Development and Psychopathology (2025), **37**, 2317–2326
doi:10.1017/S0954579424001937

CAMBRIDGE UNIVERSITY PRESS

Regular Article

Risk factor, consequence, or common cause? Linking lower self-regulation and internalizing symptoms during middle childhood in a random intercept cross-lagged panel model

Johanna Lilian Klinge[1], Petra Warschburger[2] and Annette Maria Klein[1]

[1]International Psychoanalytic University Berlin, Berlin, Germany and [2]Department of Psychology, University of Potsdam, Potsdam, Germany

Abstract

This study investigates whether lower self-regulation (SR) facets are risk factors for internalizing symptoms (vulnerability models), consequences of these symptoms (scar models), or develop along the same continuum and thus share common causes (spectrum models) during middle childhood. To analyze these models simultaneously, a random intercept cross-lagged panel model was estimated using Mplus. Data were assessed at three measurement time points in a community-based sample of $N = 1657$ (52.2% female) children in Germany, aged 6–13. Internalizing symptoms were measured via parent report by the emotional problems scale of the Strengths and Difficulties Questionnaire. Seven SR facets were assessed behaviorally, via parent report and teacher report. At the within-person level, internalizing symptoms were concurrently associated with emotional reactivity at all measurement time points, while no cross-lagged paths reached significance. At the between-person level, internalizing symptoms were associated with working memory updating ($r = -.29$, $p < .001$), inhibitory control ($r = -.29$, $p < .001$), planning behavior ($r = -.49$, $p < .001$), and emotional reactivity ($r = .59$, $p < .001$). As internalizing symptoms and SR facets were primarily associated at the between-person level, the results lend support to spectrum models suggesting common causes of internalizing symptoms and impaired SR.

Keywords: Emotional reactivity; executive functions; internalizing symptoms; middle childhood; self-regulation

(Received 3 July 2024; revised 3 December 2024; accepted 7 December 2024; first published online 10 January 2025)

Introduction

Internalizing symptoms (anxiety, depression, somatic symptoms, and withdrawal) are among the most common mental health problems in childhood and adolescence (Achenbach, 1991). In children and adolescents (aged 7–17 years) in Germany, the prevalence of clinically relevant depressive and anxiety symptoms at the time of data collection for this study was 15% each (Reiß et al., 2023). After the onset of puberty, girls have an increased risk to develop internalizing symptoms compared to boys (Hayward & Sanborn, 2002). Studies show that internalizing symptoms affect almost all areas of life (Wille et al., 2008), have negative effects on later developmental outcomes (Musliner et al., 2016), and carry a high risk of chronicity (Leadbeater et al., 2012).

Self-regulation facets

In recent years, the development of internalizing symptoms has been increasingly associated with self-regulation (SR), which can be defined as the ability to control and modulate one's cognitions, behaviors, emotions, and physiological responses to attain future benefits (Bailey & Jones, 2019; Nigg, 2017; Robson et al., 2020). As a multidimensional construct, SR encompasses various interrelated facets (review: Nigg, 2017). In this study, emotional reactivity was investigated as a basal temperament-related aspect of SR that encompasses emotional responses to events in terms of response threshold, latency, amplitude, rise time to peak intensity, and recovery time (Rothbart & Derryberry, 1981). Further SR facets examined in this study include core executive functions (EF) (Diamond, 2013; Miyake et al., 2000), which can be considered basal cognitive facets of SR: working memory updating (the ability to mentally retain and manipulate information), cognitive flexibility/set-shifting (the ability of switching between different tasks, perspectives, or rules), and inhibition (the ability to suppress primary behavioral impulses in favor of a less dominant response). Inhibitory control (the ability to curb impulsive behaviors when instructed), a closely associated but more temperament-related aspect of SR, was also assessed. Additionally, a more complex and later emerging EF, planning behavior (skills like goal setting, strategy development, and action organization), was examined. The aforementioned core EFs are often referred to as cognitive-driven "cool" EFs in contrast to emotion-driven "hot" EFs, a distinction supported by neural and behavioral evidence (Zelazo, 2020). Our study also examined one hot EF, affective decision-making, which involves integrating emotions and cognitive processes in decision-making.

Corresponding author: Johanna Lilian Klinge; Email: johanna.klinge@ipu-berlin.de

Cite this article: Klinge, J. L., Warschburger, P., & Klein, A. M. (2025). Risk factor, consequence, or common cause? Linking lower self-regulation and internalizing symptoms during middle childhood in a random intercept cross-lagged panel model. *Development and Psychopathology* **37**: 2317–2326, https://doi.org/10.1017/S0954579424001937

© The Author(s), 2025. 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/S0954579424001937 Published online by Cambridge University Press

How can relations between internalizing symptoms and SR facets be explained?

Even though numerous studies show that internalizing symptoms and SR facets are associated, the direction of these relations and their explanation remain controversial. Possible explanations can be subdivided using a taxonomy by Tackett (2006) that distinguishes between *vulnerability* (or *predisposition*) *models*, *scar* (or *complication*) *models, spectrum models, and pathoplasty* (or *exacerbation*) models. Vulnerability models propose that lower levels of SR facets increase the risk for subsequent internalizing symptoms, while scar models suggest the opposite that preexisting internalizing symptoms impair SR facets. In contrast, spectrum models view SR facets and internalizing symptoms as part of a shared continuum, attributing them to common underlying causes. Pathoplasty models posit that SR facets influence the course, characteristics, and severity of internalizing symptoms. In this study, we decided to estimate a random intercept cross-lagged panel model (RI-CLPM) (Hamaker et al., 2015), which allows to model the first three models simultaneously. The investigation of pathoplasty models, in contrast, requires person-centered analytical methods, for example, latent profile or trajectory analysis, as we employed in a previous study (Klinge et al., 2023). Accordingly, the following sections outline the research on the first three models.

