Empirical Research

Check for updates

Reciprocal Associations Between Executive Function and Academic Achievement: A Conceptual Replication of Schmitt et al. (2017)

Alexa Ellis [1], Sammy F. Ahmed [1], Selin Zeytinoglu [2], Elif Isbell [3], Susan D. Calkins [4], Esther M. Leerkes [4],

Jennie K. Grammer [5], William J. Gehring [1], Frederick J. Morrison [1], Pamela E. Davis-Kean [1]

[1] *Department of Psychology, University of Michigan, Ann Arbor, MI, USA.* [2] *Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, USA.* [3] *Department of Psychological Sciences, University of California Merced, Merced, CA, USA.* [4] *Department of Human Development and Family Studies, University of North Carolina at Greensboro, Greensboro, NC, USA.* [5] *Department of Education, University of California, Los Angeles, CA, USA.*

Abstract

The goal of the current study was to conduct a conceptual replication of the reciprocal associations between executive function (EF) and academic achievement reported in Schmitt et al. (2017, https://doi.org/10.1037/edu0000193). Using two independent samples (N (STAR) = 279, and N (Pathways) = 277), we examined whether the patterns of associations between EF and achievement across preschool and kindergarten reported in Schmitt et al. (2017) replicated using the same model specifications, similar EF and achievement measures, and across a similar developmental age period. Consistent with original findings, EF predicted subsequent math achievement in both samples. Specifically, in the STAR sample, EF predicted math achievement from preschool to kindergarten, and kindergarten to first grade. In the Pathways sample, EF at kindergarten predicted both math and literacy achievement in first grade. However, contrary to the original findings, we were unable to replicate the bidirectional associations between math achievement and EF in either of the replication samples. Overall, the current conceptual replication has revealed that bidirectional associations between EF and academic skills might not be robust to slight differences in EF measures and number of measurement occasions, which has implications for our understanding of the development EF and academic skills across early childhood. The present findings underscore the need for more standardization in both measurement and modeling approaches – without which the inconsistency of findings in published studies may continue across this area of research.

Keywords

executive function, academic achievement, early childhood, bidirectional relations

Research in the developmental and educational literatures has consistently demonstrated associations between children's executive function (EF) skills and early academic achievement (e.g., Best, Miller, & Naglieri, 2011; Bull, Espy, & Wiebe, 2008). However, the nature of these relations has become a topic of debate in recent years (see Peng & Kievit, 2020 for review). Historically, research in this area has treated EF as foundational for academic skill development given

its support of learning and adaptation in early school settings (e.g., Blair, 2002; Zelazo, Blair, & Willoughby, 2016). Specifically, children's ability to process and manipulate information, inhibit automatic and potentially inappropriate responses to the environment, and direct their attention to appropriate tasks has been shown to be particularly useful in early learning settings (e.g., Morrison, Ponitz, & McClelland, 2010). Collectively, this research suggests that strengthening children's EF skills could enhance their math and literacy development during the early and formative years of schooling (Blair & Diamond, 2008; Blair & Razza, 2007; McClelland et al., 2007). As such, large-scale interventions have been designed to target EF skills through educational practices and comprehensive curricula programs (e.g., Tools of the Mind; Bodrova & Leong, 2001; Promoting Alternative Thinking Strategies, Kusche et al., 1994; Chicago School Readiness Program; Raver et al., 2008).

Executive Function and Academic Achievement

Recent work, however, has cast doubt on the causal links between early EF skills and children's academic development (e.g., Jacob & Parkinson, 2015). Longitudinal studies, for example, examining bidirectional links between EF skills and academic achievement have challenged the predominant unidirectional perspective of EF supporting academic skill development (e.g., Cameron, Kim, Duncan, Becker, & McClelland, 2019; Fuhs, Nesbitt, Farran, & Dong, 2014; McKinnon & Blair, 2019; Meixner, Warner, Lensing, Schiefele, & Elsner, 2019; Miller-Cotto & Byrnes, 2020; Welsh, Nix, Blair, Bierman, & Nelson, 2010). This work has leveraged the availability of multiple time-points of data to describe reciprocal relations between EF and achievement over time, testing whether these constructs co-develop or are directional in nature during the early years of schooling. In many cases, autoregressive cross-lagged panel (ARCL) models were used to test whether domain general cognitive abilities, such as EF, prospectively predict domain specific abilities, such as academic achievement - or the degree to which cognitive abilities and academic skills co-develop (mutually influence each other) over time (see Peng & Kievit, 2020 for review). This research is often referred to as the theory of mutualism or co-development between EF and academic skills across time.

Findings from these studies using the concept of mutualism have been mixed. For example, Welsh and colleagues (2010) found bidirectional relations between EF and numeracy skills (but not literacy skills) during preschool (M_{age} = 4.49 years), whereas other work has shown that EF prospectively predicts math and literacy achievement from preschool to kindergarten (Fuhs et al., 2014). Additionally, studies examining reciprocal relations using more than two time-points have yielded different directional patterns across early development. For example, Schmitt, Geldhof, Purpura, Duncan, and McClelland (2017) found bidirectional relations between EF and math achievement across the preschool school year (M_{age} = 4.70 years), and unidirectional associations from EF to math achievement across kindergarten (M_{age} = 5.70 years). Conversely, McKinnon and Blair (2019) reported bidirectional associations between EF and math skills across kindergarten (M_{age} = 5.75 years), as well as from kindergarten to first grade.

Inconsistencies in Prior Work

Although the general consensus is that EF and academic abilities are linked across early development (e.g., Peng & Kievit, 2020), the inconsistencies in measurement and modeling approaches across studies may have contributed to the mixed findings in this area. For example, the skills that make up EF are frequently referred to and measured inconsistently and interchangeably, creating what has been referred to as "Conceptual Clutter" and "Measurement Mayhem" (Morrison & Grammer, 2016). Additionally, the diversity of EF measures used, within and across disciplines, has created barriers to achieving an agreed-upon definition of these important skills (e.g., Jones et al., 2016). Further, the differences in model specification across studies examining the concept of mutualism may have also contributed to the inconsistencies in findings in the current literature (Camerota, Willoughby, & Blair, 2020; Rhemtulla, van Bork, & Borsboom, 2020). For example, some studies treat EF as a composite variable, (e.g., Welsh et al., 2010) whereas others model EF as a latent variable (e.g., Schmitt et al., 2017). Finally, the developmental window during which these constructs are measured differ considerable across the studies in this area, including, but not limited to the frequency that EF and academic skills are measured in early childhood (Fuhs et al., 2014; Schmitt et al., 2017).

The Present Study

Given the inconsistencies in the literature on the role of EF and academic skills and the continued emphasis on interventions of EF to improve academic outcomes, it is important to try to replicate the findings of the co-development of these constructs. Therefore, the goal of the present study is to conduct a conceptual replication of the Schmitt et al. (2017) study that found important reciprocal relations between EF and academic achievement outcomes by leveraging data from two independent longitudinal studies. We chose to replicate Schmitt et al. (2017) for several reasons. First, the overlap in the EF and academic achievement measures used across the three samples allowed us to replicate the Schmitt et al. (2017) study using similar measures of EF and academic skills. Second, the timing of measurement (i.e., preschool through the end of kindergarten) is similar across the three samples, which will allow us to test the concept of mutualism (co-development) between EF and academic skills across a similar developmental window (i.e., 4.5 years − 6.5 years old). Finally, the two independent samples included a sufficient number of EF measures to allow for similar model specifications (e.g., latent variable modeling) as Schmitt and colleagues (2017).

