Development and Psychopathology (2026), **38**, 1438–1453
doi:10.1017/S0954579426101278

CAMBRIDGE UNIVERSITY PRESS

Regular Article

From traits to trajectories: How changes in reactive temperament relate to the development of externalizing symptoms

Henning Heinze[1,2], Florian Schmiedek[2,3], Valerie Brandt[4], Monika Daseking[1,2] and Julia Kerner auch Koerner[2,5]

[1]Helmut-Schmidt-University/University of the Armed Forces Hamburg, Germany; [2]IDeA - Center for Research on Individual Development and Adaptive Education of Children at Risk, Frankfurt, Germany; [3]DIPF Leibniz Institute for Research and Information in Education, Frankfurt, Germany; [4]Centre for Innovation in Mental Health, University of Southampton, UK and [5]University of Münster, Germany

Abstract

Temperament has been linked to the development of externalizing symptoms, but the nature of these associations remains unclear. Traditional approaches often treat early reactive temperament as static, overlooking developmental variation. This study applied a longitudinal latent change score model to examine how levels and changes in Negative Affect (NA) and Surgency from age 3 to 5 predict Conduct Problems (CP) and Attention-Deficit/Hyperactivity Disorder (ADHD) symptoms in early childhood. Data from the National Educational Panel Study ($N = 2{,}477$) were analyzed. Temperament was assessed at ages 3, 4, and 5 using the Children's Behavior Questionnaire, and CP and ADHD symptoms were measured at ages 5, 6, and 8 with the Strengths and Difficulties Questionnaire. Measurement invariance was established. Significant individual differences in developmental change emerged. Change in NA and Surgency, but not baseline levels, predicted higher latent CP and ADHD symptom levels at age 5 and further increases through age 8. These findings indicate that intraindividual change in reactive temperament can be a relevant marker of developmental risk. Temperamental risk for externalizing symptoms in early childhood is not fixed but may be shaped by both stable dispositions and developmental change, highlighting the importance of assessing temperament development to identify early emerging risk.

Keywords: ADHD; externalizing; longitudinal; preschool; temperament

(Received 6 June 2025; revised 8 December 2025; accepted 28 January 2026; first published online 9 March 2026)

Introduction

Regarding mental health problems of children and adolescents, a distinction is often made between internalizing and externalizing disorders. Internalizing disorders encompass affective and anxiety disorders, whereas externalizing disorders include impulse control problems such as attention deficit hyperactivity disorder (ADHD) and conduct problems (CP; American Psychiatric Association, 2013). In early childhood, externalizing behaviors are often developmentally typical but may reach clinically relevant levels when they persist or intensify (Wakschlag et al., 2015). Externalizing symptoms, such as aggression and rule-breaking, vary in severity and reflect a dimensional continuum rather than discrete categories (Coghill & Sonuga-Barke, 2012; Forslund et al., 2016; Wakschlag et al., 2015).

Externalizing symptoms

Externalizing symptoms can significantly affect a child's socio-emotional development and academic performance (Brennan et al., 2012; Breslau et al., 2009; Metcalfe et al., 2013). Clinically relevant externalizing disorders are associated with considerable stress and challenges for children and their families. The symptoms appear early and tend to persist (Erskine et al., 2016; McElroy et al., 2018; Ravens-Sieberer et al., 2007). Consistent with the increasing focus on a dimensional perspective (Hirjak et al., 2021), a complex interaction between interdependent individual vulnerabilities (e.g., genetic, epigenetic) and equally interdependent contextual risk factors (e.g., harsh parenting, bullying, deviant or criminal environment) is assumed to affect development trajectories of mental health symptoms in minors (Beauchaine & McNulty, 2013). Early temperament can predict later onset of those symptoms (Golm & Brandt, 2023). To understand why some children develop externalizing problems while others do not, early-emerging temperament traits provide a key explanatory framework.

Temperament

Temperament, defined by individual differences in emotional, motor, and attentional reactivity and self-regulation, plays a critical role in shaping a child's behavioral responses to the environment (Rothbart et al., 2007; Rothbart & Bates, 2006). Reactive temperament refers to early-emerging, biologically based differences in how strongly and how quickly individuals respond to positive or negative stimulation, reflecting the intensity and threshold of

Corresponding author: Henning Heinze; Email: henning.heinze@hsu-hh.de

Cite this article: Heinze, H., Schmiedek, F., Brandt, V., Daseking, M., & Kerner auch Koerner, J. (2026). From traits to trajectories: How changes in reactive temperament relate to the development of externalizing symptoms. *Development and Psychopathology* **38**: 1438–1453, https://doi.org/10.1017/S0954579426101278

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

emotional, behavioral, and physiological responses (Rothbart & Derryberry, 1981; Rothbart & Rueda, 2005).

Three broad components are central: Negative Affect (NA) and Surgency, which represent reactive dimensions, and Effortful Control (EC), which denotes the self-regulatory capacity to modulate these reactive tendencies through attention and inhibition. Because temperament is both biologically grounded and shaped by experience, it shows substantial developmental continuity, especially for reactive traits such as NA and Surgency, and greater plasticity for EC (Rothbart & Rueda, 2005).

Developmental stability

Developmental stability can be described at complementary levels that differ in focus and timescale. Mean-level stability refers to changes in the average level of a characteristic within a group over time, indicating whether a trait becomes generally stronger or weaker with age (Caspi et al., 2005). Rank-order stability describes the extent to which individuals maintain their relative position compared to others within the group, even when average levels change (Caspi & Roberts, 2001). Beyond these interindividual perspectives a more dynamic form of within-person stability captures how individuals tend to return toward or deviate from their own typical level over time, reflecting an ongoing intraindividual regulatory process (Grimm et al., 2012; McArdle, 2009; Nesselroade & Ram, 2004). Temperament is thus both stable and malleable, reflecting ongoing interactions between maturation and experience (Putnam et al., 2008; Rothbart & Rueda, 2005). The following sections review evidence for these dynamics in NA, Surgency, and EC.

Reactive temperament: negative affect

NA reflects biologically rooted but contextually shaped reactivity to negative or punishing stimuli (Rothbart & Bates, 2006). Longitudinal studies show moderate-to-high rank-order stability and small but systematic mean-level changes during early childhood. Bornstein et al. (2015) found moderate rank-order stability and curvilinear mean-level change, with NA peaking in toddlerhood and slightly declining afterward. Schmidt et al. (2025) reported comparable results in a large German sample, confirming both rank-order stability and age-related mean-level decreases between ages 3 and 6. These findings align with the view that NA reflects a reactive disposition that remains relatively stable in individual ranking but can vary in intensity across developmental stages (Caspi et al., 2005; Caspi & Roberts, 2001).

Cross-sectionally, Heinze et al. (2025) found that NA was positively associated with ADHD and CP, confirming the concurrent link between reactive temperament and externalizing behavior in preschool age. Wichstrøm et al., (2018) demonstrated that higher NA at age 4 predicted increases in ADHD and CP symptoms from preschool to middle childhood. Kostyrka-Allchorne et al. (2020) observed a similar effect, with early NA forecasting internalizing and externalizing symptoms at age 7, particularly in children with low self-regulation. Gartstein et al. (2012) likewise found that NA in toddlerhood predicted later behavior problems, especially when EC was low. More recent evidence extends this pattern into adolescence. Harris et al. (2025) reported that NA at age 4 predicted a general psychopathology factor, as well as later externalizing problems at age 15, underlining the long-term relevance of early reactivity.

NA is therefore best conceptualized as a relatively stable yet developmentally responsive trait that shapes trajectories of externalizing behavior through both its enduring rank-order position and its interaction with self-regulatory and contextual processes (Putnam et al., 2008; Rettew & McKee, 2005).

Reactive temperament: surgency

Whereas NA represents sensitivity to negative or punishing stimuli Surgency reflects approach-related tendencies toward positive stimulation and reward. Moderate levels promote exploration and engagement, but high levels may increase impulsivity and poorly regulated behavior (De Pauw & Mervielde, 2010; Nigg, 2006). Empirical findings indicates moderate-to-high rank-order stability and modest mean-level change in Surgency during early childhood. Bornstein et al. (2015) found increasing mean-level Surgency from infancy to toddlerhood and substantial rank-order stability, suggesting a developmental consolidation of approach tendencies. Schmidt et al. (2025) replicated this pattern between ages 3 and 6 in a large German sample, again showing high rank-order stability and age-related mean-level increases. These findings suggest that Surgency is a relatively stable dimension in relative ranking but shows gradual mean-level increases as children gain autonomy and social experience (Caspi et al., 2005; Caspi & Roberts, 2001).

Cross-sectionally, Heinze et al. (2025) confirmed that higher Surgency was associated with more externalizing symptoms, suggesting that approach-related tendencies foster both engagement and impulsivity. Several longitudinal studies highlight developmental continuity of Surgency in relation to externalizing problems. Wichstrøm et al. (2018) found that higher Surgency at age 4 predicted increases in ADHD and conduct symptoms through childhood. Jonas and Kochanska (2018) showed that toddlers with high Surgency developed more disruptive behavior when EC was low, indicating a regulatory moderation effect. Similarly, Gartstein et al. (2012) reported that early Surgency predicted later problem behavior, particularly when EC was weak. Dollar and Stifter (2012) extended this to social domains: high Surgency combined with poor emotion regulation predicted peer problems. These results illustrate that Surgency shows developmental continuity in its behavioral expression, maintaining approach motivation but manifesting differently depending on regulatory development and environmental context.

Surgency therefore represents a dual-sided developmental disposition that is adaptive when sufficiently regulated but can become maladaptive when environmental demands are high (Beauchaine, 2001; De Pauw & Mervielde, 2010). Taken together, these findings suggest that Surgency shows high rank-order and moderate mean-level stability, while its developmental plasticity determines whether it contributes to adaptive or maladaptive outcomes during development.

Regulatory temperament: effortful control

EC represents the self-regulatory aspect of temperament, the capacity to suppress a dominant response, sustain attention, and plan actions (Rothbart & Rueda, 2005). Conceptually, EC overlaps with executive functions, particularly inhibitory control and attention shifting, integrating affective and cognitive regulation (Heinze et al., 2025; Kälin & Roebers, 2021; Zhou et al., 2012).

Compared with reactive temperament traits, EC shows moderate rank-order stability and pronounced mean-level increases during early childhood. Kopala–Sibley et al., (2018) observed significant mean-level growth in EC from early to middle childhood alongside moderate individual stability, consistent with neurodevelopmental changes in prefrontal control systems (Rothbart et al., 2007; Zelazo, 2020). Schmidt et al. (2025) similarly reported mean-level increases in EC between ages 3 and 6 and

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

partial structural invariance across measurement occasions. These findings indicate that EC develops through maturation and experience while maintaining moderate stability in individual differences (Caspi et al., 2005; Caspi & Roberts, 2001). Eisenberg et al. (2009) found that lower EC predicted higher externalizing symptoms over time, whereas higher EC buffered the effects of NA. Similarly, Sulik et al. (2015) reported that EC moderated the link between emotional reactivity and problem behavior.

Thus, EC constitutes a moderately stable, developmentally responsive construct that underlies self-regulatory maturation and functions as a protective factor, buffering or reversing the adverse effects of reactive temperament (Belsky & Pluess, 2009). While EC highlights the regulatory processes that shape behavioral adjustment, understanding developmental pathways of psychopathology requires integrating reactive and regulatory components within a broader temporal framework.

Integrative perspective on stability and developmental change

The link between temperament and mental health symptoms has been well established at both theoretical and empirical levels. Numerous longitudinal studies demonstrate that early temperament traits predict later psychopathological symptoms, particularly externalizing problems. However, most of this research has focused on relatively stable interindividual differences rather than developmental change. As a result, it remains unclear how within-person fluctuations in temperament, alongside stable between-person tendencies, shape the emergence and persistence of externalizing symptoms. Addressing this gap requires a framework that simultaneously considers stability and development at multiple levels, mean-level, rank-order, and intraindividual, and examines how these processes relate to individual differences in symptom expression.

