Child Development, 2026, 1–14
https://doi.org/10.1093/chidev/aacaf053
Published: April 2, 2026
Empirical Article

Society for Research in Child Development OXFORD

Longitudinal profiles of executive functioning from infancy to 5 years of age—A FinnBrain Birth Cohort Study

Anniina Karonen[1,2,3,4,*,†], Pauliina Juntunen[1,2,3†], Akie Yada[2,5,6,7], Fiia Takio[1,2,5], Elisabeth Nordenswan[1,2,5], Eeva Eskola[1,2,5], Kirby Deater-Deckard[2,8], David J. Bridgett[9], Michelle Fernandes[10,11,12], Elina Mainela-Arnold[1,2,5], Hasse Karlsson[1,3,13], Linnea Karlsson[1,3,4,14], Eeva-Leena Kataja[1,3], Riikka Korja[1,2,5], and Saara Nolvi[1,2,3,5]

[1]FinnBrain Birth Cohort Study, Turku Brain and Mind Center, Department of Clinical Medicine, University of Turku, Turku, Finland
[2]Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland
[3]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland
[4]Department of Child Psychiatry, Turku University Hospital, Turku, Finland
[5]The Centre of Excellence for Learning Dynamics and Intervention Research (InterLearn), University of Turku and University of Jyväskylä, Turku and Jyväskylä, Finland
[6]Department of Psychology, Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland
[7]Department of Education, Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland
[8]Department of Psychological and Brain Sciences, University of Massachusetts, Amherst, MA, United States
[9]Department of Psychology, University of Nevada, Reno, NV, United States
[10]Department of Paediatrics, Institute of Developmental and Regenerative Medicine, University of Oxford, Oxford, United Kingdom
[11]Oxford Maternal and Perinatal Health Institute, Green Templeton College, University of Oxford, Oxford, United Kingdom
[12]Nuffield Department of Women's and Reproductive Health, John Radcliffe Hospitals, University of Oxford, Oxford, United Kingdom
[13]Department of Psychiatry, Turku University Hospital, Turku, Finland
[14]Department of Public Health, University of Turku and Turku University Hospital, Turku, Finland
*Corresponding author: FinnBrain Research, University of Turku, Kiinamyllynkatu 8-10, Medisiina A, Turku 20520, Finland. Email: ammatt@utu.fi

†Anniina Karonen and Pauliina Juntunen equally contributed to this study.

Abstract

Methodological challenges diminish the number and reliability of longitudinal studies on executive functions (EFs) starting in infancy. To address this, the current study used latent profile analysis (LPA) to examine EF task performance across three age points: 8 months, 2.5 years, and 5 years. Participants were children ($N = 830$; 55.5% boys; > 95% White) from the FinnBrain Birth Cohort Study. Three profiles were identified: constant below average EF profile (14.2%), and two average EF profiles differentiated by Spin the Pots performance (working memory) at 5 years (above average 29.8%, below average 56%). Expected associations between the below average EF profile, male sex, and lower general cognitive performance were found, further supporting the validity of the profiles.

Keywords executive functions, early childhood, longitudinal methodology

Lay summary

Executive function skills (EFs)—inhibition, working memory, and cognitive flexibility—are key for regulating behavior and emotions. Studying how these skills develop from infancy is challenging but important. This study followed 830 children from 8 months to 2.5 years to 5 years of age to track their EFs' development. Using a statistical method called latent profile analysis, the current study identified three groups based on EF task performance. One group showed consistently below-average skills, while the other two groups had average skills but differed in how they performed on a working memory task at age 5. Boys and children with lower general cognitive performance were more likely to belong to the below-average group, supporting the validity of these findings.

Handling Editor: Gigi Luk
Editor: Shauna Cooper
Received: December 28, 2024. **Revised:** October 21, 2025. **Accepted:** December 18, 2025

Executive functions (EFs) are higher-level cognitive processes essential for goal-directed behaviors and adaptive functioning, developing rapidly and dynamically from the latter half of the first year (Hendry et al., 2016). Despite the growing use of longitudinal designs, gaps remain in understanding the variety of developmental courses, particularly the stability of individual differences in EFs during infancy and toddlerhood. Methodological challenges—such as uneven performance across tasks and the need to adapt measures for young children stemming from rapid developmental changes—may have hindered progress in the field. Person-centered analysis strategies offer a potential solution by tracking individual trajectories and investigating whether all children follow the same developmental pattern. While a few studies (e.g., Patwardhan et al., 2023; Reilly et al., 2022; Willoughby et al., 2017b) have employed such approaches with longitudinal data spanning a targeted developmental period during preschool years, evidence remains especially limited from infancy across early childhood. Advancing our understanding of early EF trajectories is crucial for identifying the potential variety of developmental pathways—an essential goal for improving early detection and prevention of maladaptive outcomes (Yang et al., 2022). The present study aims to characterize this variety in EFs before age of 5.

EFs are involved in all top-down processes of self-regulation (Nigg, 2017; Ribeiro et al., 2024), which may explain their broad and enduring impact on health and overall life trajectories (e.g., Robson et al., 2020). Early life EFs contribute to later outcomes such as quality of life, health, and academic achievement (Diamond, 2013), whereas sub-optimal EF development is characteristic of many childhood-onset behavioral and psychiatric disorders. Additionally, childhood EF skills prospectively relate to a wide range of externalizing and internalizing symptoms in both nonclinical and clinical samples (Yang et al., 2022).

EFs are generally described as comprising three core functions: inhibitory control (i.e., inhibition of a prepotent response; IC), working memory (i.e., updating; WM), and set shifting (i.e., cognitive flexibility), as identified by late childhood through adolescence and into adulthood (Lehto et al., 2003; Miyake & Friedman, 2012). IC is an ability to control attention, behavior, thoughts, and emotions by suppressing prepotent responses to achieve goal-directed behavior (Diamond, 2013). WM involves the temporary storage and manipulation of information and is closely interrelated with IC (Diamond, 2013). Set shifting is an ability to switch attention between tasks or concepts, and is utilized when demands are placed on IC and WM as well as many other domains of cognitive functioning (Fiske & Holmboe, 2019). Although these three core functions have been considered distinct constructs, they are correlated both behaviorally and neuroanatomically. Together, their development provides the foundation for higher-level EFs, such as planning, reasoning, and problem solving (Diamond, 2013). However, developing EF structure is not very well understood; thus, the theorized structure of EFs in infancy and toddlerhood may differ from the structure in later childhood (Hendry et al., 2016). Moreover, not all tasks used to measure EFs are categorized in the same way among researchers regarding EF abilities they are considered to measure (Miller et al., 2023). On the basis of these identified limitations, we refer to EFs instead of different core functions.

The protracted development of the prefrontal cortex (PFC), which continues for more than two decades and reaches full maturity in adulthood, parallels improvements in behavioral performance on EF tasks (Fiske & Holmboe, 2019; Garon et al., 2008; Hendry et al., 2016). Although individual differences in EFs are highly heritable by adulthood (Friedman & Miyake, 2017), within-person changes in EFs may be malleable to environmental and sociocultural impacts such as parenting practices or structural inequalities—perhaps especially in early childhood when these skills and relevant neural circuits are developing most rapidly (Miller et al., 2023). For example, using a longitudinal, genetically informed design, Bridgett et al. (2018) showed that parenting behavior contributes to children's top-down self-regulation beyond genetic influences. Additionally, Fujisawa et al. (2017) found that developmental changes in EFs during the preschool years are related to environmental and genetic factors, whereas the stability of individual differences may be linked to shared environmental experiences.

Early aspects of EFs can be detected behaviorally as early as the latter half of the first year of life. For instance, simple response inhibition emerges around eight months, when infants may occasionally be able to withdraw from an enjoyable activity upon request (Garon et al., 2008). Performance on the "A not B" task, a frequently used measure of early EFs, also improves substantially between 7 and 12 months, reflecting advances in maintaining mental representations over longer delays (Diamond, 1985). By age 2, toddlers begin coordinating inhibitory and attentional control as well as WM (e.g., holding a rule in mind while completing another task), although multi-demand tasks (e.g., "Simon says") often remain challenging even at age 5 (Garon et al., 2008). Beyond infancy and toddlerhood, the growth of EF abilities is proposed to be more rapid between ages 3 and 4 than between 4 and 5 (Clark et al., 2013; Wiebe et al., 2012; Willoughby et al., 2012b), with slower growth from 5 to 6 years (Reilly et al., 2022). However, trajectories vary by core function. For example, average growth in IC is fastest in the preschool years before leveling off, whereas WM and set shifting seem to improve in a more linear fashion across early and middle childhood (Best & Miller, 2010).

