Bi-Directional Relations Between Behavioral Problems and Executive Function: Assessing the Longitudinal Development of Self-Regulation

Chen Li[1], Emma R. Hart[2], Robert J. Duncan[3], & Tyler W. Watts[2*]

1. University of Pennsylvania
2. Teachers College, Columbia University
3. Purdue University

* Corresponding Author:
525 W 120th St
GDH 462
New York, NY 10027
tww2108@tc.columbia.edu

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A190521 to Teachers College, Columbia University, and The Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD046160). The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences, the U.S. Department of Education, nor the National Institutes of Health. The data for this project originated at New York University. We would also like to thank Drew Bailey for his helpful comments, and the team of the Chicago School Readiness Project for their contributions to this project, including Cybele Raver,

Bi-directional relations between behavior and EF

Christine Li-Grining, Amanda Roy, Dana McCoy, Stephanie Jones, Javanna Obregon-Steeby, Hannah Ellerbeck, Alaa Khader, Michael Masucci, Deanna Ibrahim, Jill Gandhi, and Xinyu Pan.

This research was deemed exempt from committee review by the IRB for Teachers College, Columbia University (protocol # 19-474).

The authors have no conflicts of interest to declare.

Data Availability Statement

The data and analytic files needed to reproduce the analyses in this are available upon request from open ICPSR: http://doi.org/10.3886/E182562V1.

Research Highlights

- We tested associations between behavioral problems and executive function among children from low-income families, using a random intercept cross-lagged panel model.
- Our results demonstrated that bi-directional effects at the within-child level between behavioral problems and executive function were consistently small and non-significant over the course of childhood and adolescence.
- To the extent that behavioral problems and executive function were related to one another, this relation only appeared at the between-child level.
- Our results imply that behavioral problems and executive function may not be as developmentally intertwined as previous theory suggests.

Abstract

During childhood, the ability to limit problem behaviors (i.e., externalizing) and the capacity for cognitive regulation (i.e., executive function) are often understood to develop in tandem, and together constitute two major components of self-regulation research. The current study examines bi-directional relations between behavioral problems and executive function over the course of childhood and adolescence. Relying on a diverse sample of children growing up in low-income neighborhoods, we applied a random intercept cross-lagged panel model to longitudinally test associations between behavioral problems and executive function from age 4 through age 16. With this approach, which disaggregated between- and within-child variation, we did not observe significant cross-lagged paths, suggesting that within-child development in one domain did not strongly relate to development in the other. We also observed a moderate correlation between the stable between-child components of behavioral problems and executive function over time in our preferred model, suggesting that these two domains may be relatively distinct when modeled from early childhood through adolescence.

Bi-directional relations between behavior and EF

During childhood, the capacity for self-regulation is commonly understood to involve both cognitive and behavioral components (Bailey & Jones, 2019; Blair & Ursache, 2011; McClelland & Cameron, 2011). Although correlational studies have reported associations between measures of cognitive regulation (i.e., executive function) and indicators of child behavioral regulation (i.e., behavioral problems) at various ages (e.g., Kahle et al., 2018; Ogilvie et al., 2011; Sulik et al., 2015), it remains uncertain whether these two domains show coherence or divergence over the course of development. Importantly, most studies investigating this issue have relied on modeling approaches that cannot easily disentangle within- and between-child variation, making it unclear whether previously reported correlations between cognitive and behavioral regulation may actually be due to stable between-child factors (e.g., environmental resources). Indeed, recent work has stressed the importance of clarifying conceptualizations of self-regulation (see Bailey & Jones, 2019; Inzlicht et al., 2021; Morrison & Grammer, 2016). To better inform theory and practice, additional research is needed to understand the relation between executive function and behavioral problems, at both within-child and between- child levels.

The current study attempts to shed new light on the development of self-regulation by employing a novel approach that examines the longitudinal co-development of executive function and behavioral problems. We fit a random-intercept crossed-lag panel model (RI-CLPM; Hamaker et al., 2015) to disaggregate between-child and within-child variation in executive function and behavioral problems over time. This approach allowed us to estimate the bi-directional associations between the two domains using only within-child variation, while also testing the degree to which between-child, stable, variation in one domain relates to stable variation in the other. We leveraged data from the Chicago School Readiness Project (CSRP), a longitudinal study of children growing up in high poverty neighborhoods in Chicago (see Raver et al., 2009; 2011). This dataset allowed us to examine the long-run development of these two domains across three distinct developmental periods: early childhood, middle childhood, and adolescence. Thus, we examined whether within-child changes in executive function led to within-child changes in children's abilities to limit problem behaviors.

Self-Regulation Involves both Cognitive and Behavioral Control

Self-regulation is commonly understood as one's capacity to modulate and control cognition, emotion, behavior, and attention in order to engage in goal-directed pursuits across diverse situations and contexts (Bailey & Jones, 2019; Inzlicht et al., 2021; Karoly, 1993). Self-regulation has often been described as a "domain-general" capacity, for which development and activation is influenced by stress response physiology (Blair, 2016; Blair & Raver, 2015). Like most cognitive and behavioral capacities, self-regulation develops over the course of childhood and adolescence, whereby rapid changes occur during early childhood before self-regulatory skills reach relative stability in adolescence. Over the decades, many attempts have been made to provide an overarching theory of self-regulation, with definitions and developmental models spanning multiple sub-fields of Psychology (for review, see Inzlicht et al., 2021). However, across most theoretical and empirical work, and as considered in the current study, self-regulation is understood to involve some form of both cognitive and behavioral control.

The cognitive aspect of self-regulation is most commonly conceptualized as executive functioning (EF). EF has been defined as the "top-down" component of self-regulation that one employs when faced with situations that do not allow for reliance on automatic cognitive processing (Ursache et al., 2012). Thus, EF involves cognitive abilities that govern attention and allow one to adaptively function across contexts (Bailey & Jones, 2019). Operationalizations of

EF typically include three core capacities: working memory, inhibitory control, and cognitive flexibility (Miyake et al., 2000). Briefly, working memory describes one's ability to hold and manipulate pieces of information in "short term" memory while processing or responding to additional cognitive inputs or distraction (Conway et al., 2005). Inhibitory control has been defined as the capacity to resist distraction and automatic response tendencies (e.g., Eisenberg et al., 2010). Finally, cognitive flexibility, or set shifting, describes one's ability to flexibly shift attention between tasks (Blair, 2016). Some conceptualizations of EF integrate these components with temperament-based factors such as executive attention and control (Bailey & Jones, 2019; Blair & Ursache, 2011). Empirical work has documented relations between EF during childhood and markers of school success (e.g., Ahmed et al., 2019; Blair & Razza, 2007).

Self-regulation is also often thought to involve a child's ability to regulate their emotional affect to exhibit behaviors that elicit positive social interactions with parents, teachers, and peers (see Blair & Raver, 2015; Raver et al., 2009). As Campbell et al. (2016) defined it, "behavior regulation refers to one's ability to monitor his/her own behavior, including compliance to adult demands and directives, the ability to control impulsive responses, and delay engagement in specific activities" (p. 32). The conceptualization and operationalization of behavioral regulation has varied across studies due to the complexity of measuring internal behavior-relevant processes, and the dearth of measures that directly assess such processes. For example, in some research, this construct has been labeled as "self-control" (see Inzlicht et al., 2021; Nigg, 2017), which has often been measured by performance on delay tasks, such as the Marshmallow Test (Mischel et al., 1989; Watts et al., 2018). In other work, survey-based measures of behavioral problems (e.g., externalizing, hyperactive, inattentive, and impulsive behaviors) have been used as indicators of these regulatory capacities (see Hughes & Ensor, 2008; McAuley et al., 2010;

Moffitt et al., 2011). In the current study, we conceptualize the behavioral aspects of self-regulation using this behavioral problems approach.

Relations Between Executive Function and Behavioral Problems

The degree to which executive function and the ability to limit behavioral problems relate to one another throughout development has been the focus of both theoretical and empirical work. Some have theorized that EF and other dimensions of self-regulation develop in a reciprocal fashion (Blair & Ursache, 2011; Bridgett et al., 2015), and others have purported theories of self-regulation that suggest EF lays the foundation for behavioral regulation (Doebel, 2020). These theories assume that EF provides the basis for more complex behavioral regulation, as the ability to regulate thoughts and attention supports a child's capacity to regulate their behavioral response to social stimuli (see Blair & Raver, 2015).