To this end, the present study investigates developmental changes in reactive temperament, focusing on NA and Surgency between ages 3 and 5, and their joint contribution to externalizing symptoms. EC is examined as a self-regulatory factor that may modify how such changes in reactive temperament translate into behavioral adjustment. This change-oriented approach extends previous trait-based research by integrating interindividual and intraindividual perspectives, thereby capturing both enduring predispositions and dynamic developmental processes. Early childhood represents a particularly sensitive period for this investigation, as regulatory systems are rapidly maturing and behavioral tendencies remain malleable (Diamond, 2013; Rothbart & Bates, 2006). Understanding whether shifts in reactive temperament reflect adaptive adjustment or increasing vulnerability offers deeper insight into how early individual differences and developmental processes jointly construct risk for psychopathology.

Building on this framework, the present preregistered study (osf.io/5hrnu) examines how developmental changes in NA and Surgency between ages 3 and 5 predict externalizing symptoms, considering EC as a potential moderator.

Hypotheses

We hypothesize that changes in NA and surgency will be positively associated with changes in CD and ADHD throughout development, while changes in EC will be negatively associated with changes in CD and ADHD. Additionally, we propose that EC moderates the impact of NA and surgency on externalizing symptoms.

Specifically, we hypothesize that changes in NA will be positively associated with changes in symptoms of CD and ADHD (Wichstrøm et al., 2018). Similarly, we propose that changes in surgency will also be positively associated with changes in symptoms of CD and ADHD (Beauchaine, 2001; De Pauw & Mervielde, 2010; Nigg, 2006; Wichstrøm et al., 2018).

Regarding the moderating effect of EC, we hypothesize that increases in EC will predict a weaker effect of changes in NA on changes in symptoms of CD and ADHD (Gartstein et al., 2012; Santens et al., 2020; Wichstrøm et al., 2018). Similarly, we propose that increases in EC will predict a weaker effect of changes in surgency on changes in symptoms of CD and ADHD (Nigg, 2006; Zelazo, 2020).

Method

Data from the latest release of the National Educational Panel Study (NEPS; Blossfeld et al., 2019) were used for this study. The data were collected by the Institute of Educational Sciences and Longitudinal Research (INBIL) from 2009 to 2013 and by the Leibniz Institute for Educational Pathways (LIfBi) since 2014, both at the Otto-Friedrich University Bamberg, Germany. Details on survey implementation are publicly available on the NEPS website (www.neps-data.de).

Sample

This study uses data from NEPS Start Cohort 1 (Newborn), which provides longitudinal data on temperament and externalizing symptoms in early childhood. A representative sample of children born in Germany between February and July 2012 was drawn according to established sampling procedures (NEPS-Network, 2023). The sample was stratified by community size, and participants were recruited through systematic interval sampling from municipal register data. The study followed participants across survey waves using contact information such as address, telephone number, and email (NEPS-Network, 2023). The present study includes data from waves 4 (age 3 years), 5 (age 4 years), 6 (age 5 years), 7 (age 6 years), and 9 (age 8 years). Interviewers and test coordinators were experienced in working with children and were trained by NEPS staff (NEPS-Network, 2023). Participants received ten euros after each completed survey wave. For further descriptives, see Table 1.

Children's behavior questionnaire (CBQ)

To assess Temperament, the child's caregivers completed a nine item abbreviated form of the Very Short Form of the Children's Behavior Questionnaire (CBQ; Putnam & Rothbart, 2006; 36 items) when children were 3, 4, and 5 years old. The three scales NA, Surgency, and EC each contained three items (NA: "Gets quite frustrated when prevented from doing something s/he wants to do," "Seems to feel depressed when unable to accomplish some task," "Is very difficult to soothe when s/he has become upset"; Surgency: "Often rushes into new situations," "Likes rough and rowdy games," "Is full of energy, even in the evening"; EC: "Sometimes becomes absorbed in a picture book and looks at it for a long time," "When drawing or coloring in a book, shows strong concentration," "Enjoys gentle rhythmic activities such as rocking or swaying"). Caregivers rated behaviors on a 7-point Likert scale (plus a "not applicable" option) from 1 to 7, with higher scores indicating stronger expression of the trait.

McDonald's ω and Cronbach's α for internal consistency were moderate for NA and Surgency ($\omega = .53–.58$; $\alpha = .48–.62$) but low for EC ($\omega = .31–.43$; $\alpha = .26–.38$) across the three waves. These

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

Table 1. Demographic data of the study sample

		Girls		Boys		Age				Language at home		Household income Total per month (€)			Questionnaire	
Age Group	*N*	*n*	%	*n*	%	*M*	*SD*	Min	Max	German	Other	*M*	*Median*	*SD*	CBQ	SDQ
3-years	2,477	1,206	48.7	1,272	51.3	3.22	.09	3.02	3.73	1856	621	3,862.34	3,600.00	1828.69	X	
4-years	2,380	1,176	49.4	1,205	50.6	4.19	.14	3.82	4.60	1786	594	4,056.91	3,900.00	1845.62	X	
5-years	2,208	1,096	49.6	1,113	50.4	5.11	.15	4.74	5.54	1673	535	4,217.03	4,000.00	1883.51	X	X
6-years	2,115	1,048	49.5	1,068	50.5	6.17	.13	5.81	6.57	1648	467	4,521.41	4,200.00	2,111.36		X
8-years	1847	911	49.3	929	50.3	8.25	.13	7.71	8.57	1,468	379	4,955.39	4,600.00	2,239.15		X

Note. SDQ = strengths and difficulties questionnaire; CBQ = children's behavior questionnaire.

values reflect the expected limitations of very short (three-item) scales. Slightly higher but still modest reliability estimates have been reported elsewhere, with α values typically ranging between .62 and .78 in the original validation (Putnam & Rothbart, 2006) and $\alpha = .69–.70$ in a recent German sample (Heinze et al., 2025). These values are consistent with prior research on this version of the CBQ. In short scales, moderate internal consistency can be considered acceptable, as reliability indices such as Cronbach's α are sensitive to the number of items and tend to underestimate true reliability (Clark & Watson, 1995; Schmitt, 1996). Although the CBQ has not yet been formally validated in German, a German version is available on Rothbart's CBQ homepage and has been successfully used in several studies (e.g., Kerner auch Koerner et al., 2018; Licata-Dandel et al., 2021). Because scalar measurement invariance could not be established across time points, EC was excluded from the longitudinal analyses. Even after removal of one item, scalar invariance was not achieved, and the resulting two-item scale lacked conceptual breadth and statistical reliability, leading to non-convergence in the growth models (see Table 3).

Strength and difficulties questionnaire (SDQ)

The parent-rated SDQ is a brief questionnaire designed to assess behavioral problems and strengths in children and adolescents aged 2-16 years (Goodman et al., 1998). Five subscales are derived from the 25 items: Emotional Problems, Hyperactivity/Inattention, CP, Peer Problems, and Prosocial Behavior. In NEPS, the parent version included only the five CP items and five Hyperactivity/Inattention items, excluding the internalizing and prosocial subscales. Parents completed the SDQ when children were 5, 6, and 8 years old.

The CP scale includes the items "tantrums," "obedience," "fights with other children," "argumentative," and "steals." The Hyperactivity/Inattention scale, consisting of "restless," "fidgety," "easily distracted," "thinks," and "attention," captures the three symptom domains of ADHD: hyperactivity, impulsivity, and inattention. Given that it covers the core ADHD domains, this scale is referred to as ADHD symptoms in the present study. Prior research shows strong correlations between the SDQ Hyperactivity/Inattention scale and longer ADHD-specific measures, such as the Conners Rating Scale (Woerner et al., 2004) and the Child Behavior Checklist (Becker et al., 2004).

The SDQ is answered on a three-point scale: Not True, Somewhat True, and Certainly True. Exploratory and confirmatory factor analyses (CFAs) have been used to reproduce the scales in different countries (Kersten et al., 2016). McDonald's ω and Cronbach's α indicated acceptable reliability for ADHD symptoms ($\omega = .72–.73$; $\alpha = .73–.74$) and marginal reliability for CP ($\omega = .53–.54$; $\alpha = .51–.53$). These results are comparable to previous validation studies (Goodman et al., 1998; Goodman, 2011; Stone et al., 2010) and reflect typical reliability estimates for short behavioral symptom subscales. The German version is used in the NEPS study (Klasen et al., 2003).

Data analysis

To examine the factor structure of the CBQ and SDQ, CFAs were performed. Model fit was evaluated using *RMSEA* ($\leq .06$), *CFI* ($\geq .95$), *SRMR* ($\leq .08$; Hu & Bentler, 1999), and for model comparisons using the $RMSEA_D$ (Savalei et al., 2023). Where fit was suboptimal, modification indices were used to introduce within-factor residual correlations until acceptable fit was achieved. Each construct was then tested for longitudinal measurement invariance (configural, metric, scalar, strict). When full invariance could not be achieved, partial invariance was applied by freeing parameters with large modification indices. Final invariance models were fixed before the longitudinal analyses.

Subsequent analyses aimed to capture developmental stability and change across multiple levels. To this end, dual-change score models (DCSMs) were employed to capture individual differences in overall developmental trajectories as well as interval-to-interval dynamics, that is, relations of levels at one measurement occasion to changes at the next (i.e., autoproportional effects). This approach provides a developmentally sensitive alternative to traditional latent growth curve models (LGCMs), which only estimate overarching trajectories rather than dynamic change processes between adjacent assessments (Cáncer et al., 2021; Kievit et al., 2018; McArdle, 2009).

In all models, indicator loadings for latent intercept factors were fixed to 1, and item intercepts were constrained to equality across waves to ensure scalar measurement invariance. Each latent state variable at time t was regressed on its prior state ($t - 1$) with a fixed coefficient of 1. A constant change factor (gX) captured the shared, time-invariant component of developmental change across intervals (e.g., from ages 3 to 5), representing the overall rate and direction of change in the construct. The autoproportion parameter (β) was freely estimated and held equal across intervals. For model identification, residual variances of latent state variables (e.g., SUR_{Age3}, $ADHD_{Age5}$) were fixed to zero, while residuals of observed indicators were allowed to correlate within waves and were constrained to equality across time.

Three parameters describe different but complementary aspects of developmental change. The mean of the constant change factor (gX) represents the model-implied average direction and

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

magnitude of change (e.g., *gADHD* = overall change in ADHD symptoms from ages 3 to 5). A positive mean indicates that, on average, the construct increased over time, whereas a negative mean reflects a general decrease across the sample. The variance of *gX* captures interindividual variability in change, indicating how strongly children differ in their developmental trajectories. A large variance reflects heterogeneous change, whereas a small variance reflects uniform change. Finally, the autoproportion parameter (β) represents the model-implied within-person dynamic, quantifying how prior levels predict subsequent change. A negative β indicates that children with higher scores at one occasion tend to show smaller subsequent increases (or stronger decreases), while those with lower scores show stronger increases over time, reflecting convergence toward the overall group trajectory. Conversely, a positive β indicates that children with higher scores at one occasion tend to show further increases, (or smaller decreases) whereas those with lower scores exhibit weaker increases or further decreases, reflecting divergence and the amplification of individual differences (Grimm et al., 2012; McArdle, 2009).

In addition, the covariance between the intercept and the constant change factor provides a complementary indicator of developmental stability at the interindividual level. Whereas β reflects within-person dynamics (how an individual's prior level predicts their own subsequent change), the intercept – change covariance reflects between-person associations (how individuals who start higher or lower differ in their average rate of change). A negative covariance indicates that children with higher initial levels tend to show smaller average changes, suggesting convergence across individuals, whereas a positive covariance indicates that children with higher initial levels show larger average changes, reflecting divergence in developmental trajectories. The intercept and constant change factors and their covariance complement the model by capturing how stable between-person differences contribute to the overall patterns of change. A negative covariance indicates convergence across individuals, whereas a positive covariance indicates divergence, meaning that children who start higher tend to increase more over time.

In a first step, univariate DCSMs were estimated for each construct (NA, SUR, EC, ADHD, CP) to describe average developmental trends, interindividual variability in change, and proportional dynamics. Because both a constant change factor (*gX*) and an autoproportion parameter (β) were included, these models represent dual-change specifications. Constant change factors are denoted by *g* (e.g., *gNA* = change in NA between adjacent ages).