Schmitt et al. (2017) examined longitudinal relations between EF, math, and literacy using ARCL modeling. In the original investigation, Schmitt and colleagues (2017) reported bidirectional associations between EF and math achievement, but not literacy, across the preschool year. However, EF prospectively predicted literacy achievement from the spring of preschool to the fall of kindergarten. Further, they found that EF prospectively predicted math, but not literacy achievement across the kindergarten school year. Based on these findings, we expect to see bidirectional associations between EF and math and literacy achievement from preschool to kindergarten (STAR sample) and that EF would prospectively predict math from kindergarten to first grade (STAR and Pathways sample).

Method

Participants and Procedures

Schmitt et al. (2017) Dataset

Data were collected on a total of 435 children in the Pacific Northwest, U.S. The study consisted of four waves of data collection; children were assessed in the fall of preschool, spring of preschool, fall of kindergarten, and spring of kindergarten. On average, children were 4.70 years old (*SD* = 0.30) at the beginning of the study, and 51% were male. This sample consisted of 63% White children, 19% Latino/Hispanic children, 13% multiracial children, 3% Asian/Pacific Islander children, and 2% other ethnicities. In the fall of preschool, 55% of children were enrolled in Head Start and 15% were primarily Spanish speakers. At each wave, children was assessed on a battery of EF, literacy, and math measures.

Children were recruited from schools using a convenience sampling approach, such that, schools and children that were accessible and willing to participate were included in the study. Parents of children signed a written informed consent letter agreeing for their child to participate in the study. Children were assessed in two to three sessions that lasted 10 to 15 minutes each. For more information about this sample, refer to Schmitt et al. (2017).

STAR Dataset

This project was an extension of a larger longitudinal project on trajectories of early academic development. Data were collected on a total of 278 children in a Southeastern U.S. city. The study consisted of three waves of data collection; children were assessed in preschool, kindergarten, and first grade. On average, children were 4.67 years old (*SD* = 0.42) at the beginning of the study, and 55% were male. This sample consisted of 60% White children, 28% Black children, 2% Asian children, and 10% multiracial children. The sample broadly represented the region in which the children were recruited. All children had no known developmental disorders.

Children were recruited from libraries, daycare centers, and local establishments. Data collection took place as laboratory visits which lasted approximately two hours. During these visits, children participated a number of tasks that assessed cognitive and emotional development. Each child was assessed on a battery of executive function and achievement measures by a trained experimenter. Parents received monetary compensation for their time, and children

selected a small toy at the completion of the visit. All procedures were approved by the university institutional review board.

Pathways Dataset

This project was an extension of a larger longitudinal project studying the effect of schooling on executive function development. A total of 367 children participated in the larger longitudinal project; however, 88 children were recruited in either first or second grade, and were therefore excluded from the current sample. Thus, the full sample for the current study consisted of 279 children attending seven elementary schools in Midwestern U.S. cities. The sample included three cohorts of children who were assessed across the fall and winter of kindergarten and first grade. On average, children were 5.38 years old (*SD* = 0.10) when first tested, 47% were male. Although child level race, ethnicity, and socioeconomic status was not collected, all children were recruited from racially and socioeconomically diverse schools. Schools included in this sample served children from a broad range of socioeconomic backgrounds based on school-wide percentages of free or reduced-price lunch (FRPL; 2% - 71.9%).

Children in this sample were recruited from schools using a convenience sampling approach, similar to Schmitt et al. (2017), such that schools and children that were accessible and willing to participate were included in the study. Parents of children signed a written informed consent letter agreeing for their child to participate in the study. Children were individually assessed in schools outside their classrooms for a 45-minute period. During these assessments, each child was assessed on a battery of executive function and achievement measures by a trained experimenter. The order and versions of assessments were counterbalanced, as there were two different orderings of assessments and two different versions of each of the assessments. Children received a bookmark with stickers at the completion of the visit. All procedures were approved by the university institutional review board.

Measures

Executive Function

A variety of children's executive function skills were assessed in each sample. Both STAR and Pathways samples included executive function measures of working memory and inhibitory control, as well as an additional measure. The additional measure in the STAR sample included a cognitive flexibility measure, and the additional measure in the Pathways sample included one global executive function measure. See Table 1 for a summary of overlapping variables across all datasets.

Table 1

Summary of Samples and Measures Across Datasets

Variable	Schmitt et al. (2017)	STAR	Pathways
N	424	277	279
Covariates	ELL, Head Start, Age	Age	Age
Waves	Fall PK, Spring PK, Fall K, Spring K	PK, K, 1	K, 1
Achievement			
Literacy	Letter Word ID	Letter Word ID	Letter Word ID
Math	Applied Problems	Applied Problems	Applied Problems
Executive Function			
Working Memory	Auditory Working Memory	Numbers Reversed	Digit Span
Cognitive Flexibility	Card Sort (traditional)	Card Sort (computer)	
Inhibitory Control	Simon Says	Go-No/Go (d')	Zoo Go-No/Go (d')
All	HTKS		HTKS

Note. ELL = English Language Learner; PK = Preschool; K = Kindergarten; G1 = First Grade; HTKS = Head-Toes-Knees-Shoulders.

Schmitt et al. (2017) Executive Function

Auditory Working Memory — Children's working memory was measured using the Auditory Working Memory subtest from the Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock, McGrew, & Mather, 2001). Participants were instructed to repeat back to the experimenter things and numbers in a specific order. An overall accuracy score was calculated by adding children's correct responses (each correct trial = 1 point). See Schmitt et al. (2017) for more information on this task.

Simon Says — Children's inhibitory control was measured using the Simon Says task (Carlson, 2005; Strommen, 1973). The experimenter asked children to perform an action only if the experimenter says, "Simon says", otherwise the child should remain still. For more information on how this task was scored, see Schmitt et al. (2017).

Card Sort — Children's cognitive flexibility was measured using a Card Sort task similar to the traditional Dimensional Change Card Sort task (Blackwell, Cepeda, & Munakata, 2009; Frye, Zelazo, & Palfai, 1995; Zelazo, 2006). The experimenter asked children to sort colored picture cards of a dog, fish, or bird on the basis of three dimensions: color, shape, and size. See Schmitt et al. (2017) for more information on this task.

Head-Toes-Knees-Shoulders (HTKS) — The HTKS task was used to measure all of children's executive function skills through gross motor responses: working memory, inhibitory control, and cognitive flexibility (McClelland & Cameron, 2012; McClelland et al., 2014). Children were told they were going to play a game in which they must do the opposite of what the examiner's directions say, varying from touching your head, toes, knees, or shoulders. For example, if the trained examiner said, "touch your head" children were expected to touch their toes. The task grows in difficulty across three sections of questions in which the rules change. If children responded incorrectly, they were given a score of 0. If children responded correctly, they were given a score of 2, and if children self-corrected their response, they were given a score of 1. For more information on this task, see Schmitt et al. (2017).

STAR Executive Function

Numbers Reversed — Children's working memory capacity was measured using the Numbers Reversed subtest of The Woodcock-Johnson III (Woodcock et al., 2001). Participants were instructed to listen to the experimenter recite a string of numbers (beginning with two numbers and gradually increasing) and then repeat the numbers in reverse order. An overall accuracy score was calculated by adding children's correct responses (each correct trial = 1 point).

Go/No-Go — A computer-based Go/No-Go paradigm was used to assess children's inhibitory control and sustained attention. Children were asked to press a button each time they saw an animal, except for when they saw a dog (Lahat, Todd, Mahy, Lau, & Zelazo, 2010). There were a total of 144 trials (75% Go). A discriminability index (d' = Z(Correct/Hit) − Z (Incorrect/False Alarm)) was used to assess the participants' ability to distinguish signals from noise (Stanislaw & Todorov, 1999).