Although some empirical work has found that EF is more predictive of behavioral problems than vice versa, the strength of this association has been inconsistent across studies and often smaller than what theory might predict. Indeed, some studies have found that EF predicts fewer subsequent behavioral problems across early childhood (Hughes & Ensor, 2008; Hughes et al., 2020; Kahle et al., 2018), and Sulik et al. (2015) found that lower levels of behavioral problems also predicted later EF. Still other studies have observed inconsistent, and statistically non-significant, relations among these domains. For example, Blair (2003) found inconsistent cross-sectional associations between EF and teacher-rated classroom behaviors among preschoolers. Similarly, Schmitt et al. (2021) found inconsistent relations between performance on the Heads, Toes, Knees, and Shoulders task and behavioral problems across preschool and kindergarten. Relatively weak relations have been observed in middle childhood and adolescence as well (McAuley et al., 2010; Sasser et al., 2014). These equivocal results are reflected in mixed findings from meta-analytic work that has attempted to examine the convergent validity of assessments of EF and behavioral problems (e.g., Duckworth & Kern, 2011; Ogilvie et al., 2011; Saunders et al., 2018; Toplak et al., 2013).

Just as the results from correlational studies linking EF to behavioral regulation have been mixed, the application of theory relating EF and behavioral regulation in intervention development has also varied. Interestingly, several self-regulation interventions have primarily targeted behavioral regulation, with the hope that reducing behavioral problems will lead to improvements in cognitive regulation (e.g., Raver et al., 2009). The Chicago School Readiness Project (CSRP), from which the current study data were drawn, targeted teachers' approaches to behavioral management in preschool classrooms serving low-income children. Raver and colleagues (2009; 2011) reported that the program improved both teacher reports of behavioral problems and direct assessments of executive functioning. Similar programs targeting early social-emotional skills and behavioral problems have reported mixed results on measures of child outcomes (Hsueh et al., 2014; Morris et al., 2014; O'Connor et al., 2014), as programs that have attempted to target both behavioral problems and executive function directly (Blair & Raver, 2014; Diamond et al., 2019; Nesbitt & Farran, 2021). Thus, although the intervention literature demonstrates interest in targeting behavioral regulation and executive function as malleable factors during early childhood, this experimental evidence does not easily disentangle the theoretical links between EF and behavioral problems.

Current Study

Together, the previous correlational and experimental evidence paints an unclear picture of the developmental process linking behavioral problems and EF, making apparent the need for additional investigation to increase theoretical clarity. Morrison and Grammer's (2016) critique of the self-regulation empirical literature suggested the potential benefit of investigating the relations between these constructs longitudinally to better decipher the extent to which the two domains are, or are not, related (see also Bailey & Jones, 2019). The common approach to examining the co-development of two related constructs in Developmental Psychology has been the employment of the cross-lagged panel model (CLPM), which tests bi-directional relations in longitudinal data. Indeed, this model has been recently used to examine reciprocal relations between cognitive and behavioral regulation during early childhood (Wolf & McCoy, 2019), yet recent methodological work has noted the apparent limitations of the CLPM for advancing developmental theory (Berry & Willoughby, 2017; Hamaker et al., 2015). In short, the traditional CLPM conflates between and within-child variation, making the cross-lagged paths susceptible to bias due to stable factors that influence between-child differences (e.g., socioeconomic status; cognitive ability). Yet, empirically examining the co-development of behavioral problems and EF *within* children may provide more theoretical utility, as most theoretical work on the structure of self-regulation describes within-child processes that interact to influence developmental change over time (e.g., Blair & Ursache, 2011). Further, interventions targeting elements of self-regulation as malleable factors also act upon within-child processes (e.g., Raver et al., 2009).

The Random Intercept Cross-Lagged Panel Model (RI-CLPM) provides an innovative way to disaggregate between-child and within-child variation by using multi-level modeling (Bailey et al., 2020; Berry & Willoughby, 2017; Lougheed et al., 2022). In essence, this model allows us to examine how within-child change in both behavioral problems and EF influence one another over the course of childhood and adolescence, while also allowing us to observe how stable, between-child, variation in each domain relate to one another. By separating within-child and between-child variation, this approach may provide less biased, and more causally informative estimates (see Bailey et al., 2018; Hamaker et al., 2015). Indeed, longitudinal work has shown that between-child differences in self-regulatory processes are largely stable over time due to a host of genetic and environmental influences (Friedman et al., 2016; Raffaelli et al., 2005), thus emphasizing the need to account for stable variation if we hope to understand how change in each domain may contribute to development.

We examined the bi-directional relations between EF and behavioral problems using a RI-CLPM, with EF and behavioral problems measured during early childhood, middle childhood and adolescence. Leveraging data from the Chicago School Readiness Project, we extended previous work that has examined the development of self-regulation in a sample of racially and ethnically minoritized children growing up in high-poverty neighborhoods in Chicago (Li-Grining et al., 2019; McCoy et al., 2018; Raver et al., 2013; Ursache & Raver, 2015). Our study also builds on recently published work by Schmitt et al. (2021), which explored a similar model examining bi-directional relations between directly-assessed self-regulation and measures of behavioral problems and social skills during early childhood.

Following previous theoretical work suggesting that EF undergirds the development of behavioral regulation (e.g., Bailey & Jones, 2019; Ursache et al., 2012), we expected to observe significant cross-lagged paths for EF predicting reductions in behavioral problems across development. However, because models that isolate within-child effects often produce more conservative estimates, we also expected these paths to be smaller than what has been reported in previous work. Finally, given some past empirical work finding evidence of associations between EF and behavioral problems, we also expected to observe a strong correlation between the stable factors for EF and behavioral problems reflecting that at the between-child level, as children with higher EF generally demonstrate fewer behavioral problems.

Methods

Data

Data were drawn from the Chicago School Readiness Project (CSRP), a longitudinal study that was originally designed as an evaluation of an intervention that targeted the self-regulation of children enrolled in Head Start centers serving high-poverty neighborhoods in Chicago (Raver et al., 2008). As part of the study, 18 Head Start sites in Chicago were recruited for participation, with roughly half of the sites participating in cohort 1 (2004-2005) and the others participating in cohort 2 (2005-2006). Across the two cohorts, study developers recruited 602 children from the Head Start centers, and the study has followed children through adolescence. Throughout the study, researchers gathered data on participants' behavioral functioning and EF. Participants' EF and problem behaviors were measured in the fall (average age = 4.4 years) and spring of the pre-kindergarten year (i.e., directly proceeding and following CSRP intervention). In the current study, we also used follow-up data collected during elementary school (average age = 10.1 years) as well as during two adolescent waves corresponding to high school (average age = 15.3 and 16.2 years, respectively).

The current study included data from 598 children who had at least one measurement of EF and one measurement of behavioral problems across the five time points considered. Because children were sampled from high-poverty Head Start centers, children in the current sample largely grew-up in low-income homes (approximately 77% were living below the poverty line in the fall of the prekindergarten year). The current sample was racially diverse, with approximately 66% identifying as African American and 27% as Hispanic. Just over half of the participants were female (54%).

Intervention

Although the current analysis does not directly address intervention effects, some description of the intervention is necessary. The original preschool intervention targeted children’s self-regulation through a series of services offered to intervention classrooms. Head Start centers were randomly assigned to either the intervention or a control group. Teachers in the intervention group received professional development that provided them with new strategies for responding to students’ behavioral problems. Intervention classrooms were also regularly visited by mental health consultants (MHCs) who supported teachers in the implementation of the classroom management techniques introduced in the training sessions, and they provided some direct services to children. Additionally, to guard against teacher burnout, MHCs organized stress reduction workshops for teachers.

The intervention has been described at length in previous studies (see Raver et al., 2008). Results from the initial evaluations reported that the intervention had positive effects on children’s EF, behavioral regulation, and pre-academic skills (Raver et al., 2009; 2011). Follow-up work has also found some indication that the intervention may have positively affected longer-term academic achievement and EF, though the adolescent follow-up found no effects on behavioral problems (see Watts et al., 2018).