Vulnerability models: lower SR facets as risk factors for internalizing symptoms

Vulnerability models assume that lower SR facets increase the risk of later internalizing symptoms, which is supported by multiple longitudinal studies during childhood and adolescence (e.g., Kertz et al., 2015; Nelson et al., 2018). Most evidence exists regarding lower cool EF (for a meta-analysis, see Yang et al., 2022) and higher emotional reactivity (e.g., Carthy et al., 2010) or, as a related construct, negative emotionality (for a meta-analysis, see Kostyrka-Allchorne et al., 2020) as preceding higher internalizing symptoms. Only few to no associations have been reported between lower hot EF and later internalizing symptoms (Yang et al., 2022), possibly due to a smaller number of studies. However, it should be critically noted that most studies did not control for previous manifestations of internalizing symptoms. Therefore, it remains open whether SR facets actually increase the vulnerability for internalizing symptoms or whether their relation can be better explained by other models.

Scar models: lower SR facets as consequences of internalizing symptoms

In opposition to vulnerability models, scar models assume that preexisting internalizing symptoms impair SR facets (Tackett, 2006), possibly mediated through alterations of still developing brain structures (Berl et al., 2006). Although studied scarcely, supporting evidence exists. For example, Donati et al. (2021) report that internalizing symptoms in early adolescence were associated with impaired working memory in mid-to-late adolescence, but not vice versa.

Recently, some studies also corroborated bidirectional associations between constructs similar or related to those investigated in this study, supporting both the vulnerability and scar models equally. Halse et al. (2022) established bidirectional cross-lagged associations between symptoms of multiple psychiatric disorders (including affective disorders) during childhood and parent-reported EF. Additionally, bidirectional cross-lagged associations between social anxiety symptoms and behavioral inhibition have been reported (Goldsmith et al., 2021).

Spectrum models: lower SR facets and internalizing symptoms on a continuum

Contrary to the previous models, spectrum models view internalizing symptoms and lower SR facets as overlapping constructs that develop along the same continuum and may therefore be due to common causes. Such models have especially been proposed to explain the relation between temperament-related SR and psychopathology (Stifter & Dollar, 2016). Potential common causes of internalizing symptoms and SR facets include genetics (Friedman & Miyake, 2017; Musci et al., 2015) as well as environmental factors such as child maltreatment (Cichetti & Toth, 2005; Kavanaugh et al., 2017; Klein et al., 2019) or harsh parenting (Halse et al., 2019) to name a few.

(Potential) common causes that are hypothesized to equally influence development over time can be controlled for in cross-lagged panel models (CLPM) by additionally implementing random intercepts (RI) of each investigated construct (Hamaker et al., 2015; Usami et al., 2019). A correlation of the RIs of internalizing symptoms and SR facets, that is, a correlation of these variables at the between-person level, can therefore be interpreted as an indicator that these could be due to (potential) common causes, supporting spectrum models. Consistent with this, two studies report that inhibitory control (Maasalo et al., 2020) and self-control problems (Kim et al., 2022) showed no significant cross-lagged associations with internalizing symptoms but were correlated with these at the between-person level.

Research gaps

Which explanatory model best explains the relations between internalizing symptoms and SR facets is still unclear. The described results suggest that multiple models may simultaneously apply or that models differ depending on SR facets investigated. However, there is a lack of prospective studies on children and adolescents using appropriate methods to control for prior outcome values, separate within- and between-person variance, and thus allow to model possible associations according to different theoretical models simultaneously. The few existing studies using such methods focused on only one SR facet (Kim et al., 2022; Maasalo et al., 2020), the emergence of specific social anxiety (Goldsmith et al., 2022) or psychopathology in general (Halse et al., 2022) rather than investigating multiple facets of SR and their specific relations with internalizing symptoms.

Another research gap exists regarding gender differences in the relations between internalizing symptoms and SR facets. So far, these have been demonstrated only in a few individual studies. For instance, Schwartz et al. (1999) reported that the association between inhibited temperament at age two and symptoms of generalized anxiety disorder in adolescence was moderated by gender with girls showing a stronger association than boys. Rudolph et al. (2017) reported a significant three-way interaction of cognitive control, negative emotionality and gender in the prediction of depressive symptoms: deficits in cognitive control predicted higher depressive symptoms in girls with high, but not with average or low negative emotionality. However, this interaction was nonsignificant for boys. Similarly, Agoston and Rudolph (2016) found that high EF deficits combined with childhood peer stress predicted depression in early adolescent girls, but not in boys. Furthermore, Han et al. (2012) showed that boys

https://doi.org/10.1017/S0954579424001937 Published online by Cambridge University Press

with depression perform worse in affective decision-making tasks than girls with depression. Apart from these individual studies, however, most studies focus solely on gender differences in either internalizing symptoms or SR facets, but not on gender differences in their associations.

Current study and hypotheses

To address these research gaps, the current study examined the relations between internalizing symptoms and seven different SR facets over three measurement time points. To account for the possible associations between internalizing symptoms and SR facets according to vulnerability, scar, and spectrum models, a RI-CLPM (Hamaker et al., 2015) was estimated, containing all eight constructs at once. This sophisticated model distinguishes within- and between-person variance and controls for the stability and covariances of these constructs and for prior outcome values.

First, we expect higher internalizing symptoms to be concurrently associated with lower SR facets at each measurement time point. Second, we expect that lower SR facets at one measurement time point are associated with higher internalizing symptoms at a later measurement time point (according to vulnerability models). Third, we expect that higher internalizing symptoms at one measurement time point are associated with lower SR facets at a later measurement time point (according to scar models). Additionally, we will examine associations between internalizing symptoms and SR facets at the between-person level (according to spectrum models) as well as gender and age differences in association patterns exploratively. To ensure transparency and reduce publication bias, the study's hypotheses and analysis plan were preregistered; see doi:10.17605/OSF.IO/36HMV.