Dimensional Change Card Sort (DCCS) — Cognitive flexibility (also known as task shifting) was measured using a computerized version of The Dimensional Change Card Sort task (Espinet, Anderson, & Zelazo, 2012). In the pre-switch block, children were asked to sort the stimuli according to their shape (15 trials). In the post-switch block, children were asked to sort the stimuli according to color (30 trials). The post-switch was followed by a "borders" block in which children were instructed to sort stimuli on one dimension (color) if the picture had a border around it but the other dimension (shape) if the picture did not have a border (12 trials). Percent accuracy was computed for each block and weighted averages were created as follows: Preschool: 33.3% pre-switch, 66.7% post-switch; kindergarten & 1st grade: 25% pre-switch, 50% post-switch, 25% borders. Higher scores indicated greater cognitive flexibility.

Several outcome measures from this dataset were previously published, thus for further information regarding these measures, see Isbell, Calkins, Swingler, and Leerkes (2018), Isbell, Calkins, Cole, Swingler, and Leerkes (2019), Zeytinoglu, Leerkes, Swingler, and Calkins (2017) and Zeytinoglu, Calkins, and Leerkes (2019). However, this publication differs from the previous publications as Isbell and colleagues (2018, 2019) only used the Go/No-Go task and the WJ

subtests in their work, and even though Zeytinoglu and colleagues (2017, 2019) used the EF measures, they did not investigate links between EF and academic outcomes.

Pathways Executive Function

Digit Span Backward — Children's working memory was assessed using the Digit Span—Backward subtest of the McCarthy Scales for Children's Abilities (McCarthy, 1972). Participants were read a sequence of numbers (beginning with two numbers and gradually increasing), and asked to repeat the same sequence back to the examiner in reverse order.

Go/No-Go — A Go-No Go paradigm called the Zoo Game (Grammer, Carrasco, Gehring, & Morrison, 2014) was used to assess children's inhibitory control and sustained attention. Children were told to press a button each time they saw an animal, except for when they saw an orangutan. There were a total of 320 trials (75% Go). A discriminability index (d' = Z(Correct/Hit) – Z (Incorrect/False Alarm)) was used to assess participants' ability to distinguish signals from noise (Stanislaw & Todorov, 1999). Larger values of d' indicate better task performance.

Head-Toes-Knees-Shoulders (HTKS) — The Head-Toes-Knees-Shoulders (HTKS; Cameron et al., 2008) task was used to assess children's cognitive flexibility, working memory, and inhibitory control through gross motor responses (McClelland & Cameron, 2012; McClelland et al., 2014). This EF measure was the same measure used in the Schmitt et al. (2017) study.

Academic Achievement

Mathematics — The standardized Applied Problems subtest of the Woodcock-Johnson III Tests of Achievement (WJ-AP; Woodcock, McGrew, & Mather, 2001) was used to assess individual mathematical skills. The Applied Problems task assesses children on numerous early math skills such as counting, representational arithmetic, abstract arithmetic, and the ability to read a clock. Items increase in difficulty as children progress through the task, and basal and ceiling levels are determined for each student. The WJ-AP was counterbalanced (Form A or Form B) such that children would be less likely to remember questions from the year before and completed by children at all waves.

Literacy — The standardized Letter-Word Identification subtest of the Woodcock-Johnson III Tests of Achievement (WJ-LWID; Woodcock, McGrew, & Mather, 2001) was used to assess children's literacy skills. The WJ-LWID subtest assessed children's ability to read letters and words in both expressive and receptive language. Items in this task were also ranked in order of difficulty, and basal and ceiling levels were determined for each student. This task was completed by children at all waves.

Covariates

In an effort to replicate the results from the original study as closely as possible, we also considered which covariates should be included. The original analyses, Schmitt et al. (2017), included English Language Learners (ELL), Head Start enrollment, and age as covariates. The STAR dataset did not have Head Start enrollment, but there were 11 children for whom English was not the primary language spoken at home. Language spoken at home did not relate to the independent or dependent variables in our analyses (p = .09-.99) and did not predict attrition in kindergarten (χ^2 = 140, p =.93) or grade 1 assessments (χ^2 =1.03, p = .60). Based on these preliminary analyses and the small percentage of children who were ELL (4%), we did not include ELL as a covariate in our analyses. In the replication analyses, the Pathways dataset did not include ELL or Head Start enrollment. Thus, age was the only covariate common to all three datasets (STAR, Pathways, and Schmitt et al. (2017)).

Analytic Approach

Similar to Schmitt et al. (2017), all analyses were conducted in Mplus (Muthén & Muthén, 1998-2015). We used full information maximum likelihood estimation (FIML) to handle missing data to reduce potential bias in the parameter

estimates (Enders & Bandalos, 2001). This permitted the inclusion of all participants with data on one or more variables. Due to the missing data and potential departures from multivariate normality, the model was estimated using a robust maximum likelihood estimator (MLR). We used ARCL models to examine longitudinal relations between EF, math, and literacy achievement. In both datasets, we first specified an initial longitudinal confirmatory factor analysis (CFA) model, controlling for participants' age at first testing to examine the fit and factor loadings of the latent EF factors at each wave. In STAR, the EF latent factors consisted of Numbers Reversed, DCCS, and Go-No/Go. In the initial CFA models, we scaled all latent factors by fixing the latent means to zero and latent variables to one. In Pathways, the EF latent factors consisted of HTKS, Digit Span Backward, and Go-No/Go.

Next, we examined the measurement invariance of the EF construct across waves to understand if EF was measured in a consistent way across time. Consistent with Schmitt et al. (2017), we first tested weak factorial invariance (also called metric invariance) to examine the degree to which the specific EF indicators (e.g., working memory) loaded on to the EF constructs equally across time. This was tested by equating the same EF indicators' loadings across wave. Next, we tested strong factorial invariance (also called scalar invariance) to understand whether the EF construct was measured on the same interval or ratio across time. This was tested by equating the same EF indicators' intercepts across waves. Although there are different strategies for evaluating measurement invariance, we followed the approach used in Schmitt et al. (2017) given our goal of replicating their study. Thus, consistent with Schmitt and colleagues, the models were compared to the initial CFA model and were rejected if the Comparative Fit Index (CFI) decreased by more than .01 (Chen, 2007), and if full weak or full strong factorial invariance led to a decrease in model fit, partial measurement invariance was tested by freeing at most two parameters. For a further discussion, see Little (1997) and Schmitt et al. (2017).

After establishing longitudinal measurement invariance for the EF factors, we specified a longitudinal structural equation model (SEM) with math and literacy as manifest variables. The SEM includes single-lag stability regressions and single-lag cross-construct regressions. Executive function factors, math, and literacy were all allowed to covary. All models included child age at initial assessment as time-invariant covariates. Similar to Schmitt et al. (2017), model fit was adequate based on appropriate fit statistics including Comparative Fit Index (CFI) and Tucker Lewis Index (TLI) between 0.95 and 1.00 (Hu & Bentler, 1999; Kline, 2005), and the Root Mean Square Error of Approximation (RMSEA) less than 0.06 (Hu & Bentler, 1999).

Sensitivity Power Analyses

The STAR and Pathways samples are existing datasets, thus, a sensitivity power analysis was used to calculate the minimally detectable effect sizes (MDES) given the sample sizes for all statistical analyses (Cribbie, Beribisky, & Alter, 2019; Giner-Sorolla et al., 2019). This provides some context for why we see different rates of significance across the studies for given effect sizes. In the STAR dataset, with three latent variables, six observed variables, 277 participants, α = .05, and power $(1-\beta)$ = .80, the sensitivity power analysis suggested that the MDES was 0.21 (Soper, 2020). In the Pathways dataset, with two latent variables, four observed variables 279 participants, α = .05, and power $(1-\beta)$ = .80, the sensitivity power analysis suggested that the MDES was 0.18 (Soper, 2020). Whereas Schmitt et al. (2017) reported an effect size as small as .11 as significant in their sample, our sensitivity power analyses suggest that neither the STAR nor Pathways datasets are powered to detect effect sizes under .18 as significant.