It should also be noted that during high school, students were rerandomized to a Purpose-for-Learning/Growth Mindset intervention that was presented as a 30-min module during the two adolescent follow-up waves. However, this follow-up intervention had primarily null effects on both proximal and distal measures of task persistence, self-regulation and academic achievement (see Gandhi et al., 2020).

Due to the possibility of longer-term effects from the preschool intervention, we tested models that split our sample between the preschool intervention and control groups. As we detail below, we saw little indication that either intervention affected our key model parameters.

Measures

Because we pursued a secondary data analysis of a study with multiple waves of data including a diverse set of measures tapping different domains of self-regulation, we attempted to balance several priorities when selecting the measures considered in the current analysis. First, given the longitudinal nature of our analysis, we prioritized measures that were administered at multiple waves, and we attempted to select measures that tapped behavioral problems and executive function with limited overlap to related constructs like internalizing and emotional regulation. Second, we relied on previous CSRP papers (e.g., Raver et al., 2011; Watts et al., 2018) and attempted to use measure operationalizations that were consistent with previous work on this sample. As we note below, this led us to several operationalizations that we tested across our key models.

Executive Function

EF was measured using direct assessments during the fall and spring of preschool, and during middle childhood and adolescent follow-up waves. We operationalized EF in two ways: 1) using a single measure from each assessment point, and 2) using a composite of EF during early childhood, which followed previous CSRP work (Raver et al., 2011). For the composite, we averaged performance across the Balance Beam and Pencil Tap tasks at both the fall and spring of kindergarten (see Smith-Donald et al., 2007 for more details). For measures of EF in middle childhood and adolescence, we relied on the Hearts and Flowers task. The measures are described in detail below.

Balance Beam Task. For the Balance Beam task (Maccoby et al., 1965; Murray & Kochanska, 2002), participants were asked by the assessor to walk the "balance beam." First, the child was instructed to walk across a straight line demarcated on the floor as they normally would, after which the child was directed to walk the same line slowly. The assessor recoded the amount of time (in seconds) the child took to walk the line for both trials and calculated the difference between slow trial and regular trial (Smith-Donald et al, 2007). Larger time differences between the slow and regular trials indicated higher EF performance. Participants completed the balance beam task in the fall and spring of the pre-kindergarten year, and assessments were collected at Head Start sites (see Raver et al., 2011).

Pencil Tap Task. The Pencil Tap task (Blair, 2002; Diamond & Taylor, 1996) was administered as a second measure of preschool EF. During the task, participants were directed to tap a pencil two times when the assessor tapped a pencil one time, and to tap once when the assessor tapped twice (Smith-Donald et al, 2007). The assessor recoded the percent of correct response for 16 trials. Performance on the task was measured by the average percent of correct taps across all 16 trials, such that higher scores indicated greater EF. Like the balance beam task, participants completed the pencil tap task in the fall and spring of the pre-kindergarten year at their Head Start site.

Hearts and Flowers Task. During middle childhood and early adolescence, EF was measured using the Hearts and Flowers task (Davidson et al., 2006). For this task, hearts and flowers appeared on the right or left of a computer screen in random order. Participants were instructed to press a key ("Q" or "P") in congruence or incongruence with the side of the screen where the heart or flower appeared. Specifically, participants were instructed to press the congruent key when they saw a heart (e.g., press "Q" if the heart was on the left side of the screen), and the incongruent key when they saw a flower (e.g., press “Q” if the flower was on the right side of the screen). “Mixed trials” occurred when participants were shown both hearts and flowers in succession and had to switch from congruent to incongruent responses. Performance was measured as the percentage of correct responses for 33 mixed trials. Higher scores indicated stronger EF. Participants completed this task in late elementary school and during both adolescent follow-up waves, and most participants completed this task on computers in their school setting[1]. To ensure the task was developmentally appropriate, the maximum response time was set to 2000 milliseconds (ms) at the middle childhood assessment, and 750 ms for the adolescent assessment.

Behavioral Problems

Across the five waves considered here, behavioral problems were measured using various survey reports taken from parents, teachers, and adolescent participants (i.e., self-report). As with EF, we performed analyses using both single measure indicators of behavioral problems (scores from a single measure) at each assessment point, and composites incorporating broader measures and respondent reports collected at each assessment point (see Table 1). Supplementary file Table S1 provides examples of the items from the measures. This composite approach to measuring behavioral problems has been employed in other highly-cited work (see Moffitt et al., 2011). During the preschool and middle childhood waves, parents and teachers were the primary respondents on children’s behavioral problems. During adolescence, we relied on youth self-reports of their behavior.

---

[1] During middle childhood, the percent correct on mixed trials was calculated if participants had at least 75 percent non-missing data (i.e., 75% “valid trials”). In the adolescent waves, this “valid trials” rule was not applied, and accuracy was calculated across all 33 trials.

Behavior Problem Index. At the preschool and middle childhood waves, the behavioral problem index (BPI: Zill, 1990) was used to measure the frequency with which children displayed problem behaviors. Using the approach employed in previous work (NLSY79, 2000), an Externalizing Behaviors subscale was created for use in the current study. The subscale included 18 parent-report items and 17 teacher-report items collected during the preschool timepoints, and five items collected at the middle childhood timepoint. For each item on the scale, parents and teachers were asked to rank the frequency with which the child displayed various problematic behaviors using a 3-point Likert scale (0 (*not true*), 1 (*sometimes true*), and 2 (*very/often true*)). Higher BPI scores indicated higher level of behavioral problems. In the fall of the prekindergarten year, the BPI was administered to both parents and teachers. In the spring of the prekindergarten year, only teacher-report was collected. For the elementary school wave, only parent-report was collected.

Caregiver and Teacher Report Forms. At the spring preschool wave and middle childhood wave, the Caregiver and Teacher Report Forms (C-TRF: Achenbach & Rescorla, 2001; TRF: Achenbach, 1991) was administered. While this measure includes items about academic performance, adaptive functioning, and behavioral/emotional problems, in the current study, we drew on items measuring children's externalizing behavioral problems. The C-TRF is validated for use with children ages 1.5 to 5 years old (Achenbach & Rescorla, 2001). For the C-TRF, externalizing behaviors were measured through 34 items created through the combination of 2 subscales: the Attention Problems subscale (e.g., "can't sit still, restless, or hyperactive"), and the Aggressive Behavior subscale. (e.g., "gets in many fights"). The C-TRF was administered at the spring preschool wave. The TRF is validated for use among children 6-18 years old (Achenbach, 1991), and includes 32 externalizing behavior items from the Aggressive

Behavior subscale (e.g., “gets in many fights”), and Rule-Breaking subscale (e.g., “breaks school rules”). The TRF was administered during the middle childhood wave. For each item on both the C-TRF and TRF, teachers indicated the extent to which each externalizing behavior was present for a given student using a 3-point Likert scale (0 (*not true*), 1 (*somewhat or sometimes true*), and 2 (*very true or often true*)). Higher TRF scores indicated greater behavioral problems.

Barratt Impulsiveness Scale and *Behavior Rating Inventory of Executive Function.* The Barratt Impulsiveness Scale (BIS-11; Patton et al., 1995), and Behavior Rating Inventory of Executive Function (BRIEF; Gioia et al., 2000) were collected in middle childhood and adolescence. These measures assessed children’s impulsiveness and everyday disruptive behaviors (Isquith et al., 2004). In middle childhood, teacher reports on the BIS and BRIEF were collected. During adolescence, a self-report version of the scale was administered. In line with previous work (McCoy et al., 2011), BIS and BRIEF items that focused explicitly on behavioral problems were combined to form a composite behavioral problems score. For the middle childhood wave, BIS-BRIEF scores were created through combining two BIS items and 10 BRIEF items, and the resulting subscale showed strong reliability ($\alpha = 0.96$). In the adolescent waves, BIS-BRIEF scores were formed through six BIS items and eight BRIEF items measuring behavioral problems ($\alpha = 0.87$ and 0.85, respectively). For the BIS, respondents indicated the frequency with which a student displayed each behavior using a 4-point Likert scale (0 (*rarely/never*), 1 (*occasionally*), 2 (*often*), 3 (*almost always/always*)). For the BRIEF, respondents indicated frequency using a 3-point Likert scale (0 (*never*), 1 (*sometimes*), 2 (*often*)). In order to combine items across the two measures, the items from each scale were standardized by dividing BIS items by three and BRIEF items by two (McCoy, et al., 2011).

