Received: 15 July 2021 | Revised: 22 November 2021 | Accepted: 8 January 2022

DOI: 10.1111/desc.13236

Developmental Science WILEY

RESEARCH ARTICLE

Developmental trajectories of executive functions from preschool to kindergarten

Shannon E. Reilly[1] | Jason T. Downer[2] | Kevin J. Grimm[3]

[1] University of Virginia Health, Department of Neurology, Charlottesville, Virginia, USA

[2] University of Virginia, Center for Advanced Study of Teaching and Learning (CASTL), Charlottesville, Virginia, USA

[3] Arizona State University, Department of Psychology, Tempe, Arizona, USA

Correspondence
Shannon E. Reilly, University of Virginia Health, Department of Neurology, Charlottesville, VA 22908, USA.
Email: Shannon.Reilly596@gmail.com

Funding information
National Institute of Child Health and Human Development, Grant/Award Number: 2R01HD051498-06A1

Abstract

Executive functions (EF) are key predictors of long-term success that develop rapidly in early childhood. However, EF's developmental trajectory from preschool to kindergarten is not fully understood due to conceptual ambiguity (e.g., whether it is a single construct or multiple related constructs) and methodological limitations (e.g., previous work has primarily examined linear growth). Whether and how this trajectory differs based on characteristics of children and their families also remains to be characterized. In a primarily low-income, racially and ethnically diverse, typically developing, urban sample, the present study employed confirmatory factor analyses to examine the construct of EF and latent growth curve modeling to examine nonlinear growth across five time points. Results indicated that the development of a single EF construct with partial measurement invariance across time points was best characterized as nonlinear, with disproportionately more growth during the preschool year. There was individual variability in EF trajectories, such that children with higher EF at preschool entry showed relatively steeper growth during preschool compared to low-EF peers. However, children with less EF growth in preschool had steeper growth in kindergarten, attenuating the gains of high-EF preschoolers and resulting in some convergence in EF by the end of kindergarten. Findings have implications for (1) examining EF development in early childhood with more specificity in future studies, (2) informing the timing of EF interventions in early childhood, and (3) identifying children for whom such interventions might be especially beneficial.

KEYWORDS
developmental trajectories, executive functions, individual differences, preschool, school readiness

1 | INTRODUCTION

Executive functions (EF), a multidimensional cognitive skillset that guides goal-directed behavior (Baggetta & Alexander, 2016; Blair, 2016), develop rapidly in early childhood (Garon et al., 2008) and facilitate long-term academic achievement, positive social functioning, and school success (Baggetta & Alexander, 2016; Blair, 2016; Zelazo et al., 2016). Since foundational skills that emerge early on set the stage for more complex abilities (Masten & Cicchetti, 2010), it is essential that young children develop EF prior to and across the transition to kindergarten. This transition is characterized by increases in demands and expectations (Rimm-Kaufman & Pianta, 2000), which tax children's capacity to leverage EF skills (Blair et al., 2005; Jones et al., 2016) in order to regulate themselves in a classroom setting (e.g., follow directions; Bassok et al., 2016).

In recent decades, myriad studies examining EF in early childhood (e.g., Anderson & Reidy, 2012; Garon et al., 2008; Willoughby et al., 2012) have enhanced understanding of what EF looks like (Friedman & Miyake, 2017) at different ages (Hughes Ensor et al., 2009; Lee et al., 2013; Miyake et al., 2000) and how it relates to other skills (Blair et al., 2005; Clark et al., 2014). However, many studies are cross-sectional and thus do not examine changes in EF over time (Garon et al., 2008). More recently, longitudinal studies have begun to fill this gap (Hughes & Ensor, 2011; Hughes et al., 2009; Willoughby et al., 2012) but remain constrained by a limited number of time points that allow for only a rough understanding of EF's trajectory during the transition to kindergarten. Therefore, this study employs latent growth curve modeling to examine nonlinear growth across five time points during the 2 years of this transition period. Additionally, given the importance of EF development in early childhood for later success, along with work indicating that there are individual differences in EF development (Zelazo et al., 2016), it is important to understand the variability in children's developmental trajectories of EF at this key transition time from preschool to kindergarten.

RESEARCH HIGHLIGHTS

- This study examines EF development across five time points from preschool to kindergarten in a primarily low-income, ethnically and racially diverse, urban sample.
- EF was best characterized as a unitary factor, and growth from preschool to kindergarten was nonlinear, with more growth in preschool but significant individual variability.
- Children with relatively high EF at preschool entry had disproportionately steep growth during preschool, whereas children with less preschool growth had steeper growth in kindergarten.
- Some convergence in EF trajectories by kindergarten's end raises questions about how characteristics of children, families, classrooms, and schools might moderate variability in skill growth.

2 | DEVELOPMENT OF EXECUTIVE FUNCTIONS

EF has been conceptualized as a multidimensional construct comprised of three primary components—working memory, inhibitory control, and set shifting or cognitive flexibility (Baggetta & Alexander, 2016; Blair et al., 2005; Miyake et al., 2000; Zelazo et al., 2016)—that "coordinate multiple sources of information in the service of purposeful, goal-directed behavior" (Blair, 2016). *Working memory* involves keeping information in mind and working with or updating it (Diamond, 2013; Garon et al., 2008; Willoughby et al., 2012; Zelazo et al., 2016). *Inhibitory control* is the ability to "control one's attention, behavior, thoughts, and/or emotions to override a strong internal predisposition … and instead do what's more appropriate or needed" (Diamond, 2013). *Set shifting* entails "changing how we think about something" (Diamond et al., 2013), such as "considering someone else's perspective … or solving a mathematics problem in multiple ways" (Zelazo et al., 2016).

EF underpins academic achievement, social competence, and school success in the short- and long-term (Baggetta & Alexander, 2016; Blair, 2016) and enables increased benefit from learning opportunities (Morgan et al., 2018). Having well-developed EF in early childhood has been linked to growth in language, literacy, and math by the end of kindergarten (Fuhs et al., 2014; McClelland et al., 2014), academic achievement in elementary and middle school (Morgan et al., 2018; Sabol & Pianta, 2012), and young adulthood, and educational attainment by age 25 (McClelland, Acock, Piccinin, Rhea, & Stallings, 2013). Conversely, having meager EF abilities early on has been linked to maladaptive outcomes, such as learning and behavioral difficulties (Zelazo et al., 2016).

Like other emerging skills, expectations for EF are developmentally graded. For example, one might expect a preschooler to be able to keep in mind and execute the rule to use "walking feet," a 10-year-old to be able to pay attention in class despite distractions, and an adolescent to appropriately plan and execute writing a term paper instead of going out with friends. It is especially important to understand the nature of EF development during preschool and kindergarten because these years represent an important transition during which routines shift and demands and expectations increase (Rimm-Kaufman & Pianta, 2000), particularly around EF and related self-regulatory capacities (Bassok et al., 2016; Jones et al., 2016).

A large body of evidence suggests that EF hangs together as a single construct in early childhood (Hughes et al., 2009; Nelson et al., 2016; Wiebe et al., 2008; Wiebe et al., 2011; Willoughby et al., 2012) and even into middle childhood (Brydges et al., 2012), then becomes differentiated into the three component components in adolescence (Baggetta & Alexander, 2016; Best & Miller, 2010; Blair & Ursache, 2011; Miyake, 2000). From this work, it is clear in broad strokes that multiple components of EF are united early on and become increasingly refined and differentiated over time, but there is room for further precision around understanding the structure of EF within and across these early years.

3 | MODELING EF DURING A KEY DEVELOPMENTAL TRANSITION

EF research has been plagued by a "measurement impurity problem," meaning that many assessments in this domain tend to simultaneously measure multiple components of EF, and also tap other abilities such as motor skills, processing speed, and language (Anderson & Reidy, 2012; Baggetta & Alexander, 2016; Fuhs et al., 2014; Miyake et al., 2000; Zelazo et al., 2016). To address this problem, many studies use a latent variable approach to isolate an underlying EF skill that is common across multiple tasks (Miyake & Friedman, 2012; Zelazo et al., 2016). Often, the fit of the latent EF factor is stronger than associations among separate EF tasks, which tend to have relatively small correlations (Morgan et al., 2018; Wiebe et al., 2011). This indicates that EF tasks tap different aspects of a multidimensional construct that is more unified than the sum of its components.