Results

Descriptive statistics for all three datasets are presented in Tables 2 and 3. Correlation tables for both the STAR and Pathways studies can be found in the Appendix.

Table 2

Descriptive Statistics of Preschool Waves

Variable	Schmitt et al. Wave 1			STAR Wave 1				Schmitt et al. Wave 2		
	N	*M*	*SD*	*N*	*range*	*M*	*SD*	*N*	*M*	*SD*
Age	424	4.70	0.30	277	3.75 − 5.83	4.69	0.39	394	5.15	0.30
Achievement										
Literacy	408	335.65	26.59	278	276 − 494	348.98	28.87	390	349.33	26.80
Math	401	410.17	23.30	278	332 − 471	416.18	19.52	391	419.83	23.11
Executive Function										
Working Memory	400	450.30	14.80	276	403 − 489	424.50	27.22	385	456.17	17.97
Cognitive Flexibility	409	13.64	6.67	274	28.89 − 100	75.53	22.68	389	16.49	5.92
Inhibitory Control	408	0.14	0.28	264	-0.08 − 5.39	2.23	1.04	387	0.29	0.38
All	403	17.41	17.20	−	−	−		391	25.15	18.28

Note. Schmitt et al. Wave 1 = fall of preschool and Wave 2 = spring of preschool. STAR Wave 1 = preschool. Pathways dataset did not include a preschool wave. Working Memory = Auditory Working Memory in Schmitt et al., Numbers Reversed in STAR, and Digit Span in Pathways. Cognitive Flexibility = DCCS traditional in Schmitt et al., and DCCS computer in STAR. Inhibitory Control = Simon Says in Schmitt et al., Go-No/Go (d') in STAR, and Zoo Go-No/Go (d') in Pathways. All = HTKS in Schmitt et al. and Pathways.

Table 3

Descriptive Statistics of Kindergarten and First Grade Waves

Variable	Schmitt et al. Wave 3			STAR Wave 2				Pathways Wave 1			
	N	*M*	*SD*	*N*	*range*	*M*	*SD*	*N*	*range*	*M*	*SD*
Age	308	5.67	0.30	249	5.08 − 6.75	5.90	0.32	279	4.93 − 6.94	5.71	0.39
Achievement											
Literacy	305	366.00	29.14	249	326 − 519	407.78	35.07	275	283 − 507	385.57	33.96
Math	305	431.02	20.71	249	372 − 490	441.02	17.68	273	333 − 481	430.50	18.77
Executive Function											
Working Memory	303	464.60	19.21	249	403 − 516	457.02	26.81	276	0 − 8	2.10	1.65
Cognitive Flexibility	307	18.60	4.88	248	13.75 − 100	81.47	15.88	−	−	−	
Inhibitory Control	307	0.45	0.39	245	0.79 − 5.39	2.96	0.92	202	-0.41 − 3.68	1.67	.79
All	303	33.17	17.74	−	−	−		277	0 − 58	30.75	16.89

Note. Schmitt et al. Wave 1 = fall of preschool and Wave 2 = spring of preschool. STAR Wave 1 = preschool. Pathways dataset did not include a preschool wave. Working Memory = Auditory Working Memory in Schmitt et al., Numbers Reversed in STAR, and Digit Span in Pathways; Cognitive Flexibility = DCCS traditional in Schmitt et al., and DCCS computer in STAR; Inhibitory Control = Simon Says in Schmitt et al., Go-No/Go (d') in STAR, and Zoo Go-No/Go (d') in Pathways; All = HTKS in Schmitt et al. and Pathways.

Table 3

Cont.

Variable	Schmitt et al., Wave 4			STAR Wave 3				Pathways Wave 2			
	N	*M*	*SD*	*N*	*range*	*M*	*SD*	*N*	*range*	*M*	*SD*
Age	299	6.17	0.29	257	6.17 − 8.00	6.96	0.35	168	6.14 − 8.05	6.84	.34
Achievement											
Literacy	295	400.24	35.21	240	374 − 530	455.78	31.25	164	348 − 525	448.70	33.36
Math	295	442.09	19.29	240	423 − 507	460.35	15.84	165	409 − 502	457.82	16.46
Executive Function											
Working Memory	294	473.18	19.90	239	403 − 522	475.16	20.93	167	0 − 8	3.47	1.35
Cognitive Flexibility	295	19.78	3.88	240	35.42 − 100	88.71	10.76	–	–	–	
Inhibitory Control	294	0.54	0.38	239	1.02 − 5.39	3.41	0.93	124	-0.35 − 3.99	2.26	0.81
All	296	39.20	16.00	–	–	–		166	0 − 59	44.29	11.31

Note. Schmitt et al. Wave 1 = fall of preschool and Wave 2 = spring of preschool. STAR Wave 1 = preschool. Pathways dataset did not include a preschool wave. Working Memory = Auditory Working Memory in Schmitt et al., Numbers Reversed in STAR, and Digit Span in Pathways; Cognitive Flexibility = DCCS traditional in Schmitt et al., and DCCS computer in STAR; Inhibitory Control = Simon Says in Schmitt et al., Go-No/Go (d') in STAR, and Zoo Go-No/Go (d') in Pathways; All = HTKS in Schmitt et al. and Pathways.

Confirmatory Factor Analyses and Measurement Invariance

In both datasets, the initial CFA of the EF variables fit the data well (see Tables 4 and 5), such that all factor loadings were above 0.40 (Stevens, 1992), and all factor loadings were statistically significant for all indicators at each wave (all *p*s < .05). The initial tests of weak factorial invariance substantially decreased model fit in both datasets (STAR: Δ CFI = .02; Pathways: Δ CFI = .03). Thus, for each dataset, we assessed partial weak and partial strong factorial invariance across EF factors.

In the STAR dataset, freely estimating the numbers reversed factor loading for wave 1 resulted in a model that supported partial weak invariance (Δ CFI = -.00; Δ BIC = 12.99). Although freeing indicators other than numbers reversed could also result in partial weak invariance, one reason why we chose to free this indicator was because its standardized factor loading seemed to be smaller (β = .48) than the factor loadings at the subsequent waves (β = .68 & .55) in the unconditional model (see Table 4). This was likely because there was less variability in the distribution of numbers reversed in the first wave compared to the subsequent two waves, given the difficulty of this task for some preschoolers. Moreover, the numbers reversed and DCCS intercepts were freely estimated across waves, resulting in partial strong invariance (Δ CFI = -.00; Δ BIC = 13.22). Numbers reversed was freely estimated to be consistent with the weak invariance decision. In addition to numbers reversed, we chose to freely estimate DCCS intercepts because the DCCS task at the second and third waves also included the "borders" block, whereas this block was not included in the first wave and thus this change has likely affected the scale of the latent variable across time. In the Pathways dataset, freely estimating the Zoo Go/No-Go factor loading for wave 1 resulted in a model that supported partial weak invariance (Δ CFI = .00; Δ BIC = 4.70). Similar to the STAR sample, we chose to free this indicator because its standardized factor loading seemed to be smaller (β = .41) than the factor loadings at the subsequent waves (β = .82 & .51) in the unconditional model (see Table 4). Moreover, we freely estimated the Zoo Go/No-Go intercept across waves to be consistent with the weak invariance decision, thus resulting in partial strong invariance (Δ CFI = .01; Δ BIC = 11.74). Thus, both STAR and Pathways samples demonstrated partial weak and partial strong measurement invariance in EF latent construct across time, suggesting that the EF constructs showed an acceptable level of measurement equivalence across time.

