *Child Development, 91*, 596–619. https://doi.org/10.1111/cdev.13194

Williams, R., Allison, P. D., & Moral-Benito, E. (2018). Linear Dynamic Panel-data Estimation Using Maximum Likelihood and Structural Equation Modeling. *The Stata Journal, 18*(2), 293–326. https://doi.org/10.1177/1536867X1801800201

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

Wolf, S., & McCoy, D. C. (2019). The role of executive function and social-emotional skills in the development of literacy and numeracy during preschool: A cross-lagged longitudinal study. *Developmental science, 22*(4), e12800. https://doi.org/10.1111/desc.12800

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

Wrulich, M., Brunner, M., Stadler, G., Schalke, D., Keller, U., & Martin, R. (2014). Forty years on: Childhood intelligence predicts health in middle adulthood. *Health Psychology, 33*, 292–296. https://doi.org/10.1037/a0030727

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.

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). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78*(4), 16–33. https://doi.org/10.1111/mono.12032

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., Conway, K. P., et al. (2014). NIH Toolbox Cognition Battery (CB): Validation of executive function measures in adults. *Journal of the International Neuropsychological Society: JINS, 20*(6), 620–629. https://doi.org/10.1017/S1355617714000472

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., et al. (2020). From Data to Causes I: Building A General Cross-Lagged Panel Model (GCLM). *Organizational Research Methods, 23*(4), 651–687. https://doi.org/10.1177/1094428119847278