In early childhood, EF has most consistently been found to be a single latent factor (Nelson, Sheffield et al., 2016; Wiebe et al., 2008, 2011; Willoughby, Blair et al., 2012; Zelazo et al., 2016), which aligns with the notion that EF is relatively undifferentiated early on and becomes more distinguished with age (Lehto et al., 2003; Miyake et al., 2000). However, a few early childhood studies (Lee, Ho, & Bull, 2013; Miller et al., 2012; Usai et al., 2014) have found evidence for two EF factors that disaggregate inhibitory control from working memory and set shifting. Understanding the structure of EF in early childhood is important because it can clarify conceptual and methodological ambiguity (Jones et al., 2016), guide avenues for future research, and inform interventions (Zelazo et al., 2016).

Much of the early work examining the structure of EF development has been cross-sectional (Garon et al., 2008), but it is important to examine skill growth longitudinally because EF rapidly develops in early childhood (Blair, 2016). An important consideration when tracking the functional form of EF longitudinally is the extent to which tasks consistently measure the same underlying construct over time (i.e., measurement invariance). Given the rapid progression of EF during early childhood, its developmentally graded nature, and its relation to other foundational cognitive abilities (e.g., language) that are simultaneously coming online, it is not surprising that prior research on measurement invariance of EF in early childhood has been mixed. Whereas some studies (e.g., Hughes & Ensor, 2011; Hughes et al., 2009; Willoughby, Wirth et al., 2012) have identified invariance adequately sufficient to conduct longitudinal growth models, other studies (e.g., Clark et al., 2014; Nelson, James et al., 2016; Nelson, Sheffield et al., 2016) have not found evidence of even partial measurement invariance. Researchers in the latter category have hypothesized that this could theoretically be related to the assertion that earlier on, EF tasks more strongly tap other foundational skills (e.g., processing speed, language) that are more widely present across the sample (and thus less differentiating) at older ages, and/or more logistically associated with initial floor and later ceiling effects of EF measures across this important development time period (e.g., Clark et al., 2014; Nelson, Sheffield et al., 2016). As such, assessment of measurement invariance in the context of longitudinal tracking of EF is essential for interpreting the findings and contextualizing them in prior literature.

Although there is a dearth of longitudinal studies examining EF development over time (Zelazo et al., 2016), several exceptions exist. Linear EF growth has been demonstrated within the preschool year in a racially and ethnically diverse sample (Fuhs & Day, 2011) and between ages 4 and 6 in a low-income, primarily white sample (Hughes et al., 2009; Hughes & Ensor, 2011). Moreover, early EF growth predicted academic and behavioral outcomes (Hughes & Ensor, 2011), such that slow growth was related to higher teacher-reported externalizing and internalizing behaviors in first grade and fast growth was linked with higher academic competence at age 6. This provides further evidence that differing EF trajectories have long-term implications for children's success in school.

More recent work has used additional time points to analyze nonlinear development in early childhood (Clark et al., 2013; Montroy et al., 2016; Wiebe et al., 2012; Willoughby, Wirth et al., 2012). Altogether, these studies have found evidence for nonlinear EF development from ages 3 to 7, with accelerated early development that slows over time. Specifically, in a racially and ethnically diverse, low-income, rural sample, Willoughby, Wirth, and colleagues (2012) found that 60% of EF growth occurred during ages 3–4, with slower improvement from 4 to 5. Two studies of a predominantly white, relatively socioeconomically at-risk sample (Clark et al., 2013; Wiebe et al., 2012) found more growth from ages 3 to 4 than ages 4–5 when looking at one specific component of EF. In predominantly a white, mixed-income sample, growth in EF (again assessed by a single measure) was found to be exponential, with faster growth in preschool that decelerated in elementary school (Montroy et al., 2016). This study identified a need to further understand the functional form of EF in the transition from preschool to elementary school with measures that attempt to capture the multidimensional nature of EF more comprehensively (Montroy et al., 2016).

4 | THE PRESENT STUDY

The present study aimed to replicate and extend prior research on EF in early childhood by examining whether EF is best conceptualized as a unitary latent construct and modeling the trajectory of EF. Its main contribution to the literature is that it investigates nonlinear growth in EF across five time points during a key transition point in early childhood within a low-income, racially and ethnically diverse, urban sample; past work has analyzed linear growth, examined fewer time points, and/or used a low-income rural or primarily white sample (e.g., Hughes et al., 2009; Wiebe et al., 2011; Willoughby, Wirth et al., 2012).

It was hypothesized that a unitary construct of EF would best fit the data and that this construct would be psychometrically equivalent across the five time points (i.e., that measurement invariance would be established). EF growth was expected to be nonlinear, with more growth occurring in preschool than in kindergarten (Montroy et al., 2016; Wiebe et al., 2012; Willoughby, Wirth et al., 2012). We also aimed to explore whether children's initial levels of EF were related to rates of EF growth across this transition period.

5 | METHOD

5.1 | Sample

Data for the present study were collected as part of the Understanding the Power of Preschool for Kindergarten Success (P2K) project, an observational study of children's experiences from preschool through kindergarten. A total of 104 publicly funded preschool classrooms participated across two cohorts; children then matriculated into 227 kindergarten classrooms in 35 schools, across two school districts. Preschool teachers were eligible for participation if they served children who would matriculate into kindergarten the following school year; in inclusion classrooms, the general education teacher was selected for participation. In cohort 1, 51 preschool teachers from the 59 eligible teachers who consented were randomly selected to participate. One child transitioned to a different classroom during the year and one teacher was replaced with a new teacher due to an extended leave, so a total of 53 preschool teachers participated from 51 classrooms. In cohort 2, 51 eligible teachers consented and were selected to participate; three children transferred to a different classroom during the year, and two teachers who moved or took an extended leave were replaced with new teachers, resulting in a total of 55 preschool teachers from 54 classrooms. Therefore, in total, 108 teachers within 104 preschool classrooms participated across the two cohorts.

Children in participating preschool classrooms were eligible for the study. Up to eight consented children per classroom were randomly selected to participate, blocked by gender. When fewer than eight children were consented, all consented children in the classroom were chosen. A total of 758 children participated in the study across both cohorts (cohort 1 $n = 380$; cohort 2 $n = 378$) and were followed into their kindergarten classrooms when possible. Consent was then obtained from kindergarten teachers of continuing study children. Fall of kindergarten data for both cohorts were collected on all children in attendance. However, in spring of the kindergarten year (i.e., the final time point) for cohort 1, there was a planned missingness design strategy in 50% of direct assessment data, guided by Little and Rhemtulla (2013). Specifically, children were randomly selected to participate in the EF direct assessment prior to this data collection window. In addition, due to COVID-19 disruptions, EF direct assessment data were not collected in spring 2020 (i.e., spring of kindergarten for cohort 2). This resulted in significantly more missing data in the final time point compared to the earlier four time points, a limitation which is discussed further below.

Preschool teachers were 98.04% female with a mean age of 43.2 years ($SD = 11.2$); 65.3% were white, non-Hispanic, 29.7% were Black or African American, and 3% were Hispanic or Latino/a. Teachers had an average of 15.1 years of experience at their current facility ($SD = 8.6$); 49.5% had a master's degree or higher. Approximately 94% of preschool classrooms were in public schools, and 6% were housed in Head Start centers. Class sizes ranged from 16 to 21 children, with a mean of 18 ($SD = 0.83$). On average there were 7.45 study children in preschool classrooms ($SD = 1.51$, range = 1–10) and 2.89 in kindergarten classrooms ($SD = 2.28$, range = 1–22).

Across the 104 preschool classrooms, 758 eligible and consented children participated in the present study. The sample was 49.4% female; 48.8% were Black/African American, 21.9% were white, non-Hispanic, 12.5% were Hispanic/Latino/a, and 14.4% were of another racial or ethnic background (e.g., Asian American, American Indian or Alaska Native) or multiple racial identities. At the beginning of preschool, children had a mean age of 52.63 months ($SD = 3.60$). Mothers had on average 13.41 years of education ($SD = 1.85$). Families had an average income-to-needs ratio of 1.45 ($SD = 1.06$), meaning that the average family in the study had an annual income at 145% of the federal poverty level for their household size (e.g., $35,478 for a family of four; U.S. Health & Human Services Department, 2016). Important to note is that these racial/ethnic proportions are well aligned with the general student population in the two school districts involved in this study, as is the high rate of economic disadvantage for families; therefore, the sample appears to be representative of the communities from which it was drawn.