Risky Behavior. At the middle childhood and adolescent waves, the Child Health Risk Behavior Scale (CHRBS; Riesch et al., 2006) and Middle School Youth Risks and Behavior Survey (MS-YRBS; CDC, 2015) were administered to measure the prevalence of health risk behaviors (i.e., violent activities that may cause physical harm). For the purposes of this study, an Externalizing Risk scale was created. In middle childhood the scale included six student-report items, and in adolescence the scale included eight student-report items. For each item, respondents indicated the prevalence or lack thereof of each behavior. Higher scores indicated more risky behaviors.

Composite scores. For models that relied on composite measures, we averaged the standardized behavior subscales within a given wave to form a composite behavioral problems score. Table 1 reflects the measures that were combined to generate the behavioral composite at each wave. For the fall of pre-kindergarten behavioral problems score, an aggregate score was created by averaging parent- and teacher-reported behavioral problems using the BPI ($\alpha = 0.89$). For the spring of pre-kindergarten, teacher-reported behavioral problems using the BPI and C-TRF were aggregated ($\alpha = 0.97$). For middle childhood, teacher-reported behavioral problems using the TRF, BIS, BRIEF, parent-reported problems using the BPI, and student self-report on the MS-YRBS were aggregated to form a behavioral problems score ($\alpha = 0.96$). For the two adolescent waves, student self-report on the BIS, BRIEF, and MS-YRBS were aggregated ($\alpha = 0.84$ for both years).

Family Background

Information regarding family background was collected in the fall of the pre-kindergarten year. Parent-reported background characteristics were used as covariates in the present analyses, which included child gender, age during preschool, ethnicity, and parent education status. We calculated family income-to-needs ratio in preschool based on reported total family income from the previous year divided by that same year's federal poverty threshold for the number of adults and children in the family (descriptive characteristics for family covariates can be found in supplementary file Table S2).

Analytic Plan

Our key analytic results were generated from fitting a RI-CLPM model (Hamaker et al., 2015) to our full analytic sample (*n* = 598), as depicted in Figure 1. As Figure 1 reflects, we began with a model using MLR estimates that included no controls where EF and behavioral problems were each modeled as a latent random intercept, with paths for each of the five waves constrained to be equal. We then modeled cross lagged paths between EF and behavioral problems at each wave, while also including auto-regressive paths for both domains over time and correlated error terms at each wave. Following the Hamaker et al. (2015) RI-CLPM, the auto-regressive and cross-lagged paths between measures of EF and behavioral dysregulation were modeled using occasion-specific latent variables (see also Bailey et al., 2020).

We employed two approaches to operationalizing behavioral problems and EF. Based on the argument that that the BRIEF may capture EF, and the Balance Beam task may rely on behavioral control, our first analysis made use of single measure indicators of EF and behavioral problems that were the most conceptually differentiated at each assessment point. Thus, we tested single measure models that did not include the Balance Beam task or BRIEF.

Next, given previous work that has relied on broader operationalizations of EF (Brock et al., 2009; Willoughby et al., 2012) and behavioral problems (Moffitt et al., 2011), we executed a model that incorporated composite measures of EF and behavioral regulation. Items used from the BRIEF were conceptually similar to items on other behavior problem measures used in previous studies (Hughes & Ensor 2008; Kahle et al., 2018; Moffit et al., 2011; Sulik et al., 2015). Of note, the BRIEF has shown to be more related to such behavioral problem measures than to direct EF assessments (Mahone & Hoffman, 2006; McAuley et al., 2010), and previous work on the current sample has suggested that specific items from the BRIEF load onto a "behavioral dysregulation" factor (McCoy et al., 2011).

We then tested a model that included controls, with each of the respective random intercepts for EF and behavioral problems regressed on cohort, age at preschool, gender, and race. For the random intercepts, we also controlled for assignment to the preschool intervention in order to account for any sustained intervention effects on the respective domains. Finally, to control for time-sensitive intervention effects, we controlled for the preschool treatment on the spring of preschool measures of EF and behavioral problems (i.e., posttest for the preschool intervention), and we controlled for the adolescent mindset intervention on the final adolescent wave measures of EF and behavioral problems (i.e., follow-up for the mindset intervention).

Additional supplementary analyses were performed to check if our models differed across key subgroups: 1) boys and girls, 2) higher poverty risk vs. lower poverty risk; and 3) preschool treatment versus control. Details on these analyses and results are reported in the online supplementary file. All analyses were performed using RI-CLPM syntax (Mulder & Hamaker, 2021) in Mplus 8.2 (Muthén & Muthén). Model input syntax has been included in the online supplementary file.

Results

Descriptive Findings

Table 1 presents descriptive statistics for each subscore included in our EF and behavioral problems composites. Table 2 presents correlations among the aggregated EF scores and behavioral problems measures across each wave (supplementary file Table S3 presents correlations among the subscales used in our measures). As expected, we observed positive correlations among the EF measures across waves, and the correlations suggested some modest stability in EF over time. For example, the fall-of-preschool EF measure was moderately correlated with spring-of-preschool EF (r [425] = 0.57). At subsequent timepoints, the association between fall-of-preschool EF and later EF dropped and then remained fairly consistent across the middle childhood and adolescent waves ($r = 0.21 – 0.28$). We also observed positive correlations among the measures of behavioral problems over time. As compared with the EF measures, we observed less longitudinal stability. The fall-of-preschool behavioral problems composite strongly correlated with the spring measure (r [514] = 0.44), but the fall measure was less correlated with late elementary school behavioral problems (r [471] = 0.31), and even less correlated with adolescent behavior (r [446] = 0.11; and r [413] = 0.14). Finally, as expected, we observed negative concurrent associations between EF and measures of behavioral problems ($r = -0.17 – -0.02$; p-values ranged from < 0.001 to 0.65).

Key Results

First, we investigated the relations between EF and behavioral problems using the RI-CLPM with no controls (see Figure 1). Column 1 of Table 3 presents path estimates and fit statistics for the RI-CLPM with no controls, using single measure indicators of both EF and behavioral problems at each wave. Model fit statistics suggested that the RI-CLPM fit the data well ($CFI = 0.97$; $RMSEA = 0.04$). Here, we observed positive and statistically significant auto-regressive paths between measures of behavioral problems over time (standardized βs ranged from 0.17 to 0.62; p ranged from < 0.001 to 0.01). For EF, we observed two positive and statistically significant auto-regressive paths ($\beta = 0.43$; $p < 0.001$; $\beta = 0.33$, $p < 0.001$), though the path from spring-of-preschool to middle childhood ($\beta = -0.04$; $p = 0.54$), and from middle childhood to the first adolescent wave ($\beta = 0.10$; $p = 0.25$) were not statistically significant. Standardized factor loadings were moderate in size for EF ($\lambda = 0.51 – 0.53$, $p < 0.001$) and smaller for behavioral problems ($\lambda = 0.24$, $p < 0.01$). Importantly, we observed small and non-statistically significant cross-lagged paths between behavioral problems and EF suggesting that within-child changes in EF ability at one timepoint were not predictive of changes in behavioral problems at the subsequent timepoint, or vice versa. Of note, there was one marginally statistically significant cross-lagged path in the opposite direction than we hypothesized, showing that EF within-child positive variation in the spring of preschool were predictive of more behavioral problems in middle childhood ($\beta = 0.12$, $p < 0.10$). The latent random intercepts for EF and behavioral problems were strongly correlated ($\beta = -0.76$, $p < 0.05$) suggesting that, at the between-child level, children with higher EF were likely to have fewer behavioral problems, and vice versa.