Methods

Sample and procedure

Data were collected between 2011 and 2015 at three measurement time points (t1, t2, t3) by multiple informants (children, parents, teachers) and by using multiple methods (behavioral measures, questionnaires) in a large community-based longitudinal study on intrapersonal developmental risk factors in childhood and adolescence conducted at the University of Potsdam, Germany (for an overview of the whole study, see study protocol by Warschburger et al., 2023). Participants and their families were recruited from 120 classes in 33 public primary schools from rural and urban areas in the Federal State of Brandenburg, Germany, to obtain a most representative sample. To ensure economic data collection by examining several participants at the same location, the entire primary school classes were recruited wherever possible. 1657 children (52.2% girls), aged between 6 and 11 years ($M_{age} = 8.36$, $SD = 0.95$), and 1340 (80.7%) parents participated at t1. On average, 9.14 months ($SD = 1.80$) after t1, 1612 children (51.9% girls; $M_{age} = 9.11$ years, $SD = 0.93$; 97.5% retention rate) and 1197 (72.1%) parents participated again at t2. On average, 23.83 months ($SD = 1.66$) after t2, 1534 children (51.7% girls; $M_{age} = 11.06$ years, $SD = 0.92$; 92.6% retention rate) and 1070 (64.5%) parents participated again at t3. For each child, one parent questionnaire was completed at each measurement time point by mothers (70.9%–79.2%), by both parents together (12.5%–20.8%), by fathers (7.1%–7.4%), or by other caregivers (0.8%–1.5%). To account for missing data, we implemented full information maximum likelihood estimation as proposed by Little and Rubin (1987), allowing us to include children who provided data at least once. Table 1 presents further sociodemographic characteristics of the sample and their parents at t1.

Table 1. Sociodemographic characteristics of the sample at t1 ($n = 1340$, parent report)

	n	(%)
Domestic situation		
Lives with both parents	979	(73.1%)
Lives with one parent w/wo new partner	105	(7.8%)
Another social situation	9	(0.7%)
Missing	247	(18.4%)
Highest education degree of both parents		
No degree/special school	14	(1.0%)
Lower secondary school	34	(2.5%)
Upper secondary school	422	(31.5%)
High school diploma	257	(19.2%)
University degree	607	(45.3%)
Missing	6	(0.4%)
Household income (measured at t2)		
less than €2000/month	220	(16.4%)
€2000 to €5000/month	735	(54.9%)
more than €5000/month	179	(13.4%)
Missing	206	(15.4%)
Language spoken at home		
Only German	1239	(92.5%)
German and at least one other language	77	(5.7%)
Another language than German	13	(1.0%)
Missing	11	(0.8%)

Participation in the study was voluntary, and parents gave informed consent. Assessments were approved by the Research Ethics Board at the University of Potsdam and the Ministry of Education, Youth, and Sport in Brandenburg, Germany. Children completed standardized tests in two 1-hr sessions and received small gifts afterward. Parents and teachers filled out online or printed questionnaires. Due to economic reasons, each measure was assessed by one informant and one method. Teachers received a €5 contribution to the class fund for each completed questionnaire.

Measures

Internalizing symptoms were reported by parents on the emotional problems scale of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001) at t1, t2, and t3. Consisting of five items (e.g., "Often unhappy, down-hearted, or tearful"), this scale can be rated on a 3-point scale from *not true* (0) to *certainly true* (2). A sum score was calculated with higher values indicating higher internalizing symptoms. The parent-reported emotional problems scale of the SDQ demonstrates strong concurrent validity with the internalizing problems scale of the Child Behavior Check List ($r = .77$) and satisfactory criterion validity (AUC = .69) with respect to emotional disorders according to the ICD-10 criteria assessed by psychiatrists in a sample of 5- to 17-year-olds (Becker

https://doi.org/10.1017/S0954579424001937 Published online by Cambridge University Press

et al., 2004). Adequate reliability was confirmed at all measurement time points (α = .66–.72).

Emotional reactivity was reported by parents on 10 items of the emotional control scale of the Behavior Rating Inventory of Executive Function (Gioia et al., 2000) at t1, t2, and t3. Items could be scored on an adapted 5-point scale from *never* (1) to *always* (5). A mean score was calculated with higher values indicating higher emotional reactivity and thus lower SR. High reliability was confirmed at all measurement time points (α = .91–.92).

Working memory updating was measured behaviorally by the sum of correctly answered sequences (max. 16) in the digit span backward task (ZN-R) of the German version of the Wechsler Intelligence Scale for Children – Fourth Edition (Petermann & Petermann, 2008) at t1, t2, and t3. In this task, children were asked to repeat digit spans backward, increasing the span lengths from 2 to 8 in a maximum of eight trials, each consisting of two equal-length sequences. When a child incorrectly repeated both sequences during one trial, the test was stopped. The task shows adequate retest reliability ($r = .62$; Piovesana et al., 2014).

Cognitive flexibility/set-shifting was measured behaviorally by the number of correct responses on randomly occurring switch trials (22 out of 46 trials) of the Cognitive Flexibility Task (Roebers & Kauer, 2009) at t1 and t2. The task shows an acceptable split-half reliability ($r = .64$; Röthlisberger et al., 2013). At t3, a computerized version of the Dimensional Change Card Sort Task (Qu et al., 2013) was used, also measuring cognitive flexibility/set-shifting by the number of correct responses on randomly occurring switch trials (12 out of 48). The task shows a good internal consistency (α = .85–.90; Friedman et al., 2016).