5.2 | Procedure

5.2.1 | Recruitment

Preschool administrators in two urban areas in the southeastern United States were contacted to participate. Administrators and teachers were invited to attend study recruitment sessions. Interested, eligible teachers gave informed consent. Parents or guardians of children in participating classrooms were given a letter explaining the study and an informed consent. Of consented children, up to eight from each classroom were randomly selected to participate (blocked by gender). When children matriculated to kindergarten, their teachers were asked to consent to participate.

5.2.2 | Data collection

Parents and teachers completed demographic surveys in the fall of preschool and kindergarten. Trained data collectors administered direct child assessments in the fall, winter, and spring during preschool and in fall and spring during kindergarten. Again, for cohort 2, the COVID-19 pandemic disrupted data collection in spring 2020 (kindergarten).

5.3 | Measures

5.3.1 | Executive functions

EF Touch is a computer-administered battery of executive function tasks that has been shown to have good reliability and validity in an early childhood sample (Willoughby, Blair, et al., 2012; Willoughby et al., 2013). Three subtests from the battery were used in preschool. In the inhibitory control *Animal Go/No-Go* (Pig) task (Cronbach's $\alpha = 0.86$), farm animals flash quickly on the screen and children are asked to touch all animals except for the pig. In the working memory *Pick the Picture* (PtP) task (Cronbach's $\alpha = 0.60$), children are asked to consistently choose pictures from a set that they have not chosen before, holding in mind those they have already picked. In the set-shifting *Something's the Same* (StS) task (Cronbach's $\alpha = 0.76$), the screen displays two pictures that are similar on one dimension (e.g., color, size), and then adds another picture that is the same as one of the first pictures in a different way. Children are asked to choose which of the first two pictures is similar to the new picture.

In kindergarten, a fourth EF Touch subtest measuring inhibitory control, *Spatial Conflict* (Arrows), was added because of ceiling effects on the preschool inhibitory control (Pig) task. In previous studies (Willoughby, Wirth et al., 2012), Arrows has been added at follow-up time points to reflect the developmentally graded nature of EF. In this task, children are first asked to push the button aligned with the direction in which an arrow points. Then, in a second trial, they are asked to push the button in the opposite direction of the arrow. Two subscales—Arrows Congruent (trials in which the button and the arrow direction align; Cronbach's $\alpha = 0.87$) and Arrows Switch (trials in which they are discrepant; Cronbach's $\alpha = 0.92$)—were included as separate variables in analyses. All other EF Touch subtests remained the same in the kindergarten year. Proportion correct was calculated for each of the subtests at each time point (three subtests in preschool, five subtests in kindergarten).

Head-Toes-Knees-Shoulders (HTKS) is a measure of executive functioning that requires inhibitory control, working memory, and attention focusing (Ponitz et al., 2009). Children are asked to learn simple commands (i.e., "touch your head," "touch your toes") then do the opposite of what the assessor said (e.g., touch their toes when asked to touch their head). An advanced trial incorporating the same task with knees and shoulders is given to children who do not reach a ceiling on the first set of items. Children receive two points for each correct response and one point for a self-corrected response; scores ranged from 0 to 60 across 30 trials. HTKS has been shown to have adequate reliability and concurrent and predictive validity in a preschool sample (Ponitz et al., 2009).

Inhibitory control was also measured by the Pencil Tap subtest of the *Preschool Self-Regulation Assessment* (PSRA; Smith-Donald, Raver, Hayes, & Richardson, 2007). Children are asked to tap their pencil once when the examiner taps twice and twice when the examiner taps once. Scores represent percentage of correct responses. This test has shown acceptable concurrent and construct validity (Smith-Donald et al., 2007).

These measures (*EF Touch* subtests, HTKS, and Pencil Tap) were included in a CFA to determine whether EF was best characterized by a single latent factor in these data (see Analytic Plan for further details). Table 1 provides descriptive statistics for as well as bivariate correlations between all key study variables.

5.4 | Analytic plan

5.4.1 | Clustering

Intraclass correlation coefficients (ICCs) from unconditional regression models and design effects (Mass & Hox, 2004) were examined to determine the need to control for nestedness of children within preschool and/or kindergarten classrooms. The design effect takes into account the ICC and the average cluster (in this case, classroom) size (design effect = 1+[average cluster size-1] · ICC). This second step of assessing nestedness was important for this study because of the small cluster sizes in kindergarten (mean number of study children in a kindergarten classroom = 2.89, $SD = 2.28$). ICCs greater than 0.10 indicate that a given type of nestedness should be considered in clustering (Raudenbush & Bryk, 2002). Design effects greater than or equal to 2 indicate that multilevel models are necessary; if they are less than 2, then multilevel modeling is not necessary (Maas & Hox, 2004).

ICCs ranged from 0.01 to 0.17, and two out of 29 ICCs were above the 0.10 cutoff (one in preschool, one in kindergarten). However, all design effects (ranging from 1.01 to 1.75, with average cluster sizes of 7.45 for preschool and 2.89 for kindergarten) were less than the cutoff of 2 (Maas & Hox, 2004). Thus, multilevel models were not fit. However, to provide a more conservative, robust estimate of results given the dependency of data when children are nested in classrooms, all analyses used TYPE = COMPLEX in M*plus*, which uses maximum likelihood estimation with robust standard errors.

5.4.2 | Missing data

Missingness was fairly minimal during the preschool assessments with missing data rates under 10%. The fall kindergarten assessment variables had a higher level of missingness, with rates between 15% and 21%, due to reasons such as children moving out of the district, being lost to follow-up, and kindergarten teachers declining to participate in the study. The spring kindergarten assessment variables had the highest level of missingness with a rate of 80%. The large percentage of data missing at the kindergarten spring data collection window can be considered MCAR because children in cohort 1 were randomly selected to participate at this time point, and inability to collect direct assessment data with children in cohort 2 at this time point because of COVID-19 school closures.

Covariate-dependent missingness was assessed to determine whether study variables (i.e., components of EF at each time point) were significantly predicted by child characteristics (e.g., gender, race, ethnicity). Correlations between missing data indicators and child characteristics were estimated. Correlations were weak, with the strongest correlation being 0.11. Importantly, the likelihood of missing values was more strongly related to the other study variables (e.g., missingness of Pencil Tap at the first assessment correlated with the other assessments of EF); however, these correlations were moderate at best (e.g., $r = 0.30$). Thus, the effect of missingness on results is expected to be minimal (Collins et al., 2001), and data can be considered MAR (Li, 2013). As such, full information maximum likelihood estimation was implemented to retain cases missing data on study variables.

5.4.3 | Confirmatory factor analysis

A CFA was conducted in Mplus Version 8.3 (Muthén & Muthén, 1998–2015) on the HTKS, Pencil Tap, and *EF Touch* subtests to determine whether a single unitary factor of EF fit the data. The hypothesized one-factor solution (in which all measures load onto a single latent factor) was compared to a two-factor solution that separated inhibitory