PsychOpen GOLD

Table 4

Factor Loadings for Unconditional CFA Models

Construct	Standardized Loading (*SE*)		
Indicator	Schmitt et al.	STAR	Pathways
Wave 1 EF			
Working Memory	.34* (.04)	.48* (.07)	
HTKS	.45* (.06)		
Cognitive Flexibility	.59* (.04)	.68* (.06)	
Inhibitory Control	.45* (.04)	.55* (.08)	
Wave 2 EF			
Working Memory	.40* (.03)		
HTKS	.58* (.04)		
Cognitive Flexibility	.55* (.05)		
Inhibitory Control	.52* (.04)		
Wave 3 EF			
Working Memory	.41* (.03)	.55* (.08)	.51* (.08)
HTKS	.63* (.04)		.82* (.11)
Cognitive Flexibility	.52* (.05)	.77* (.10)	
Inhibitory Control	.52* (.05)	.44* (.08)	.41* (.08)
Wave 4 EF			
Working Memory	.54* (.04)	.51* (.09)	.47* (.12)
HTKS	.66* (.04)		.68* (.13)
Cognitive Flexibility	.59* (.04)	.63* (.11)	
Inhibitory Control	.54* (.04)	.38* (.09)	.55* (.13)

Note. Working Memory = Auditory Working Memory in Schmitt et al., Numbers Reversed in STAR, and Digit Span in Pathways; Cognitive Flexibility = DCCS traditional in Schmitt et al., and DCCS computer in STAR; Inhibitory Control = Simon Says in Schmitt et al., Go-No/Go (d') in STAR, and Zoo Go-No/Go (d') in Pathways; All = HTKS in Schmitt et al. and Pathways.
*p < .05.

Table 5

Model Fit for Unconditional CFA Models

Dataset	χ2	*df*	RMSEA	CFI	TLI
STAR	40.28	21	0.06	0.96	0.92
Pathways	16.44	9	0.05	0.96	0.91

Autoregressive Cross-Lagged Models

ARCL models were tested and examined using the EF latent variables and math and literacy variables. Syntax is available at https://osf.io/5twgv/. The structural component and standardized results of the final models are presented in Figure 1 and 2. A synopsis of standardized coefficients for both autoregressive and cross-lagged paths are summarized across all datasets in Table 6. For the structural component and standardized results of the original article, see Figure B.3 in Schmitt et al. (2017).

Table 6

Summary of ARCL Associations Between Executive Function and Achievement Across All Datasets

	Schmitt et al.		STAR			Pathways		
Coefficient Type	*ß*	**Coefficient Type**	*ß*	*SE*	*p*	*ß*	*SE*	*p*
Cross-lagged coefficients								
$EF_{PKFall} \rightarrow Math_{PKSpring}$	0.33***							
$EF_{PKFall} \rightarrow Literacy_{PKSpring}$	-0.08							
$Math_{PKFall} \rightarrow EF_{PKSpring}$	0.22**							
$Literacy_{PKFall} \rightarrow EF_{PKSpring}$	0.03							
$Math_{PKFall} \rightarrow Literacy_{PKSpring}$	0.16*							
$Literacy_{PKFall} \rightarrow Math_{PKSpring}$	0.03							
$EF_{PKSpring} \rightarrow Math_{KFall}$	0.20*	$EF_{PK} \rightarrow Math_{K}$	0.84*	.39	.03			
$EF_{PKSpring} \rightarrow Literacy_{KFall}$	0.16*	$EF_{PK} \rightarrow Literacy_{K}$	0.30	.25	.25			
$Math_{PKSpring} \rightarrow EF_{KFall}$	0.25**	$Math_{PK} \rightarrow EF_{K}$	0.18	.27	.50			
$Literacy_{PKSpring} \rightarrow EF_{KFall}$	-0.06	$Literacy_{PK} \rightarrow EF_{K}$	0.12	.11	.26			
$Math_{PKSpring} \rightarrow Literacy_{KFall}$	-0.08	$Math_{PK} \rightarrow Literacy_{K}$	-0.01	.19	.96			
$Literacy_{PKSpring} \rightarrow Math_{KFall}$	0.06	$Literacy_{PK} \rightarrow Math_{K}$	0.03	.09	.77			
$EF_{KFall} \rightarrow Math_{KSpring}$	0.39***	$EF_{K} \rightarrow Math_{G1}$	0.39**	.15	.01	0.66***	.17	< .001
$EF_{KFall} \rightarrow Literacy_{KSpring}$	0.08	$EF_{K} \rightarrow Literacy_{G1}$	0.12	.11	.30	0.33*	.13	.01
$Math_{KFall} \rightarrow EF_{KSpring}$	0.04	$Math_{K} \rightarrow EF_{G1}$	0.18	.23	.44	0.27	.22	.22
$Literacy_{KFall} \rightarrow EF_{KSpring}$	0.02	$Literacy_{K} \rightarrow EF_{G1}$	0.04	.08	.61	-0.04	.11	.69
$Math_{KFall} \rightarrow Literacy_{KSpring}$	0.16*	$Math_{K} \rightarrow Literacy_{G1}$	0.05	.10	.58	-0.01	.11	.92
$Literacy_{KFall} \rightarrow Math_{KSpring}$	0.11*	$Literacy_{K} \rightarrow Math_{G1}$	0.19**	.06	.002	0.24*	.10	.02
Autoregressive coefficients								
$EF_{PKFall} \rightarrow EF_{PKSpring}$	0.75***							
$Math_{PKFall} \rightarrow Math_{PKSpring}$	0.50***							
$Literacy_{PKFall} \rightarrow Literacy_{PKSpring}$	0.74***							
$EF_{PKSpring} \rightarrow EF_{KFall}$	0.76***	$EF_{PK} \rightarrow EF_{K}$	0.67*	.33	.04			
$Math_{PKSpring} \rightarrow Math_{KFall}$	0.62***	$Math_{PK} \rightarrow Math_{K}$	-0.03	.32	.93			
$Literacy_{PKSpring} \rightarrow Literacy_{KFall}$	0.77***	$Literacy_{PK} \rightarrow Literacy_{K}$	0.49***	.07	< .001			
$EF_{KFall} \rightarrow EF_{KSpring}$	0.86***	$EF_{K} \rightarrow EF_{G1}$	0.74**	.25	.003	0.69**	.22	.002
$Math_{KFall} \rightarrow Math_{KSpring}$	0.40***	$Math_{K} \rightarrow Math_{G1}$	0.30*	.14	.03	-0.05	.15	.74
$Literacy_{KFall} \rightarrow Literacy_{KSpring}$	0.69***	$Literacy_{K} \rightarrow Literacy_{G1}$	0.71***	.05	< .001	0.54***	.07	< .001

Note. N(Schmitt et al.) = 424, N(STAR) = 277, N(Pathways) = 279; EF = Executive Function; PK = Preschool, K = Kindergarten, G1 = First Grade.
*p < .05. **p < .01. ***p < .001.

STAR ARCL

The longitudinal ARCL fit the data well, χ^2 = 106.60, df = 69, CFI = .98, TLI = .96, RMSEA = .04. First, within wave correlations demonstrated a similar pattern as mentioned above in which the first wave correlations between math, literacy, and EF were large and statistically significant. In particular, the correlation between EF and math achievement was very large (r = .87). However, the later within-wave correlations were smaller. Second, the factor stabilities for literacy and EF were all significant and above β = .49. The factor stability for math was close to zero from wave one to wave two, but then moderate and statistically significant from wave two to wave three (β = .30, SE = .14, p = .03). Third, the cross-lagged paths demonstrated that higher executive functioning predicted higher math achievement from wave one to wave two, as well as wave two to wave three. Higher executive functioning was not associated with higher literacy achievement at either wave. Furthermore, higher literacy achievement at the second wave predicted higher math achievement at the third wave, but not at the wave prior. Finally, higher math was not significantly associated with higher executive functioning at wave two or wave three.