Regarding the broader literature on EF development, it is noteworthy that most of the evidence has been based on cross-sectional designs. Furthermore, longitudinal studies typically exclude infants and toddlers because of the methodological challenges of assessing EFs at these very young ages—although samples including infants have increased in recent years (Miller et al., 2023). A handful of relevant longitudinal studies show mixed evidence for the stability of individual differences in EFs in the first few years of life. Broomell and Bell (2022) measured EFs at 5, 10, 24, 36, 48, and 108 months by using several behavioral tasks at each time point except at 5 and 10 months. They reported that variance in infant EF scores did not predict individual differences in subsequent EFs. Instead, the stability of individual differences in EFs emerged from 2 years onward, suggesting that toddler (but not infant) EF scores can be used as an early predictor of later EF scores. In another paper based on the same longitudinal study, it was found that 10-month EF predicted EF variation at 36 months, which in turn predicted EF scores at 48 months, which subsequently predicted EF scores at 72 months (Blankenship et al., 2019). However, while Broomell and Bell (2022) and Blankenship et al. (2019) both employed the same single EF task in infancy (i.e., the looking version of the A-not-B task), in subsequent age points, the two studies differed not only in the specific EF measures they used but also in the number of tasks administered. Moreover, differences in sample sizes and analytical approaches may have further contributed to the divergence in their findings. In another longitudinal study, Johansson et al. (2016) found that performance on a prohibition task (i.e., withholding a response) at 12 months predicted performance on multiple EF tasks at 36 months but not at 24 months. The authors concluded that IC in infancy may be a crucial foundation of more complex EF components later in early childhood. Additional support for stability of individual differences from infancy comes from research on early attentional abilities for subsequent EF development (Garon et al., 2008). For instance, connections between 5-month attention and EFs at 10 months and 3 years have been reported (Blankenship et al., 2019; Kraybill et al., 2019).

Moreover, recent studies suggest that the change and pace of development in early childhood may be of significance for subsequent development of EFs and other outcomes. For instance, Reilly et al. (2022) and Hughes et al. (2010) found that from 4 to 6 years, less growth in EFs in one period may be compensated by steeper growth in a subsequent period (e.g., catch-up). In addition, small improvements in EFs during the transition to middle childhood, and delayed EF development prior to first grade, increases the risk for poorer academic performance and externalizing and internalizing behaviors in kindergarten and grade school years (Hughes & Ensor, 2011; Patwardhan et al., 2023; Willoughby et al., 2017b).

Accordingly, it is essential to investigate the longitudinal stability and change of EFs not only during the preschool years, but even earlier across infancy and toddlerhood. To date, as noted, a few studies have addressed longitudinal EF development during these early periods, and their findings remain inconsistent, likely reflecting the considerable methodological challenges of assessing EFs in very young children. First, infants and toddlers can only complete a few tasks at a time, and these tasks are quite different from more complex EF tasks that are more readily available with children older than 3 years. More specifically, the limitations in the attentional control and verbal abilities of infants and toddlers set constraints on staying "on task" and comprehending and following the task rules, causing "noise" in the data (Carlson, 2005). Second, the rapid development of diverse skills from infancy through preschool years necessitates changes in assessment tasks, including introduction of new tasks and modifications of earlier ones when assessing the same children longitudinally. Such changes may capture EFs in comparable ways across ages, or in qualitatively different ways depending on developmental stages. Altogether, these issues introduce measurement error and constrain both the variety and number of tasks that can be administered, thereby limiting construct validity and reliability and reducing the capacity for detecting replicable estimates of stability and change in individual differences in EFs. However, considering the restraints in measuring early childhood EFs, it is noteworthy that a recent meta-analysis found that composite measures of several EF tasks were not as strongly related to child outcomes than single established lab-based measures of early childhood EF tasks, such as A-not-B task (Stucke & Doebel, 2024). This lends cautious support to the suggestion that a single-task approach is suitable for measuring early childhood EFs given the practical limitations related to the assessment of very young children using multiple tasks.

To address challenges in measuring and operationalizing EFs in infancy and toddlerhood, data-driven methods can be applied to identify latent factors that capture developmental pathways of EFs, using multiple behavioral tasks that are suitable for repeated administration across developmental periods (Broomell & Bell, 2022). However, the modelling approaches required are still rarely used. In the current research, we address this gap by highlighting and applying two such approaches. First, we applied item response theory (IRT) to enhance detection of individual differences and within-person developmental change by combining information about each child's latent EF ability with the psychometric properties of task items (e.g., difficulty level, discrimination between constructs). IRT yields estimates of underlying EF abilities while minimizing measurement error (Willoughby et al., 2011; Willoughby et al., 2012b). Unlike summed scores, which treat all items equally, IRT accounts for differences in item difficulty and provides test information curves that show the precision of measurement across the full ability range (Embretson, 1996; Houts et al., 2022). This makes IRT particularly valuable for identifying how well tasks capture variation in children's performance in developmental research. Second, we applied latent profile analysis (LPA) for identifying underlying subgroups of children based on latent EF abilities when the number and type of tasks changes with child age. This article focuses on the LPA, given that the IRT analyses have been reported in detail elsewhere (Yada et al., 2025).

To our knowledge, prior research using LPA has mostly been limited in cross-sectional studies spanning a limited period of early childhood development. One previous study identified, for instance, three distinct latent performance profiles in 4- to 5-year-olds from low-income families (low, moderate, and high EF performance across IC, WM, and set shifting tasks) (Williams & Bentley, 2021) while another study identified five profiles (low, moderate, and high EF performance profiles as well as two profiles reflecting contradictory EF performance across WM versus set shifting tasks in 5- and 6-year-olds) (Litkowski et al., 2020). Longitudinal group-based modeling approaches have provided evidence for typical and atypical groups of EF development from 3 to 12 years of age (Lynch et al., 2024), suggesting that such groups might be found already in early childhood. However, none of these studies or any studies we are aware of have applied both IRT and LPA to longitudinal EF data in early childhood, and even in older (preschool- or school-aged) children, such literature is limited.

To summarize the aims and the hypotheses of the present study, we applied two person-centered methods in the following order, IRT and LPA, to shed light on trajectories of EF development across infancy, toddlerhood, and the preschool years. Our aim was to identify longitudinal latent profiles of EF task performance from 8 months to 30 months (2.5 years) to 60 months (5 years) of age by using several behavioral tasks that are thought to measure core EFs. We applied IRT (Willoughby et al., 2011) factor scores in LPA to identify EF task performance trajectory groups. Finally, we examined how certain variables generally recognized as important correlates of EFs including socio-economic status (SES; measured as maternal education and maternal financial satisfaction), duration of gestation including prematurity, child's official sex at birth, and child's general cognitive performance at 30 months (2.5 years) and 60 months (5 years) were related to latent EF profiles. The profile analysis was exploratory, but confirmatory analyses were also conducted to test associations between the profiles and covariates. We hypothesized to identify distinct latent profiles of EF task performance across age points. Given that previous cross-sectional studies typically reported three latent profiles of EFs, we expected a similar result reflecting below average,

average, and above average performance longitudinally. However, it was plausible that our exploratory approach instead would reveal mixed or nonlinear EF task performance across age points that may indicate changing or nonlinear, rather than stable linear, EF development. Nevertheless, we did not have a hypothesis on EFs' stability from infancy to later childhood given the prior mixed evidence (Miller et al., 2023). However, we expected finding evidence of stability in development from 2.5 years to 5 years (Broomell & Bell, 2022), suggestive of an underlying continuum of stable individual differences already from toddlerhood (Blankenship et al., 2019; Johansson et al., 2016). Additionally, on the basis of prior research (e.g., Friedman & Miyake, 2017; Olson et al., 2021; Sandoval et al., 2022; Silverman, 2021), we expected higher SES, longer duration of gestation, female sex at birth, and better general cognitive development at 2.5 and 5 years of age to be related to membership in a latent profile group with better EF task performance.

Method

Participants

The participants ($N = 830$), mainly White, were a subgroup of children from the FinnBrain Birth Cohort Study ($N = 3{,}808$ families), a population-based cohort locating in Southwest Finland (Karlsson et al., 2018). Participants were recruited to the cohort study at the gestational week (gw)-12 ultrasound visit by research nurses between December 2011 and April 2015. A mother, and a spouse if participating in the visit, was invited to participate in the cohort study if they had a normal ultrasound screening result, lived in Southwest Finland or the Åland Islands, and were able to participate in either Finnish or Swedish, the official languages in Finland. For the current study, families were recruited from the baseline cohort separately for each research visit.