TABLE 1 Descriptive statistics

	01.	02.	03.	04.	05.	06.	07.	08.	09.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.
Mean	6.96	8.76	6.69	10.39	5.60	7.27	9.19	7.30	16.99	7.20	7.33	9.31	7.84	22.82	7.69	7.72	9.56	8.40	33.50	8.63	8.11	7.13	7.95	9.64	8.81	40.46	9.17	8.59	7.85
Standard Deviation	1.13	1.32	1.32	14.62	3.48	1.12	1.01	1.52	18.28	3.04	1.16	0.91	1.45	19.64	2.65	1.01	0.69	1.35	19.01	2.04	2.02	2.98	0.91	0.55	1.20	17.18	1.66	1.80	2.72
Correlation Matrix																													
1. PTP Fall PK	1.00																												
2. PIG Fall PK	0.26	1.00																											
3. STS Fall PK	0.24	0.24	1.00																										
4. HTKS Fall PK	0.27	0.31	0.36	1.00																									
5. PT Fall PK	0.29	0.44	0.34	0.46	1.00																								
6. PTP Winter PK	0.35	0.29	0.26	0.25	0.29	1.00																							
7. PIG Winter PK	0.31	0.53	0.22	0.26	0.39	0.30	1.00																						
8. STS Winter PK	0.31	0.31	0.48	0.37	0.43	0.39	0.33	1.00																					
9. HTKS Winter PK	0.27	0.34	0.02	0.65	0.47	0.31	0.33	0.46	1.00																				
10. PT Winter PK	0.27	0.41	0.24	0.32	0.62	0.31	0.47	0.38	0.43	1.00																			
11. PTP Spring PK	0.28	0.24	0.21	0.26	0.28	0.42	0.27	0.32	0.32	0.27	1.00																		
12. PIG Spring PK	0.17	0.40	0.16	0.23	0.36	0.22	0.48	0.27	0.26	0.38	0.31	1.00																	
13. STS Spring PK	0.28	0.34	0.45	0.38	0.46	0.38	0.35	0.61	0.44	0.41	0.41	0.36	1.00																
14. HTKS Spring PK	0.26	0.32	0.01	0.57	0.54	0.28	0.32	0.42	0.68	0.44	0.37	0.36	0.53	1.00															
15. PT Spring PK	0.25	0.42	0.23	0.32	0.58	0.30	0.46	0.40	0.37	0.70	0.36	0.46	0.48	0.46	1.00														
16. PTP Fall K	0.24	0.26	0.20	0.28	0.38	0.36	0.29	0.34	0.28	0.33	0.39	0.31	0.34	0.29	0.39	1.00													
17. PIG Fall K	0.23	0.29	0.15	0.18	0.27	0.14	0.33	0.19	0.20	0.32	0.21	0.40	0.27	0.23	0.38	0.30	1.00												
18. STS Fall K	0.26	0.32	0.37	0.36	0.44	0.27	0.34	0.53	0.39	0.37	0.33	0.29	0.60	0.43	0.44	0.36	0.36	1.00											
19. HTKS Fall K	0.30	0.42	0.32	0.43	0.51	0.26	0.39	0.40	0.49	0.50	0.26	0.28	0.42	0.55	0.48	0.33	0.33	0.52	1.00										
20. PT Fall K	0.17	0.41	0.18	0.23	0.47	0.21	0.37	0.26	0.28	0.57	0.22	0.33	0.33	0.30	0.57	0.31	0.33	0.32	0.42	1.00									
21. AC Fall K	0.07	0.13	0.03	0.03	0.14	0.07	0.17	0.08	0.07	0.17	0.12	0.20	0.11	0.12	0.24	0.11	0.20	0.18	0.15	0.16	1.00								
22. AS Fall K	0.02	0.12	0.11	0.12	0.16	0.03	0.16	0.15	0.15	0.17	0.07	0.13	0.14	0.14	0.19	0.13	0.15	0.14	0.19	0.18	0.41	1.00							
23. PTP Spring K	0.21	0.23	0.23	0.15	0.26	0.34	0.15	0.21	0.19	0.28	0.39	0.25	0.25	0.20	0.24	0.40	0.40	0.35	0.26	0.30	0.17	0.14	1.00						
24. PIG Spring K	0.11	0.18	0.15	0.13	0.34	0.21	0.16	0.23	0.14	0.29	0.24	0.22	0.26	0.18	0.29	0.23	0.42	0.26	0.25	0.35	0.32	0.24	0.38	1.00					
25. STS Spring K	0.24	0.38	0.39	0.27	0.45	0.26	0.38	0.50	0.30	0.41	0.30	0.33	0.61	0.36	0.40	0.27	0.25	0.62	0.45	0.35	0.20	0.18	0.42	0.26	1.00				
26. HTKS Spring K	0.24	0.42	0.32	0.36	0.47	0.37	0.46	0.34	0.41	0.49	0.35	0.29	0.43	0.44	0.47	0.33	0.37	0.47	0.70	0.41	0.08	0.14	0.34	0.34	0.47	1.00			
27. PT Spring K	0.02	0.30	0.20	0.21	0.28	0.19	0.20	0.14	0.21	0.30	0.09	0.24	0.17	0.29	0.34	0.16	0.26	0.16	0.31	0.41	−0.07	0.03	0.15	0.28	0.16	0.47	1.00		
28. AC Spring K	−0.01	0.06	0.08	0.01	0.06	0.05	0.20	0.00	0.08	0.21	0.08	0.16	0.10	−0.01	0.21	0.06	0.22	0.18	0.11	0.22	0.45	0.18	0.09	0.26	0.09	0.19	0.11	1.00	
29. AS Spring K	−0.04	0.10	0.11	0.05	0.14	0.18	0.19	0.06	0.07	0.22	0.11	0.13	0.17	0.07	0.24	0.11	0.25	0.16	0.12	0.21	0.20	0.21	0.13	0.19	0.15	0.25	0.18	0.45	1.00

Abbreviations: PTP, EF Touch Pick the Picture; PIG, EF Touch Animal Go/No-Go; STS, EF Touch Something's the Same; HTKS, Head Toes Knees Shoulder; PT, Pencil Tap; AC, EF Touch Arrows Congruent; AS, EF Touch Arrows Switch.

control from working memory and set shifting (following Miller et al., 2012; Usai et al., 2014). In the two-factor model, Pencil Tap and Pig subtests were indicators of the Inhibitory Control factor, whereas *Pick the Picture* and *Something's the Same* subtests loaded on the Working Memory/Set Shifting factor. HTKS, which assesses aspects of all three EF components (Ponitz et al., 2009), loaded on both factors. For the kindergarten time points, the Arrows Congruent and Arrows Switch were regressed on the single latent factor in the unitary model and on the Inhibitory Control factor in the two-factor model. A three-factor model was not examined because (1) most prior literature indicates that EF is a unitary factor in early childhood, with a small subset finding two factors, and (2) there were not enough measures specific to each of the three components to enable an examination of each separately. In CFA models, the variance of the latent factor was constrained to 1 for identification. A unique covariance between the two Arrows subscales (Congruent and Switch) was modeled. In the two-factor model, the latent factors were allowed to covary with one another.

Model fit was assessed using Bentler's comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the χ^2 statistic. Because the latter is sensitive to relatively few degrees of freedom (Kenny et al., 2015; Perry et al., 2015), the Satorra-Bentler χ^2 difference test was used to account for the non-normality of data and complex analytical models (Satorra & Bentler, 2001). Models with CFI greater than or equal to 0.95 and RMSEA less than or equal to 0.05 are considered to exhibit "good fit," whereas models meeting just one of these criteria or models with CFI greater than or equal to 0.90 and RMSEA less than 0.08 are considered to have "adequate fit" (Fuhs & Day, 2011; Hu & Bentler, 1999; Willoughby, Wirth et al., 2012). SRMR less than 0.08 is indicative of good fit (Chen et al., 2005).

5.4.4 | Measurement invariance across time

Once the best-fitting model (one- vs. two-factor) was determined, a longitudinal CFA model was fit for each of the five time points to determine whether measurement invariance of the EF construct over time could be established. Establishing invariance indicates that the EF construct is psychometrically equivalent at each time point, which is a prerequisite before moving forward with latent growth curve modeling (Fuhs & Day, 2011; Meredith & Horn, 2001; Willoughby, Wirth et al., 2012). Testing measurement invariance involves fitting a set of increasingly restrictive models (Fuhs & Day, 2011; Schmitt et al., 2017) and comparing their model fit (Cheung & Rensvold, 2002).

In the first, least restrictive model (configural invariance), factor loadings, intercepts, and unique factor variances were freely estimated over time. In this model, the latent variable means were fixed to 0 and the latent variable variances were fixed to 1 at all time points for identification (Schmitt et al., 2017; Willoughby, Wirth et al., 2012). The latent factors were allowed to covary across time points, and the unique factors of the same indicator were allowed to covary over time.

Next, metric (weak) invariance was tested by setting the factor loadings to be equal across time. In this model, the latent variable variance at the first time point was fixed at 1 and the remaining latent variable variances were estimated. The fit of the metric invariance model was compared to fit of the configural invariance model using the likelihood ratio test (LRT), as well as absolute and global fit indices. If the metric invariance model does not fit significantly worse than the configural invariance model, then the factors are measuring the same construct across time. Next, we fit the scalar (strong) invariance model, in which the factor loadings and measurement intercepts were constrained to be equal across time points. The means and variances of the latent variables were freely estimated, with the exception of the first measurement occasion. Scalar invariance indicates that the factors are measuring the same construct in the same scale over time, and is the minimum level of measurement invariance necessary to examine change in the factor with growth models (Grimm et al., 2017; Meredith & Horn, 2001). Finally, strict invariance was tested by additionally constraining the unique factor variances to be equal across time. Partial invariance models, where some of the measurement parameters are equal over time, were considered when constraining all factor loadings, intercepts, or residual variances led to a severe reduction in model fit.