Inhibition was measured behaviorally by the interference score of the Fruit-Stroop task (Roebers et al., 2011), an adapted version of the Stroop task (Archibald & Kerns, 1999) at t1, t2, and t3. In this task, four pages of 25 stimuli were presented, each under different conditions (pages 1 and 2: no interference, pages 3 and 4: interference). Children were asked to name the colors of all 25 stimuli per page. Based on the time (in seconds) they needed, an interference score was calculated (Roebers et al., 2011). To ensure that higher values indicate better inhibition, the interference score was z-standardized and inverted. The task shows a good split-half reliability ($r = .78$; Röthlisberger et al., 2011).

Inhibitory control was reported by parents on six items of the inhibitory control scale of the Temperament in Middle Childhood Questionnaire (Simonds et al., 2007) at t1, t2, and t3. Items could be scored on a 5-point scale from *not at all* (1) to *a lot* (5). A mean score was calculated with higher values indicating higher inhibitory control. Adequate reliability was confirmed at all measurement time points (α = .67–.71).

Planning behavior was reported by teachers on eight items of the planning/organizing scale of the Behavior Rating Inventory of Executive Function (Gioia et al., 2000) at t1, t2, and t3. Items could be scored on an adapted 5-point scale ranging from *never* (1) to *always* (5). A mean score was calculated with higher values indicating better planning behavior. High reliability was confirmed at all measurement time points (α = .93–.95).

Affective decision-making was measured behaviorally by the net-score difference between advantageous and disadvantageous choices in the Hungry Donkey Task (Crone & van der Molen, 2004), a computer-based, age-appropriate version of the Iowa Gambling Task (Bechara et al., 1994) at t1, t2, and t3. In this task, children performed 50 test trials in which they could choose between four doors with different win/lose contingencies. The task shows adequate reliability (α = .75–.78; Orm et al., 2022).

Statistical analyses

To test our hypotheses, we estimated a RI-CLPM (Hamaker et al., 2015) with single indicators using Mplus (version 8.8; Muthén & Muthén, 1998–2012). The maximum likelihood estimator with robust standard errors was applied. Missing data were addressed by full information maximum likelihood estimation. The RI-CLPM is an extension of the traditional CLPM and allows the distinction between between- and within-person variance by computing a RI for each variable (Hamaker et al., 2015). RIs load equally across all measurement occasions of the corresponding variable and thus represent a child's individual mean value for that variable over time. Consequently, RIs account for the stability of the construct and control all time-invariant confounders/common causes assumed to influence development consistently across time points (Usami et al., 2019). By doing so, RIs capture trait-like differences at the between-person level (Hamaker et al., 2015). The stable relations between constructs that persist across multiple time points – and that might be explained by shared time-invariant confounders/common causes – are reflected in the correlations among RIs. However, the variance of RIs cannot be statistically decomposed into specific variables that could represent time-invariant confounders/common causes, as they are an aggregate measure of all (unobserved) confounders/common causes at once.

Variance in the manifest variables that is not bound by the RIs feeds into latent within-person factors, which are connected by autoregressive and cross-lagged paths as well as covariances of their residuals. This part of the model appears to correspond to traditional CLPMs, but with the important difference that it represents within-person dynamics. Autoregressive paths therefore do not reflect the stability of the ranking of participants from one measurement time to the next as they would in traditional CLPMs, but rather within-person carry-over effects (Hamaker et al., 2015). For example, if the autoregressive parameter in a RI-CLPM is positive, this means that children who scored above their mean score at one time point are also likely to score above their mean score at the following time point. As a result, autoregressive effects are typically smaller in RI-CLPMs than in traditional CLPMs (Mulder & Hamaker, 2021). Similarly, the cross-lagged parameters in a RI-CLPM provide information on how children's deviation from their own mean score on internalizing symptoms at a given time point is influenced by their deviation from their own mean score on a particular SR facet at an earlier time point, and vice versa.

Model fit was evaluated by the comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and root mean square residual (SRMR). A model is considered showing a good fit when CFI ≥ .95, TLI ≥ .95, RMSEA ≤ .05, and SRMR ≤ .10 (Beauducel & Wittmann, 2005). The Satorra–Bentler chi-square difference test (Satorra & Bentler, 2010) was used to determine whether additional model constraints led to a significant deterioration of the model fit.

In all models, grand means of manifest variables were constrained to be equal as they were z-standardized in advance, and gender was included as a covariate. Models for subsamples (age and gender groups) were compared using multigroup analyses.

Before estimating the RI-CLPM, weak factorial measurement invariance was confirmed separately for all questionnaire-based measures (see Supplementary Table S1). The test of measurement invariance was not feasible for behavioral measures as these measures are not test-equivalent due to the varying test procedure (e.g., termination criteria for the digit span backward test) at

https://doi.org/10.1017/S0954579424001937 Published online by Cambridge University Press

Table 2. Means and standard deviations of internalizing symptoms and self-regulation facets at all measurement points

	t1			t2			t3		
	n	*M*	(*SD*)	*n*	*M*	(*SD*)	*n*	*M*	(*SD*)
Internalizing symptoms	1320	1.70	(1.80)	1180	1.65	(1.81)	1056	1.60	(1.85)
Working memory updating	1641	6.18	(1.47)	1607	6.63	(1.51)	1495	7.37	(1.63)
Cognitive flexibility/set-shifting	1641	15.58	(4.68)	1586	18.14	(3.92)	1481	9.77	(1.87)
Inhibition[1]	1644	24.95	(8.78)	1605	20.55	(6.90)	1490	16.72	(5.31)
Inhibitory control	1311	3.52	(0.66)	1174	3.59	(0.63)	1055	3.75	(0.68)
Planning behavior	1371	3.71	(0.89)	1127	3.64	(0.90)	1090	3.68	(0.96)
Affective decision-making	1643	5.49	(11.43)	1589	8.45	(12.94)	1481	9.56	(13.69)
Emotional reactivity	1294	2.20	(0.71)	1166	2.13	(0.68)	1030	2.24	(0.73)

[1]Higher values indicate lower inhibition (non-inverted interference score).

different measurement time points. However, since we work with single indicators to reduce model complexity – that is, each latent within-person variable is assigned exactly one observed indicator at each measurement time point – the test of measurement invariance is not a prerequisite (Hamaker et al., 2015).