Inclusion criterion for the present study was that a child had participated in EF measurement at least at one age point; at 8 months, 2.5 years, or 5 years. Children belonging to the Focus Cohort (high vs. low prenatal distress exposure) (Karlsson et al., 2018) as well as children with moderate prenatal distress exposure were invited in initial measurements at 8 months (Kataja et al., 2020). In the subsequent age points, the sample was further enriched by other children from the cohort with an aim to represent baseline population of the cohort and maximize overlap between different research visits. Participants were invited to the laboratory visits by their age calculated from due date. The data was gathered between June 2013 and June 2016 (8 months), between April 2015 and June 2018 (2.5 years), and between October 2017 and December 2020 (5 years). The study was approved by The Ethics Committee of The Hospital District of Southwest Finland (ETMK: 57/180/2011; ETMK: 59/1801/2013). The parents gave informed written consent on behalf of themselves and their child. Demographic characteristics of the sample are shown in Table 1. No children in the sample had known major central nervous system abnormalities. By age 5, on the basis of parental report, four participants had different clinical diagnoses of developmental disorders, such as autism spectrum disorder that may have contributed to EF performance. As most of such disorders are undiagnosed in early childhood, suggesting that there might be unidentified disorders within the participants, we included these children in analyses to maximize the true variation in data. For 128 participants, information of developmental diagnoses was not available by the age of 5.

In total, complete EF measurements were obtained from 363 children at 8 months, 414 children at 2.5 years, and 435 children at 5 years. The number of the participants who had participated in at least one measurement at all three age points was 150. The detailed patterns of sample sizes and missing data are presented in the Supplementary material (Table S1). Children participating either in one, two, or three measurements from the age of 8 months to 2.5 years to 5 years did not differ in terms of sex, prematurity, maternal education, or EF task performance at any age point (ANOVA, $p > .05$). However, they did differ in terms of maternal financial satisfaction ($F(2, 816) = 4.08$, $p = .017$) and duration of gestation ($F(2, 827) = 3.14$, $p = .044$); maternal financial satisfaction was higher among those mothers whose children participated in all three measurements, and children of the mothers with shorter duration of gestation participated more often in only one measurement. Mothers of participating children at 8 months were younger (t (823.08) = 2.22, $p = .027$) and had longer duration of gestation (t (826.79) = −2.26, $p = .024$), and these mothers had more often vocational tertiary or applied university degree (χ^2 (2, $N = 819$) = 6.95, $p = .031$) than mothers of infants who did not participate. The participating and nonparticipating children at 2.5 years did not differ in terms of any background factors. At 5 years of age, the participating children's mothers were more highly educated (χ^2 (2, $N = 819$) = 14.56, $p < .001$) and more satisfied with their financial situation (t (798) = −3.06, $p = .002$) than the mothers of nonparticipating children.

Procedures

Across the age points, i.e., 8 months, 2.5 years, and 5 years, EF measurements were carried out during a laboratory visit at the FinnBrain Child Development and Parental Functioning Lab. The measurements were conducted by a clinical psychologist or a trained master's student in psychology. Laboratory visits also included other neurocognitive, temperament, and parent–child interaction measurements. The duration of the entire visit was a maximum of 90 min for the 8-month-olds and the 2.5-year-olds, and 2 hours for the 5-year-olds. Variables of maternal age, infant sex, and length of gestation were acquired from the Wellbeing Services County of Southwest Finland (Varha) records. Maternal education was obtained from questionnaires completed during pregnancy at gw 14. Mother's financial satisfaction was obtained from questionnaires completed during pregnancy at gw 14 and again at child age of 5 years.

Measures

Child EFs

As described later in "Statistical analysis," child responses in each trial of each EF task were rescored with a binary scale (pass = 1, fail = 0), for IRT analysis, except for the Snack Delay task. Here, theoretical range of each task score is described.

Child EFs at 8 months

Infant EFs were assessed using one task, a modified A-not-B (AB) procedure measuring attention, IC, and WM (e.g., Diamond, 1985; Sun et al., 2009). The task was a delayed response task; that is, it

Table 1 Sample characteristics ($N=830$).

Characteristic	Range	*M* (*SD*)	No. (%)
Child characteristics			
Child sex			
Boy			461 (55.5)
Girl			369 (44.5)
Child age (months, from the birth date)			
At M8 visit	6.9–9.4	8.1 (0.39)	
At Y2.5 (M30) visit	29.0–32.2	30.1 (0.45)	
At Y5 (M60) visit	58.7–64.5	60.1 (0.93)	
Duration of gestation (weeks)	29.3–42.4	39.8 (1.7)	
Children born before 37 weeks of gestation			47 (5.7)
Birth weight (g)	1,330–5,470	3,549 (524.0)	
APGAR score 5 min	4–10	9.0 (0.8)	
APGAR score at 5 min, < 7			12 (1.4)
INTER-NDA cognitive score	19–52	45.5 (5.2)	
WPPSI PIQ score	61–146	102.3 (16.3)	
Maternal characteristics			
Maternal age at delivery	18–45	31.0 (4.5)	
Financial satisfaction, early 2nd trimester (0–10)	0–10	5.9 (2.4)	
Financial satisfaction at child age of 5 y (0–15)	1–15	10.0 (3.3)	
Average financial satisfaction (standardized)[a]	−2.46 to 1.70	0.0 (0.1)	
Maternal education, highest known at 5 y (%)[b]			
High school/vocational education or lower			201 (24.2)
Vocational tertiary/applied university			250 (30.1)
University degree			368 (44.3)
Maternal ethnicity[c]			
White			(96.4)
Other			(<1)
Missing			(3.5)

Note. M = month; Y = year; APGAR score = a clinical indicator of newborn's condition after birth ranging between 0 and 10. A score of 7 or more at 5-min age indicates good adaptation to environment, while a score below 7 indicates complications. INTER-NDA = INTERGROWTH-21st Neurodevelopmental Assessment, WPPSI PIQ = Wechsler Preschool and Primary Scale of Intelligence, Performance IQ. Attrition analysis is described in more detail in the Participants section.
[a] Standardized mean of the two time points (gw 14 and 5 years).
[b] Most recent/highest known information on maternal education.
[c] Paternal ethnicity was available from 65% of participants, all reporting belonging to White ethnicity.

required maintaining and updating the hiding location in working memory and inhibiting a prepotent response. In this modified AB task (Nolvi et al., 2018), a toy was hidden under two identical hiding locations (L = left or R = right) and the infant was asked to search for the toy. The hiding order and delay between hiding and searching were modified. The infant reaches for the location of a toy under two hiding locations in a fixed pattern on three delay levels (0, 2, and 4 s). Each delay level included six trials, and the entire AB procedure includes 18 trials. For example, in the first delay level, the series of trials was R-L-L-R-R-L. If the infant scored correctly in half of the six trials, they proceeded to the next delay level. In each trial, infant's reaching for a location was scored either as correct or incorrect. All reaching behavior that was initiated within 8 ss was considered reaching. Infant performance was scored by the experimenter and verified later by a trained coder from video. The coding reached adequate interrater reliability (81% agreement across trials). The maximum score in the task is 18, with higher scores reflecting better EF performance.

Child EFs at 2.5 years

Toddler EFs were measured with two tasks, Spin the Pots (SP) (WM) (Hughes & Ensor, 2005) and Snack Delay (IC) (Kochanska et al., 2000). In Spin the Pots task, child searched for stickers (six) hidden in front of the child in eight visually distinct pots on a lazy Susan tray. Before each search, the tray was covered with an opaque scarf and rotated 180 degrees. The maximum number of trials was 16. The task was terminated when the maximum number of trials is reached, or the child finds all the stickers. The maximum score in this task was 16 (16 minus the number of incorrect trials), with higher scores reflecting better performance on WM task.

In Snack Delay task, the child held their hands on a table while waiting for a bell to ring before taking either candy or raisin from under a transparent cup. The delay before the bell ring was modified from 10 to 60 seconds in length during the six trials. The trials were coded on a scale from 0 to 4 (0 = Child eats the snack before the researcher touched the bell and rings it, 4 = Child does not touch the cup or the bell before the bell has rung). Moreover, a maximum of 2 extra points was given on the basis of the child's ability to hold their hands on the mat that was placed on the table (0 = Not able to hold hands on the mat, 1 = At least one hand on the mat during the trial, 2 = Both hands on the mat during the trial) (Spinrad et al., 2007). The maximum score in this task was 36, with higher scores indicating higher IC ability. As described later in "Statistical analysis," child responses in each trial of this task were rescored using an ordinal scale (0 to 6) for further analyses.

Child EFs at 5 years

Preschoolers' EFs were measured using five tasks: Spin the Pots (WM), Delay of Gratification (IC), and three tasks from the EF Touch battery: Farmer (WM), Arrows (IC), and Pig (IC). Spin the Pots procedure was identical to the task presented at 2.5 years, but the task included ten stickers and 12 pots. In addition, the pots were visually more like each other than at 2.5 years, to make the task more challenging. The maximum number of trials was 20. The maximum score in this task was 20 (20 minus the number of incorrect trials). Delay of Gratification task was introduced to the child as a game, in which the child got nine trials to choose a smaller, immediate reward (0 points) or a larger, delayed reward (1 point) (Beck et al., 2011). Rewards were either stickers, edible treats, or 5-cent coins, which varied in a standardized order. In accordance with the child's choice, the child received the rewards either immediately or they were put in an envelope and given to the child when leaving the visit. The maximum score was 9, with a higher score representing better IC.