5.4.5 | Longitudinal growth analysis

Second-order growth models (McArdle, 1988) were fit to examine changes in EF across grade. Second-order growth models examine changes in the first-order EF latent variable as long as a form of strong invariance can be established (e.g., full strong invariance, partial strong invariance). There are several advantages of fitting second-order growth models compared to first-order growth models (i.e., growth models specified for an observed variable; see von Oerzen et al., 2010). Different functional forms of change were examined including the (1) intercept only, (2) linear growth, (3) latent basis growth, and (4) bilinear spline growth model, with the knot or transition point between preschool and kindergarten. The intercept-only model was fit to examine whether systematic change was evident (i.e., if the intercept-only model is rejected, then systematic change is present). The linear and latent basis growth models examine whether there is a single change process for EF across preschool and kindergarten, with the latent basis model allowing for non-constant change within each student. The bilinear spline growth model was fit to examine whether the change process for EF was different during the preschool year compared to the kindergarten year.

6 | RESULTS

6.1 | Confirmatory factor analyses

The CFA models were fit to the data from each measurement occasion separately using the preschool or kindergarten classroom as the cluster variable with `TYPE = COMPLEX` in M*plus*. Table 2 contains the model fit statistics for the one- and two-factor confirmatory factor models fit to the EF indicators at each of the five measurement occasions. The fit of

TABLE 2 Confirmatory factor analysis model fit statistics

Measurement occasion	Fall PreK	Winter PreK	Spring PreK	Fall K	Spring K
One factor models					
Sample size	755	736	717	650	154
Fit statistics					
CFI	0.987	0.950	0.987	0.981	0.946
RMSEA	0.041	0.087	0.042	0.035	0.062
SRMR	0.022	0.034	0.022	0.024	0.047
$\chi^2(df)$	11.324(5)	32.904(5)	11.282(5)	23.590(13)	20.592(13)
Two factor models					
Fit statistics					
CFI	1.000	1.000	1.000	0.980	0.966
RMSEA	0.000	0.031	0.000	0.040	0.054
SRMR	0.013	0.015	0.009	0.0222	0.041
$\chi^2(df)$	2.888(3)	5.064(3)	1.785(3)	22.261(11)	15.851(11)

the one-factor model was at least adequate (i.e., CFI $\geq$ 0.90 and RMSEA $\leq$ 0.08) at all time points, with the exception of the winter assessment in preschool; however, the SRMR was less than 0.08 in all models, suggesting good model fit. The two-factor model demonstrated good fit in preschool and adequate fit at the kindergarten measurement occasions. Although the two-factor model fit statistically better than the one-factor model in preschool fall ($\Delta\chi^2(2) = 8.06$, $p < 0.05$), winter ($\Delta\chi^2(2) = 25.97$, $p < 0.01$), and spring ($\Delta\chi^2(2) = 8.38$, $p < 0.01$), there were no significant differences between the two models in the fall of kindergarten ($\Delta\chi^2(2) = 1.81$, p = ns) and the spring of kindergarten ($\Delta\chi^2(2) = 4.02$, p = ns). Moreover, the two factors in the two-factor models were always highly correlated ($r \sim 0.80$), indicating that the two factors were not very distinguishable. Given the preference for the one-factor model and the adequate overall model fit of the one-factor model at most time points, the single factor model of EF was retained in all subsequent analyses.

6.2 | Measurement invariance across time

The longitudinal CFA models were fit using the preschool classroom as the cluster variable with `TYPE = COMPLEX`. With `TYPE = COMPLEX`, the standard chi-square difference testing is not appropriate, so chi-square difference testing was carried out following Satorra and Bentler (2010). The configural invariance model fit the data well ($\chi^2(313) = 460.600$, $p < 0.01$, CFI = 0.973, RMSEA = 0.025, SRMR = 0.055). Constraining the factor loadings (metric invariance) resulted in significantly worse fit ($\Delta\chi^2(18) = 198.487$, $p < 0.01$); however, the model showed adequate model fit (CFI = 0.930, RMSEA = 0.039, SRMR = 0.125). The Pencil Tap variable was found to be a major cause for the lack of metric invariance. Freely estimating this factor loading across time led to good model fit (CFI = 0.960, RMSEA = 0.029, SRMR = 0.124). Next, the measurement intercepts were constrained to be equal and the mean of the factor was allowed to change over time. This partial scalar invariance model fit significantly worse than the partial metric invariance model ($\Delta\chi^2(14) = 67.582$, $p < 0.01$); however, the overall model fit remained adequate (CFI = 0.949, RMSEA = 0.033, SRMR = 0.113). Next, we constrained the unique factor variances to be equal over time (with the exception of Pencil Tap). This partial strict invariance model fit significantly worse than the partial scalar invariance model ($\Delta\chi^2(18) = 113.885$, $p < 0.01$) and most of the additional model misfit was primarily due to the equality constraint on the unique factor variance for the Pig subtest of the *EF Touch*. Relaxing this constraint led to a model in which there remained a significant increase in model misfit compared to the partial scalar invariance model ($\Delta\chi^2(14) = 53.629$, $p < 0.01$); however, the overall model fit was adequate (CFI = 0.939, RMSEA = 0.035, SRMR = 0.139).

The factor means increased at each occasion with a fairly linear increase across the three preschool assessments (0, 7.652, 13.543 for the fall, winter, and spring assessments, respectively) and somewhat slower increases during kindergarten (22.677, 28.824 for the fall and spring assessments). The factor variances increased over the preschool year (98.994, 188.092, 220.407 for the fall, winter, and spring assessments, respectively), but the variances were smaller in kindergarten (173.380, 129.606 for the fall and spring assessments). The EF factors were highly correlated over time. The smallest correlation was 0.770, between the EF factor in the winter of preschool and the EF factor in the spring of kindergarten.

6.3 | Second-order latent growth curve models

The second-order growth models were built upon the first-order confirmatory factor model with partial strict invariance. Given that our partial strict invariance factor model serves as the basis for

TABLE 3 Model fit statistics for second-order growth models

Fit statistics	Intercept only	Linear	Latent basis	Bilinear spline
CFI	0.645	0.906	0.920	0.930
RMSEA	0.087	0.045	0.042	0.039
SRMR	0.299	0.192	0.203	0.148
$\chi^2(df)$	2517.097(371)	936.952(368)	845.509(365)	789.467(364)

Note. Bilinear spline growth model had an inadmissible parameter estimate.

TABLE 4 Growth parameter estimates for the latent basis growth model fit to the executive function factor

Parameter	Estimate	Standard error
Slope factor loading		
Fall preschool	0.000	–
Winter preschool	4.980	0.276
Spring preschool	8.686	0.325
Fall kindergarten	14.217	0.408
Spring kindergarten	18.000	–
Means		
Intercept	0.000	–
Slope	1.597	0.054
Variances		
Intercept	114.445	10.737
Slope	0.027	0.036
Executive function factor residual	20.820	2.247
Covariance		
Intercept-slope	1.552	0.480

Note. '–' indicates that the parameter was fixed and does not have a standard error.

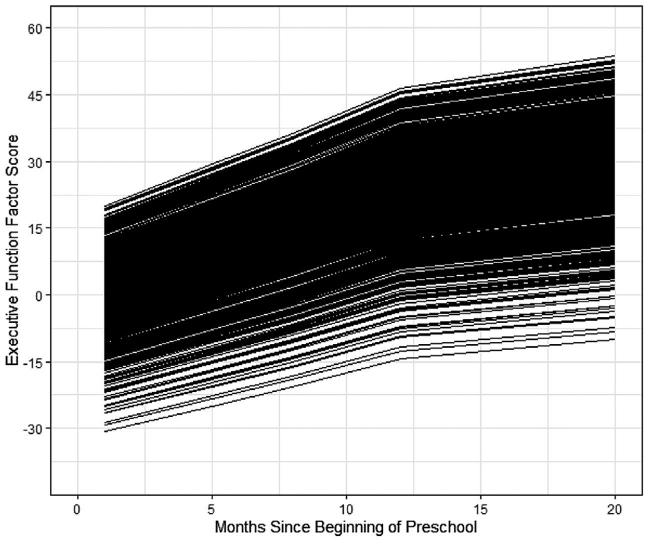

FIGURE 1 Predicted EF factor scores over time from the latent basis model

examining change, the growth models and their parameters are predominantly determined by the indicators whose model parameters are invariant over time.