In a second step, bivariate DCSMs were specified to examine whether temperament predicted externalizing symptoms. In these models, both the baseline level factors of externalizing symptoms (e.g., $ADHD_{Age5}$) and their constant change factors (e.g., *gADHD*) were regressed on the baseline level and constant change factors of temperament (e.g., SUR_{Age3}, *gSUR*). This specification tested developmental predictions of both symptom level and symptom change from initial temperament levels and temperament change. All structural relations were estimated jointly within a single model for each outcome, capturing parallel developmental dynamics across domains and allowing for simultaneous analysis of proportional and constant components of developmental change.

Temperament at age 5 was not included as a separate predictor within the same model because its variance is already decomposed into the baseline level at age 3 and the preceding constant change factor (*gTemperament*). Including age 5 alongside these predictors would therefore introduce redundancy and multicollinearity, as it is not statistically independent of them. For comparison, additional models included either age-3 or age-5 temperament levels as predictors of baseline symptoms and subsequent change.

Concurrent cross-lagged effects could not be estimated because temperament and externalizing symptoms only had one concurrent measurement (temperament: ages 3–5; symptoms: ages 5–9). We therefore modeled temporally ordered associations using DCSMs. To maintain identification, limit the number of free parameters, and preserve statistical power, we estimated separate DCSMs for each pairing (NA → ADHD, NA → CP, SUR → ADHD, SUR → CP). This specification also avoids cross-domain multicollinearity and facilitates transparent interpretation of coupling paths.

Figure 1 illustrates the general structure of the DCSM used in this study. All models were estimated in Mplus 8.4 using the robust full-information maximum-likelihood (FIML) estimator.

Results

Sample characteristics, including measurement time points, sample sizes, and demographic information at each wave, are presented in Table 1. These data provide an overview of the study population and data availability across assessment periods. Descriptive statistics for the main study variables, including latent means, standard deviations, and correlations among temperament and externalizing symptoms, are provided in ESM Table 1.

Construct validity

To assess the construct validity of the SDQ and CBQ, CFAs were conducted, examining the factor structure at each measurement time point. For the CBQ, one- and three-factor models were tested at age 3, 4, and 5, and for the SDQ, one- and two-factor models were tested at age 5, 6, and 8 (see Table 2).

The two-factor model provided a better fit for the SDQ data than the one-factor model, confirming that CP and ADHD are best represented as distinct constructs. Similarly, for the CBQ, the three-factor model separating NA, Surgency, and EC demonstrated superior model fit, supporting the differentiation of temperament dimensions. Overall, the results for both the CBQ and the SDQ indicate that models separating the constructs into distinct, yet correlated, factors have a better fit than single-factor models. This supports the construct validity of NA, Surgency, and EC for the CBQ and of CP and ADHD for the SDQ.

Measurement invariance

Measurement invariance was tested separately for each construct. Configural and metric invariance were supported for all CBQ and SDQ scales, indicating that the same factor structures and loadings were maintained across waves. To achieve scalar invariance, a few intercepts per scale were freed. The resulting partial scalar invariance models showed acceptable to good fit, supporting valid latent mean comparisons over time even though strict equality of all intercepts could not be established (see Table 3).

Strict invariance, which requires residual variances to remain constant across time points, could not be fully achieved. Nevertheless, model fit indices remained acceptable in most cases, suggesting approximate comparability over time. For EC, scalar invariance could not be achieved even after removing one item. The resulting two-item solution lacked conceptual coverage and showed convergence problems in change models; therefore, EC was excluded from further analyses.

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

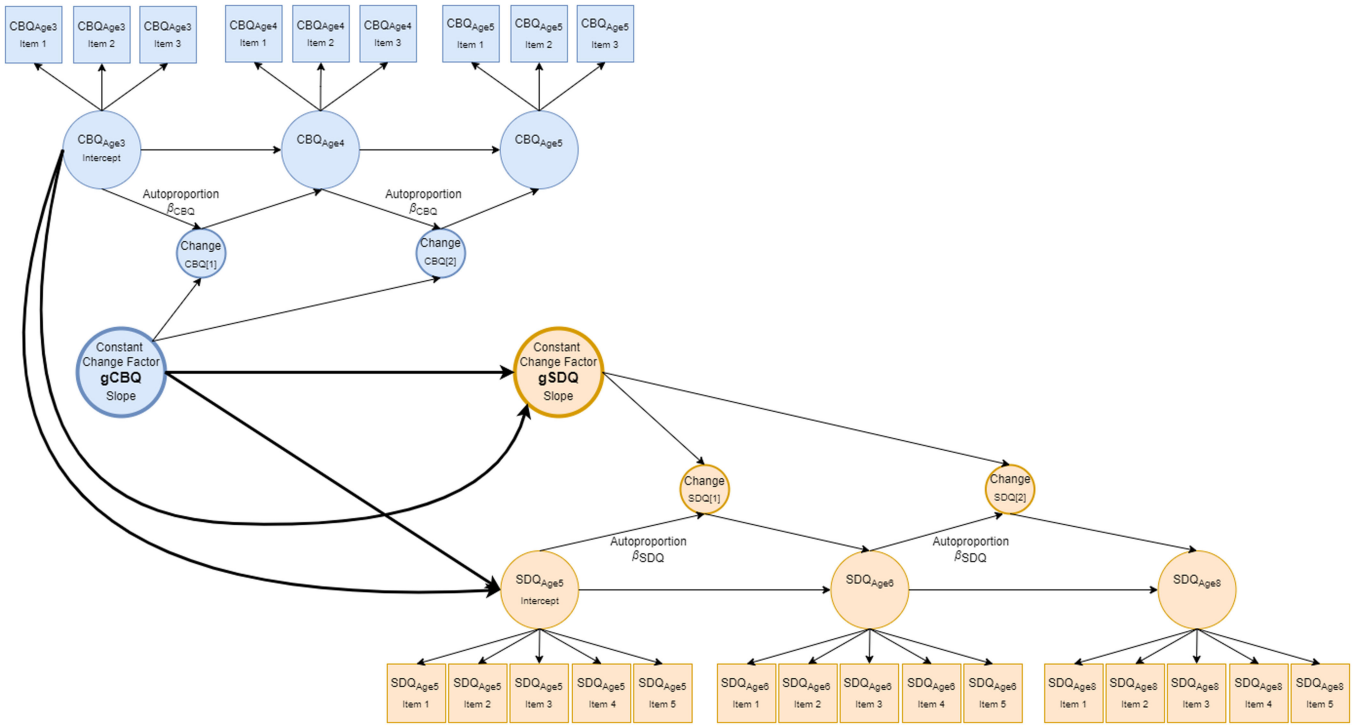


Figure 1. Dual change score model of the development of reactive temperament and externalizing symptoms. *Note.* Squares represent observed indicators and circles represent latent variables; Latent change score factors capture short-term changes between adjacent measurement occasions (*Change*), while the *Constant Change Factor* (*gCBQ*, *gSDQ*) represents the average rate of change across intervals; Autoproportion parameters (β) quantify proportional, level-dependent dynamics and were constrained to equality across intervals to represent constant proportional change over time; The variances of the latent change factors were likewise constrained to equality, and their loadings on the Constant Change Factor were fixed to 1 for model identification; Cross-domain relations (from temperament to symptoms) are shown with thicker arrows to emphasize the primary developmental couplings tested conceptually; CBQ = Children's Behavior Questionnaire; SDQ = Strengths and Difficulties Questionnaire.

Table 2. *Fit indices for confirmatory factor analysis models for SDQ and CBQ*

Age	Model	χ^2 (*df*)	RMSEA	CFI	SRMR	$\Delta\chi^2$	$RMSEA_D$	*ΔCFI*	*ΔSRMR*
SDQ									
5	1-Factor	475.989 (33)	0.080	0.868	0.053	263.089 (1)***	0.345	0.078	0.019
5	2-Factor	212.900 (32)	0.052	0.946	0.034				
6	1-Factor	542.741 (33)	0.088	0.838	0.059	263.348 (1)***	0.352	0.083	0.017
6	2-Factor	279.393 (32)	0.062	0.921	0.042				
8	1-Factor	538.122 (33)	0.091	0.841	0.055	236.817 (1)***	0.357	0.074	0.011
8	2-Factor	301.305 (32)	0.068	0.915	0.044				
CBQ									
3	1-Factor	468.596 (26)	0.083	0.686	0.056	368.138 (3)***	0.222	0.259	0.030
3	3-Factor	100.458 (23)	0.037	0.945	0.026				
4	1-Factor	596.913 (26)	0.101	0.691	0.066	433.345 (3)***	0.255	0.233	0.035
4	3-Factor	163.568 (23)	0.053	0.924	0.031				
5	1-Factor	516.128 (26)	0.092	0.710	0.060	401.064 (3)***	0.268	0.236	0.029
5	3-Factor	115.064 (23)	0.043	0.946	0.031				

Note. SDQ = strengths and difficulties questionnaire; CBQ = children's behavior questionnaire; χ^2 (*df*) = Chi Square Value (degrees of freedom); RMSEA = root mean Square Error of Approximation; $RMSEA_D$ = RMSEA associated with chi-square difference test comparing successive models (Savalei et al., 2023); CFI = comparative fit index; SRMR = standardized root mean square residual; *** $p < .001$, ** $p < .01$, * $p < .05$ (two-tailed).

Analysis of change

To examine intraindividual change in NA, Surgency, ADHD, and CP from age 3 to age 5, DCSMs were applied. As shown in Table 4, all constructs exhibited significant Constant Change Factors (*gX*), indicating that the constructs changed systematically over time rather than remaining static. Variances of the Constant Change

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

Table 3. *Fit indices for measurement invariance analysis for SDQ and CBQ*

Description	Loosened constraints on intercepts	RMSEA	$RMSEA_D$	CFI	ΔCFI	SRMR	ΔSRMR	χ^2 (df)
Configural								
SDQ ADHD	/	0.060	/	0.941	/	0.044	/	649.300 (70)
SDQ CP	/	0.028	/	0.972	/	0.029	/	199.632 (72)
Metric								
SDQ ADHD	/	0.058	0.031	0.939	0.002	0.048	0.004	675.091 (78)
SDQ CP	/	0.027	0.025	0.969	0.003	0.031	0.002	218.787 (80)
Scalar								
SDQ ADHD	[AHfidg9]; [AHdist9];	0.059	0.075	0.931	0.008	0.050	0.002	758.350 (84)
SDQ CP	/	0.040	0.102	0.926	0.043	0.038	0.007	419.212 (88)
Strict								
SDQ ADHD	/	0.069	0.119	0.891	0.040	0.060	0.010	1,165.271 (96)
SDQ CP	/	0.041	0.045	0.916	0.010	0.044	0.006	476.101 (98)
Configural								
CBQ EC	/	0.013	/	0.998	/	0.011	/	21.305 (15)
CBQ NA	/	0.015	/	0.999	/	0.009	/	18.642 (12)
CBQ SUR	/	0.022	/	0.996	/	0.013	/	34.620 (15)
Metric								
CBQ EC	/	0.018	0.032	0.996	0.002	0.016	0.005	34.564 (19)
CBQ NA	/	0.017	0.025	0.997	0.002	0.014	0.005	28.625 (16)
CBQ SUR	/	0.023	0.028	0.995	0.001	0.018	0.005	45.709 (19)
Scalar								
CBQ EC	[EC_conc6]	0.030	0.072	0.986	0.010	0.021	0.005	73.544 (22)
CBQ NA	[NA_fail4]	0.055	0.140	0.967	0.030	0.031	0.017	168.147 (19)
CBQ SUR	/	0.046	0.105	0.976	0.019	0.031	0.013	150.800 (23)
Strict								
CBQ EC	[EC_conc6]	0.071	0.142	0.897	0.089	0.103	0.082	408.033 (29)
CBQ NA	[NA_fail4]	0.054	0.049	0.964	0.003	0.036	0.005	187.944 (22)
CBQ SUR	/	0.057	0.092	0.954	0.022	0.075	0.044	273.541 (29)

Note. SDQ = strengths and difficulties questionnaire; CBQ = children's behavior questionnaire; ADHD = attention deficit hyperactivity disorder; CP = conduct problems; EC = effortful control; NA = negative affectivity; SUR = surgency; RMSEA = root mean square error of approximation; CFI = comparative fit Index; SRMR = standardized root mean square residual; χ^2 = Chi-square statistic; df = degrees of freedom; $RMSEA_D$ = RMSEA associated with chi-square difference test comparing successive models (Savalei et al., 2023); Loosened constraints digit following item name denotes the measurement wave (e.g., NA _fail4 = Wave 4);.