EF Touch is a computerized test battery for children from 3 to 5 years of age with constantly established reliability and validity (e.g., Willoughby et al., 2010 , 2012a), utilizing a touch screen monitor for tracking children's responses to the task (Willoughby et al., 2017a). In the Farmer task including 36 trials, "farmer's fields," a four-by-four grid of squares, was presented to the child. The child was told that one of the farmer's animals runs away and the child must help the farmer remember which fields the animal has visited to help the farmer bring it back. The child was shown a series of fields that were highlighted one by one. After a 1,500-ms pause, they must touch the fields on the screen in the same order that the highlights appeared (i.e., the order in which the animal visited them). The task became more difficult as the animal visited more fields, and the child was asked to remember longer sequences (2–4) of the highlighted fields. The maximum accuracy (a proportion correct) of a child's responses was 1, with a higher score representing better WM.

Arrows is a spatial conflict task with 36 trials which instructed the child to touch the button indicating which direction the arrow was pointing. One button was on the right side and the other on the left side of the screen. Initially, arrows appeared congruently with the buttons (i.e., above the button in which arrows were pointing). As the task progressed, the location and the direction of the arrows were presented incongruently (i.e., arrows appeared above the opposite button to which they were pointing), which required the child to inhibit the prepotent response. Three sets of 12 trials included (1) congruent, (2) incongruent, and (3) mixed conditions. Only incongruent trials from incongruent and mixed conditions were included in this study as they require IC (Willoughby et al., 2012a). The maximum accuracy (a proportion correct) of a child's responses was 1, with a higher score representing better IC.

Pig, a standard go–no-go task capturing IC, included 40 trials with varying difficulty in no-go responses. The difficulty was varied by the number of go responses preceding a no-go response in standard order. In this task, children were instructed to touch a large green button on the screen as fast as they can every time they saw an animal (the go response) except when the animal was a pig (the no-go response). The maximum accuracy (a proportion correct) of a child's responses was 1, with a higher score representing better IC.

Demographic factors and main predictors

Sociodemographic and birth factors

SES was outlined on the basis of maternal educational level and financial satisfaction. Maternal education was originally measured on a scale from one to nine which was transformed to a three-class variable (1 = low [secondary education or high school or lower]; 2 = middle [applied university or polytechnics]; 3 = high [university degree]). Highest education level reached when the child was 5 years old was used. Maternal financial satisfaction was indicated on a continuous scale from one to 10 in pregnancy, or from one to 15 at the child's age of 5 years. Standardized mean of the two time points (i.e., gw 14 and 5 years) was used in analyses. Data on child official sex at birth (1 = boy, 2 = girl) and on the length of gestation were acquired from the Wellbeing Services County of Southwest Finland (Varha) records after the child was born. For separate sensitivity analysis, a variable indicating a premature birth was constructed by recoding the length of gestation as 1, if < 37, and as 0, if ≥ 37. Child age at the EF measurements was calculated as experiment date minus the date of birth.

General cognitive performance

At the age of 2.5 years, general cognitive performance was measured using the cognitive score from the INTERGROWTH-21st Neurodevelopmental Assessment (INTER-NDA), which is an international standardized mixed-methodology neurodevelopmental evaluation for children aged 22–30 months (Fernandes et al., 2020). Its norms are the first international standards of early child development constructed according to WHO Multicenter Growth Reference Study's prescriptive guidelines (Fernandes et al., 2020). At the age of 5 years, Wechsler Preschool and Primary Scale of Intelligence – Third Edition (Wechsler, 2009) Performance IQ (calculated on the basis of the subtests Block Design and Matrix Reasoning) was utilized as a measure of general cognitive performance.

Statistical analyses

Descriptive statistics were calculated using SPSS software (version 28). To take item level difficulty and discrimination properties into account (Willoughby et al., 2011), an IRT analysis was conducted in Mplus (version 8.6) (Muthén & Muthén, 1998-2017) for all the EF tasks to form factor scores for latent profile analysis (for detailed description of IRT analysis see Yada et al., 2025). For the IRT analysis, child responses in each trial of each EF task were rescored with a binary scale (pass = 1, fail = 0), except for the Snack Delay task, which was recorded by using an ordinal scale (0 to 6) and analyzed with a graded response model, which accounts for the multiple response categories. Following the IRT analysis, factor scores for latent profile analysis were estimated for each child on the basis of their item response patterns and the corresponding model parameters. Specifically, discrimination and difficulty parameters were included for all tasks. An additional guessing parameter was estimated for the Spin the Pots task, because the task format (multiple pots, some empty) creates a possibility of initial success through random search, which could otherwise distort estimates of latent ability. Residualized factor scores, which controlled for age in months at the research visits, were used for Spin the Pots tasks at 2.5 years and 5 years, as well as for Farmer task at 5 years, because they showed significant bivariate correlations with age.

First, LPA was conducted in Mplus to identify different profiles of EF. The LPA enables researchers to group individuals with similar performance into subgroups (latent profiles). Essentially, variations in individual responses to observed items are attributed to their membership in a particular latent profile, with each profile exhibiting a unique, subgroup-specific pattern (Geiser, 2012). The proportion of missing values for EF tasks varied from 34.8% to 56.3%. A Missing Completely At Random (MCAR; [Little, 1988]) test showed statistically nonsignificant results for the EF tasks (χ^2 (207) = 238.46, p = .066) indicating that the data was missing at random. The Full Information Maximum Likelihood (FIML) method was used to handle the missing values in this analysis. For choosing the profile solution that best fit our data, we relied on established goodness-of-fit indexes such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and sample-size adjusted BIC (SABIC). Smaller values in the AIC, BIC, and SABIC indicate better model fit. In addition, the results of the Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (VLMR), the Lo–Mendell–Rubin adjusted likelihood ratio test (LMR-A), and the bootstrapped likelihood ratio test (BLRT), comparing a class solution against a solution with −1 number of classes, were examined. Significant p values in those tests mean that the model with $k-1$ classes is rejected in favor of the model with k classes. Further, average class assignment probabilities were assessed, with values > 0.80 indicating well-specified classes (Geiser, 2012). Given the recommendation for a large set of random starting values to ensure profile validity, 500 starting values were utilized in this study.

Second, we examined whether there were significant associations between the assigned profile group and covariates. Potential covariates included demographic variables such as duration of gestation (separate sensitivity analysis conducted by using prematurity, i.e., children born before gw 37), sex, general cognitive performance, and SES measured as maternal educational level and financial satisfaction—all of which have been shown to be correlated with EF task performance in childhood (e.g., Lawson et al., 2018). For demographic variables, the proportion of missing values varied from 0% to 44.1% and the result of the MCAR test was statistically significant (χ^2 (36) = 51.50, p = .045). Since the missingness seemed to relate to SES variables, i.e., maternal education and financial satisfaction, we assumed that the data was missing at random (Enders, 2011). Moreover, given the possible effect of SES on EFs, we wanted to retain the possibility to investigate the associations between SES variables and longitudinal EF profiles. Thus, SES variables were not considered in the imputation. Multinomial logistic regression analyses were conducted to analyze the association of the latent profiles and covariates. The Vermunt's (2010) three-step estimation approach, using the R3STEP command in Mplus, was employed to address classification-error associated with the most likely class membership, which is recommended for covariate analysis (Asparouhov & Muthen, 2021). Missing data (besides SES) were handled by using multiple imputation in this step because the FIML method is not available for covariates with R3STEP. Logistic regression coefficients near zero indicate that the demographic variables are less influential. Further, we examined odds ratios (OR) for each demographic variable. Variables that increase the logit have ORs above 1.0, those with no effect have ORs of 1.0, and variables that decrease the logit have ORs below 1.0 (Gray & Kinnear, 2012). It is considered that if the 95% confidence interval of the OR includes 1, then there is no statistically significant effect.

Results

Descriptive statistics for the EF tasks across ages (raw scores before conducting an IRT analysis) are presented in the Supplementary material (Table S2). Correlations between different EF measurements are reported in Table 2. Since child age at measurement only correlated with some of the outcome variables (Spin the Pots task at 2.5 and 5 years, and the Farmer task at 5 years), child age at the respective time point was regressed out of those EF scores in subsequent analyses. However, it was checked that a comparable latent profile solution and associations between the profiles and main predictors could also be outlined without residualized variables.

First, utilizing latent profile analysis, we explored longitudinal latent profiles of EFs. Starting with a one-class solution, we added the number of possible classes one by one until the highest possible profile-solution that could be identified with the data was reached. Fit statistics for one through five class solutions are reported in Table 3. The VLMR, LMR-A, and BLRT showed that a two-class solution fit the data better than a one-class solution (p < .001) and a three-class solution fit better than a two-class solution (p < .001). However, a four-class solution did not fit better than a three-class solution (VLMR and LMR-A p values = .050). Thus, the model with the three profiles was considered as the best-fitting model.