The intercept only, linear growth, latent growth, and bilinear spline growth models were specified to examine changes in the first-order EF factor. Model fit statistics for the second-order growth models are contained in Table 3. The latent basis growth model fit the data best (χ^2 (365) = 845.509, $p < 0.001$, RMSEA = 0.042, CFI = 0.920; SRMR = 0.203) and did not have convergence issues. Parameter estimates of the growth parameters are contained in Table 4. The first (fall preschool) and last (spring kindergarten) factor loadings for the slope were fixed at 0.000 and 18.000 to indicate the number of months between the assessments. The estimated slope factor loadings were 4.980 (winter preschool), 8.686 (spring preschool), and 14.217 (fall kindergarten). These slope loadings are used to construct the shape of changes and indicate that children's average gains in preschool were more than twice as large as their average gains in kindergarten. This can be seen in Figure 1, which is a plot of the predicted EF factor scores over time from the latent basis model. There was significant variance in the intercept, suggesting that children differed from one another in their EF in the fall of preschool; however, the slope variance was not significant, suggesting a fairly homogenous rate of change in EF. Moreover, children who had higher predicted EF factor scores in the fall of preschool tended to show more change in EF during preschool and kindergarten ($r = 0.876$).

7 | DISCUSSION

EF is a foundational ability that develops rapidly in early childhood (Garon et al., 2008) and has been linked to school success in the short- and long-term (Baggetta & Alexander, 2016; Blair, 2016; Zelazo et al., 2016). The present study leveraged data across five time points in a primarily low-income, racially and ethnically diverse sample of children to investigate the latent construct of EF and the trajectory of its growth from preschool to kindergarten.

Both a unitary and a two-factor EF construct fit the data well. As hypothesized, the one-factor model, which demonstrated adequate fit across the five time points, was selected because data indicated that the one-factor model provided a superior fit at the majority of time points, based on prior literature (e.g., Nelson, Sheffield et al., 2016), and for parsimony. There *was* evidence of partial measurement invariance, justifying the use of a longitudinal latent growth model; however, consistent with prior work (e.g., Willoughby, Wirth et al., 2012) *strong* measurement invariance was not obtained, indicating some developmental differences in the underlying EF construct over time. As expected, children demonstrated significant, non-linear growth in EF from preschool to kindergarten, with more substantial growth occurring during preschool. Interestingly, children who started preschool with higher EF showed the greatest growth in EF prior to kindergarten; however, those children who ended up entering preschool with lower EF actually had steeper growth in EF during the kindergarten year, compared to initially high-EF peers.

7.1 | Adequate support for EF as a unitary factor across time

The present study supports prior research that has found EF to be a unitary construct (Nelson, Sheffield et al., 2012; Wiebe et al., 2008, 2011; Willoughby, Blair et al., 2012; Zelazo et al., 2016). However, the fact that both the unitary and bipartite constructs of EF fit the data well does not discount prior work that focused on a bifactor (Nelson, James et al., 2016) or two-factor model of EF (Lee et al., 2013; Miller et al., 2012; Usai et al., 2014) and indicates that more longitudinal EF measurement work is needed, especially across this key early childhood transition period. It will continue to be important to understand the developmentally graded structure of EF to further clarify conceptual and methodological ambiguity (Baggetta & Alexander, 2016; Jones et al., 2016; Zelazo et al., 2016).

Although findings were generally consistent with hypotheses, it was unexpected that when the two-factor model did fit significantly better, it was during preschool rather than later in development, when the EF construct has been hypothesized and found to become more differentiated (Lee et al., 2013). This finding calls into question whether the significant difference between the one- and two-factor model early in preschool and the subsequent superiority of the unitary construct through kindergarten reflected actual developmental changes in the construct of EF or whether it was primarily related to limitations in measurement. Previous work has suggested that we may not yet have the tools to be able to precisely assess and adequately differentiate between EF subcomponents at different ages (Jones et al., 2016).

Relatedly, in this study Pencil Tap was not found to be invariant over time, indicating that this task may be assessing a different construct at different time points between preschool and kindergarten. It is possible that early in preschool this measure may load more on certain skills such as motor control and auditory comprehension, and that it is more strongly based on inhibitory control later in preschool and in kindergarten. This is consistent with a previous finding by Clark and colleagues (2014) that early in childhood (i.e., age 3), a latent executive functioning factor was indistinguishable from processing speed and only became differentiated from this foundational cognitive skill at subsequent time points. Within the current study, there was a bimodal distribution for Pencil Tap in the fall of preschool, such that the two most common performances were close to 0% correct and close to 100% correct (i.e., characterized by both floor and ceiling effects), whereas at all subsequent time points, the distribution was unimodal and negatively skewed. The implications of this for future research are that perhaps Pencil Tap may be a more appropriate tool for assessing inhibitory control when children on average have adequate motor control, attention, and comprehension. Future work should continue to explore age effects on Pencil Tap performance, given that it is a commonly used early childhood measure.

Prior to interpreting LGCM results, it is essential to recognize that the fit of measurement invariance and LGCMs was generally adequate, but not strong. Moreover, it is important to consider potential conceptual and technical reasons for less than ideal fit and only partial invariance. Research on the construct of EF has been plagued by a "measurement impurity problem" given its complexity, such that measures purporting to assess EF also tend to assess other cognitive functions (e.g., motor skills, processing speed, and language) and often assess multiple subcomponents of EF simultaneously (Anderson & Reidy, 2012; Baggetta & Alexander, 2016; Fuhs et al., 2014; Miyake et al., 2000; Zelazo et al., 2016). This concept, that loadings of various skills onto performance of a single task can change over time (e.g., as was hypothesized for Pencil Tap above), as well as the role of psychometric properties (e.g., floor and ceiling effects) have been well described in recent work by Nelson, James, and colleagues (2016) and Nelson, Sheffield, and colleagues (2016). In the current study, ceiling effects were indeed observed for Pencil Tap (kindergarten spring $M = 0.91$, $SD = 0.17$, possible range = 0–1), as well as for another measure, *EF Touch* Pig, by spring of kindergarten ($M = 0.96$, $SD = 0.05$, possible range = 0–1). It is arguable that rather than a measurement limitation, these ceiling effects may reflect an underlying developmental change in the construct of EF (e.g., as described by Nelson, Sheffield et al., 2016), such that it needs to be assessed with different, more complex tasks over time. To account for this and make the construct more developmentally graded, this study introduced two additional inhibitory control subscales during the kindergarten time points. An argument against this hypothesis, however, is that *EF Touch* has been found to have adequate measurement invariance over time (Willoughby et al., 2012). Notably, however, the present study used a smaller number of *EF Touch* subtests and added the HTKS (with Pencil Tap being freely estimated). For this reason and because of demographically distinct samples, measurement invariance statistics are not directly comparable across this and the Willoughby et al. (2012) study.

7.2 | The trajectory of EF development from preschool to kindergarten

Consistent with hypotheses, there was significant, nonlinear growth in EF, with steeper development in preschool than in kindergarten. This aligns with prior work examining the functional form of EF in early childhood (Hughes et al., 2009; Hughes & Ensor, 2011; Wiebe et al., 2012; Willoughby, Wirth et al., 2012). Specifically, Willoughby and colleagues (2012) followed children from ages 3 to 5 and found that 60 percent of growth in EF occurred between ages 3 and 4; two additional studies found similarly higher growth rates from 3 to 4 than 4 to 5 (Clark et al., 2013; Wiebe et al., 2012). In this study, which followed children from 4 to 6, significantly more growth occurred between ages 4 and 5 than between ages 5 and 6. Together, this indicates a trend that children make relatively more gains earlier in their development (when compared to a year later). This is consistent with the notion that EF develops rapidly in early childhood (Blair, 2016; Garon et al., 2008) and suggests a sensitive period for EF development during that time. Alternatively or in addition, slower growth toward the end of kindergarten could represent a failure of current measures to adequately capture continued growth as children's EF abilities develop (e.g., the role of ceiling effects on some measures in this study).