Factors were also significant, showing that the magnitude and direction of change differed meaningfully between children.

Significant and negative autoproportion parameters (β) indicated that higher prior levels were followed by smaller subsequent changes, meaning that children with initially higher scores tended to show less further increase, whereas those with initially lower scores showed stronger increases. In contrast, positive Intercept – Constant Change Factor covariances were observed for all constructs. These covariances refer to between-person differences in average change rather than individual fluctuations. They show that, on average, children who started at higher levels also exhibited greater overall increases across the study period.

Together, these results confirm that all constructs exhibited systematic developmental change and individual variability in change dynamics. Building on these findings, the next analyses examined whether these developmental processes in reactive temperament were associated with subsequent changes in externalizing symptoms.

Associations of reactive temperament and externalizing symptoms

As hypothesized, baseline levels of NA and SUR at age three were not significantly related to later levels or changes in CP or ADHD once intraindividual change was modeled (Table 5). In contrast, increases in both NA and SUR from ages three to five were consistently associated with higher levels of CP and ADHD at age five and with further increases in these symptoms through age eight. All significant coupling parameters were positive, indicating that intraindividual increases in reactive temperament predicted corresponding increases in externalizing behavior over time. This pattern was highly consistent across both temperament traits and

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

Table 4. Analysis of constant change factors and autoproportion effects

				Intercept		Constant Change Factor (gX)		Covariance	
	Baseline χ^2 (*df*)	χ^2(gX) Mean = 0	χ^2(gX) Variance = 0	Mean (*SE*)	Variance (*SE*)	Mean (*SE*)	Variance (*SE*)	Intercept gX	Auto-Proportion β (SE)
ADHD	791.674 (86)	957.633	843.511	0.456 (0.012)	0.217 (0.010)	0.142 (0.010)	0.008 (0.002)	0.032 (0.005)	−0.383 (0.030)
$\Delta\chi^2$ (Δdf)		165.959 (1)***	51.837 (2)***						
CP	419.212 (88)	427.517	438.457	0.321 (0.009)	0.039 (0.004)	0.292 (0.045)	0.034 (0.010)	0.028 (0.005)	−0.959 (0.134)
$\Delta\chi^2$ (Δdf)		8.305 (1)**	19.245 (2)***						
EC	16.683 (9)	64.490	N/A	4.569 (0.027)	0.169 (0.025)	1.050 (0.145)	N/A	N/A	N/A
$\Delta\chi^2$ (Δdf)		47.807 (1)***	N/A						
NA	190.561 (21)	583.844	584.422	4.199 (0.028)	1.047 (0.076)	5.122 (0.136)	1.767 (0.127)	1.116 (0.072)	−1.386 (0.034)
$\Delta\chi^2$ (Δdf)		393.283 (1)***	393.861 (2)***						
SUR	170.216 (25)	434.442	504.389	4.901 (0.026)	0.532 (0.044)	6.525 (0.173)	1.302 (0.121)	0.724 (0.062)	−1.426 (0.037)
$\Delta\chi^2$ (Δdf)		264.226 (1)***	334.173 (2)***						

Note. ADHD = attention-deficit/hyperactivity disorder symptoms; CP = conduct problems; EC = effortful control; NA = negative affectivity; SUR = surgency; χ^2 = chi-square statistic; df = degrees of freedom; Baseline χ^2 (df) refers to the freely estimated dual-change model for each construct; gX Mean = 0 = Constant Change Factor (gX) mean fixed to 0; gX Variance = 0 = Constant Change Factor/Slope variance fixed to 0; $\Delta\chi^2$ = likelihood-ratio χ^2 difference from the baseline model; Δdf = corresponding difference in degrees of freedom (Diff). Means, variances, covariance and β-coefficients are unstandardized; standard errors are in parentheses; *** $p < .001$, ** $p < .01$, * $p < .05$ (two-tailed); All Constant Change Factors (means and variances) and Auto-Proportion parameters (β) were statistically significant ($p < .001$) unless otherwise indicated; Critical χ^2 values for $p < .001$: Δdf = 1 → 10.83, 2 → 13.82, 3 → 16.27, 4 → 18.47, 5 → 20.52.

both externalizing domains. The corresponding structural diagrams for all four dual-change models (NA–ADHD, SUR–ADHD, NA–CP, and SUR–CP) are presented in ESM Figures 1a–1d.

Observed trajectories (Figure 2) illustrate this pattern descriptively. Mean increases in ADHD and CP symptoms were marginal, while level differences between tertiles of NA and SUR were visible across all time points. Children in the high-change tertiles consistently showed higher average symptom levels than those in the middle or low tertiles, but temporal increases within each group remained small, suggesting that variability was primarily between individuals rather than within individuals.

Model-implied trajectories (predicted symptom paths generated from the latent change model, rather than observed scores; Figure 3) illustrate the modeled pattern more clearly. Starting from a shared baseline, the predicted curves converge over time, consistent with negative auto-proportion parameters (Table 4). These parameters reflect within-person dynamics, meaning that individual deviations from the mean tend to diminish across time. In contrast, the Constant Change Factors (*gNA, gSUR*) represent stable between-person differences: children who showed stronger increases in NA or SUR also showed higher predicted ADHD and CP levels across time. Together, these parameters characterize the modeled developmental dynamics, indicating that intraindividual change in temperament tends to co-occur with changes in externalizing symptoms, while between-person variability is modeled as gradually stabilizing.

To complement these primary models, post hoc analyses using temperament levels at ages three and five as predictors of concurrent and later symptoms (ESM Tables 2 and 3) showed that both age-specific temperament levels were associated with symptom levels measured at the same or later time points. However, when latent change was included, these associations were no longer significant. This indicates that predictive effects of temperament were largely explained by intraindividual change rather than by static differences in baseline levels.

Discussion

This study examined how changes in reactive temperament contribute to the development of externalizing symptoms in early childhood. The results showed that early temperament and externalizing symptoms are not fixed traits but evolve dynamically during the preschool years. In particular, increases in NA and Surgency between ages 3 and 5 were closely tied to the emergence and growth of CP and ADHD symptoms, whereas baseline levels were far less predictive. These findings indicate that developmental change, rather than early trait levels alone, is a more sensitive marker of emerging risk.

The analyses further revealed that developmental processes operate at two complementary levels. The Constant Change Factors (*gNA, gSUR, gADHD, gCP*) captured intraindividual change and its covariation across domains, whereas the autoproportion parameters (β) described proportional dynamics within each construct. The positive couplings between *gNA/gSUR* and *gADHD/gCP* indicate that children who increased more strongly in reactive temperament also showed stronger increases in externalizing symptoms. The negative β-coefficients, in contrast, represent modeled convergence within constructs, indicating that individual deviations from the mean are estimated to diminish across time. Together, these parameters illustrate that modeled change in temperament is associated with modeled change in symptoms, while overall trajectories are modeled as converging at the group level, a pattern consistent with theoretical accounts emphasizing preschool as a period of both plasticity and emerging self-regulation.

Construct validity

CFA confirmed the factor structure of temperament (three factors: NA, Surgency, EC) and externalizing symptoms (two factors: CP, ADHD) that have been found in previous studies (temperament with CBQ: e.g., Putnam & Rothbart, 2006; Sleddens et al., 2011, 2012; de la Osa et al., 2014; externalizing symptoms with SDQ: Goodman et al., 1998; Kiel et al., 2018; Stone et al., 2010, 2015; see

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

Table 5. Dual change score model: associations of reactive temperament (age 3 and change) and problem behavior (age 5 and change)

Relationship	Estimate	*SE (Est./SE)*	*p*-value
NA and CP			
$NA_{Age3} \rightarrow CP_{Age5}$	−0.121	0.080 (−1.512)	0.130
gNA → CP_{Age5}	0.873	0.075 (11.601)	0.000
NA_{Age3} → gCP	−0.130	0.075 (−1.743)	0.081
gNA → gCP	0.723	0.075 (9.644)	0.000
Goodness-of-fit:	χ^2 (244) = 1,289.949, RMSEA = 0.040, CFI = 0.895, SRMR = 0.060 $R^2_{CP\ AGE\ 5/gCP}$ = 0.610/0.392		
SUR and CP			
$SUR_{Age3} \rightarrow CP_{Age5}$	−0.035	0.121 (−0.293)	0.769
gSUR → CP_{Age5}	0.465	0.116 (4.014)	0.000
SUR_{Age3} → gCP	−0.094	0.115 (−0.819)	0.413
gSUR → gCP	0.435	0.111 (3.916)	0.000
Goodness-of-fit:	χ^2 (244) = 768.369, RMSEA = 0.029, CFI = 0.947, SRMR = 0.035 $R^2_{CP\ AGE\ 5/gCP}$ = 0.189/0.127		
NA and ADHD			
$NA_{Age3} \rightarrow ADHD_{Age5}$	−0.050	0.077 (−0.646)	0.518
gNA → $ADHD_{Age5}$	0.534	0.073 (7.281)	0.000
NA_{Age3} → gADHD	−0.218	0.146 (−1.492)	0.136
gNA → gADHD	0.715	0.148 (4.832)	0.000
Goodness-of-fit:	χ^2 (238) = 1,214.633, RMSEA = 0.039, CFI = 0.934, SRMR = 0.044 $R^2_{ADHD\ AGE\ 5/gADHD}$ = 0.243/0.300		
SUR and ADHD			
$SUR_{Age3} \rightarrow ADHD_{Age5}$	−0.057	0.100 (−0.573)	0.566
gSUR → $ADHD_{Age5}$	0.671	0.095 (7.089)	0.000
SUR_{Age3} → gADHD	−0.148	−0.148 (0.188)	0.432
gSUR → gADHD	0.662	0.187 (3.544)	0.000
Goodness-of-fit:	χ^2 (242) = 1,145.547, RMSEA = 0.038, CFI = 0.942, SRMR = 0.044 $R^2_{ADHD\ AGE\ 5/gADHD}$ = 0.387/0.290		

Note. ADHD = attention deficit hyperactivity disorder; CP = conduct problems; NA = negative affect; SUR = surgency; gX = constant change factor/slope; RMSEA = root mean square error of approximation; CFI = comparative fit index; SRMR = standardized root mean square residual; χ^2 (*df*) = Chi-square statistic (degrees of freedom); R^2 = coefficient of determination.

Table 2). Measurement invariance was adequate for all constructs except EC (see Table 2). Because EC did not meet invariance requirements, hypotheses involving this construct were not tested. The lack of invariance suggests that observed score differences over time may partly reflect changes in measurement properties or informant perceptions rather than true developmental change.