Column “No Covariates (Composite)” of Table 3 presents path estimates for the same model using composite measures. Overall, the model fit the data well ($CFI = 0.98$; $RMSEA = 0.04$), and the estimates were similar to the model with the single measures of EF and behavioral problems. One notable difference, however, was that the factor loadings for the composite measure of behavioral problems were larger in this model ($\lambda = 0.42 – 0.43$, $p < 0.001$) than in the single measure model, suggesting that the composite behavioral problems measure showed more inter-individual stability over time than the single measures. In this model, the path between spring of preschool EF and middle childhood behavioral problems was no longer marginally significant. However, there was a marginally significant path between fall or preschool EF and spring of preschool behavioral problems in the hypothesized direction (i.e., improvements in EF predicted reductions in behavioral problems; *ß* = - 0.08, *p* = 0.09). Additionally, there was a smaller correlation between the latent random intercepts (*ß* = -0.33, *p* < .01) in this model than in the single measure model.

Next, we tested the composite measure model with covariates for fall of preschool family and demographic characteristics (the coefficients for control variables can be found in the supplementary file Table S4). Importantly, this model also controlled for treatment status (see Table 3 note). Here, model fit indices were slightly worse, though the model still fit the data well (*CFI* = 0.89; *RMSEA* = 0.05). Path estimates were similar to those observed for the model with no controls. With the controlled model, we observed a statistically significant correlation between the latent factors for EF and behavioral problems, at a magnitude very similar to the model without controls (*ß* = -0.32, *p* = 0.03).

Subgroups

Supplementary analyses were performed to test for model differences among key subgroups. Details of these analyses are discussed in the online supplementary file. Subgroup models for gender and substantial poverty are shown in Table S5, while the subgroup model for intervention status is shown in supplementary file Table S6. The cross-lagged estimates for these models were very similar to the primary models. Of note, some differences were observed in the correlations between latent random intercepts such that female children (*ß* = -0.50, *p* = 0.001) showed a stronger correlation than male children (*ß* = -0.18, *p* = 0.30), and children experiencing substantial poverty (*ß* = -0.47, *p* = 0.004) showed a stronger correlation than children experiencing less substantial poverty (*ß* = -0.26, *p* = 0.11). In addition, we observed similar correlations for preschool treatment group (*ß* = -0.38, *p* = 0.003) and control group (*ß* = -0.32, *p* = 0.17).

Sensitivity tests

In Table S7, we present results from two alternative approaches to modeling cross-lagged paths, which were executed to test the sensitivity of our results. First, we tested if our results were robust to a latent state-trait model, which also uses a latent variable approach to disaggregating stable and time-varying effects (see Bailey et al., 2020). With the state-trait model, results were largely consistent to those shown with the RI-CLPM. Although we detected several cross-lagged-paths around the margins of statistical significance during preschool, the magnitudes of these effects were nearly identical to those reported in the main text. With the traditional CLPM, which does not control for stable between-child effects, we detected some statistically significant cross-lagged paths between EF and behavioral problems in both directions (i.e., EF predicting behavioral problems and behavioral problems predicting EF). These cross-lagged effects were relatively small in magnitude ($\beta < 0.11$).

Discussion

Self-regulation research often involves consideration of development in both cognitive and behavioral domains (e.g., Bailey & Jones, 2019; Blair & Raver, 2015). However, previous studies investigating the connections between EF and behavioral problems have provided mixed results, and few have longitudinally examined the co-development of the two domains over the course of childhood and adolescence. The current study aimed to examine how EF and the ability to limit behavioral problems co-develop from early childhood through adolescence (i.e., age 4 through age 16). We employed a RI-CLPM, which allowed us to disaggregate between- and within-child variation in each domain of self-regulation over the course of approximately 12 years of development. Perhaps surprisingly, we found little evidence to support the widely embraced theory that stronger EF skills lay the foundation for reduced behavioral problems.

Bi-directional relations between behavior and EF

Crossed-lagged paths predicting behavioral problems from EF were small and statistically non-significant. Thus, at the within-child level, time-specific variations in EF and behavioral problems did not drive changes in the other. This is similar to what was recently reported by Schmitt et al. (2021), as they also found small cross-lagged relations between measures of self-regulation and behavioral skills during preschool and kindergarten when random intercepts were included to account for stable variation. As we discuss below, the lack of significant relations at the intra-individual level could be partially due to measurement error, but it could simply be the case that within-child fluctuations in each domain operate relatively independently of one another. In other words, if a child begins showing less problem behaviors at a given point in time, changes in their EF are not likely to be a main culprit behind such behavioral improvement. This may indicate that some theories of self-regulation require revision, as self-regulation does not appear to strongly tie these two domains together in such a way that changes in one domain necessarily cause changes in the other. However, both the single measure and composite models produced negative latent correlations between EF and behavioral problems over time that were statistically significant, suggesting that children with stronger EF showed fewer behavioral problems at the between-child level. This correlation implies that children who tend to be more persistently poorly behaved in the long-term are also likely to be observed as having lower stable levels of EF. This between-child level relation is not surprising, and could be driven by a host of environmental and personal factors (e.g., parenting, family resources, temperament) that would lead to inter-individual stability in behavioral and cognitive regulation, regardless of time-specific changes.

These results align with accumulating evidence that the development of EF and behavioral problems may be less intertwined than is often theoretically conceptualized. Amidst the wide array of self-regulation theories, the idea that EF lays the foundation for behavioral regulation has played a prominent role (Bailey & Jones, 2019; Barkley, 2001; Diamond, 2016; Nigg, 2017, Ursache et al, 2012). However, our findings provide additional support for a growing body of studies finding small, or null, relations between EF and measures of behavioral functioning (e.g., Saunders et al., 2018; Schmitt et al., 2021; Schoemaker et al., 2013; Toplak et al., 2013), and provide a point of contrast from other studies that have found evidence of convergence between the two domains (e.g., Hughes et al., 2020; Kahle et al., 2018; Ogilvie et al., 2011).

The gender and poverty subgroup provide some potential implications (see supplementary file). It should be noted that these subgroup analyses were largely exploratory, and we did not find an overall better model fit for these models than the main models. Thus, results should be interpreted with caution. However, we found some indication that behavioral problems and EF were more strongly related at the between-child level for girls than boys. This suggests that inter-individual differences driving behavioral problems in boys may have less do with their cognitive regulatory capacity, whereas behavioral problems may tend to coincide with reduced cognitive regulatory skills for girls. Indeed, this finding may reflect the fact that boys tend to demonstrate more behavioral problems than girls (e.g., Deković et al., 2004; Raffaelli et al., 2005), and our results suggest that the factors that drive behavioral problems for boys may be largely different from factors that drive their EF.

Results split by income-to-needs ratio in early childhood were similar to the gender findings, showing that children living in substantial poverty (defined as below 50% of the poverty line) showed a stronger correlation between the random intercepts for behavioral problems and EF. This suggests that the stable factors associated with more severe exposure to poverty may exert consistent effects on the development of both EF and behavior over time. Indeed, the poverty-related findings should be tested in samples with more heterogeneity in socioeconomic status as the sample was comprised of children living in disadvantaged communities (77% were below the poverty line during preschool). Finally, we also split results by treatment status, and found no indication that the model differed between treatment and control groups. We also found small or null relations between the pre-k treatment and the stable components of each domain in our model that incorporated controls (see Table S4). Further, we found some indication of more behavioral problems for the intervention group in the time-varying component. However, it should be noted that our model was not designed to test for treatment effects and it differs from previously-published impact analyses in key ways (e.g., using multiple waves of data to fit the random intercept; no controls for site-level factors; see Raver et al., 2009; Watts et al., 2018). Further, although initial evaluations reported positive treatment effects at the end of preschool (Raver et al., 2009; 2011), follow-up work has been more mixed (Watts et al., 2018), which further diminishes the possibility of finding positive treatment impacts on the stable components of EF or behavioral problems (i.e., the random intercepts).