Figure 1

Path Diagram for STAR Data Final Structural Replication Model, Controlling for Age (Standardized Coefficients)
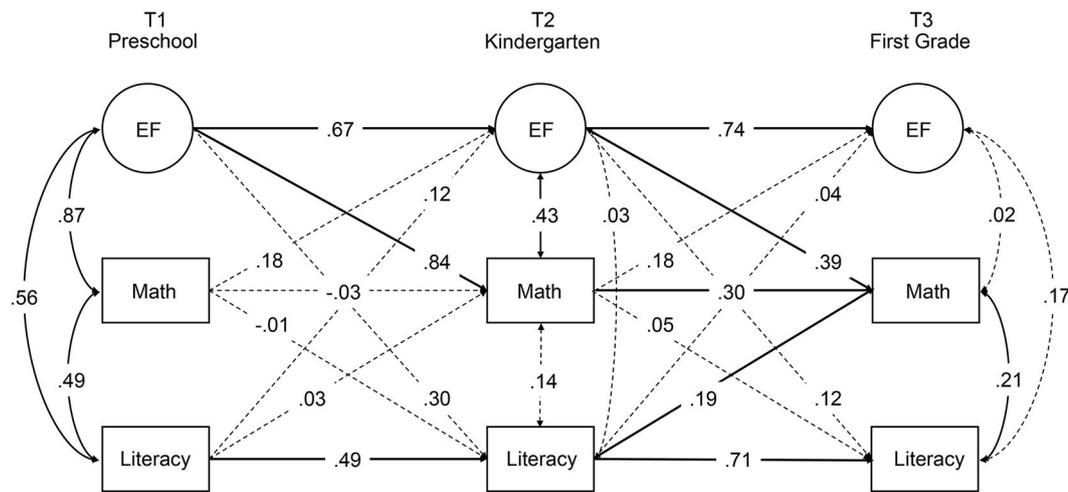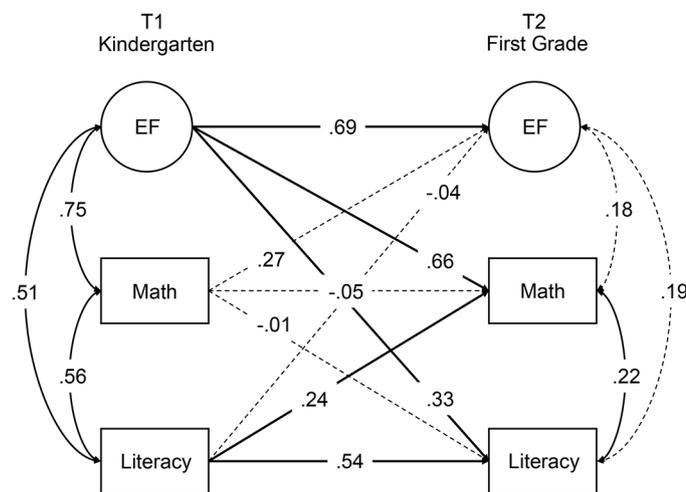

Note. Solid lines indicate significant coefficients ($p < .05$), dashed lines indicate non-significant coefficients. $\chi2 = 106.60$, $df = 69$, CFI = .98, TLI = .96, RMSEA = .04.

Pathways ARCL

Results for the Pathways sample also suggested the longitudinal ARCL model fit the data well, $\chi^2 = 51.94$, $df = 27$, CFI = .96, TLI = .93, RMSEA = .06. First, results suggest the first wave correlations between math, literacy, and EF were large and statistically significant. However, the second wave correlations were much smaller. Second, the factor stabilities for literacy and EF were above .50 for both factors, whereas the stability for math was close to zero. Third, when considering the cross-lagged paths, higher executive functioning at wave one predicted higher literacy and math achievement at wave two. Further, higher literacy achievement at wave one predicted higher math achievement at wave two. In contrast, higher math at the first wave was essentially unrelated to literacy achievement at wave two ($\beta = -.01$, $SE = .11$, $p = .92$), and not significantly related to wave two executive functioning.

Figure 2

Path Diagram for Pathways Data Final Structural Replication Model, Controlling for Age (Standardized Coefficients)

Note. Solid lines indicate significant coefficients ($p < .05$), dashed lines indicate non-significant coefficients. $\chi2 = 51.94$, $df = 27$, CFI = .96, TLI = .93, RMSEA = .06.

Discussion

The goal of the present study was to provide a conceptual replication of the ARCL findings reported by Schmitt and colleagues (2017). The original study revealed partial measurement invariance of the EF latent variables over time, strong autoregressive paths across all constructs, and found bidirectional relations between EF and math and literacy achievement from preschool to kindergarten, and unidirectional relations from EF to math across fall and spring of kindergarten. Consistent with Schmitt et al. (2017), we found that the EF latent variables demonstrated partial measurement invariance over time across both replication datasets. Further, the autoregressive estimates across EF latent and observed literacy variables across all datasets was moderate in strength, suggesting longitudinal construct stability.

However, the stability of math achievement from preschool to kindergarten (STAR study) and from kindergarten to first grade (Pathways study) was close to zero in the ARCL models, which is a significant departure from the stability estimates reported in the original study. Additionally, across both samples, we failed to replicate the bidirectional pattern of findings between EF, math, and literacy achievement reported in the original study. Specifically, in the STAR sample, we found unidirectional associations from EF to math achievement, such that preschool EF prospectively predicted math (but not literacy) achievement at kindergarten, and kindergarten EF predicted math (but not literacy) achievement at first grade. In the Pathways sample, we replicated the observed unidirectional relations found in Schmitt et al. (2017), such that kindergarten EF prospectively predicted math and literacy achievement at the beginning of first grade. Math and literacy skills, however, did not predict the EF latent factor in either sample.

The inconsistent pattern of findings in our replication study mimic the inconsistencies that are found in the EF and academic skills literature during early childhood. In addition to the results reported by Schmitt et al. (2017), several recent studies have also found evidence of the co-development of EF and academic achievement (e.g., Cameron et al., 2019; McKinnon & Blair, 2019; Meixner et al., 2019; Miller-Cotto & Byrnes, 2020; Welsh et al., 2010). However, others (Fuhs et al., 2014; Willoughby et al., 2019), including the current replication have demonstrated that EF prospectively predicts academic skills, which is inconsistent with the theory of mutualism between cognitive and academic skills across early development (e.g., Peng & Kievit, 2020).

There are several potential reasons for the inconsistency of findings across this body of research and the present replication study in particular. First, it is not clear the degree to which EF constructs measured across studies are capturing the same underlying skills (see Morrison & Grammer, 2016, for review). For example, although many studies adopt a tripartite model of EF, consisting of inhibitory control, working memory/updating, and cognitive flexibility/shifting (e.g., Diamond, 2013; Miyake et al., 2000), others include a broader range of EF-related constructs, such as impulsivity, inattention, and behavioral self-control (e.g., Fuhs et al., 2014) or do not include full coverage of subcomponents considered part of the broader EF umbrella (e.g., Miller-Cotto & Byrnes, 2020; Willoughby et al., 2019). Although we relied on a similar set of EF measures used in Schmitt et al. (2017) for the present replication study, there was not a complete overlap in the measures used across the three samples. For example, the Schmitt et al. (2017) study used the Simon Says task, whereas the STAR and Pathways studies included two child friendly versions of a Go/No-Go task to measure inhibitory control. Therefore, it is possible that the lack of uniformity of EF measurement approaches may partially explain the inconsistent findings in the present study, which is also considered a noted limitation in the area of early childhood EF research more broadly (see Morrison & Grammer, 2016).