The three profiles identified were (1) "The below average performers" (14.2%), (2) "The average performers with lower Spin the Pots (SP) (5 y) performance" (56.0%), and (3) "The average performers with higher SP (5 y) performance (29.8%)." That is, with latent profile analysis, we could identify longitudinal continuity in EF task performance during early childhood by identifying a group of children performing continuously below average, and two average performing groups that were differentiated by SP task performance at age 5. The three profiles are shown in Figure 1.

Associations with main predictors

Differences between the EF profiles in terms of the main predictors are shown in Table 4. Official sex at birth was associated with the likelihood of belonging to Profile 1 "the below average performers" compared to Profile 3 "the average performers with higher SP (5 y) performance," indicating that in comparison to girls, boys were more likely to belong to the below average profile (p = .017. OR = 0.427. 95% CI = 0.212–0.861). In addition, general cognitive performance at age 5 was negatively associated with the likelihood of belonging to Profile 1 "The below average performers" compared to Profile 3 "The average performers with higher SP (5 y) performance," which suggests that lower general cognitive performance

Table 2 Bivariate correlations between tasks.

	AB	SP, 2.5 y	Snack	Arrows	Pigs	Farmer	DG	SP, 5 y	Age, 8 m	Age, 2.5 y	Age, 5 y	Cognition, 2.5 y	Cognition, 5 y	GW	Maternal education
SP, 2.5 y	.139*														
Snack	.025	.238**													
Arrows	.077	.154*	.155*												
Pigs	.039	.146*	.241**	.141**											
Farmer	−.074	−.011	.091	.207**	.242**										
DG	.032	.068	.031	.070	.070	.174**									
SP, 5 y	−.089	−.052	.050	.040	.183**	.048	.041								
Age, 8 m	.024	−.175	−.073	.044	.015	.115	.083	−.018							
Age, 2.5 y	.088	.140**	.076	−.003	.086	−.087	−.008	−.055	.022						
Age, 5 y	.004	−.041	.016	.021	.069	.133**	.060	.108*	.271**	.042					
Cognition, 2.5 y	.074	.098*	.069	.163**	.026	.083	.041	.024	.023	.070	−.008				
Cognition, 5 y	.104	.176**	.096	.145**	.155**	.193**	.099*	.112*	.056	−.001	.299	.134*			
GW	−.059	.166**	.052	−.061	.057	.052	.059	.056	−.822*	−.077	−.014	−.039	.050		
Maternal education	.001	.117	.033	.081	.108*	.056	.115**	−.009	.024	.181**	.045	.181**	.161**	−.042	
Financial satisfaction	−.030	.053	.055	.030	.079	.113*	.072	−.041	−.024	.082	.060	.128**	.093*	−.044	.210**

Note. AB = A not B, SP = Spin the Pots, Snack = Snack Delay, DG = Delay of Gratification, GW = gestational weeks. *$p < .05$; **$p < .01$. Spearman's rho for maternal education, Pearson correlations for other variables.

Table 3 Fit statistics for latent profile analysis for one through five class solutions.

	AIC	BIC	SABIC	Entropy	VLMR *p* value	LMR-A *p* value	BLRT *p* value	Class proportions	Average latent class probabilities
1 class	9885.566	9961.109	9910.298	1	N/A	N/A	N/A	1.000	1.000
2 classes	9718.062	9836.098	9756.707	0.692	.0002	.0002	.0000	.170/.830	.875/.920
3 classes	**9609.182**	**9769.71**	**9661.738**	**0.594**	**.0002**	**.0002**	**.0000**	**.142/.560/.298**	**.871/.755/.932**
4 classes	9554.516	9757.537	9620.984	0.585	.0501	.0524	.0000	.130/.420/.121/.329	.815/.737/.832/.719
5 classes	9494.403	9739.917	9574.783	0.563	.0150	.0163	.0000	.109/.242/.348/.174/.127	.852/.729/.613/.777/.747

Note. SABIC = sample-size adjusted BIC; VLMR = Vuong-Lo-Mendell-Rubin likelihood ratio test; LMR-A = Lo–Mendell–Rubin adjusted likelihood ratio test; BLRT = bootstrapped likelihood ratio test; N/A = not applicable. Age-relatedly residualized variables for Spin the Pots, 2.5 y; Spin the Pots, 5 y; and Farmer, 5 y. The best-fitting profile solution is bolded.

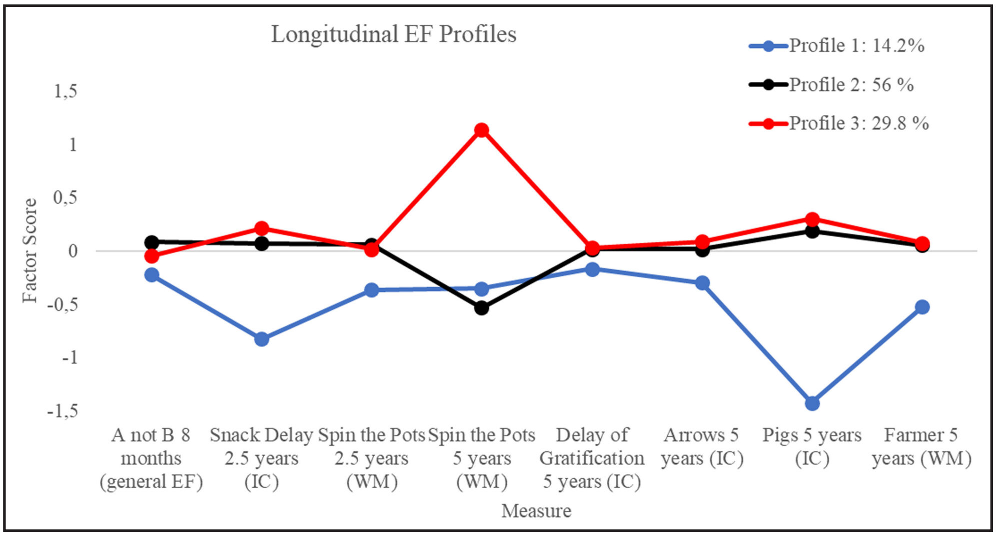

Fig. 1. The profiles of EFs across tasks (N = 830). Note: Profile 1 = "The below average performers", Profile 2 = "The average performers with lower Spin the Pots (5 y) performance", Profile 3 = "The average performers with higher Spin the Pots (5 y) performance".

Table 4 Comparisons among the profiles for the background variables .

	Profile 1 (Profile 3 as a reference)			Profile 2 (Profile 3 as a reference)			Profile 1 (Profile 2 as a reference)		
Factor	Estimate (*SE*)	*p*	OR (95% CI)	Estimate (*SE*)	*p*	OR (95% CI)	Estimate (*SE*)	*p*	OR (95% CI)
Child sex (female)	**−0.850 (0.357)**	**.017**	**0.427 (0.212–0.861)**	−0.334 (0.249)	.180	0.716 (0.439–1.167)	0.516 (0.358)	.149	1.675 (0.831–3.378)
Cognition, 2.5 years[a]	−0.021 (0.041)	.605	0.979 (0.903–1.061)	−0.012 (0.031)	.715	0.989 (0.929–1.051)	0.010 (0.041)	.810	1.010 (0.933–1.094)
Cognition, 5 years[b]	**−0.033 (0.011)**	**.003**	**0.968 (0.947–0.989)**	**−0.018 (0.009)**	**.051**	**0.982 (0.965–1.000)**	0.015 (0.011)	.180	1.015 (0.993–1.037)
Duration of gestation	−0.091 (0.103)	.381	0.913 (0.746–1.119)	0.003 (0.066)	.969	1.003 (0.880–1.142)	0.093 (0.107)	.385	1.098 (0.890–1.354)
Maternal education	−0.243 (0.221)	.271	0.784 (0.508–1.209)	0.081 (0.163)	.620	1.084 (0.788–1.492)	0.324 (0.216)	.133	1.383 (0.906–2.110)
Financial satisfaction	−0.214 (0.197)	.279	0.808 (0.548–1.189)	−0.040 (0.142)	.777	0.961 (0.727–1.269)	0.173 (0.190)	.363	1.189 (0.819–1.727)

Note. Profile 1: "The below average performers"; Profile 2: "The average performers with lower Spin the Pots (5 y) performance"; Profile 3: "The average performers with higher Spin the Pots (5 y) performance." *SE* = standard error of the coefficient; OR = odds ratio; 95% CI = 95% confidence interval. The significant or indicative differences are bolded.
[a] INTER-NDA: cognitive score.
[b] WPPSI-III: PIQ .

was associated with likelihood to belong to the below average profile ($p = .003$. OR = 0.968. 95% CI = 0.947–0.989). Similar association was detected between lower general cognitive performance at 5 years and belonging to Profile 2 “the average-performers with lower SP (5 y) performance” when compared to Profile 3 “with higher SP (5 y) performance,” but the result was just above a statistical significance ($p = .051$. OR = 0.982. 95% CI = 0.965–1.000). There were no differences between the EF profiles in terms of maternal education, maternal financial satisfaction, general cognitive performance at 2.5 years, or duration of gestation. A separate sensitivity analysis using prematurity instead of duration of gestation was performed with no differences in the estimates of the models.