Results

Unstandardized means and standard deviations of all examined variables are depicted in Table 2. Bivariate correlations between internalizing symptoms and SR facets are shown in Supplementary Table S2.

Model fitting of the RI-CLPM

First, a baseline RI-CLPM was estimated that included internalizing symptoms and seven SR facets assessed at three measurement time points. All parameters were estimated freely, resulting in a very good model fit: $\chi^2(60) = 97.19$, $p = .002$, CFI = .996, TLI = .981, RMSEA = .019, 90% CI [0.01, 0.03], SRMR = .013. The model fit of an alternative CLPM without RIs was significantly worse compared to the baseline model ($\chi^2(124) = 487.64$, $p < .001$, CFI = .963, TLI = .910, RMSEA = .042, 90% CI [0.04, 0.05], SRMR = .024; $\Delta\chi^2 = 370.35$, $df = 64$, $p < .001$) showing the implementation of RIs was justified.

Aiming for parsimony, we then constrained covariances of the residuals of all possible pairs of variables at t2 to be equal to their respective covariances at t3. The model showed no significant deterioration compared to the baseline model ($\chi^2(88) = 134.20$, $p < .001$, CFI = .995, TLI = .984, RMSEA = .018, 90% CI [0.01, 0.02], SRMR = .015; $\Delta\chi^2 = 38.24$, $df = 28$, $p = .094$) and thus served as the final model for further analyses.

Parameter estimates of the final RI-CLPM

The results of the final RI-CLPM are depicted in Figure 1. Parameter estimates of all included paths are provided in Supplementary Table S3. At the within-person level, internalizing symptoms were concurrently moderately positively associated with emotional reactivity at all measurement time points ($rs = .26$–$.29$, $ps < .001$). Cross-lagged paths were significant only between individual SR facets, predominantly concerning EFs between t1 and t2, but not between SR facets and internalizing symptoms. However, there were two marginally significant cross-lagged paths from internalizing symptoms at t2 to inhibitory control at t3 ($b = -.13$, $p = .052$) and from inhibition at t2 to internalizing symptoms at t3 ($b = -.11$, $p = .083$). At the between-person level, the RI of internalizing symptoms was moderately negatively associated with the RIs of working memory updating ($r = -.29$, $p < .001$) and inhibitory control ($r = -.29$, $p < .001$), strongly negatively associated with the RI of planning behavior ($r = -.49$, $p < .001$), and strongly positively associated with the RI of emotional reactivity ($r = .59$, $p < .001$).

Multigroup analyses

In multigroup analyses investigating age and gender differences in association patterns, both models did not converge due to insignificant variance in one or more RIs of one of the subgroups. Therefore, the multigroup analyses were conducted without RIs, as the distinction of within- and between-person variance is not necessary when investigating group differences.

Age differences

Based on a median split of age at t1 ($Md = 8.40$), the total sample was separated into a younger (range: 6.23–8.39) and an older age group (range: 8.40–11.33). A model in which associations could differ between these age groups showed a very good model fit: $\chi^2(232) = 685.67$, $p < .001$, CFI = .953, TLI = .877, RMSEA = .049, 90% CI [0.04, 0.05], SRMR = .029. However, a model in which parameters were constrained to be equal did not fit significantly worse: $\chi^2(424) = 892.08$, $p < .001$, CFI = .951, TLI = .931, RMSEA = .037, 90% CI [0.03, 0.04], SRMR = .040; $\Delta\chi^2 = 217.72$, $df = 192$, $p = .098$. According to the principle of parsimony, no age differences can be assumed.

Gender differences

A model in which associations could differ between gender groups obtained a very good model fit: $\chi^2(200) = 585.07$, $p < .001$, CFI = .959, TLI = .887, RMSEA = .048, 90% CI [0.04, 0.05], SRMR = .028. However, a model in which parameters were constrained to be equal between both groups did not fit significantly worse: $\chi^2(384) = 760.41$, $p < .001$, CFI = .960, TLI = .943, RMSEA = .034, 90% CI [0.03, 0.04], SRMR = .036; $\Delta\chi^2 = 182.03$, $df = 184$, $p = .527$. Therefore, no gender differences can be assumed.

https://doi.org/10.1017/S0954579424001937 Published online by Cambridge University Press