Another potential reason we were unable to replicate the bidirectional cross-lagged associations between EF and academic achievement reported in the original paper could be due to the number of time-points across the three samples. Specifically, while children in Schmitt et al. (2017) were assessed within a narrow window during the fall and spring of their preschool and kindergarten years, children in the replication samples were tested once a year from preschool to first grade (STAR) and during kindergarten and first grade (Pathways). This difference in timing could have also contributed to our inability to replicate the bidirectional cross-lagged effects from preschool to kindergarten, as children in the Schmitt et al. (2017) study were sampled across the entire school year. The more fine-grained sampling procedure in Schmitt et al. (2017) allowed them to identify changes in the relations between EF and math achievement (i.e., mutual relations at preschool, EF → math at the end of kindergarten), which they attribute to changes in the complexity of math instruction during the kindergarten school year. Our sampling approach did not allow for a direct

test of this hypothesis, as neither the STAR nor Pathways study included fall and spring testing occasions across preschool and kindergarten. These inconsistent findings suggest that directional patterns of relations between EF and academic skills might, in part, depend on the number of measurement occasions studied across early development.

Finally, differences in sample size and sample characteristics could also explain our inability to replicate the Schmitt et al. (2017) findings. It is possible that subtle effects were not detected due to a lack of statistical power in our replication studies. Specifically, our sensitivity analyses suggested that effect sizes under .21 were not detectable in the STAR dataset due to our sample size, which may have affected our interpretations of the bidirectional relations between EF and math achievement. Future research using larger replication samples is needed to understand whether our inability to replicate the bidirectional cross-lagged associations reported in the Schmitt et al. (2017) study is due to sample size restrictions.

Further, differences in sample characteristics across the three samples might have also contributed to the inconsistent replication findings. It could be that the Schmitt et al. (2017) sample included children with systematically different background characteristics. Given the importance of individual, demographic and family-level influences during early childhood, and their associations with EF and academic outcomes (e.g., Hackman et al., 2015; Sarsour et al., 2011). Thus, there may be possible untested moderators, or confounding variables, across samples that could explain the mechanisms involved in the co-development of EF and achievement.

Furthermore, the non-significant autoregression estimates of math achievement across the first two time-points was surprising given that the early math measures (WJ-AP) used in the current study have been extensively validated, age normed, and show excellent test-retest reliability across development (Woodcock, McGrew, & Mather, 2001). Both the Schmitt et al. (2017) study and the current replication included a working memory measure that involved verbal numerical tasks. The use of a working memory measure that includes numerical naming may introduce a confounding variable in these studies and may play a role in the associations between EF and math. However, the different and unequal sources of measurement error across latent and observed variables does not permit a fair comparison between the EF and math achievement variables in the current study, as latent factor variance is considered independent from residual measurement error, whereas observed variables include both true score and error variance (Bollen, 2002; Rhemtulla et al., 2020). The large correlations observed within waves indicate that there is considerable shared variance among the latent EF factor and math achievement (STAR $r = .87$; Pathways $r = .75$). It is possible that when EF and math were modeled simultaneously, the EF latent variable accounted for the variance in math at the next time-point, thus contributing to the decrease in the autoregression of math achievement. Thus, the non-significant math achievement autoregression estimate could be due to the presence of the numerical working memory measures, or the EF latent variable in the ARCL model and raises questions about the utility of modeling EF as a latent when examining bidirectional associations using manifest math achievement variables.

In sum, the results of the present conceptual replication were mixed. We replicated the results of the EF measurement model and longitudinal stability estimates of EF and literacy across two independent samples. However, we could not replicate the cross-lagged pattern of findings reported in the original study. The lack of bidirectional relations between EF and math achievement in both replication samples does not lend support to the theory of mutualism between these two constructs. However, the current conceptual replication has also revealed that bidirectional associations between EF and academic skills might not be robust to slight differences in EF measurement and number of measurement occasions, which might have contributed to the mixed findings in the literature, and has implications for our understanding of the development EF and academic skills across early childhood. Although this study cannot shed light on the best way to characterize associations between EF and academic achievement across early development, these findings underscore the need for more standardization in both measurement and modeling approaches – without which the inconsistency of findings may continue across this area of research.

Supplementary Materials

The Supplementary Materials contain the syntax and materials used for this manuscript (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

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). *Supplementary materials to "Reciprocal associations between executive function and academic achievement: A conceptual replication of Schmitt et al. (2017)"* [Syntax and materials]. OSF. https://osf.io/5twgv/

Journal of Numerical Cognition. (Ed.). (2021). *Supplementary materials to "Reciprocal associations between executive function and academic achievement: A conceptual replication of Schmitt et al. (2017)"* [Open peer-review]. PsychOpen GOLD. https://doi.org/10.23668/psycharchives.5225

References

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

Blackwell, K. A., Cepeda, N. J., & Munakata, Y. (2009). When simple things are meaningful: Working memory strength predicts children's cognitive flexibility. *Journal of Experimental Child Psychology, 103*, 241-249. https://doi.org/10.1016/j.jecp.2009.01.002

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

Blair, C., & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology, 20*(3), 899-911. https://doi.org/10.1017/S0954579408000436

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

Bodrova, E., & Leong, D. J. (2001). *Tools of the mind: A case study of implementing the Vygotskian approach in American early childhood and primary classrooms.* (Innodata Monographs – 7). Geneva, Switzerland: International Bureau of Education.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53*, 605-634. https://doi.org/10.1146/annurev.psych.53.100901.135239

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205-228. https://doi.org/10.1080/87565640801982312

Cameron, C. E., Kim, H., Duncan, R. J., Becker, D. R., & McClelland, M. M. (2019). Bidirectional and co-developing associations of cognitive, mathematics, and literacy skills during kindergarten. *Journal of Applied Developmental Psychology, 62*, 135-144. https://doi.org/10.1016/j.appdev.2019.02.004

Cameron, C. E., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*(2), 141-158. https://doi.org/10.1016/j.ecresq.2007.01.004

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

PsychOpen GOLD

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

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Cribbie, R., Beribisky, N., & Alter, U. (2019, June 17). *A multi-faceted mess: A review of statistical power analysis in psychology journal articles.* PsyArXiv. https://doi.org/10.31234/osf.io/3bdfu

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

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

Espinet, S. D., Anderson, J. E., & Zelazo, P. D. (2012). N2 amplitude as a neural marker of executive function in young children: An ERP study of children who switch versus perseverate on the Dimensional Change Card Sort. *Developmental Cognitive Neuroscience, 2*, S49-S58. https://doi.org/10.1016/j.dcn.2011.12.002

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483-527. https://doi.org/10.1016/0885-2014(95)90024-1

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. https://doi.org/10.1037/a0036633

Giner-Sorolla, R., Aberson, C. L., Bostyn, D. H., Carpenter, T., Conrique, B. G., Lewis, N. A., Jr., . . . Soderberg, C. (2019). *Power to detect what? Considerations for planning and evaluating sample size* [Preprint]. OSF. https://osf.io/jnmya/

Grammer, J. K., Carrasco, M., Gehring, W. J., & Morrison, F. J. (2014). Age-related changes in error processing in young children: A school-based investigation. *Developmental Cognitive Neuroscience, 9*, 93-105. https://doi.org/10.1016/j.dcn.2014.02.001

Hackman, D. A., Gallop, R., Evans, G. W., & Farah, M. J. (2015). Socioeconomic status and executive function: Developmental trajectories and mediation. *Developmental Science, 18*(5), 686-702. https://doi.org/10.1111/desc.12246

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

Isbell, E., Calkins, S. D., Cole, V. T., Swingler, M. M., & Leerkes, E. M. (2019). Longitudinal associations between conflict monitoring and emergent academic skills: An event-related potentials study. *Developmental Psychobiology, 61*(4), 495-512. https://doi.org/10.1002/dev.21809

Isbell, E., Calkins, S. D., Swingler, M. M., & Leerkes, E. M. (2018). Attentional fluctuations in preschoolers: Direct and indirect relations with task accuracy, academic readiness, and school performance. *Journal of Experimental Child Psychology, 167*, 388-403. https://doi.org/10.1016/j.jecp.2017.11.013

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

Jones, S. M., Zaslow, M., Darling-Churchill, K. E., & Halle, T. G. (2016). Assessing early childhood social and emotional development: Key conceptual and measurement issues. *Journal of Applied Developmental Psychology, 45*, 42-48. https://doi.org/10.1016/j.appdev.2016.02.008

Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation.* Thousand Oaks, CA, USA: SAGE.