Developmental change in temperament and externalizing symptoms

Our findings underscore that early temperament is shaped by development rather than reflecting fixed traits. Significant variability in NA and Surgency trajectories challenges static trait-based conceptions of early temperament. The observed heterogeneity in trajectories, reflected in significant variance of the Constant Change Factors, indicates that children do not follow a uniform developmental path. Instead, temperament unfolds in diverse directions, likely shaped by maturational and contextual influences (Beauchaine & Cicchetti, 2019). NA and Surgency exhibited substantial heterogeneity, with some children intensifying, others declining, and some remaining stable. This aligns with Caspi et al. (2005), who noted that rank-order stability can coexist with substantial intraindividual change. Strong autoproportion effects (β) suggested that higher initial levels were followed by smaller gains, while lower initial levels often preceded greater increases, patterns consistent with normative regulatory processes (see ESM Figure 4). In the context of our models, the mean of the Constant Change Factor corresponds to mean-level change at the group level, whereas the autoproportion parameter reflects proportional or within-person change and the intercept – change covariance captures how between-person differences at baseline relate to between-person differences in average change. Taken together, these results indicate that NA and Surgency are not static traits but evolving dispositions shaped by both interindividual variability and intraindividual dynamics. Additionally, Dyson et al. (2015) emphasize heterotypic continuity, meaning that different behaviors may reflect the same

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

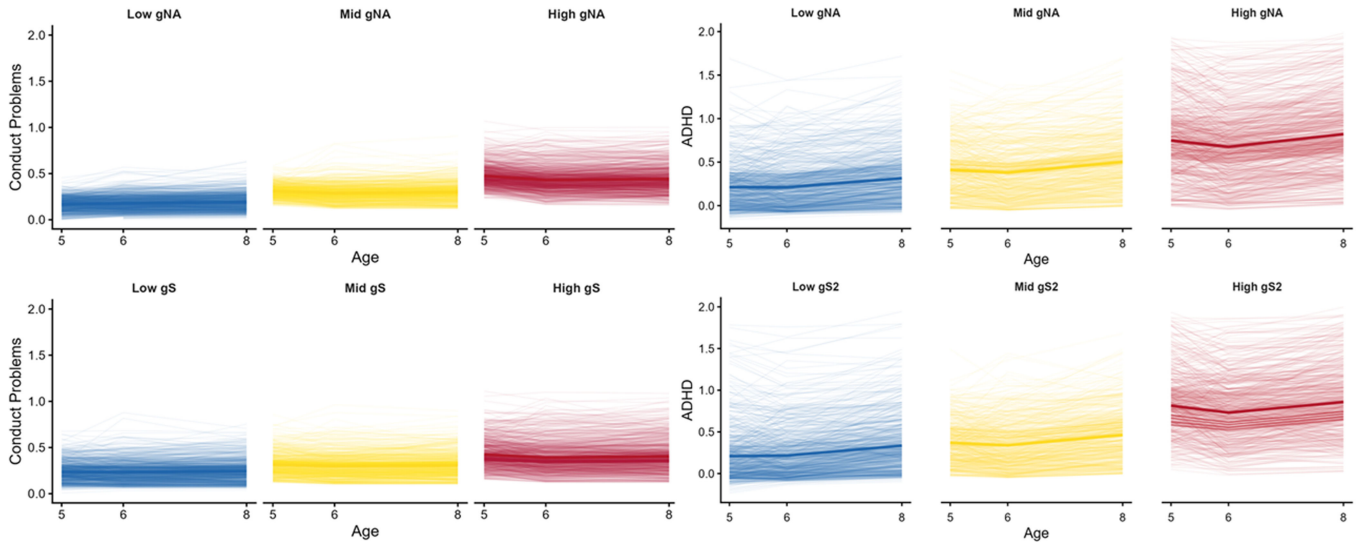


Figure 2. Observed trajectories of conduct problems and ADHD. Symptoms by change in reactive temperament. *Note.* Observed trajectories of Conduct Problems (left) and ADHD symptoms (right) are shown across three measurement points, separated by tertiles of the Constant Change Factor for Negative Affect (*gNA*; top) and Surgency (*gS/gS*$_2$; bottom); Thin lines represent individual observed values; bold lines indicate mean trajectories within each tertile; Color gradients (blue, yellow, red) correspond to low, medium, and high levels of the Constant Change Factor; Trajectories were generated from observed SDQ scores without wavewise standardization; Grouping was based on tertiles of latent Constant Change Factor scores estimated in the latent change models; For each group, individual trajectories were plotted as semi-transparent lines, and group mean trajectories were overlaid using smoothed line estimates; Colors were harmonized across panels for comparability.

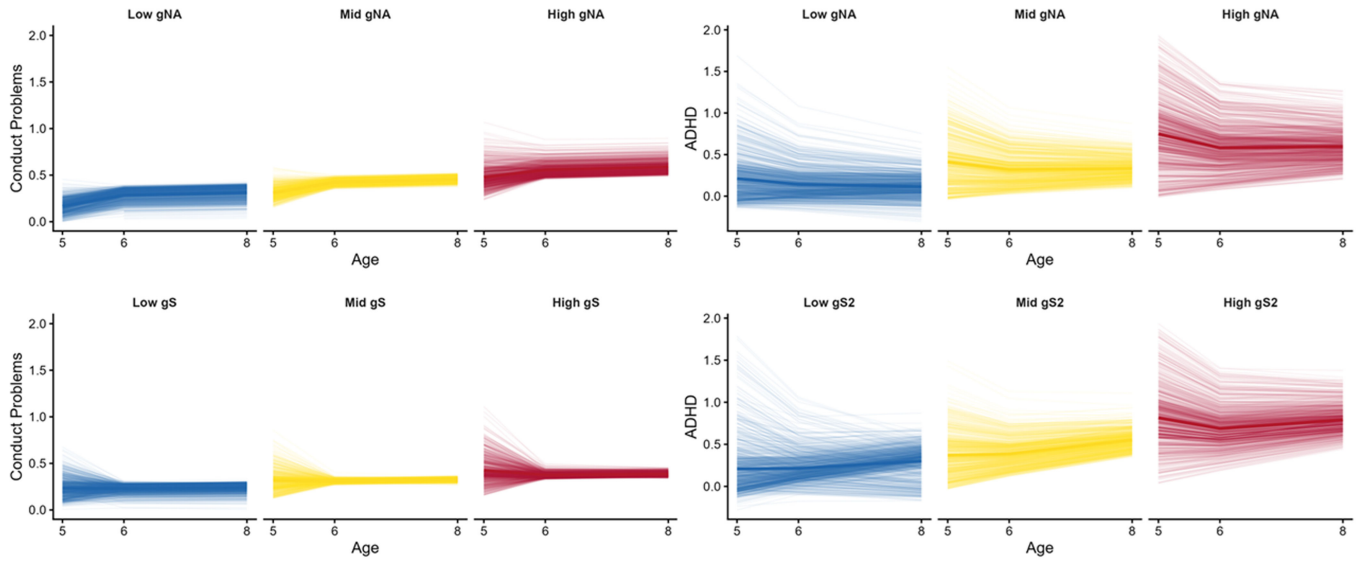


Figure 3. Model-implied trajectories of conduct problems and ADHD symptoms by change in reactive temperament. *Note. Model-implied trajectories of Conduct Problems (left) and ADHD symptoms (right) are shown across three measurement points, separated by tertiles of the Constant Change Factor for Negative Affect (gNA; top) and Surgency (gS/gS*$_2$*; bottom); Thin lines represent individual model-implied trajectories derived from the latent change score model; bold lines indicate mean trajectories within each tertile; Color gradients (blue, yellow, red) correspond to low, medium, and high levels of the Constant Change Factor; Predicted trajectories were computed using unstandardized Mplus parameters obtained from the latent change score model; For each individual, latent estimates of initial status and change (Constant Change Factor and proportional change parameters) were used to calculate model-implied ADHD and CP values at ages 5, 6, and 8; The resulting trajectories were plotted using the same tertile grouping, color scheme, and scaling as the observed data to ensure comparability between observed and model-implied figures.*

underlying trait at different ages. This is particularly relevant to our reliance on parent-reported CBQ data, where observed changes may reflect both genuine developmental shifts and heterotypic manifestations of temperament. Integrating this perspective helps contextualize our findings: increases in NA and Surgency could partly signal maturational changes in how these traits are expressed and perceived during the preschool years, in line with the broader literature on multiple types of stability in early childhood.

Externalizing symptoms also showed dynamic properties. Initial ADHD and CP levels predicted subsequent change, but children with higher initial scores tended to show smaller increases or even reductions, while those with lower initial scores often increased more. This reflects self-correcting tendencies and supports developmental cascade models in which early symptoms serve as starting points rather than fixed pathways (Diamond, 2013; Frick & Viding, 2009; Van Meter et al., 2024). These results

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

suggest that externalizing problems in early childhood can shift in intensity and expression as children mature. For example, elevated ADHD or CP at preschool age may decline if regulatory capacities strengthen or if environmental supports are introduced, whereas initially low levels can intensify under adverse conditions. This pattern is consistent with maturational processes and the emergence of regulatory capacities (Ágrez et al., 2025; Diamond, 2013). Rather than reflecting stable, trait-like vulnerabilities, externalizing symptoms appear embedded in developmental systems that can both escalate and attenuate over time depending on individual and contextual conditions (Van Meter et al., 2024). The pattern also supports models suggesting that early symptoms of CP can initiate but not determine developmental trajectories, which remain malleable in the presence of protective or exacerbating influences (Blair et al., 2018, 2020; Frick & Viding, 2009).

When temperament and externalizing symptoms are considered together, the findings suggest parallel developmental dynamics. Both NA/Surgency and CP/ADHD showed heterogeneity in trajectories and evidence of autoregressive or self-correcting processes. This indicates that reactive temperament and externalizing behaviors may co-develop in ways that reinforce each other: shifts in NA or Surgency can set the stage for symptom escalation, while early symptom levels may in turn shape subsequent temperament change. Such reciprocal influences support transactional and cascade models of development, emphasizing that vulnerabilities are constructed through ongoing interactions rather than fixed traits. This interpretation is consistent with Beauchaine and McNulty's (2013) ontogenic model, which highlights how comorbidities and continuities emerge through reciprocal developmental processes that accumulate over time. It also resonates with developmental cascade perspectives (Masten et al., 2005), in which functioning in one domain of behavior spreads to other domains in lasting ways. Strong autoproportion effects emerged for NA and Surgency, clearly surpassing those observed for ADHD and CP. Such patterns may indicate normative regulatory processes, including maturational shifts and emerging self-regulation (Perry et al., 2018). Importantly, the variance in Constant Change Factors demonstrates that children differ meaningfully in their developmental trajectories, significant mean changes indicate group-level trends, autoproportion parameters may capture self-correcting tendencies within individuals, and intercept–change covariances show how starting levels shape subsequent growth. Together, these parameters provide quantitative evidence for the developmental trajectories of stability and change observed between temperament and externalizing symptoms.

Associations of reactive temperament and externalizing symptoms

When modeled jointly with developmental change, baseline levels of NA and Surgency at age 3 were no longer associated with CP or ADHD outcomes at age 5 or with subsequent changes through age 8. In contrast, increases in NA and Surgency between ages 3 and 5 were consistently related to higher CP and ADHD symptom levels at age 5 and to further symptom growth thereafter (see Table 5, ESM Figures 1a–1d). These findings suggest that intraindividual change in reactive temperament provides additional information beyond early trait levels. Similar patterns were reported by Wichstrøm et al. (2018), who demonstrated that within-person increases in negative affectivity and surgency predicted subsequent growth in externalizing problems, and by Nielsen et al. (2019), who found that changes in temperament across preschool years, assessed using a performance-based temperament task, were prospectively associated with externalizing trajectories. Recent work also links temperament-based pathways to heterogeneity in ADHD (Karalunas & Nigg, 2020; Nigg, 2022), underscoring the developmental significance of reactive traits.

Concurrent links between temperament and behavior problems are well established across cross-sectional and longitudinal research (Dougherty et al., 2010; Martel & Nigg, 2006; Nielsen et al., 2019; Oldehinkel et al., 2004; Wichstrøm et al., 2018). It is therefore plausible that same-wave assessment of age-5 temperament and symptoms contributed to the strength of concurrent effects. Our modeling approach extends this work by explicitly estimating intraindividual change, thereby distinguishing developmental variation from time-specific covariance and providing a developmentally sensitive perspective on early risk.