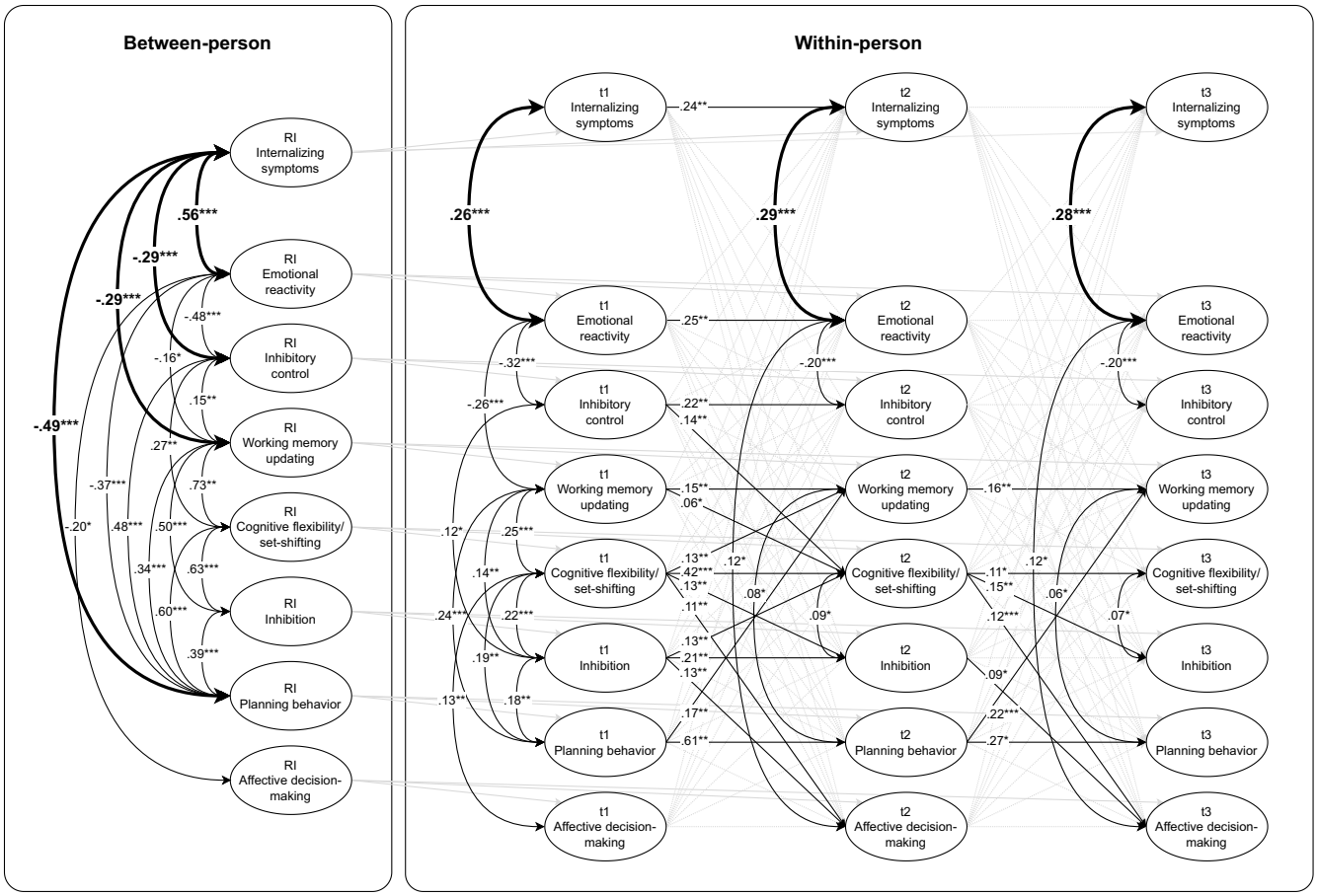


Figure 1. Results of the final RI-CLPM showing associations at the between-person and within-person level of internalizing symptoms and multiple SR facets.
Note: All coefficients are standardized. Significant associations are indicated by black lines. Non-significant associations and fixed factor loadings of RIs are indicated by grey lines. To ensure clarity, manifest single-indicator variables, residuals of the within-person factors, non-significant covariances, and the covariate gender are not depicted. *p < .05, **p < .01, ***p < .001.

Discussion

In this large community-based longitudinal study, we estimated a RI-CLPM containing internalizing symptoms and seven SR facets assessed at three measurement time points in middle childhood to examine their associations according to vulnerability, scar, and spectrum models. Our first hypothesis was partially confirmed, as we found internalizing symptoms and emotional reactivity to be moderately concurrently associated at all measurement time points. However, our second and third hypotheses were not supported, as no cross-lagged paths between internalizing symptoms and SR facets reached significance. Instead, we found significant negative associations between internalizing symptoms and working memory updating, inhibitory control, planning behavior, and most strongly and positively with emotional reactivity at the between-person level.

Our results are consistent with two studies also reporting no significant cross-lagged associations, but instead finding associations between the RIs of internalizing symptoms and lower inhibitory control (Maasalo et al., 2020) or self-control problems (Kim et al., 2022). Contrary to our primary findings, but corresponding with our two marginally significant cross-lagged paths, Halse et al. (2022) reported significant cross-lagged paths in both directions between affective disorders and four different EFs, including inhibition. However, they measured core EFs using parent and teacher reports that have little overlap with performance-based tests (Toplak et al., 2013) as used in our study. In addition, focusing on psychopathology in general, they included other psychiatric disorders characterized by externalizing symptoms (e.g., attention-deficit/hyperactivity disorder) and constrained cross-lagged paths of the same direction to be equal. Since externalizing symptoms usually show stronger associations with SR (Robson et al., 2020), this could also explain the differences to our results. We did not identify any age or gender differences in association patterns, which is consistent with Halse et al. (2022) who investigated a similarly aged sample.

Are lower SR facets risk factors or consequences of internalizing symptoms, or do both have common causes?