Kusche, C. A., Greenberg, M. T., & Anderson, L. A. (1994). *The PATHS curriculum: Promoting alternative thinking strategies.* Seattle, WA, USA: Developmental Research & Programs.

Lahat, A., Todd, R., Mahy, C. E., Lau, K., & Zelazo, P. D. (2010). Neurophysiological correlates of executive function: A comparison of European-Canadian and Chinese-Canadian 5-year-olds. *Frontiers in Human Neuroscience, 3*, Article 72. https://doi.org/10.3389/neuro.09.072.2009

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*, 53-76. https://doi.org/10.1207/s15327906mbr3201_3

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities (MSCA).* Cleveland, OH, USA: Psychological Corporation.

McClelland, M. M., & Cameron, C. E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives, 6*(2), 136-142. https://doi.org/10.1111/j.1750-8606.2011.00191.x

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*, Article 599. https://doi.org/10.3389/fpsyg.2014.00599

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

Meixner, J. M., Warner, G. J., Lensing, N., Schiefele, U., & Elsner, B. (2019). The relation between executive functions and reading comprehension in primary-school students: A cross-lagged-panel analysis. *Early Childhood Research Quarterly, 46*, 62-74. https://doi.org/10.1016/j.ecresq.2018.04.010

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*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

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). https://doi.org/10.1037/14797-015

Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2010). *Self-regulation and academic achievement in the transition to school.* In S. D. Calkins & M. A. Bell (Eds.), *Human brain development: Child development at the intersection of emotion and cognition* (pp. 203–224). American Psychological Association. https://doi.org/10.1037/12059-011

Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus user's guide* (7th ed.) Los Angeles, CA, USA: Authors.

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

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Champion, K. M., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. *Early Childhood Research Quarterly, 23*, 10-26. https://doi.org/10.1016/j.ecresq.2007.09.001

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

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: The roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society, 17*(1), 120-132. https://doi.org/10.1017/S1355617710001335

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

Soper, D. S. (2020). A-priori Sample Size Calculator for Structural Equation Models [Software]. Available from http://www.danielsoper.com/statcalc

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. https://doi.org/10.3758/BF03207704

Stevens, J. P. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, NJ, USA: Erlbaum.

Strommen, E. A. (1973). Verbal self-regulation in a children's game: Impulsive errors on "Simon Says". *Child Development, 44*(4), 849-853. https://doi.org/10.2307/1127737

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. https://doi.org/10.1037/a0016738

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

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

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

Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). Executive Function: Implications for Education (NCER 2017-2000). *National Center for Education Research.* https://eric.ed.gov/?id=ED570880

Zeytinoglu, S., Calkins, S. D., & Leerkes, E. M. (2019). Maternal emotional support but not cognitive support during problem-solving predicts increases in cognitive flexibility in early childhood. *International Journal of Behavioral Development, 43*(1), 12-23. https://doi.org/10.1177/0165025418757706

Zeytinoglu, S., Leerkes, E. M., Swingler, M., & Calkins, S. D. (2017). Pathways from maternal effortful control to child self-regulation: The role of maternal emotional support. *Journal of Family Psychology, 31*, 170-180. https://doi.org/10.1037/fam0000271

Appendix

Table A.1

Correlation Table for STAR Study Variables

Variable	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Age	278	–															
2. Female	278	-.11	–														
Executive Function																	
3. Working Memory Time 1	276	.29***	.02	–													
4. Cognitive Flexibility Time 1	274	.20**	.05	.34***	–												
5. Inhibitory Control Time 1	264	.25***	.20**	.25***	.34***	–											
6. Working Memory Time 2	249	.20**	.02	.49***	.36***	.16*	–										
7. Cognitive Flexibility Time 2	248	.06	.13*	.25***	.38***	.30***	.39***	–									
8. Inhibitory Control Time 2	245	.15*	.25***	.19**	.19**	.56***	.19**	.35***	–								
9. Working Memory Time 3	239	.19**	.02	.35***	.35***	.16*	.45***	.37***	.09	–							
10. Cognitive Flexibility Time 3	240	.05	.12	.21**	.25***	.21*	.31***	.48***	.28***	.27***	–						
11. Inhibitory Control Time 3	239	.14*	.20**	.17**	.15*	.42***	.17*	.22**	.57***	.11	.33***	–					
Achievement																	
12. Applied Problems Time 1	278	.33***	.03	.53***	.50***	.42***	.52***	.46***	.26***	.41***	.42***	.23***	–				
13. Applied Problems Time 2	249	.34***	-.09	.50***	.49***	.35***	.56***	.48***	.24***	.42***	.48***	.22**	.73***	–			
14. Applied Problems Time 3	240	.28***	.10	.45***	.42***	.21**	.49***	.41***	.15*	.40***	.37***	.20**	.64***	.73***	–		
15. Letter Word ID Time 1	278	.20**	.08	.41***	.28***	.29***	.41***	.27***	.20*	.24***	.27***	.20*	.52***	.49***	.53***	–	
16. Letter Word ID Time 2	249	.29***	.04	.41***	.31***	.27***	.41***	.26***	.22**	.33***	.28***	.16*	.53***	.56***	.57***	.68***	–
17. Letter Word ID Time 3	240	.20**	.03	.35***	.22**	.21*	.39***	.31***	.11	.37***	.27***	.16*	.46***	.52***	.60***	.60***	.79***

Note. Working Memory = Numbers Reversed; Cognitive Flexibility = DCCS computer; Inhibitory Control = Go-No/Go (d').

*p < .05. **p < .01. ***p < .001.

Table A.2

Correlation Table for Pathways Study Variables

Variable	N	1	2	3	4	5	6	7	8	9	10	11
1. Age	279	–										
2. Female	281	-0.01	–									
Executive Function												
3. Working Memory Time 1	276	0.13*	-0.03	–									
4. Inhibitory Control Time 1	202	-0.02	0.17*	0.14	–								
5. HTKS Time 1	277	0.18**	0.21***	0.41***	0.33***	–							
6. Working Memory Time 2	167	-0.03	-0.01	0.41***	0.19*	0.28***	–						
7. Inhibitory Control Time 2	124	0.14	0.14	0.24**	0.57***	0.39***	0.35***	–					
8. HTKS Time 2	166	0.09	0.00	0.34***	0.11	0.41***	0.34***	0.35***	–				
Achievement												
9. Applied Problems Time 1	278	0.26***	0.09	0.54***	0.18*	0.48***	0.39***	0.38***	0.38***	–			
10. Applied Problems Time 2	168	0.02	-0.17*	0.53***	0.17	0.32***	0.43***	0.39***	0.29***	0.56	–		
11. Letter Word ID Time 1	281	0.27***	0.02	0.43***	0.15*	0.22***	0.27**	0.30**	0.26**	0.57***	0.49***	–	
12. Letter Word ID Time 2	168	0.01	-0.22**	0.46***	0.04	0.24**	0.34***	0.36***	0.23**	0.50***	0.57***	0.66***

Note. Working Memory = Digit Span; Inhibitory Control = Zoo Go/No-Go (d'); HTKS = Head-Toes-Knees-Shoulders.

*p < .05. **p < .01. ***p < .001.