Discussion

The main aim of the present study was to investigate the largely unexplored variation in longitudinal latent profiles of early EF development from infancy through the preschool years. To our knowledge, the current research is the first longitudinal data on early EF trajectories utilizing person-centered, data-driven approaches (IRT, LPA) from infancy to toddlerhood to preschool. Interestingly, by using methods that can handle the challenges of assessing at varying ages and different measures and task modalities, we identified three distinct latent profiles of EF task performance across infancy through early childhood. We identified a group of children performing continuously below average (14.2%), and two average performing groups that were differentiated by SP task performance at 5 years (lower SP 56.0%; higher SP 29.8%). We also found expected associations with sex and general cognitive performance in early childhood, further supporting the external validity of the identified latent profiles.

Our findings resemble results from previous cross-sectional studies of older children that identified low, moderate, and high EF performance groups (e.g., Litkowski et al., 2020; Williams & Bentley, 2021). Some previous studies also identified nonlinear development of EFs and changes in developmental trajectories such as descending or ascending delays in EFs (e.g., Lynch et al., 2024; in 3–12-year-olds). However, we did not find evidence of nonlinear developmental profiles in our study. Importantly, we did identify one group with continuously low performance across multiple different tasks and at different ages. Identifying such children is of clinical relevance for recognizing children at risk of delayed EF development as early as possible for prevention of later negative outcomes. EF delays or deficits are an identified transdiagnostic risk factor for psychopathology and other atypical developmental outcomes, and EF skills are a promising target for preventive interventions early in development (e.g., Yang et al., 2022).

We also found an important distinguishing role of performance on SP working memory task at 5 years, for two profile groups (“average performers,” 56.0%, and “above average performers,” 29.8%), who otherwise were indistinguishable across other tasks and prior measurement time points. The IRT analyses revealed a ceiling effect on some of the tasks that might have contributed not to finding a high performing profile as in prior cross-sectional studies. The IRT analyses also showed that the SP task seemed to best differentiate above-average performers from average performers, and to not efficiently differentiate below-average from average performers. Interestingly, at the same age point, performance on the Pig task also revealed differences between the profiles, with below-average performers showing particularly poor performance on the task. In contrast to the SP task, IRT analysis revealed that Pig measures below-average levels of EF ability with high reliability, whereas it did not seem to efficiently differentiate above-average from average performers. Thus, Pig seems to reliably detect variation among children belonging to below-average profile. This may explain why the SP task at 5 years of age stood out in terms of the two average performing groups, but not the low performers, and why the Pig task stood out only for the below-average profile. The different discriminative capacities of administrated EF tasks may also stem from differing abilities required by the tasks. For example, SP has been originally described as a broad and multifaceted EF measure (Hughes & Ensor, 2005), which cannot be considered a pure WM measure although involving a WM component (Morra et al., 2021), whereas, another 5-year measure, Farmer, may provide a more direct measure of visuospatial WM (Willoughby et al., 2017a). The detection of high performers in the person-administered SP task may also partly reflect potential effects of administration differences, such as the child’s engagement with the task, compared to the tablet-based WM task Farmer, which showed more pronounced ceiling effects. When choosing between tablet-based and person-administered tasks, which possibly involves finding a rewarding object as does SP, researchers should consider the age of the participants, as younger children may remain more motivated and engaged in the context of supportive examiner interaction. One possible insight from a practical application perspective might be whether more complex EF tasks tapping into multiple domains of EF could better capture individual differences even within the group of high-performing children. Future studies in the same sample of children should continue exploring whether these specific tasks and the identified profiles have significance for future outcomes, including later EF development.

Regarding the sociodemographic predictors, results were partially in line with previous research and suggested further evidence for external validity of the latent profiles. As expected, male sex and lower general cognitive performance at 5 years were most common for the “below average performers” group. Girls tend to outperform boys on average in various aspects of EFs (e.g., Silverman, 2021). The association with sex may be nonlinear such that even if average EF scores are similar for boys and girls, boys are overrepresented in the lowest percentiles of the distribution of scores (Robson et al., 2020). Turning to general cognitive performance, it is not surprising that we found an expected association at 5 years, given its overlap with EF task performance (Grobe et al., 2024). However, we did not find an association at 2.5 years. This may be because this association emerges only later in childhood when the validity and reliability of task performance measures become strong enough to detect the association.

Contrary to our hypotheses, neither SES (i.e., maternal education and financial satisfaction) nor the length of gestation and prematurity were associated with EF profiles. There is a broad literature on the influence of SES on EFs in offspring, favoring children in the higher quartiles of SES distribution (e.g., Cuartas et al., 2022). However, earlier studies have shown that the evidence of covariation between SES and EFs is mixed prior to 2 years (Miller et al., 2023), with SES serving as a proxy for these factors in some studies but not in others. It is noteworthy that in the sample of the current study, the overall SES level is rather high, with 44.3% of mothers having a university degree. This together with the societal context where SES differences are compensated by health care and social services may contribute to both a reduction in variance in SES as well as mitigation of potential effects of living in a lower-SES household. In the same way, high SES level and comprehensive welfare services may ameliorate effects of prematurity. These scenarios could make it more difficult to detect any link between SES or length of gestation, even at the preterm birth level, and EF profiles in the current study (e.g., Stoolmiller, 1999; Widding-Havneraas & Pedersen, 2020). In summary, the observed associations between background factors both strengthen the validity of the profiles and may identify factors that can help to identify children in need of support already early in development.

Importantly, our findings indicate that by using person-centered analysis methods, it is possible to identify separable latent profiles of longitudinal development of EFs from infancy onward even with varying number and types of tasks administered at different ages. Our findings also lend some support to the idea that individual differences in EFs are showing a detectable degree of stability already starting in infancy into toddlerhood at the latent profile group level (even if not at the individual level). The current study, to our knowledge, is the first to demonstrate stable individual differences in EF task performance across the entire developmental span from infancy through toddlerhood to preschool years. Previous evidence for stability of individual differences across transitions from infancy to toddlerhood to preschool years has been contradictory, with most evidence documenting stability from infancy to 36 months without bridging the developmental progression from infancy through toddlerhood into the preschool period (Blankenship et al., 2019; Johansson et al., 2016) or showing that some stability is evident from toddlerhood onwards (Broomell & Bell, 2022; Miller et al., 2023) This would be well in line with the current knowledge on dynamic development of skills through the first 2 years of life, especially those complex skills that contribute to variation in EFs (D'souza et al., 2017; Iverson, 2021). However, also in our study, there are significant uncertainties such as the use of only one single task in infancy. Still, the findings of the present study are promising in terms of the possibility to describe and detect EF differences from very early on, and future studies should continue exploring the development of EF starting from infancy using data-driven and person-centered methods of analysis even when the number and types of tasks change with each child age and measurement time point.

The current research has several strengths. First, our sample may be one of the few datasets that can chart EF development in the same children starting in infancy, at 8 months of age, across early childhood and beyond. Second, our study design leans on previous findings that recommend the use of several task-based (as opposed to parent-reported) measures at each age point to produce larger effect sizes (e.g., Broomell & Bell, 2022; Robson et al., 2020). Above all, we used recommendable—though still in the field of longitudinal EF research rarely used—analysis methods. Our analytical approach combined, to our knowledge for the first time, IRT and LPA analyses (Willoughby et al., 2011). This approach aims at obtaining more precise information about inter-individual differences in ability level by capturing estimates of children's underlying EF abilities while minimizing random measurement error, and maximizing the use of the item-level information in each task (Willoughby et al., 2011). By using these multistage methods, we were able to identify longitudinal latent profiles of EF across early childhood even though the rank-order stability of the ability is low at this age. Lastly, because the current study was conducted in a context with exceptionally comprehensive support structures in health, education, and welfare, it provides a unique example of EF development under favorable developmental conditions, which may be valuable for future cross-cultural comparisons.