Together with the model-implied trajectories showing that increases in NA and Surgency predicted higher ADHD and CP symptom levels over time (see ESM Figures 2–3 and Figures 2 and 3), these findings indicate that developmental change in reactive temperament captures meaningful within-person variation linked to emerging externalizing risk rather than reflecting stable between-person differences. Figure 3 illustrates that children with stronger increases in reactive temperament (red tertile) showed higher symptom levels and greater variability across waves, whereas the lower-change groups (blue tertile) remained comparatively stable over the same developmental window. The results are consistent with cascade and transactional perspectives, which propose that variation in one domain, in this case reactive temperament, can influence other domains of behavior and thereby shape the development of externalizing problems (Moilanen et al., 2010; Nigg, 2006). Moreover, the results align with dimensional approaches to psychopathology that conceptualize early behavioral tendencies such as irritability and high approach as continuous with later externalizing risk (Wakschlag et al., 2015, 2019). Examining intraindividual change therefore complements static assessments of temperament by providing a developmentally sensitive means of identifying children at elevated risk for persistent behavioral problems. This perspective highlights the value of studying temperament not only as a set of early traits but as an evolving aspect of socioemotional development (Diamond, 2013; Zelazo & Carlson, 2012). The convergence of findings from our analyses and prior longitudinal studies underscores that early temperament change is a sensitive indicator of developmental risk, aligning with dimensional models of psychopathology and highlighting opportunities for early, context-sensitive interventions.

Limitations

Although this study has several strengths, there are also limitations that should be acknowledged. First, the EC construct posed notable measurement challenges: one item had to be removed and measurement invariance was not met across waves (Widaman et al., 2010). These issues highlight difficulties in reliably assessing this dimension and precluded analyses of change, associations, or moderation effects involving EC. These concerns are consistent with our construct validity analyses, which already indicated instability in the EC factor.

Second, the CBQ used in the panel study includes only three items per subscale, limiting content coverage and leading to relatively low internal consistency, particularly for EC. This raises broader concerns about the reliability and content validity of

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

abbreviated measures (Aiken, 1980). Nevertheless, measurement invariance across waves was achieved for NA and Surgency, suggesting that despite their relatively low reliability coefficients, these constructs demonstrated sufficient temporal factor stability in measurement properties. Although our findings converge with prior research, the abbreviated short scales reduce comprehensiveness. Moreover, Surgency and NA were not modeled together, preventing tests of interactions between these traits. Future studies should employ longer and more comprehensive instruments to capture temperament dimensions and their interactions.

Third, all measures of temperament and externalizing symptoms relied exclusively on parent ratings. Shared method variance and rater biases may have inflated associations, and parental perceptions likely shift with children's age, contributing to apparent changes in NA and EC (De Los Reyes et al., 2015, 2023). Future research should integrate multimethod approaches, including observational and performance-based tasks, and multiple informants to better capture children's behaviors across contexts.

Fourth, the timing of assessments is a limitation. Temperament was assessed at ages 3 and 5, whereas externalizing symptoms were assessed at ages 5 and 9. This non-simultaneous design complicates the interpretation of whether associations reflect concurrent developmental processes or lagged effects. Although additional models with age-5 temperament provided some insight, true synchrony could not be tested.

Fifth, our models did not include covariates such as socioeconomic status, sex, parenting, parental psychopathology, or time-varying influences such as negative life events. Environmental and contextual factors are well-documented predictors of both temperament and externalizing outcomes (Belsky & Pluess, 2009; Sameroff, 2010), and their omission may have limited the specificity of our findings.

Sixth, although the latent change score approach offers advantages for modeling dynamic processes, it also entails assumptions. While measurement error is partly accounted for through the use of latent variables, and regression-to-the-mean effects are formally modeled via the autoproportion parameter rather than treated as statistical artifacts, unmeasured third-variable influences cannot be excluded (e.g. Grimm et al., 2012). Both Grimm et al. (2012) and Kievit et al. (2018) emphasize that the timing of assessments is critical: if lags are not developmentally appropriate, conclusions about change processes may be distorted. Moreover, as illustrated by the comparison between observed and model-implied trajectories (Figures 2 and 3), even well-fitting models can smooth over nonlinearities or individual differences in the empirical data. Consequently, predictions involving externalizing symptom change should be interpreted with some caution, as model-imposed structure may not fully capture the underlying developmental dynamics.

Finally, our sample was drawn from a population-based cohort, supporting generalizability to community samples. However, findings may not extend to clinical populations with more severe or comorbid presentations. The absence of diagnostic data and reliance on dimensional measures may underestimate associations at the extreme end of symptom severity, particularly given high comorbidity rates in clinical populations (e.g. Angold et al., 1999). Moreover, although our sample was population-based, it still constitutes a form of convenience sampling. As noted by Jager and colleagues (2017), homogeneous convenience samples are common in developmental research and may offer clearer generalizability than heterogeneous ones, yet they still fall short of probability samples. These considerations highlight the importance of replicating findings across both community and clinical samples. Together, these limitations underscore the need for future studies to employ multi-informant, multimethod assessments, ensure stronger measurement properties, incorporate key covariates and time-varying influences, and replicate findings across diverse contexts.

Implications

For research, our results demonstrate that intraindividual changes in reactive temperament, not only early trait levels, are important predictors of externalizing symptoms. This shows that questionnaire-based assessments capture not only dispositional reactivity but also its developmental unfolding in interaction with maturation and environmental input. Such findings further shift the perspective from viewing temperament as a fixed trait to considering it as a dynamic process. This implies that future research should focus explicitly on developmental change as a key construct, not just as error variance around stable traits. Building on this, dismantling studies are needed to identify the modifiable mechanisms of change, such as parenting behaviors, peer dynamics, and executive function development (Eisenberg et al., 2009; Zelazo, 2020). Multimethod and context-sensitive assessments, including observational paradigms, EMA, and physiological measures (e.g. Gagne et al., 2011; Gartstein & Rothbart, 2003; Putnam et al., 2008; Wass & Goupil, 2022), will be of importance to disentangle biological predispositions from experience-driven adaptations. Furthermore, because the predictive patterns were consistent across CP and ADHD, our results support transdiagnostic perspectives (e.g. Beauchaine & Cicchetti, 2019; Cuthbert & Insel, 2013; Kotov et al., 2021; Lahey et al., 2005, 2022).

For practice, our results demonstrate that reactive traits do not represent a static "risk temperament" but function as indicators of susceptibility. Highly reactive children showed elevated risk when their reactivity increased over time, which highlights plasticity rather than fixed vulnerability. This means that such traits should be understood as markers of sensitivity to context, consistent with the differential susceptibility model (Belsky & Pluess, 2009). The practical implication is that interventions should not treat temperament as destiny but as potential with space for flourishing, catch-up-effects etc., depending on the goodness-of-fit between child and environment (Thomas & Chess, 1977), parenting (Bornstein et al., 2018; Kiff et al., 2011) and tailored, reasonable interventions (Colizzi et al., 2020; Shah et al., 2022). Particularly in preschool, when children are most malleable, professionals should recognize that the same reactivity that confers vulnerability can also be harnessed for positive development. Supporting parents and educators in providing structured, responsive, and supportive environments can therefore redirect trajectories toward adaptive outcomes (Cicchetti, 2016; Masten et al., 2005; Sameroff, 2010).

Theoretically, our results demonstrate that reactive temperament and externalizing symptoms development are associated. These findings align with transactional and cascade models of development (Masten et al., 2005; Moilanen et al., 2010) and contribute empirical evidence that reinforces dimensional frameworks such as HiTOP (Kotov et al., 2021) and RDoC (Cuthbert & Insel, 2013; Ostlund et al., 2021). The implication is that models of psychopathology should integrate reactivity and regulation within

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

unified frameworks (Nigg, 2006; Santens et al., 2020; Zelazo, 2020) as we originally intended with the integration of the EC factor into our analyses. Importantly, the evidence that ages 3 to 5 are a sensitive window for developmental change means that early interventions timed to these shifts may be especially effective. Programs that target self-regulation, socio-emotional learning, or parenting during periods of heightened reactivity may not only prevent maladaptive outcomes but also leverage plasticity to promote resilience and thriving.

Conclusion

This study highlights the importance of a developmental perspective on temperament and its role in predicting emotional and behavioral difficulties. Our findings demonstrate that individual changes in NA and Surgency during early childhood are more informative than early trait levels alone when predicting later externalizing symptoms. This highlights the relevance of temperament as a dynamic, evolving construct rather than a static trait. The results suggest that early shifts in temperament may offer a more actionable window for prevention and intervention than static, cross-sectional assessments. Recognizing temperament change as a meaningful indicator of developmental trajectories can help refine early identification strategies and support individualized, adaptive responses to emerging mental health risks.

Future research should build on these insights by integrating multimethod assessments and exploring the mechanisms that drive temperament change, including environmental, relational, and maturational factors. Such work will be essential in shaping effective developmental models and translating findings into timely support strategies for children at risk.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0954579426101278.

Data availability statement. The study follows Transparency and Openness Promotion (TOP) guidelines as far as permitted by legal and ethical constraints.

Acknowledgements. This research was made possible through institutional support provided by the participating universities. Data used in this study were accessed via the NEPS (National Educational Panel Study) network. We thank the Leibniz Institute for Educational Trajectories for providing access to the data.

Funding statement. This research was supported by institutional funding from the participating universities.

Competing interests. The authors declare no competing interests.

Pre-registration statement. This study was preregistered prior to data analysis.

Active link. https://osf.io/5hrnu.

Deviations. Due to a lack of longitudinal measurement invariance, Effortful Control could not be modeled consistently across measurement occasions. As a result, the preregistered moderation hypotheses involving Effortful Control could not be tested. This deviation from the preregistered analysis plan is fully reported and discussed in the manuscript.

Availability of data. The data used in this study are from the National Educational Panel Study (NEPS), Start Cohort 1 (Newborns). Due to legal and ethical restrictions, the data cannot be shared publicly. Access to the data can be requested via the NEPS Network (https://www.neps-data.de).

Availability of code. All analysis code and the preregistration are publicly available on the Open Science Framework at https://osf.io/5hrnu.

Availability of methods/materials. All methods and materials are described in detail in the manuscript. Supplementary materials are available from the corresponding author upon reasonable request.

AI statement. No AI tools were used in the preparation of this manuscript beyond basic spelling and grammar checks (DEEPL).

References

Ágrez, K., Vakli, P., Weiss, B., Vidnyánszky, Z., & Bunford, N. (2025). Assessing the association between ADHD and brain maturation in late childhood and emotion regulation in early adolescence. *Translational Psychiatry*, *15*(1), 185. https://doi.org/10.1038/s41398-025-03411-6

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

American Psychiatric Association (Hrsg.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). American Psychiatric Association.

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry*, *40*(1), 57–87. https://doi.org/10.1111/1469-7610.00424

Beauchaine, T. (2001). Vagal tone, development, and gray's motivational theory: Toward an integratedmodel of autonomic nervous system functioning in psychopathology. *Development and Psychopathology*, *13*(2), 183–214. https://doi.org/10.1017/S0954579401002012

Beauchaine, T. P., & Cicchetti, D. (2019). Emotion dysregulation and emerging psychopathology: A transdiagnostic, transdisciplinary perspective. *Development and Psychopathology*, *31*(3), 799–804. https://doi.org/10.1017/S0954579419000671

Beauchaine, T. P., & McNulty, T. (2013). Comorbidities and continuities as ontogenic processes: Toward a developmental spectrum model of externalizing psychopathology. *Development and Psychopathology*, *25*(4pt2), 1505–1528. https://doi.org/10.1017/S0954579413000746

Becker, A., Woerner, W., Hasselhorn, M., Banaschewski, T., & Rothenberger, A. (2004). Validation of the parent and teacher SDQ in a clinical sample. *European Child & Adolescent Psychiatry*, *13*(Suppl 2), ii11–ii16. https://doi.org/10.1007/s00787-004-2003-5

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, *135*(6), 885–908. https://doi.org/10.1037/a0017376

Blair, R. J. R., Bashford-Largo, J., Zhang, R., Lukoff, J., Elowsky, J. S., Leibenluft, E., Hwang, S., Dobbertin, M., & Blair, K. S. (2020). Temporal discounting impulsivity and its association with conduct disorder and irritability. *Journal of Child and Adolescent Psychopharmacology*, *30*(9), 542–548. https://doi.org/10.1089/cap.2020.0001

Blair, R. J. R., Veroude, K., & Buitelaar, J. K. (2018). Neuro-cognitive system dysfunction and symptom sets: A review of fMRI studies in youth with conduct problems. *Neuroscience & Biobehavioral Reviews*, *91*, 69–90. https://doi.org/10.1016/j.neubiorev.2016.10.022

Blossfeld, H. P., Roßbach, H. G., & von Maurice, J. (Hrsg.) (2019). *Education as a lifelong process: The german national educational panel study (NEPS)* (Bd. 3). Springer Fachmedien Wiesbaden.