It should also be noted that measures of both EF and behavioral problems showed some degree of stability and change in development over the years considered here. We found that EF measures tended to have stronger latent factor loadings than auto-regressive paths, and we found that measures of behavioral problems had lower latent factor loadings when compared with EF, especially in the single measure model (see similar findings for social-emotional skills in Soland et al., 2019). Yet, for both domains, auto-regressive paths were largely statistically significant and substantively important in magnitude, suggesting that time-varying factors can influence development in each domain. It should be noted that we found small, statistically non-significant, auto-regressive paths between end-of-preschool EF and middle childhood EF, and the same was also true for the path from middle childhood EF to adolescent EF. This could be due to the fact that the measures changed across waves, but could also indicate that within-child fluctuations in EF do not strongly predict longer-term changes in EF. These two paths constituted the largest time span between points included in our study, and it may not be surprising that within-domain effects are weaker at longer time intervals at the within-child level. Thus, correlations between longitudinal measures of EF appear to be largely driven by inter-individual stability.

Our results further underscore that the field could stand to improve alignment between constructs and measures, a point that has been made repeatedly in the past (e.g., Morrison & Grammer, 2016). One interpretation of our findings could simply be that measurement modality dictates how these constructs appear to relate with one another (i.e., we observed small cross-lagged paths between EF and behavioral problems because EF was measured directly and behavioral problems via survey). If measurement modality is the most salient reason why we find only weak relations between behavioral problems and EF in our empirical work, then we are left with two possibilities. First, the measures could simply be too unreliable to capture the underlying constructs of behavioral problems and EF. Yet, this does not seem to account for the entirety of the problem given that the latent variables should be devoid of measurement error and we observed a relatively weak relation between the latent measures of EF and behavioral problems in our composite measure model, which should be less prone to measurement error than the single-measure model. The second possibility could be that the constructs do not relate strongly to one another, and we should reconsider theories regarding the interconnected nature of these developmental domains. If this is the case, then we need better explication of constructs in this area. Of course, these two possibilities are not mutually exclusive, and our progress theoretically will inevitably depend on sound measurement. Nevertheless, it is clear that the measurement issue should be a priority for research in this area, as theoretical progress will be difficult if our constructs are confounded by measurement modality.

Improved clarification of these relations carries real-world implications for intervention. Theoretical assertions regarding EF and behavioral regulation have informed the creation of interventions aimed to improve children's self-regulation (e.g., Nesbitt & Farran, 2021; Raver et al., 2009). The current findings provide little evidence to suggest that interventions narrowly aimed at improving EF will be efficacious in improving behavioral regulation (or vice versa), as our within-child effects were null. Moreover, while our data were drawn from intervention work, we saw little indication that random assignment to the early intervention focused on improving child self-regulation affected the developmental model (see Table S4).

Taken together, these findings highlight the need for future research to further clarify the relation between the various domains typically categorized under the self-regulation umbrella. As Morrison and Grammer (2016) noted, the "conceptual clutter" surrounding the EF and behavioral regulation literature has produced substantial confusion. Our results provide some clarity on the developmental structure of self-regulation, and suggest the possibility that future empirical work should potentially consider EF and behavioral problems as relatively independent developmental domains. However, it should be noted that our work has important limitations. Indeed, although we used measures of EF generally accepted in the field, many of our timepoints included proficiency on only one EF task (Hearts and Flowers), as did our single-measure models, making our cross-lagged paths susceptible to measurement error. Indeed, for many of our measures, we know of few studies reporting test-retest reliability, and it remains possible that low reliability could bias downward the within-child effects reported here. Relatedly, our sample was not especially large (*n* = 598), and several within-child cross-lagged paths were not significant, though the estimated coefficients were potentially meaningful in magnitude (i.e., ~ 0.10). Consequently, more work with larger samples may be needed to more precisely detect the effects that were beyond the level of precision possible in the current study. Finally, our model included varying time periods between waves, and measures changed across development. Thus, we cannot totally rule out measurement confounding as the lack of cross-lagged paths could be partly due to the changing nature of the measures over time.

In conclusion, the present study suggests that EF and behavioral problems may not be as developmentally intertwined as suggested by theory. Results demonstrated that to the extent that the two domains do relate to one another, the relation appears to exist primarily at the between-child level. In contrast, we found that within-child changes in EF had only limited effects on within-child changes in behavioral problems, and vice versa.

References

Achenbach, T. M. (1991). *Manual for the Teacher's Report Form and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. (2001). *ASEBA school-age forms & profiles*.

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

Bailey, D. H., Duncan, G. J., Watts, T., Clements, D., & Sarama, J. (2018). Risky business: Correlation and causation in longitudinal studies of skill development. *American* Ps*ychologist, 73*(1), 81-94.

Bailey, D. H., Oh, Y., Farkas, G., Morgan, P., & Hillemeier, M. (2020). Reciprocal effects of reading and mathematics? Beyond the cross-lagged panel model. *Developmental Psychology*, 56(5), 912-921.

Bailey, R., Jones, S. M. (2019). An integrated model of regulation for applied settings. *Clinical Child and Family Psychology Review, 22*, 2-23.

Barkley, R. A. (2001). The executive functions and self-regulation: An evolutionary neuropsychological perspective. *Neuropsychology review, 11*(1), 1-29.

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development, 88*(4), 1186-1206.

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.

Blair, C. (2003). Behavioral inhibition and behavioral activation in young children: Relations with self-regulation and adaptation to preschool in children attending Head Start. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 42*(3), 301-311.

Blair, C. (2016). The development of executive functions and self-regulation: A bidirectional psychobiological model. In Vohs, K. D., & Baumeister, R. F., *Handbook of self-regulation: Research, theory, and applications.* (pp. 417–439). New York: Guilford Press.

Blair, C., & Raver, C. C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *Plos One, 9*(11), 1-13.

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology, 66*, 711-731.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663.

Blair, C., & Ursache, A. (2011). *A bidirectional model of executive functions and self-regulation*. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, Theory, and Applications* (p. 300–320). New York: Guilford Press.

Bridgett, D. J., Burt, N. M., Edwards, E. S., & Deater-Deckard, K. (2015). Intergenerational transmission of self-regulation: A multidisciplinary review and integrative conceptual framework. *Psychological Bulletin, 141*(3), 602-654.

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot'and 'cool'executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly, 24*(3), 337-349.

Campbell, S. B., Denham, S. A., Howarth, G. Z., Jones, S. M., Whittaker, J. V., Williford, A. P., ... & Darling-Churchill, K. (2016). Commentary on the review of measures of early childhood social and emotional development: Conceptualization, critique, and recommendations. *Journal of Applied Developmental Psychology*, *45*, 19-41.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078.

Deković, M., Buist, K. L., & Reitz, E. (2004). Stability and changes in problem behavior during adolescence: Latent growth analysis. *Journal of Youth and Adolescence, 33*(1), 1-12.

Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive Function in Preschool-age Children: Integrating Measurement, Neurodevelopment, and Translational Research* (pp. 11–43). American Psychological Association

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental Psychobiology*, *29*(4), 315-334.

Doebel, S. (2020). Rethinking executive function and its development. *Perspectives in Psychological Science*, *15*(4), 942–956.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*(3), 259-268.

Eisenberg, N., Spinrad, T. L., & Eggum, N. D. (2010). Emotion-related self-regulation and its relation to children's maladjustment. *Annual Review of Clinical Psychology*, *6*, 495-525.

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*(2), 326.

Gandhi, J., Watts, T. W., Masucci, M. D., & Raver, C. C. (2020). The effects of two mindset interventions on low-income students' academic and psychological outcomes. *Journal of Research on Educational Effectiveness*, *13*(2), 351-379.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Test review behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235-238.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102.

Hsueh, J., Lowenstein, A. E., Morris, P., Mattera, S. K., & Bangser, M. (2014). *Impacts of social‐emotional curricula on three‐year‐olds: Exploratory findings from the Head Start CARES demonstration* (OPRE Report 2014-78). U.S. Department of Health and Human Services.

Hughes, C., & Ensor, R. (2008). Does executive function matter for preschoolers' problem behaviors? *Journal of Abnormal Child Psychology*, *36*, 1-14.

Hughes, C., Devine, R. T., Mesman, J., & Blair, C. (2020). Understanding the terrible twos: A longitudinal investigation of the impacts of early executive function and parent-child interactions. *Developmental Science*, *23*, 1-12.

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating models of self-regulation. *Annual Review of Psychology*, *72*, 319-345.

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology*, *26*(1), 403-422.