There are also limitations in the current study. First, there was relatively low overlap between the samples, with only 150 children having some data from all time points (maximum likelihood methods were used to address the missing data to minimize biased estimates). Further, the entropy values in the solutions were generally low, indicating low classification certainty. This may be due to insufficiently distinct profiles, particularly between Profile 2 and Profile 3—"average performers" and "above average performers". Further research using indicators that better differentiate children's EF skills is necessary. Second, we had only one EF task at 8 months, which stems from the very limited selection of age-appropriate tasks available in infancy and infants' limited capacity for task performance. This challenge has also led to using only one infant EF measure in previous research (e.g., Blankenship et al., 2019; Broomell & Bell, 2022) and needs to be addressed in future studies. Similarly, at 5 years, only one task appeared to differentiate between better performers and poorer performers, and use of larger variety of tasks with better task performance distribution might have shed more light on the differences between average performing profiles. Also, other tasks had uneven abilities for differentiation between the profiles (e.g., Delay of Gratification showing little differences across profiles), which should be examined more closely in future studies. Third, within our sample, there were very few children with different clinical diagnoses of developmental disorders, such as autism spectrum disorder, that may have contributed to EF performance. However, since most such disorders are unidentified in early childhood, exclusion of these children was considered unnecessary and might have contributed to less true variation in the data. Thus, the association between developmental diagnoses and EF profiles needs to be further examined at school age. Finally, the children in the cohort come from relatively homogeneous and high SES backgrounds, which may further limit generalizability to high-risk populations. Replication studies with more diverse and representative sampling are needed to improve the cross-cultural generalizability of findings.

To conclude, using data-driven and person-centered analysis methods, we described the variation of early longitudinal development in EFs by identifying three separable latent profiles from infancy to preschool age. We also found expected associations between the profiles and sex and general cognitive performance in early childhood, which further supports the validity of the profiles and may help identifying children in need of support early in development. Moreover, our findings give future studies a promising insight into the possibility of detecting EF differences starting from infancy, even when using different tasks across ages to characterize the distribution of EF performance.

Supplementary material

Supplementary material is available at *Child Development* online.

Data availability

The data necessary to reproduce the analyses presented here are not publicly accessible. The data protection legislation in Finland and ethics regulations of the FinnBrain Research do not permit open distribution of data, but data could be delivered by special requests (which includes formal collaboration and material transfer agreements). Investigators interested in research collaboration and obtaining access to the data are encouraged to contact FinnBrain board (finnbrain-board@lists.utu.fi). Contact information of the board members or FinnBrain Principal Investigators are listed on the project website: https://sites.utu.fi/finnbrain/en/contact/

The analytic code necessary to reproduce the analyses presented in this paper is not publicly accessible but is available from the corresponding author upon reasonable request.

The materials necessary to attempt to replicate the findings presented here are not publicly accessible.

The analyses presented here were not preregistered.

Author contributions

Anniina Karonen (Conceptualization, Data curation, Investigation, Methodology, Writing—review & editing [equal], Formal analysis [supporting], Writing—original draft [lead]), Pauliina Juntunen (Conceptualization, Formal analysis, Methodology, Writing—review & editing [equal], Visualization [lead], Writing—original draft [supporting]), Akie Yada (Formal analysis, Methodology, Supervision, Writing—review & editing [equal]), Fiia Takio (Conceptualization, Writing—review & editing [equal]), Elisabeth Nordenswan (Data curation, Investigation, Methodology, Writing—review & editing [equal], Project administration [supporting]), Eeva Päivikki Eskola (Data curation, Investigation, Methodology, Writing—review & editing [equal], Project administration [supporting]), Kirby Deater-Deckard (Conceptualization, Methodology, Supervision, Writing—review & editing [equal]), David J. Bridgett (Methodology, Supervision, Writing—review & editing [equal]), Michelle Fernandes (Methodology, Writing—review & editing [equal]), Elina Mainela-Arnold (Funding acquisition, Writing—review & editing [equal]), Hasse Karlsson (Funding acquisition, Supervision, Writing—review & editing [equal], Project administration [lead]), Linnea Karlsson (Funding acquisition, Methodology, Supervision, Writing—review & editing [equal], Project administration [lead]), Eeva-Leena Kataja (Investigation, Methodology, Supervision, Writing—review & editing [equal], Project administration [supporting]), Riikka Korja (Conceptualization, Funding acquisition, Supervision, Writing—review & editing [equal], Project administration [lead]), and Saara Nolvi (Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Supervision, Writing—review & editing [equal], Project administration [lead])

Funding

This research was supported by the Emil Aaltonen Foundation, Finnish Cultural Foundation, Finnish State Grants for Clinical Research (VTR), Jane and Aatos Erkko Foundation, Jenny and Antti Wihuri Foundation, Juho Vainio Foundation, Olvi Foundation, Research Council of Finland (253270, 264363, 308176, 308252, 308589, 314390, 332444, 346121, 346790), Signe and Ane Gyllenberg Foundation, Strategic Research Council (SRC) within the Research Council of Finland (352648, 352655), and the Yrjö Jahnsson Foundation.

Conflicts of interest

No conflict of interest was reported by the authors.

Acknowledgments

Anniina Karonen and Pauliina Juntunen equally contributed to this study.

We are grateful to the participating families for their commitment. We also thank the research personnel and students involved in data collection for their invaluable assistance in processing and disseminating this research. Additionally, we thank Dr. Asko Tolvanen for his statistical expertise and consultation throughout the analysis process.

References

Asparouhov, T., & Muthen, B. (2021). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. *Mplus Web Notes*, *21*, 1–22.

Beck, D. M., Schaefer, C., Pang, K., & Carlson, S. M. (2011). Executive function in preschool children: Test-retest reliability. *Journal of Cognition and Development*, *12*, 169–193. https://doi.org/10.1080/15248372.2011.563485

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

Blankenship, T. L., Slough, M. A., Calkins, S. D., Deater-Deckard, K., Kim-Spoon, J., & Bell, M. A. (2019). Attention and executive functioning in infancy: Links to childhood executive function and Reading achievement. *Developmental Science*, *22*, e12824. https://doi.org/10.1111/desc.12824

Bridgett, D. J., Ganiban, J. M., Neiderhiser, J. M., Natsuaki, M. N., Shaw, D. S., Reiss, D., & Leve, L. D. (2018). Contributions of mothers' and fathers' parenting to children's self-regulation: Evidence from an adoption study. *Developmental Science*, *21*, e12692. https://doi.org/10.1111/desc.12692

Broomell, A. P. R., & Bell, M. A. (2022). Longitudinal development of executive function from infancy to late childhood. *Cognitive Development*, *63*, 101229. https://doi.org/10.1016/j.cogdev.2022.101229

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

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

Cuartas, J., Hanno, E., Lesaux, N. K., & Jones, S. M. (2022). Executive function, self-regulation skills, behaviors, and socioeconomic status in early childhood. *PLoS One*, *17*, e0277013. https://doi.org/10.1371/journal.pone.0277013

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development*, *56*, 868–883. https://doi.org/10.2307/1130099

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

D'souza, D., D'souza, H., & Karmiloff-Smith, A. (2017). Precursors to language development in typically and atypically developing infants and toddlers: The importance of embracing complexity. *Journal of Child Language*, *44*, 591–627. https://doi.org/10.1017/S030500091700006X

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, *8*, 341–349. https://doi.org/10.1037/1040-3590.8.4.341

Enders, C. K. (2011). Analyzing longitudinal data with missing values. *Rehabilitation Psychology*, *56*, 267–288. https://doi.org/10.1037/a0025579

Fernandes, M., Villar, J., Stein, A., Staines Urias, E., Garza, C., Victora, C. G., Barros, F. C., Bertino, E., Purwar, M., Carvalho, M., Giuliani, F., Wulff, K., Abubakar, A. A., Kihara, M., Cheikh Ismail, L., Aranzeta, L., Albernaz, E., Kunnawar, N., Di Nicola, P., … Kennedy, S. (2020). INTERGROWTH-21st Project international INTER-NDA standards for child development at 2 years of age: An international prospective population-based study. *BMJ Open*, *10*, e035258. https://doi.org/10.1136/bmjopen-2019-035258

Fiske, A., & Holmboe, K. (2019). Neural substrates of early executive function development. *Developmental Review*, *52*, 42–62. https://doi.org/10.1016/j.dr.2019.100866

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

Fujisawa, K. K., Todo, N., & Ando, J. (2017). Genetic and environmental influences on the development and stability of executive functions in children of preschool age: A longitudinal study of Japanese twins. *Infant and Child Development*, *26*, 1–27. https://doi.org/10.1002/icd.1994

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

Geiser, C. (2012). *Data analysis with Mplus*. Guilford Press.

Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS Statistics 19 made simple* (pp. xiv, 671). Psychology Press.