Bornstein, M. H., Putnick, D. L., Gartstein, M. A., Hahn, C., Auestad, N., & O'Connor, D. L. (2015). Infant temperament: Stability by age, gender, birth order, term status, and socioeconomic status. *Child Development*, *86*(3), 844–863. https://doi.org/10.1111/cdev.12367

Bornstein, M. H., Putnick, D. L., & Suwalsky, J. T. D. (2018). Parenting cognitions → parenting practices → child adjustment? The standard model. *Development and Psychopathology*, *30*(2), 399–416. https://doi.org/10.1017/S0954579417000931

Brennan, L. M., Shaw, D. S., Dishion, T. J., & Wilson, M. (2012). Longitudinal predictors of school-age academic achievement: Unique contributions of toddler-age aggression, oppositionality, inattention, and hyperactivity. *Journal of Abnormal Child Psychology*, *40*(8), 1289–1300. https://doi.org/10.1007/s10802-012-9639-2

Breslau, J., Miller, E., Breslau, N., Bohnert, K., Lucia, V., & Schweitzer, J. (2009). The impact of early behavior disturbances on academic achievement

in high school. *Pediatrics*, *123*(6), 1472–1476. https://doi.org/10.1542/peds.2008-1406

Cáncer, P. F., Estrada, E., Ollero, M. J. F., & Ferrer, E. (2021). Dynamical properties and conceptual interpretation of latent change score models. *Frontiers in Psychology*, *12*, 696419. https://doi.org/10.3389/fpsyg.2021.696419

Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, *12*(2), 49–66. https://doi.org/10.1207/S15327965PLI1202_01

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*(1), 453–484. https://doi.org/10.1146/annurev.psych.55.090902.141913

Cicchetti, D. (2016). Socioemotional, personality, and biological development: Illustrations from a multilevel developmental psychopathology perspective on child maltreatment. *Annual Review of Psychology*, *67*(1), 187–211. https://doi.org/10.1146/annurev-psych-122414-033259

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Coghill, D., & Sonuga-Barke, E. J. S. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders - implications of recent empirical study: Categories and dimensions. *Journal of Child Psychology and Psychiatry*, *53*(5), 469–489. https://doi.org/10.1111/j.1469-7610.2011.02511.x

Colizzi, M., Lasalvia, A., & Ruggeri, M. (2020). Prevention and early intervention in youth mental health: Is it time for a multidisciplinary and trans-diagnostic model for care? *International Journal of Mental Health Systems*, *14*(1), 23. https://doi.org/10.1186/s13033-020-00356-9

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. https://doi.org/10.1186/1741-7015-11-126

de la Osa, N., Granero, R., Penelo, E., Domènech, J. M., & Ezpeleta, L. (2014). The short and very short forms of the children's behavior questionnaire in a community sample of preschoolers. *Assessment*, *21*(4), 463–476. https://doi.org/10.1177/1073191113508809

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858–900. https://doi.org/10.1037/a0038498

De Los Reyes, A., Epkins, C. C., Asmundson, G. J. G., Augenstein, T. M., Becker, K. D., Becker, S. P., Bonadio, F. T., Borelli, J. L., Boyd, R. C., Bradshaw, C. P., Burns, G. L., Casale, G., Causadias, J. M., Cha, C. B., Chorpita, B. F., Cohen, J. R., Comer, J. S., Crowell, S. E., Dirks, M. A. . . . Youngstrom, E. A., (2023). Editorial Statement About *JCCAP* 's 2023 Special Issue on Informant Discrepancies in Youth Mental Health Assessments: Observations, Guidelines, and Future Directions Grounded in 60 Years of Research. *Journal of Clinical Child & Adolescent Psychology*, *52*(1), 147–158. https://doi.org/10.1080/15374416.2022.2158842

De Pauw, S. S. W., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry & Human Development*, *41*(3), 313–329. https://doi.org/10.1007/s10578-009-0171-8

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

Dollar, J. M., & Stifter, C. A. (2012). Temperamental surgency and emotion regulation as predictors of childhood social competence. *Journal of Experimental Child Psychology*, *112*(2), 178–194. https://doi.org/10.1016/j.jecp.2012.02.004

Dougherty, L. R., Klein, D. N., Durbin, C. E., Hayden, E. P., & Olino, T. M. (2010). Temperamental positive and negative emotionality and children's depressive symptoms: A longitudinal prospective study from age three to age ten. *Journal of Social and Clinical Psychology*, *29*(4), 462–488. https://doi.org/10.1521/jscp.2010.29.4.462

Dyson, M. W., Olino, T. M., Durbin, C. E., Goldsmith, H. H., Bufferd, S. J., Miller, A. R., & Klein, D. N. (2015). The structural and rank-order stability of temperament in young children based on a laboratory-observational measure. *Psychological Assessment*, *27*(4), 1388–1401. https://doi.org/10.1037/pas0000104

Eisenberg, N., Valiente, C., Spinrad, T. L., Cumberland, A., Liew, J., Reiser, M., Zhou, Q., & Losoya, S. H. (2009). Longitudinal relations of children's effortful control, impulsivity, and negative emotionality to their externalizing, internalizing, and co-occurring behavior problems. *Developmental Psychology*, *45*(4), 988–1008. https://doi.org/10.1037/a0016213

Erskine, H. E., Norman, R. E., Ferrari, A. J., Chan, G. C. K., Copeland, W. E., Whiteford, H. A., & Scott, J. G. (2016). Long-term outcomes of attention-deficit/Hyperactivity disorder and conduct disorder: A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(10), 841–850. https://doi.org/10.1016/j.jaac.2016.06.016

Forslund, T., Brocki, K. C., Bohlin, G., Granqvist, P., & Eninger, L. (2016). The heterogeneity of attention-deficit/hyperactivity disorder symptoms and conduct problems: Cognitive inhibition, emotion regulation, emotionality, and disorganized attachment. *British Journal of Developmental Psychology*, *34*(3), 371–387. https://doi.org/10.1111/bjdp.12136

Frick, P. J., & Viding, E. (2009). Antisocial behavior from a developmental psychopathology perspective. *Development and Psychopathology*, *21*(4), 1111–1131. https://doi.org/10.1017/S0954579409990071

Gagne, J. R., Van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological Assessment*, *23*(2), 337–353. https://doi.org/10.1037/a0021746

Gartstein, M. A., Putnam, S. P., & Rothbart, M. K. (2012). Etiology of preschool behavior problems: Contributions of temperament attributes in early childhood. *Infant Mental Health Journal*, *33*(2), 197–211. https://doi.org/10.1002/imhj.21312

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behavior and Development*, *26*(1), 64–86. https://doi.org/10.1016/S0163-6383(02)00169-8

Golm, D., & Brandt, V. (2023). The longitudinal association between infant negative emotionality, childhood maltreatment, and ADHD symptoms: A secondary analysis of data from the fragile families and child wellbeing study. *Development and Psychopathology*, *36*, 1–8. https://doi.org/10.1017/S0954579423000457

Goodman, R. (2011). Strengths and difficulties questionnaire [Dataset]. https://doi.org/10.1037/t00540-000

Goodman, R., Meltzer, H., & Bailey, V. (1998). The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry*, *7*(3), 125–130. https://doi.org/10.1007/s007870050057

Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(2), 268–292. https://doi.org/10.1080/10705511.2012.659627

Harris, J. L., LeBeau, B., & Petersen, I. T. (2025). Reactive and control processes in the development of internalizing and externalizing problems across early childhood to adolescence. *Development and Psychopathology*, *37*(2), 836–858. https://doi.org/10.1017/S0954579424000713

Heinze, H., Daseking, M., Gawrilow, C., Gunzenhauser, C., Karbach, J., Ulitzka, B., & Kerner Auch Koerner, J. (2025). Reactive temperament and self-regulation in preschool: Unique constellations impacting psychopathology. *Applied Neuropsychology: Child*, 1–17. https://doi.org/10.1080/21622965.2025.2519862

Heinze, H., Daseking, M., Gawrilow, C., Karbach, J., & Kerner Auch Koerner, J. (2025). Self-regulation in preschool: Are executive function and effortful control overlapping constructs? *Developmental Science*, *28*(1), e13595. https://doi.org/10.1111/desc.13595

Hirjak, D., Schwarz, E., & Meyer-Lindenberg, A. (2021). Zwölf Jahre research domain criteria in der psychiatrischen Forschung und Praxis: Anspruch und Wirklichkeit. *Der Nervenarzt*, *92*(9), 857–867. https://doi.org/10.1007/s00115-021-01174-1

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

Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). II. More than just convenient: the scientific merits of homogeneous convenience samples.

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

Monographs of the Society for Research in Child Development, *82*(2), 13–30. https://doi.org/10.1111/mono.12296

Jonas, K., & Kochanska, G. (2018). An imbalance of approach and effortful control predicts externalizing problems: Support for extending the dual-systems model into early childhood. *Journal of Abnormal Child Psychology*, *46*(8), 1573–1583. https://doi.org/10.1007/s10802-018-0400-3

Kälin, S., & Roebers, C. M. (2021). Self-regulation in preschool children: Factor structure of different measures of effortful control and executive functions. *Journal of Cognition and Development*, *22*(1), 48–67. https://doi.org/10.1080/15248372.2020.1862120

Karalunas, S. L., & Nigg, J. T. (2020). Heterogeneity and subtyping in attention-deficit/Hyperactivity disorder—Considerations for emerging research using person-centered computational approaches. *Biological Psychiatry*, *88*(1), 103–110. https://doi.org/10.1016/j.biopsych.2019.11.002

Kerner auch Koerner, J., Gust, N., & Petermann, F. (2018). Developing ADHD in preschool: Testing the dual pathway model of temperament. *Applied Neuropsychology: Child*, *7*(4), 366–373. https://doi.org/10.1080/21622965.2017.1347509

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the strengths and difficulties questionnaire. *International Journal of Behavioral Development*, *40*(1), 64–75. https://doi.org/10.1177/0165025415570647

Kiel, N., Bruckdorfer, R., Petermann, F., & Reinelt, T. (2018). Temperament in der frühen Kindheit und die Entwicklung externalisierender Störungen: Implikationen für die klinische Diagnostik. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, *66*(3), 177–186. https://doi.org/10.1024/1661-4747/a000354

Kievit, R. A., Brandmaier, A. M., Ziegler, G., Van Harmelen, A.-L., De Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, *33*, 99–117. https://doi.org/10.1016/j.dcn.2017.11.007

Kiff, C. J., Lengua, L. J., & Zalewski, M. (2011). Nature and nurturing: Parenting in the context of child temperament. *Clinical Child and Family Psychology Review*, *14*(3), 251–301. https://doi.org/10.1007/s10567-011-0093-4

Klasen, H., Woerner, W., Rothenberger, A., & Goodman, R. (2003). Die deutsche Fassung des Strengths and Difficulties Questionnaire (SDQ-Deu)—Übersicht und Bewertung erster Validierungs- und Normierungsbefunde [The German version of the Strengths and Difficulties Questionnaire (SDQ-Deu)—Overview over first validation and normative studies]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *52*(7), 491–502. https://doi.org/10.23668/psycharchives.11726

Kopala–Sibley, D. C., Olino, T., Durbin, E., Dyson, M. W., & Klein, D. N. (2018). The stability of temperament from early childhood to early adolescence: A multi–Method, multi–Informant examination. *European Journal of Personality*, *32*(2), 128–145. https://doi.org/10.1002/per.2151

Kostyrka-Allchorne, K., Wass, S. V., & Sonuga-Barke, E. J. S. (2020). Research review: Do parent ratings of infant negative emotionality and self-regulation predict psychopathology in childhood and adolescence? A systematic review and meta-analysis of prospective longitudinal studies. *Journal of Child Psychology and Psychiatry*, *61*(4), 401–416. https://doi.org/10.1111/jcpp.13144