Kahle, S., Utendale, W. T., Widaman, K. F., Hastings, P. D. (2018). Parasympathetic regulation and inhibitory control predict the development of externalizing problems in early childhood. *Journal of Abnormal Child Psychology*, *46*, 237-249.

Karoly, P. (1993). Mechanisms of self-regulation: A systems view. *Annual Review of Psychology*, *44*(1), 23-52.

Li-Grining, C. P., McKinnon, R. D., & Raver, C. C. (2019). Self-Regulation in Early and Middle Childhood as a Precursor to Social Adjustment Among Low-Income, Ethnic Minority Children. *Merrill-Palmer Quarterly*, *65*(3), 265-293.

Lougheed, J. P., Duncan, R. J., Keskin, G., & Marceau, K. (2022). Longitudinal associations between mother-child conflict and child internalizing problems in mid-childhood. *Development and Psychopathology*, *34*(1), 263-272. https://doi.org/10.1017/S0954579420000863

Maccoby, E. E., Dowley, E. M., Hagen, J. W., & Degerman, R. (1965). Activity level and intellectual functioning in normal preschool children. *Child Development*, *36*(3), 761-770.

Mahone, E. M., & Hoffman, J. (2007). Behavior ratings of executive function among preschoolers with ADHD. *The Clinical Neuropsychologist, 12*(4), 569-586.

McAuley, T., Chen, S., Goos, L., Schachar, R., & Corsbie, J. (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function? *Journal of International Neuropsychological Society*, *16*, 495-505.

McClelland, M. M., & Cameron, C. E. (2011). Self-regulation and academic achievement in elementary school children. *New Directions for Child and Adolescent Development*, *2011*(133), 29-44.

McCoy, D. C., Jones, S., Roy, A., & Raver, C. C. (2018). Classifying trajectories of social–emotional difficulties through elementary school: Impacts of the Chicago School Readiness Project. *Developmental Psychology*, *54*(4), 772-787.

McCoy, D. C., Raver, C. C., Lowenstein, A. E., & Tirado-Strayer, N. (2011). Assessing self-regulation in the classroom: Validation of the BIS-11 and the BRIEF in low-income, ethnic minority school-age children. *Early Education & Development*, *22*(6), 883-906.

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*(4907), 933–938.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (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.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the national Academy of Sciences*, *108*(7), 2693-2698.

Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact* *findings from the Head Start CARES demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence* (OPRE Report 2014-44). U.S. Department of Health and Human Services.

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* (p. 327–348). American Psychological Association.

MS-YRBS: Centers for Disease Control and Prevention. (2015) Youth Risk Behavior Survey.

Murray, K. T., & Kochanska, G. (2002). Effortful control: Factor structure and relation to externalizing and internalizing behaviors. *Journal of Abnormal Child Psychology*, *30*(5), 503-514.

Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 638-648.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus: The comprehensive modeling program for applied researchers: User's guide*. Muthén & Muthén.

Nesbitt, K. T., & Farran, D. C. (2021). Effects of prekindergarten curricula: tools of the mind as a case study. *Monographs of the Society for Research in Child Development*, *86*(1), 7-119.

Nigg, J. T. (2017). Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *58*(4), 361-383.

O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G. (2014). An examination of the efficacy of INSIGHTS in enhancing the academic and behavioral development of children in early grades. *Journal of Educational Psychology*, *106*(4), 1156-1169.

Ogilvie, J. M., Stewart, A. L., Chan, R. C., & Shum, D. H. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology*, *49*(4), 1063-1107.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768-774.

Raffaelli, M., Crockett, L. J., & Shen, Y. L. (2005). Developmental stability and change in self-regulation from childhood to adolescence. *The Journal of Genetic Psychology*, 166(1), 54-76.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' preacademic skills: self-regulation as a mediating mechanism. *Child Development*, *82*(1), 362-378.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *77*(2), 302-316.

Raver, C. C., McCoy, D. C., Lowenstein, A. E., & Pess, R. (2013). Predicting individual differences in low-income children's executive control from early to middle childhood. *Developmental Science*, *16*(3), 394-408.

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*(1), 10–26.

Riesch, S. K., Anderson, L. S., Angresano, N., Canty-Mitchell, J., Johnson, D. L., Krainuwat, K. (2006). Evaluating content validity and test-retest reliability of the children's health risk behavior scale. *Public Health Nursing, 23*(4), 366-72.

Sasser, T. R., Bierman, K. L., & Heinrichs, B. (2015). Executive functioning and school adjustment: The mediational role of pre-kindergarten learning-related behaviors. *Early Childhood Research Quarterly*, *20*, 70-79.

Saunders, B., Milyavskaya, M., Etz, A., Randles, D., & Inzlicht, M. (2018). Reported self-control is not meaningfully associated with inhibition-related executive function: A Bayesian approach. *Collabra: Psychology*, *39*, 2-16.

Schmitt, S. A., Finders, J. K., Duncan, R. J., Korucu, I., Bryant, L. M., Purpura, D. J., & Elicker, J. G. (2021). Examining transactional relations between behavioral self-regulation and social-emotional functioning during the transition to kindergarten. *Developmental Psychology, 57*(12), 2093-2105.

Schoemaker, K., Mulder, H., Deković, M., & Matthys, W. (2013). Executive functions in preschool children with externalizing behavior problems: A meta-analysis. *Journal of Abnormal Child Psychology*, *41*, 457-471.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, *22*(2), 173-187.

Soland, J., Zamarro, G., Cheng, A., & Hitt, C. (2019). Identifying naturally occurring direct assessments of social-emotional competencies: The promise and limitations of survey and assessment disengagement metadata. *Educational Researcher*, *48*(7), 466-478.

Sulik, M. J., Blair, C., Berry, D., Mills-Koonce, R., Greenberg, M. (2015). Early parenting and the development of externalizing behavior problems: Longitudinal mediation through children's executive function. *Child Development*, *86*(5), 1588-1603.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131-143.

Ursache, A., & Raver, C. C. (2015). Iowa Gambling Task performance and executive function predict low-income urban preadolescents' risky behaviors. *Personality and Individual Differences*, *79*, 1-6.

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, *6*(2), 122-128.

Watts, T. W., Gandhi, J., Ibrahim, D. A., Masucci, M. D., & Raver, C. C. (2018). The Chicago School Readiness Project: Examining the long-term impacts of an early childhood intervention. *PloS one*, *13*(7), e0200144.

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

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.

Zill, N. (1990). *Behavior problems index based on parent report*. Child Trends.

Figure 1

RI-CLPM model

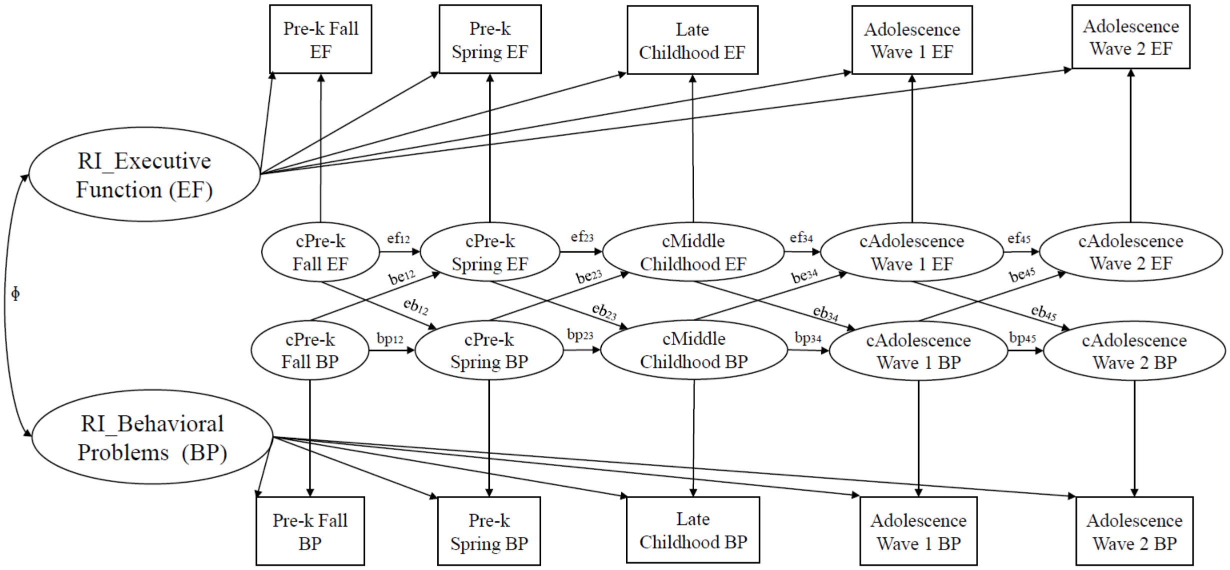

Table 1
Descriptive Characteristics of Key Executive Function and Behavioral Problem Measures