Associations primarily emerged at the between-person level, while there were almost none at the within-person level (cross-lagged paths and concurrent associations). This indicates that stable between-person differences may be better suited to explain the relations between SR facets and internalizing symptoms than within-person dynamics. Therefore, our results support spectrum models, according to which internalizing symptoms and SR facets may develop along the same continuum and thus potentially share common causes. This result is also compatible with studies identifying several SR facets, for example, EF deficits and negative affectivity, as transdiagnostic factors for psychopathology (Lynch et al., 2021).

https://doi.org/10.1017/S0954579424001937 Published online by Cambridge University Press

Spectrum models have been discussed in the existing literature especially in the context of temperament-related SR facets (Stifter & Dollar, 2016), which is also confirmed in our study, as two temperament-related SR facets (emotional reactivity and inhibitory control) show significant between-person correlations with internalizing symptoms. However, it should be noted that both temperament-related SR facets and internalizing symptoms were reported by parents. Shared informant and method variance could therefore have contributed to the strength of their association. Furthermore, we were able to show that spectrum models may also explain the association of internalizing symptoms with two cognitive SR facets that were assessed behaviorally and by teacher report, respectively: working memory updating and planning behavior. This is a new finding, as most studies assume relations in terms of vulnerability models (e.g., Yang et al., 2022) without considering different explanatory models or controlling for prior outcome values.

Although our results primarily support spectrum models, it should be noted that we also found two marginally significant cross-lagged paths from internalizing symptoms at t2 to inhibitory control at t3 and from inhibition at t2 to internalizing symptoms at t3. There are several reasons why these and other cross-lagged paths that would have supported vulnerability and scar models may not have become significant. For example, it is possible that changes in internalizing symptoms and SR facets may influence each other over shorter intervals of time (e.g., weeks or days) than were examined in this study. The concurrent within-person associations between internalizing symptoms and emotional reactivity that emerged in addition to their strong between-person associations could reflect such short-term within-person reciprocal effects. Kim et al. (2022), who found a similar pattern in their study on internalizing symptoms and self-control, also discussed this possibility. Additionally, the marginal or nonsignificance of cross-lagged parameters could indicate the need for higher power to detect these effects. Considering Cohen (1994), who warns against equating *p* with the probability of rejecting the null hypothesis, our results do not suggest that vulnerability or scar models should be discarded, especially since they are also supported by preexisting literature (e.g., Halse et al., 2022). Further studies are needed to disentangle the effects between internalizing symptoms and SR facets according to different theoretical models. We also recommend investigating potential short-term dynamics between SR facets and internalizing symptoms by implementing experience sampling methods that involve several daily surveys over a period of 1–2 weeks (Boemo et al., 2022).

Why are some SR facets significantly associated with internalizing symptoms while others are not?

In addition to comparing explanatory models, our study raises the question of why some SR facets are significantly associated with internalizing symptoms while others are not. As shown in Table 2, SR facets with significant associations with internalizing symptoms had stable means across the three measurement time points. In this regard, they resembled internalizing symptoms, which also showed stable mean values despite high variance indicating interindividual heterogeneity. This suggests, in accordance with spectrum models, that key developmental processes related to both internalizing symptoms and significantly related SR facets may have already been completed. In contrast, nonsignificant SR facets show more change in mean scores toward improvement in SR over time, possibly requiring longer developmental trajectories to exhibit clear associations with internalizing symptoms. This is in accordance with evidence that SR facets mature at different rates (for a meta-review, see Wesarg-Menzel et al., 2023; for a systematic review on cool and hot EF, see García et al., 2021). For example, García et al. (2021) establish that interference inhibition, which is measured by the Stroop task (as used in our study), seems to continue to improve beyond the age of 12. The authors argue that this type of inhibition requires a higher cognitive load than response inhibition, which is often measured by Go/No-go tasks and reaches a performance level similar to adults at the age of 12. This perspective furthermore aligns with meta-analyses showing that EF deficits in individuals with depression symptoms are not detectable in adolescence and early adulthood (Goodall et al., 2018; Dotson et al., 2020), but in adulthood (Rock et al., 2014; Snyder, 2013). Consequently, future research should examine how the maturation processes of specific SR facets influence their relations to internalizing symptoms during different developmental periods. Regardless of this, it is also possible that the significant SR facets – especially emotional reactivity – show greater relevance and construct proximity with internalizing symptoms than the nonsignificant ones.

Further research and clinical implications

While our results indicate that higher internalizing symptoms and lower SR facets may share common causes, our study design does not allow us to identify these as RIs represent an aggregate measure of all (unobserved) confounders/common causes. Next to genetic factors, plenty of environmental factors could be influential such as child maltreatment (Cichetti & Toth, 2005; Kavanaugh et al., 2017; Klein et al., 2018), harsh parenting (Halse et al., 2019), or parents' insufficient containment of their infants affects (Bion, 1962). Future studies are needed to understand which specific common causes can possibly account for the stable between-person associations of internalizing symptoms and lower SR facets to develop tailored prevention and intervention measures.

The results of our study suggest that prevention and intervention measures should, where possible, target *both* internalizing symptoms and the enhancement of SR, with emotional reactivity playing a significant role. Given that environmental factors can be considered common causes of higher internalizing symptoms and lower SR, it is also advisable to develop measures that support families and school professionals in managing children characterized by a higher symptom burden, more intense and prolonged negative emotions, and lower SR skills. In particular, the use of adaptive emotion regulation strategies and increased co-regulation of children by their parents could be helpful in regulating their heightened emotional reactivity.

Strengths and limitations

This study examined associations between internalizing symptoms and multiple SR facets by simultaneously modeling vulnerability, scar, and spectrum models. The large community-based sample showed very high retention rates and allowed us to employ a RI-CLPM containing eight constructs at once. This sophisticated model distinguished within- and between-person effects and controlled for the stability of constructs, the covariances of multiple SR facets, and prior outcome values. Data were collected from different informants and by using multiple methods including well-validated behavioral measures and questionnaires.