One practical implication of this relatively slower growth in kindergarten is that it may indicate a potentially beneficial time for intervention aimed at promoting children's EF, particularly given the shifts in routines and expectations and increases in demands that occur during the transition to kindergarten (Bassok et al., 2016; Jones et al., 2016; Rimm-Kaufman & Pianta, 2000). Although findings from prior studies (Clark et al., 2013; Wiebe et al., 2012; Willoughby et al., 2012) indicate that relatively more growth occurred between ages 3 and 4 than between 4 and 5, the kindergarten year likely remains a beneficial time and setting for intervention given that 79 percent of 3- to 5-year-old children were enrolled in full-day kindergarten programs as of 2017 (National Center for Education Statistics, 2019), compared to 42 percent of 3-year-olds and 66 percent of 4-year-olds enrolled in preprimary education in 2016 (McFarland et al., 2018). Future work examining the relative efficacy of EF interventions in preschool and kindergarten contexts could further elucidate whether either year may provide a relatively more beneficial time period to bolster the development of this foundational skillset.

Findings from the present study also bring additional nuance to understanding the trajectory of EF development across this pivotal transition by describing how individual variability in EF growth during preschool is associated with growth during kindergarten. Results indicate that although children who start preschool advantaged in EF make more growth during that year, it is children with shallower preschool growth trajectories who show disproportionately more EF development in kindergarten. That is to say, there seems to be an EF "catch-up" effect in kindergarten leading to more equitable EF abilities across children on average prior to first grade. This aligns with prior work on the convergence of academic skills over time for children who did versus did not participate in early childhood education programs (Bassok et al., 2015; Lipsey et al., 2015; Puma et al., 2010). Again, however, it is important to acknowledge that challenges in obtaining strong measurement invariance could argue against this interpretation in favor of the notion that the underlying EF construct may be at least somewhat different by the end of kindergarten from its representation at the beginning of preschool.

At least for academic skills, it has been hypothesized that this convergence, or attenuation of earlier skill advantage, may be related to true individual differences in children's trajectories or learning experiences and/or divergence due to varied classroom-level quality or focus in subsequent years (Ansari & Purtell, 2018; Bailey et al., 2017). For example, it may be that kindergarten classrooms are supporting EF at the same, more basic level as in preschool, such that there is a "ceiling" on EF support for high-EF preschoolers entering kindergarten, and/or that kindergarten teachers disproportionately focus on building the EF skills of those who enter kindergarten with lower EF, given the need for increased self-regulatory capacity at that time (Bassok et al., 2016; Jones et al., 2016). Understanding these moderating factors in children's skill trajectories has implications not only for cultivating students' growth but also larger educational funding and social policy decisions (Abenavoli, 2019; Winsler & Mumma, 2021). As such, as such, future work should seek to understand characteristics of children, families, teachers, classrooms, and schools that may be associated with divergent trajectories specifically in EF over time. For instance, relative disadvantages in early EF trajectories have been found for boys (Conway et al., 2018; Wiebe et al., 2008, 2012), children from low-SES or impoverished backgrounds (Conway et al., 2018), and relatively older children in mixed-age classrooms compared to children with same-age peers (Ansari, 2017).

7.3 | Limitations and future directions

It is essential to interpret the contributions of this study's finding within the context of a number of limitations. Although the sample was racially and ethnically diverse, children were primarily from low-income households (and approximately half were in poverty). The sample demographics did closely match proportions of poverty and racial/ethnic diversity in the local community, increasing the external validity of these findings and helping to understand the experiences of this population. However, this also limits generalizability of results to children from economically disadvantaged families and confounds racial and ethnic identity with socioeconomic status and preschool attendance. Future research might leverage a racially and ethnically diverse, mixed-income sample with varying preschool participation rates to understand the trajectory of EF development and its demographic correlates more broadly.

Another threat to generalizability was the fact that over half of the sample was missing data in the spring of kindergarten because of planned missingness (in cohort 1), and no data were collected at this time point for cohort 2 because of the COVID-19 disruption. Given that these data were conceptualized as being missing completely at random and there was rich information about children who were not assessed at that time point, full information maximum likelihood estimation was applied. However, because of the abovementioned limitations, results can only offer suggestions about what EF looks like at the end of kindergarten and about the trajectory of nonlinear growth from preschool entry to the end of kindergarten. This is an important limitation in the context of extant EF literature, because there is not yet consensus about nonlinear growth in EF within the preschool and kindergarten years. However, it was possible to obtain a nuanced understanding of the trajectory of EF development during preschool because of the three data collection windows in that year. In future studies, it would be helpful to have two consecutive years with fall, winter, and spring data for all children to assess whether the within-year growth trajectory looks similar in kindergarten.

One measurement limitation was that there were not enough distinct measures of the three conceptual subcomponents of EF (i.e., inhibitory control, working memory, and set shifting) to be able to examine whether a three-factor model of EF fit the data well. Although this would be unexpected in early childhood, it would be beneficial going forward to compare a three-factor model to the unitary and bipartite models, particularly given continued debate in the field about the construct of EF in early childhood. The adequate model fit for CFAs and LGCMs requires that interpretations be made with caution. Specifically, these findings may overstate the extent to which EF is definitively a unitary construct in early childhood, particularly given that the bipartite model also fit the data well. It may also overstate the specificity of the nonlinear trajectory of EF development from preschool to kindergarten (e.g., accelerated and then decelerated growth), particularly given that more specific nonlinear functional forms were not tested.

Finally, using the bilinear spline necessitates choosing where the "knot" or transition point is. For the purposes of this study, it made sense to position the knot at the point between preschool and kindergarten to assess differences between the 2 years. Given that this functional form was not preferred compared to the latent basis model in this study, future work might consider adjusting the transition point in the bilinear spline model to better understand stability versus gains in EF over the summer between preschool and kindergarten, when it is less likely that children would be engaged in formal learning opportunities. For instance, EF may be less subject to "summer learning loss" (Stewart, Watson, & Campbell, 2018, p. 517)—which disproportionately affects children from low-income backgrounds who may have less access to learning opportunities outside of school than more affluent peers—than academic skills, perhaps because EF is less dependent on formal instruction. It would be useful to have information about children's summer activities and/or to compare the nonlinear growth of EF and academic skills to be able to test this hypothesis.

8 | CONCLUSION

Despite its limitations, the present study makes multiple contributions to the literature by examining EF development across five time points in a primarily low-income, ethnically and racially diverse sample. Specifically, findings provide further evidence that a unitary construct of EF develops nonlinearly in early childhood, as is hypothesized in the majority of prior work (e.g., Hughes & Ensor, 2009; Willoughby et al., 2012). Although measurement invariance was adequate to assess longitudinal growth, findings from this study add to prior literature indicating that the developmentally graded nature of EF could be somewhat qualitative rather than solely quantitative (Nelson, James et al., 2016). Practically, findings point to kindergarten as a potentially beneficial time for intervention to continually bolster EF skills, given decelerated growth found during this time period. However, the fact that children entering preschool with lower EF tended to show less growth during the preschool year also suggests that attention to EF skills for this subgroup earlier on may be helpful. Finally, the present study addresses the "measurement impurity problem" of EF research in the context of its findings and limitations, and suggests ways future research can further elucidate this key foundational school readiness skill.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Child Health and Human Development of the National Institutes of Health (grant number 2R01HD051498-06A1).

CONFLICT OF INTEREST

The co-authors state that there are no conflicts of interest.

ETHICAL STATEMENT

The University of Virginia's Institutional Review Board provided an exempt review of this study protocol and determined that it met the qualifications for approval as described in 45 CFR 46.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

DECLARATIONS

The opinions expressed are those of the authors and do not represent views of the funding agency. Special thanks to the participating teachers, children, and families.

ORCID

Shannon E. Reilly https://orcid.org/0000-0003-2279-0595
Jason T. Downer https://orcid.org/0000-0001-8827-242X
Kevin J. Grimm https://orcid.org/0000-0002-8576-4469

REFERENCES

Abenavoli, R. M. (2019). The mechanisms and moderators of "fade-out": Towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin*, *145*(12), 1103. https://doi.org/10.1037/bul0000212

Ansari, A. (2017). Multigrade kindergarten classrooms and children's academic achievement, executive function, and socioemotional development. *Infant and Child Development*, *26*, e2036. https://doi.org/10.1002/icd.2036

Ansari, A., & Purtell, K. M. (2018). Commentary: What happens next? Delivering on the promise of preschool. *Early Childhood Research Quarterly*, *45*, 177–182. https://doi.org/10.1016/j.ecresq.2018.02.015

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

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

Bassok, D., Gibbs, C. R., & Latham, S. (2015). *Do the benefits of early childhood interventions systematically fade? Exploring variation in the association between preschool participation and early school outcomes.* EdPolicyWorks Working Paper Series,(36). Charlottesville, VA: University of Virginia.