Grobe, S. E., Könen, T., David, C., Grüneisen, L., Dörrenbächer-Ulrich, L., Perels, F., & Karbach, J. (2024). The factorial structure of executive functions in preschool and elementary school children and relations with intelligence. *Journal of Experimental Child Psychology*, *246*, 106014. https://doi.org/10.1016/j.jecp.2024.106014

Hendry, A., Jones, E. J. H., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review*, *42*, 1–33. https://doi.org/10.1016/j.dr.2016.06.005

Houts, C. R., Savord, A., & Wirth, R. J. (2022). Overview of modern measurement theory and examples of its use to measure execution function in children. *Journal of Pediatric Neuropsychology*, *8*, 1–14. https://doi.org/10.1007/s40817-021-00117-7

Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology*, *28*, 645–668. https://doi.org/10.1207/s15326942dn2802_5

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. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, *35*, 20–36. https://doi.org/10.1080/87565640903325691

Iverson, J. M. (2021). Developmental variability and developmental cascades: Lessons from motor and language development in infancy. *Current Directions in Psychological Science*, *30*, 228–235. https://doi.org/10.1177/0963721421993822

Johansson, M., Marciszko, C., Brocki, K., & Bohlin, G. (2016). Individual differences in early executive functions: A longitudinal study from 12 to 36 months. *Infant and Child Development*, *25*, 533–549. https://doi.org/10.1002/icd.1952

Karlsson, L., Tolvanen, M., Scheinin, N. M., Uusitupa, H.-M., Korja, R., Ekholm, E., Tuulari, J. J., Pajulo, M., Huotilainen, M., Paunio, T., & Karlsson, H.; FinnBrain Birth Cohort Study Group. (2018). Cohort profile: The FinnBrain Birth Cohort Study (FinnBrain). *International Journal of Epidemiology*, *47*, 15–16. https://doi.org/10.1093/ije/dyx173

Kataja, E.-L., Karlsson, L., Leppänen, J. M., Pelto, J., Häikiö, T., Nolvi, S., Pesonen, H., Parsons, C. E., Hyönä, J., & Karlsson, H. (2020). Maternal depressive symptoms during the pre- and postnatal periods and infant attention to emotional faces. *Child Development*, *91*, e475–e480. https://doi.org/10.1111/cdev.13152

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*, 220–232. https://doi.org/10.1037/0012-1649.36.2.220

Kraybill, J. H., Kim-Spoon, J., & Bell, M. A. (2019). Infant attention and age 3 executive function. *The Yale Journal of Biology and Medicine*, *92*, 3–11.

Lawson, G. M., Hook, C. J., & Farah, M. J. (2018). A meta-analysis of the relationship between socioeconomic status and executive function performance among children. *Developmental Science*, *21*, e12529. https://doi.org/10.1111/desc.12529

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

Litkowski, E. C., Finders, J. K., Borriello, G. A., Purpura, D. J., & Schmitt, S. A. (2020). Patterns of heterogeneity in kindergarten children's executive function: Profile associations with third grade achievement. *Learning and Individual Differences*, *80*, 101846. https://doi.org/10.1016/j.lindif.2020.101846

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Lynch, J. D., Xu, Y., Yolton, K., Khoury, J. C., Chen, A., Lanphear, B. P., Cecil, K. M., Braun, J. M., & Epstein, J. N. (2024). Environmental predictors of children's executive functioning development. *Child Neuropsychology*, *30*, 615–635. https://doi.org/10.1080/09297049.2023.2247603

Miller, S. E., Galvagno, L. G., & Elgier, Á. (2023). Universality and context-specificity in early executive function development. *Infant Behavior & Development*, *71*, 101841. https://doi.org/10.1016/j.infbeh.2023.101841

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

Morra, S., Gandolfi, E., Panesi, S., & Prandelli, L. (2021). A working memory span task for toddlers. *Infant Behavior & Development*, *63*, 101550. https://doi.org/10.1016/j.infbeh.2021.101550

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide* (8th ed.). Muthén & Muthén.

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, and Allied Disciplines*, *58*, 361–383. https://doi.org/10.1111/jcpp.12675

Nolvi, S., Pesonen, H., Bridgett, D. J., Korja, R., Kataja, E.-L., Karlsson, H., & Karlsson, L. (2018). Infant sex moderates the effects of maternal pre- and postnatal stress on executive functioning at 8 months of age. *Infancy*, *23*, 194–210. https://doi.org/10.1111/infa.12206

Olson, L., Chen, B., & Fishman, I. (2021). Neural correlates of socioeconomic status in early childhood: A systematic review of the literature. *Child Neuropsychology*, *27*, 390–423. https://doi.org/10.1080/09297049.2021.1879766

Patwardhan, I., Gordon, C., & Mason, W. A. (2023). Trajectories of cognitive flexibility through kindergarten and first grade: Implications for externalizing and internalizing behavior problems in the second grade. *Developmental Psychology*, *59*, 1794–1806. https://doi.org/10.1037/dev0001597

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

Ribeiro, M., Yordanova, Y. N., Noblet, V., Herbet, G., & Ricard, D. (2024). White matter tracts and executive functions: A review of causal and correlation evidence. *Brain: A Journal of Neurology*, *147*, 352–371. https://doi.org/10.1093/brain/awad308

Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*, *146*, 324–354. https://doi.org/10.1037/bul0000227

Sandoval, C. C., Gaspardo, C. M., & Linhares, M. B. M. (2022). The impact of preterm birth on the executive functioning of preschool children: A systematic review. *Applied Neuropsychology: Child*, *11*, 873–890. https://doi.org/10.1080/21622965.2021.1915145

Silverman, I. W. (2021). Gender differences in inhibitory control as assessed on simple delay tasks in early childhood: A meta-analysis. *International Journal of Behavioral Development*, *45*, 533–544. https://doi.org/10.1177/01650254211020385

Spinrad, T. L., Eisenberg, N., & Gaertner, B. M. (2007). Measures of effortful regulation for young children. *Infant Mental Health Journal: Infancy and Early Childhood*, *28*, 606–626. https://doi.org/10.1002/imhj.20156

Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior–genetic adoption studies. *Psychological Bulletin*, *125*, 392–409. https://doi.org/10.1037/0033-2909.125.4.392

Stucke, N. J., & Doebel, S. (2024). Early childhood executive function predicts concurrent and later social and behavioral outcomes: A review and meta-analysis. *Psychological Bulletin*, *150*, 1178–1206. https://doi.org/10.1037/bul0000445

Sun, J., Mohay, H., & O'Callaghan, M. (2009). A comparison of executive function in very preterm and term infants at 8 months corrected age. *Early Human Development*, *85*, 225–230. https://doi.org/10.1016/j.earlhumdev.2008.10.005

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*, 450–469. https://doi.org/10.1093/pan/mpq025

Wechsler, D. (2009). *WPPSI-III Wechsler Preschool and Primary Scale of Intelligence–Third Edition. Administration and scoring manual*. Psykologien Kustannus Oy, Helsinki.

Widding-Havneraas, T., & Pedersen, S. H. (2020). The role of welfare regimes in the relationship between childhood economic stress and adult health: A multilevel study of 20 European countries. *SSM—Population Health*, *12*, 100674. https://doi.org/10.1016/j.ssmph.2020.100674

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

Williams, K. E., & Bentley, L. A. (2021). Latent profiles of teacher-reported self-regulation and assessed executive function in low-income community preschools: Relations with motor, social, and school readiness outcomes. *Frontiers in Psychology*, *12*, 708514. https://doi.org/10.3389/fpsyg.2021.708514

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012a). 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., Blair, C. B., Wirth, R. J., & Greenberg, M.; The Family Life Project Investigators. (2010). The measurement of executive function at age 3 years: Psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment*, *22*, 306–317. https://doi.org/10.1037/a0018708

Willoughby, M. T., Kuhn, L. J., Blair, C. B., Samek, A., & List, J. A. (2017a). The test-retest reliability of the latent construct of executive function depends on whether tasks are represented as formative or reflective indicators. *Child Neuropsychology*, *23*, 822–837. https://doi.org/10.1080/09297049.2016.1205009

Willoughby, M. T., Magnus, B., Vernon-Feagans, L., & Blair, C. B. (2017b). Developmental delays in executive function from 3 to 5 years of age predict kindergarten academic readiness. *Journal of Learning Disabilities*, *50*, 359–372. https://doi.org/10.1177/0022219415619754

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2011). Contributions of modern measurement theory to measuring executive function in early childhood: An empirical demonstration. *Journal of Experimental Child Psychology*, *108*, 414–435. https://doi.org/10.1016/j.jecp.2010.04.007

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

Yada, A., Deater-Deckard, K., Takio, F., Nordenswan, E., Eskola, E., Karlsson, H., Karlsson, L., Korja, R., Nolvi, S., & Tolvanen, A. (2025). 'Using item response theory to address challenges with measuring executive function in infancy and early childhood (v96ea_v1)', PsyArXiv, preprint: not peer reviewed. https://doi.org/10.31234/osf.io/v96ea_v1

Yang, Y., Shields, G. S., Zhang, Y., Wu, H., Chen, H., & Romer, A. L. (2022). Child executive function and future externalizing and internalizing problems: A meta-analysis of prospective longitudinal studies. *Clinical Psychology Review*, *97*, 102194. https://doi.org/10.1016/j.cpr.2022.102194