	Obs	Mean	Std. Dev.	Min	Max	Alpha
Executive Function						
Fall of Preschool						
Balance Beam (difference in seconds)	500	0.49	2.51	-14.50	14	0.50
Pencil Tap (prop. correct)	498	0.36	0.27	0	0.94	
Spring of Preschool						
Balance Beam (difference in seconds)	505	1.27	3.07	-10.00	22.00	0.45
Pencil Tap (prop. correct)	504	0.51	0.34	0	1	
Middle Childhood						
Hearts and Flowers (prop. correct)	384	0.85	0.16	0	1	
Adolescence Wave 1						
Hearts and Flowers (prop. correct)	460	0.66	0.19	0.06	1	
Adolescence Wave 2						
Hearts and Flowers (prop. correct)	401	0.72	0.18	0.18	1	
Behavioral Problems						
Fall of Preschool						
Externalizing (BPI - Parent)	512	0.36	0.30	0	1.94	0.89
Externalizing (BPI - Teacher)	529	0.32	0.32	0	1.72	
Spring of Preschool						
Externalizing (BPI - Teacher)	545	0.23	0.25	0	1.33	0.97
Externalizing (C-TRF - Teacher)	545	0.23	0.28	0	1.74	
Middle Childhood						
Behavior Dysregulation (BIS-BRIEF - Teacher)	350	0.24	0.28	0	1	0.96
Externalizing (BPI - Parent)	495	0.35	0.37	0	2	
Externalizing (TRF - Teacher)	360	0.25	0.32	0	1.69	
Externalizing (Risk - Student)	387	0.35	0.27	0	1	
Adolescence Wave 1						
Behavior Dysregulation (BIS-BRIEF - Student)	320	0.29	0.19	0	1	0.84
Externalizing (Risk - Student)	460	0.45	0.26	0	1	
Adolescence Wave 2						
Behavior Dysregulation (BIS-BRIEF-Student)	434	0.27	0.17	0	0.92	0.84
Externalizing (Risk - Student)	435	0.42	0.27	0	1	

Note. When multiple measures are listed for a given domain and wave, subscores were standardized and averaged to create a composite (e.g., Balance Beam and Pencil Tap at fall of preschool were individually standardized and then averaged together). Cronbach's alpha scores are presented for the behavior problem composite measures. BPI - Behavior Problem Index; C-TRF - Caregiver-Teacher Report Form; TRF - Teacher Report Form; Risk - Risky Behavior; BIS-BRIEF - Barratt Impulsiveness Scale and Behavior Rating Inventory of Executive Function.

Table 2

Correlations Among Key Composite Measures

		1	2	3	4	5	6	7	8	9	10
Executive Function											
1	Fall of Preschool	1									
2	Spring of Preschool	0.57	1								
3	Middle Childhood	0.23	0.20	1							
4	Adolescence Wave 1	0.28	0.25	0.36	1						
5	Adolescence Wave 2	0.21	0.29	0.37	0.54	1					
Behavioral Problem											
6	Fall of Preschool	-0.11	-0.14	-0.12	-0.07^	-0.14	1				
7	Spring of Preschool	-0.13	-0.12	-0.11	-0.16	-0.05^	0.44	1			
8	Middle Childhood	-0.08^	-0.12	-0.17	-0.07^	-0.13	0.31	0.35	1		
9	Adolescence Wave 1	-0.04^	-0.04^	-0.11^	-0.07^	-0.13	0.11	0.13	0.29	1	
10	Adolescence Wave 2	-0.07^	-0.07^	-0.07^	-0.02^	-0.09^	0.14	0.20	0.34	0.63	1

Note. n = 598. Pairwise correlations are presented for non-imputed data, so the sample size of each correlation differs within the total sample of 598 students.

^ denotes correlations that were **not** statistically significant (i.e., $p > 0.05$)

Table 3

Reciprocal Relations between Executive Function and Behavioral Problems

	Single Measures (Covariates)		**Composite (No Covariates)**		**Composite (Covariates)**	
	β	SE	β	SE	β	SE
Factor loadings						
EF	0.51***~0.53***		0.51***~0.52***		0.51***~0.55***	
BP	~ 0.24**		0.42***~0.43***		0.43***~0.45***	
Auto-regressive path						
cEF1 -> cEF2	0.43***	(0.04)	0.41***	(0.05)	0.35***	(0.05)
cEF2 -> cEF3	-0.04	(0.07)	-0.10	(0.08)	-0.11	(0.08)
cEF3 -> cEF4	0.10	(0.09)	0.11	(0.09)	0.19*	(0.08)
cEF4 -> cEF5	0.33***	(0.06)	0.36***	(0.06)	0.42***	(0.05)
cBP1 -> cBP2	0.51***	(0.05)	0.32***	(0.07)	0.29***	(0.07)
cBP2 -> cBP3	0.17*	(0.07)	0.18**	(0.07)	0.14*	(0.07)
cBP3 -> cBP4	0.30***	(0.06)	0.14*	(0.06)	0.13*	(0.06)
cBP4 -> cBP5	0.62***	(0.04)	0.56***	(0.04)	0.57***	(0.04)
Cross-lagged path						
cBP1 -> cEF2	-0.06	(0.05)	-0.08	(0.05)	-0.08	(0.05)
cBP2 -> cEF3	-0.01	(0.07)	-0.04	(0.08)	-0.03	(0.07)
cBP3 -> cEF4	0.06	(0.07)	0.01	(0.07)	0.00	(0.06)
cBP4 -> cEF5	0.00	(0.06)	-0.07	(0.06)	-0.08	(0.06)
cEF1 -> cBP2	-0.04	(0.04)	-0.08+	(0.05)	-0.08+	(0.05)
cEF2 -> cBP3	0.12+	(0.06)	-0.05	(0.06)	-0.04	(0.06)
cEF3 -> cBP4	0.01	(0.06)	-0.04	(0.06)	-0.06	(0.06)
cEF4 -> cBP5	0.06	(0.05)	0.05	(0.05)	0.05	(0.05)
Relation between EF & BP						
EF with BP (random intercepts)	-0.76*	(0.35)	-0.33**	(0.12)	-0.32*	(0.14)
cEF1 with cBP1	-0.07	(0.05)	-0.04	(0.05)	-0.05	(0.05)
cEF2 with cBP2	-0.05	(0.05)	0.00	(0.06)	0.01	(0.06)
cEF3 with cBP3	0.01	(0.06)	-0.12	(0.07)	-0.11	(0.08)
cEF4 with cBP4	0.06	(0.06)	-0.01	(0.06)	-0.03	(0.06)
cEF5 with cBP5	0.01	(0.06)	0.00	(0.06)	0.00	(0.06)
Model fit						
RMSEA	0.04		0.04		0.05	
CFI	0.97		0.98		0.89	
Observations	598		598		598	

+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Note. Models 1and 2 included no controls. In model 3, the latent factors for EF and BP were regressed on controls for child gender, ethnicity, age during preschool, parent education level, and preschool intervention status. In models 1 and 2, EF2 and BP2 were also regressed on preschool intervention status (i.e., posttest) and EF5 and BP5 were each regressed on high school intervention status (i.e., posttest). Time point 1 = fall preschool; time point 2 = spring preschool; time point 3 = middle childhood; time point 4= adolescence wave 1; time point 5 = adolescence wave 5. EF = executive function; BP = behavioral problems. "cEF1" refers to the latent variable for fall-of-preschool EF (see Figure 1).