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade?. *AERA Open*, *2*, 2332858415616358. https://doi.org/10.1177/2332858415616358

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*, 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Blair, C. (2016). Executive function and early childhood education. *Current Opinion in Behavioral Sciences*, *10*, 102–107. https://doi.org/10.1016/j.cobeha.2016.05.009

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*, (Vol., 2, pp. 300–320). New York, NY: Guilford Press.

Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. *Developmental Neuropsychology*, *28*, 561–571. https://doi.org/10.4324/9780203764244.

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

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, *12*, 471–492. https://doi.org/10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Clark, C. A. C., Nelson, J. M., Garza, J., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2014). Gaining control: Changing relations between executive control and processing speed and their relevance for mathematics achievement over course of the preschool period. *Frontiers in Psychology*, *5*, 107. https://doi.org/10.3389/fpsyg.2014.00107

Clark, C. A. C., Sheffield, T. D., Chevalier, N., Nelson, J. M., Wiebe, S. A., & Espy, K. A. (2013). Charting early trajectories of executive control with the shape school. *Developmental Psychology*, *49*, 1481–1493. https://doi.org/10.1037/a0030578

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351. https://doi.org/10.1037/1082-989X.6.4.330

Conway, A., Waldfogel, J., & Wang, Y. (2018). Parent education and income gradients in children's executive functions at kindergarten entry. *Children and Youth Services Review*, *91*, 329–337. https://doi.org/10.1016/j.childyouth.2018.06.009

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

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

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at Head Start. *Developmental Psychology*, *47*, 404–416. https://doi.org/10.1037/a0021065

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*, 1698–1709. https://doi.org/10.1037/a0036633

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

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel modeling approaches.* New York, NY: Guilford Publications.

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

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychology*, *108*, 663–676. https://doi.org/10.1016/j.jecp.2010.06.005

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2009). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, *35*, 20–36. https://doi.org/10.1080/87565640903325691

Jones, S. M., Bailey, R., Barnes, S. P., & Partee, A. (2016). *Executive function mapping project: Untangling the terms and skills related to executive function and self regulation in early childhood* (No. 2016–88). Cambridge, MA: U.S. Department of Health and Human Services, Office of Planning, Research, and Evaluation.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*, 486–507. https://doi.org/10.1177/0049124114543236.

Lee, K., Bull, R., & Ho, R. M. (2013). Developmental changes in executive functioning. *Child Development*, *84*, 1933–1953. https://doi.org/10.1111/cdev.12096

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*, 59–80. https://doi.org/10.1348/026151003321164627

Li, C. (2013). Little's test of missing completely at random. *Stata Journal*, *13*, 795–809. https://doi.org/10.1177/1536867X1301300407

Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). *A randomized control trial of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade: Research Report*. Peabody Research Institute.

Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, *7*(4), 199–204. https://doi.org/10.1111/cdep.12043

Maas, C. J., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *58*, 127–137. https://doi.org/10.1046/j.0039-0402.2003.00252.x

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In (J. R. Nesselroade & R. B. Cattell Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). Boston, MA: Springer.

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Forrest Cataldi, E., & Bullock Mann, F. (2018). *The condition of education 2018 (NCES 2018-144).* Washington, DC: National Center for Education Statistics, U.S. Department of Education. Retrieved from: https://nces.ed.gov/pubs2018/2018144.pdf

Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In: (L. M. Collins & A. G. Sayer Eds.), *New methods for the analysis of change* (pp. 203–240). American Psychological Association.

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, *13*, 395–423. https://doi.org/10.1080/15248372.2011.585478

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

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

Montroy, J. J., Bowles, R. P., Skibbe, L. E., McClelland, M. M., & Morrison, F. J. (2016). The development of self-regulation across early childhood. *Developmental Psychology*, *52*, 1744–1762https://doi.org/10.1037/dev0000159

Morgan, P. L., Farkas, G., Hillemeier, M. M., Pun, W. H., & Maczuga, S. (2018). Kindergarten children's executive functions predict their second-grade academic achievement and behavior. *Child Development*. https://doi.org/10.1111/cdev.13095

Nelson, J. M., James, T. D., Choi, H. J., Clark, C. A. C., Wiebe, S. A., & Espy, K. A. (2016). III. Distinguishing executive control from overlapping foundational cognitive abilities during the preschool period. *Monographs of the Society for Research in Child Development*, *81*(4), 47–68. https://doi.org/10.1111/mono.12270

Nelson, J. M., Sheffield, T. D., Chevalier, N., Clark, C. A. C., & Espy, K. A. (2016). Structure, measurement and development of preschool executive control. In (J. A. Griffin, L. Freund & P. McCardle Eds.), *Executive function in preschool children: Current knowledge and research opportunities*. New York: APA Press.

Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in Physical Education and Exercise Science*, *19*, 12–21. https://doi.org/10.1080/1091367X.2014.952370

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, *45*, 605–619. https://doi.org/10.1037/a0015365

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., Shapiro, G., Broene, P., Jenkins, F., Fletcher, P., Quinn, L., Friedman, J., Ciarico, J., Rohacek, M., Adams, G., Spier, E., & Spier, E. (2010). Head Start Impact Study. Final Report. *Administration for Children & Families*.

Sabol, T. J., & Pianta, R. C. (2012). Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. *Child Development*, *83*, 282–299. https://doi.org/10.1111/j.1467-8624.2011.01678.x

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514. https://doi.org/10.2139/ssrn.199064

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*, 1120–1140. https://doi.org/10.1037/edu0000193

U.S. Health and Human Services Department. (2016). Annual update of the HHS poverty guidelines. Retrieved from: https://www.federalregister.gov/documents/2016/01/25/2016-01450/annual-update-of-the-hhs-poverty-guidelines

Usai, M. C., Viterbori, P., Traverso, L., & De Franchis, V. (2014). Latent structure of executive function in five-and six-year-old children: A longitudinal study. *European Journal of Developmental Psychology*, *11*, 447–462. https://doi.org/10.1080/17405629.2013.840578

von Oerzen, T., Hertzog, C., Lindenberger, U., & Ghisletta, P. (2010). The effect of multiple indicators on the power to detect inter-individual differences in change. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 627–646

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

Wiebe, S. A., Sheffield, T. D., & Espy, K. A. (2012). Separating the fish from the sharks: A longitudinal study of preschool response inhibition. *Child Development*, *83*, 1245–1261. https://doi.org/10.1111/j.1467-8624.2012.01765.x

Wiebe, S. A., Sheffield, T. D., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-old children. *Journal of Experimental Child Psychology*, *108*, 436–452. https://doi.org/10.1016/j.jecp.2010.08.008

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M., & the Family Life Project Investigators (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment*, *24*, 226–239. https://doi.org/10.1037/a0025361

Willoughby, M. T., Pek, J., & Blair, C. B., & Family Life Project. (2013). Measuring executive function in early childhood: A focus on maximal reliability and the derivation of short forms. *Psychological Assessment*, *25*, 664–670. https://doi.org/10.1037/a0031747

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment*, *24*, 418–431. https://doi.org/10.1037/a0025779

Winsler, A., & Mumma, K. (2021). Understanding long-term preschool "fade-out" effects—be careful what you ask for: Magical thinking revisited. In (S. Ryan, M. E. Graue, V. L. Gadsden, & F. J. Levine Eds.), *Advancing Knowledge and Building Capacity for Early Childhood Research* (pp. 123–146). Washington, DC: American Educational Research Association.

Zelazo, P. D., Blair, C. B., & Willoughby, M. T. (2016). *Executive function: Implications for education (NCER 2017-2000)*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

How to cite this article: Reilly, S. E., Downer, J. T., & Grimm, K. J. (2022). Developmental trajectories of executive functions from preschool to kindergarten. *Developmental Science*, *25*, e13236. https://doi.org/10.1111/desc.13236