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweatt, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. C. (2021). The hierarchical taxonomy of psychopathology (HiTOP): A quantitative nosology based on consensus of evidence. *Annual Review of Clinical Psychology*, *17*, 83–108. https://doi.org/10.1146/annurev-clinpsy-081219-093304

Lahey, B. B., Pelham, W. E., Loney, J., Lee, S. S., & Willcutt, E. (2005). Instability of the DSM-IV subtypes of ADHD from preschool through elementary school. *Archives of General Psychiatry*, *62*(8), 896–902. https://doi.org/10.1001/archpsyc.62.8.896

Lahey, B. B., Tiemeier, H., & Krueger, R. F. (2022). Seven reasons why binary diagnostic categories should be replaced with empirically sounder and less stigmatizing dimensions. *JCPP Advances*, *2*(4), e12108. https://doi.org/10.1002/jcv2.12108

Licata-Dandel, M., Wenzel, A. S., Kristen-Antonow, S., & Sodian, B. (2021). Predicting child problem behaviour at school age: The role of maternal sensitivity, child temperament and theory of mind. *Infant and Child Development*, *30*, 1–12. https://doi.org/10.1002/icd.2264

Martel, M. M., & Nigg, J. T. (2006). Child ADHD and personality/temperament traits of reactive and effortful control, resiliency, and emotionality. *Journal of Child Psychology and Psychiatry*, *47*(11), 1175–1183. https://doi.org/10.1111/j.1469-7610.2006.01629.x

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradović, J., Riley, J. R., Boelcke-Stennes, K., & Tellegen, A. (2005). Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental Psychology*, *41*(5), 733–746. https://doi.org/10.1037/0012-1649.41.5.733

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*(1), 577–605. https://doi.org/10.1146/annurev.psych.60.110707.163612

McElroy, E., Belsky, J., Carragher, N., Fearon, P., & Patalay, P. (2018). Developmental stability of general and specific factors of psychopathology from early childhood to adolescence: Dynamic mutualism or *p* -differentiation? *Journal of Child Psychology and Psychiatry*, *59*(6), 667–675. https://doi.org/10.1111/jcpp.12849

Metcalfe, L. A., Harvey, E. A., & Laws, H. B. (2013). The longitudinal relation between academic/cognitive skills and externalizing behavior problems in preschool children. *Journal of Educational Psychology*, *105*(3), 881–894. https://doi.org/10.1037/a0032624

Moilanen, K. L., Shaw, D. S., & Maxwell, K. L. (2010). Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Development and Psychopathology*, *22*(3), 635–653. https://doi.org/10.1017/S0954579410000337

NEPS-Network, N. (2023). National educational panel study, scientific use file of starting cohort newborns. Leibniz Institute for Educational Trajectories (LIfBi). https://doi.org/10.5157/NEPS:SC1

Nesselroade, J., & Ram, N. (2004). Studying intraindividual variability: What we have learned that will help us understand lives in context. *Research in Human Development*, *1*(1), 9–29. https://doi.org/10.1207/s15427617rhd0101&2_3

Nielsen, J. D., Olino, T. M., Dyson, M. W., & Klein, D. N. (2019). Reactive and regulatory temperament: Longitudinal associations with internalizing and externalizing symptoms through childhood. *Journal of Abnormal Child Psychology*, *47*(11), 1771–1784. https://doi.org/10.1007/s10802-019-00555-0

Nigg, J. T. (2006). Temperament and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *47*(3-4), 395–422. https://doi.org/10.1111/j.1469-7610.2006.01612.x

Nigg, J. T. (2022). Parsing ADHD with temperament traits. *Current Directions in Psychological Science*, *31*(4), 324–332. https://doi.org/10.1177/09637214221098079

Oldehinkel, A. J., Hartman, C. A., De Winter, A. F., Veenstra, R., & Ormel, J. (2004). Temperament profiles associated with internalizing and externalizing problems in preadolescence. *Development and Psychopathology*, *16*(02), 421–440. https://doi.org/10.1017/S0954579404044591

Ostlund, B., Myruski, S., Buss, K., & Pérez-Edgar, K. E. (2021). The centrality of temperament to the research domain criteria (RDoC): The earliest building blocks of psychopathology. *Development and Psychopathology*, *33*(5), 1584–1598. https://doi.org/10.1017/S0954579421000511

Perry, N. B., Calkins, S. D., Dollar, J. M., Keane, S. P., & Shanahan, L. (2018). Self-regulation as a predictor of patterns of change in externalizing behaviors from infancy to adolescence. *Development and Psychopathology*, *30*(2), 497–510. https://doi.org/10.1017/S0954579417000992

Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood. *Infant and Child Development*, *17*(4), 387–405. https://doi.org/10.1002/icd.582

Putnam, S., & Rothbart, M. K. (2006). Development of short and very short forms of the children's behavior questionnaire. *Journal of Personality Assessment*, *87*(1), 102–112.

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press

Ravens-Sieberer, U., Wille, N., Bettge, S., & Erhart, M. (2007). Psychische gesundheit von kindern und jugendlichen in deutschland: Ergebnisse aus der BELLA-studie im kinder- und jugendgesundheitssurvey (KiGGS). *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, *50*(5-6), 871–878. https://doi.org/10.1007/s00103-007-0250-6

Rettew, D. C., & McKee, L. (2005). Temperament and its role in developmental psychopathology. *Harvard Review of Psychiatry*, *13*(1), 14–27. https://doi.org/10.1080/10673220590923146

Rothbart, M. K., & Derryberry, D. (1981). Development of individual difference in temperament. In M. E. Lamb, & A. L. Brown (Eds.), *Advances in Developmental Psychology* (pp. 37–86). Lawrence Erlbaum Associates.

Rothbart, M. K., & Bates, J. F. (2006). Temperament in children's development. In W. Damon, R. Lerner, & N.Eisenberg (Hrsg.). (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 99–166). Wiley,

Rothbart, M. K., & Rueda, M. R., (2005). The development of effortful control. In U. Mayr, E. Awh, & S. W. Keele (Hrsg.) (Eds.), *Developing individuality in the human brain: A tribute to Michael I. Posner*. (Vol. *1*, pp. 167–188). American Psychological Association, https://doi.org/10.1037/11108-009

Rothbart, M. K., Sheese, B. E., & Posner, M. I. (2007). Executive attention and effortful control: Linking temperament, brain networks, and genes. *Child Development Perspectives*, *1*(1), 2–7. https://doi.org/10.1111/j.1750-8606.2007.00002.x

Sameroff, A. (2010). A unified theory of development: A dialectic integration of nature and nurture. *Child Development*, *81*(1), 6–22. https://doi.org/10.1111/j.1467-8624.2009.01378.x

Santens, E., Claes, L., Dierckx, E., & Dom, G. (2020). Effortful control—A transdiagnostic dimension underlying internalizing and externalizing psychopathology. *Neuropsychobiology*, *79*(4-5), 255–269. https://doi.org/10.1159/000506134

Savalei, V., Brace, J. C., & Fouladi, R. T. (2023). We need to change how we compute RMSEA for nested model comparisons in structural equation modeling. *Psychological Methods*, *29*, 480–493. https://doi.org/10.1037/met0000537

Schmidt, J. A., Aschersleben, G., & Henning, A. (2025). Early-life temperament: A longitudinal investigation of stability and structure. *International Journal of Behavioral Development*, *49*(3), 214–227. https://doi.org/10.1177/01650254241290999

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Shah, J. L., Jones, N., Van Os, J., McGorry, P. D., & Gülöksüz, S. (2022). Early intervention service systems for youth mental health: Integrating pluripotentiality, clinical staging, and transdiagnostic lessons from early psychosis. *The Lancet Psychiatry*, *9*(5), 413–422. https://doi.org/10.1016/S2215-0366(21)00467-3

Sleddens, E. F. C., Hughes, S. O., O'Connor, T. M., Beltran, A., Baranowski, J. C., Nicklas, T. A., & Baranowski, T. (2012). The children's behavior questionnaire very short scale: Psychometric properties and development of a one-item temperament scale. *Psychological Reports*, *110*(1), 197–217. https://doi.org/10.2466/08.10.21.PR0.110.1.197-217

Sleddens, E. F. C., Kremers, S. P. J., Candel, M. J. J. M., De Vries, N. N. K., & Thijs, C. (2011). Validating the children's behavior questionnaire in dutch children: Psychometric properties and a cross-cultural comparison of factor structures. *Psychological Assessment*, *23*(2), 417–426. https://doi.org/10.1037/a0022111

Stone, L. L., Janssens, J. M. A. M., Vermulst, A. A., Van Der Maten, M., Engels, R. C. M. E., & Otten, R. (2015). The strengths and difficulties questionnaire: Psychometric properties of the parent and teacher version in children aged 4-7. *BMC Psychology*, *3*(1), 4. https://doi.org/10.1186/s40359-015-0061-8

Stone, L. L., Otten, R., Engels, M., R., C., Vermulst, A. A., & Janssens, J. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review*, *13*, 254–274. https://doi.org/10.1007/s10567-010-0071-2

Sulik, M. J., Blair, C., Mills-Koonce, R., Berry, D., Greenberg, M., & **The Family Life Project Investigators**. (2015). Early parenting and the development of externalizing behavior problems: Longitudinal mediation through children's executive function. *Child Development*, *86*(5), 1588–1603. https://doi.org/10.1111/cdev.12386

Thomas, A., & Chess, S. (1977). *Temperament and development*. Brunner/Mazel.

Van Meter, A. R., Sibley, M. H., Vandana, P., Birmaher, B., Fristad, M. A., Horwitz, S., Youngstrom, E. A., Findling, R. L., & Arnold, L. E. (2024). The stability and persistence of symptoms in childhood-onset ADHD. *European Child & Adolescent Psychiatry*, *33*(4), 1163–1170. https://doi.org/10.1007/s00787-023-02235-3

Wakschlag, L. S., Estabrook, R., Petitclerc, A., Henry, D., Burns, J. L., Perlman, S. B., Voss, J. L., Pine, D. S., Leibenluft, E., & Briggs-Gowan, M. L. (2015). Clinical implications of a dimensional approach: The normal: Abnormal spectrum of early irritability. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(8), 626–634. https://doi.org/10.1016/j.jaac.2015.05.016

Wakschlag, L. S., Roberts, M. Y., Flynn, R. M., Smith, J. D., Krogh-Jespersen, S., Kaat, A. J., Gray, L., Walkup, J., Marino, B. S., Norton, E. S., & Davis, M. M. (2019). Future directions for early childhood prevention of mental disorders: A road map to mental health. *Earlier. Journal of Clinical Child & Adolescent Psychology*, *48*(3), 539–554. https://doi.org/10.1080/15374416.2018.1561296

Wass, S. V., & Goupil, L. (2022). Studying the developing brain in real-world contexts: Moving from castles in the air to castles on the ground. *Frontiers in Integrative Neuroscience*, *16*, 896919. https://doi.org/10.3389/fnint.2022.896919

Wichstrøm, L., Penelo, E., Rensvik Viddal, K., Osa, N., & Ezpeleta, L. (2018). Explaining the relationship between temperament and symptoms of psychiatric disorders from preschool to middle childhood: Hybrid fixed and random effects models of Norwegian and Spanish children. *Journal of Child Psychology and Psychiatry*, *59*(3), 285–295. https://doi.org/10.1111/jcpp.12772

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x

Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgalarrondo, P., Lui, M., & Tannock, R. (2004). The strengths and difficulties questionnaire overseas: Evaluations and applications of the SDQ beyond Europe. *European Child & Adolescent Psychiatry*, *13*(Suppl 2), II47–II54. https://doi.org/10.1007/s00787-004-2008-0

Zelazo, P. D. (2020). Executive function and psychopathology: A neurodevelopmental perspective. *Annual Review of Clinical Psychology*, *16*, 431–454. https://doi.org/10.1146/annurev-clinpsy-072319-024242

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354–360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives*, *6*(2), 112–121. https://doi.org/101111/j.1750-8606.2011.00176.x.

https://doi.org/10.1017/S0954579426101278 Published online by Cambridge University Press