Nonetheless, some limitations must be noted. First, due to economic reasons, each measure was assessed by only one

https://doi.org/10.1017/S0954579424001937 Published online by Cambridge University Press

informant and one method. As internalizing symptoms and temperament-related SR facets were assessed via parent report, shared informant and method variance may have contributed to the strength of their associations. However, we also found significant associations between internalizing symptoms and two cognitive SR facets, which were measured behaviorally and by teacher report, respectively. Further studies are needed to examine how these associations may be influenced by measurement methods and whether they persist when other informants and methods are included. This could be particularly important regarding core EFs, as parent and teacher reports are known to show little overlap with behavioral measures (Toplak et al., 2013). It can be assumed that the former primarily measures the average, everyday performance, while the latter measures the potential maximum performance of a child.

Second, internalizing symptoms were only assessed via parent reports. Since all children at t1 and the majority at t2 were younger than 11 years, the self-rating version of the SDQ was not yet suitable. In addition, the RI-CLPM is based on covariances between constructs, so potential underestimation of internalizing symptoms, which has been observed in parent reports during adolescence (Klasen et al., 2016), would not have significantly affected the parameters.

Third, cross-lagged paths did not reach significance, but two marginal significant cross-lagged effects in the direction of vulnerability and scar models have been found. This may have been due to the long intervals between measurement time points, or it may have required more power to detect them. As Cohen (1994) cautioned, *p*-values should not be conflated with the probability of rejecting the null hypothesis. Therefore, it remains possible that both vulnerability or scar models, alongside spectrum models, are valid explanations for the associations between internalizing symptoms and SR facets, requiring further studies on this topic.

Fourth, our results cannot be generalized due to the homogeneity of our sample. Further studies should investigate whether our results can be replicated in samples with greater ethnic and socioeconomic diversity to avoid bias in interpretation.

Finally, time-variant confounders such as stressful life events (Grant et al., 2006) could not be controlled for, as this would have required at least one additional measurement time point for model identification (Usami et al., 2019).

Conclusion

While we identified concurrent associations between internalizing symptoms and emotional reactivity at all measurement time points, we did not observe any significant cross-lagged associations between internalizing symptoms and SR facets. However, at the between-person level, we established that internalizing symptoms were moderately associated with lower working memory updating and lower inhibitory control and strongly associated with lower planning behavior and higher emotional reactivity. These findings suggest that not only temperament-related but also cognitive SR facets may develop along the same continuum as internalizing symptoms and therefore may share common causes, as assumed by spectrum models. Nonetheless, vulnerability and scar models may be simultaneously valid alongside spectrum models as indicated by two marginal significant cross-lagged effects and previous research. Prevention and intervention measures should, where possible, target *both* internalizing symptoms and the enhancement of SR, with emotional reactivity playing a significant role. Further studies are needed to explore potential common causes, investigate bidirectional effects at shorter time intervals, account for different maturation rates of SR facets and examine the influence of time-variant confounders.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0954579424001937.

Acknowledgments. We thank all participating children and families.

Funding statement. This work is funded by the German Research Foundation (KL 2338/2-1) as part of a research group (FOR 5034; grant number 426314138). The first three measurement points used in this study were funded by the German Research Foundation as part of the research training group GRK 1668.

Competing interests. None.

References

Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18 and 1991 profile*. Department of Psychiatry, University of Vermont.

Agoston, A. M., & Rudolph, K. D. (2016). Interactive contributions of cumulative peer stress and executive function deficits to depression in early adolescence. *The Journal of Early Adolescence*, *36*(8), 1070–1094. https://doi.org/10.1177/0272431615593176

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology*, *5*(2), 115–129. https://doi.org/10.1076/chin.5.2.115.3167

Bailey, R., & Jones, S. M. (2019). An integrated model of regulation for applied settings. *Clinical Child and Family Psychology Review*, *22*(1), 2–23. https://doi.org/10.1007/s10567-019-00288-y

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*(1), 41–75. https://doi.org/10.1207/s15328007sem1201_3

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*(1-3), 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

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*(S2), ii11–ii16. https://doi.org/10.1007/s00787-004-2003-5

Berl, M. M., Vaidya, C. J., & Gaillard, W. D. (2006). Functional imaging of developmental and adaptive changes in neurocognition. *Neuroimage*, *30*(3), 679–691. https://doi.org/10.1016/j.neuroimage.2005.10.007

Bion, W. R. (1962). *Learning from experience*. Heinemann.

Boemo, T., Nieto, I., Vazquez, C., & Sanchez-Lopez, A. (2022). Relations between emotion regulation strategies and affect in daily life: A systematic review and meta-analysis of studies using ecological momentary assessments. *Neuroscience and Biobehavioral Reviews*, *139*, 104747. https://doi.org/10.1016/j.neubiorev.2022.104747

Carthy, T., Horesh, N., Apter, A., & Gross, J. J. (2010). Patterns of emotional reactivity and regulation in children with anxiety disorders. *Journal of Psychopathology and Behavioral Assessment*, *32*(1), 23–36. https://doi.org/10.1007/s10862-009-9167-8

Cicchetti, D., & Toth, S. L. (2005). Child maltreatment. *Annual Review of Clinical Psychology*, *1*(1), 409–438. https://doi.org/10.1146/annurev.clinpsy.1.102803.144029

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Crone, E. A., & van der Molen, M. W. (2004). Developmental changes in real life decision making: Performance on a gambling task previously shown to depend on the ventromedial prefrontal cortex. *Developmental Neuropsychology*, *25*(3), 251–279. https://doi.org/10.1207/s15326942dn2503_2

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Donati, G., Meaburn, E., & Dumontheil, I. (2021). Internalising and externalising in early adolescence predict later executive function, not the

other way around: A cross-lagged panel analysis. *Cognition and Emotion*, *35*(5), 986–998. https://doi.org/10.1080/02699931.2021.1918644

Dotson, V. M., McClintock, S. M., Verhaeghen, P., Kim, J. U., Draheim, A. A., Syzmkowicz, S. M., Gradone, A. M., Bogoian, H. R., & Wit, L. (2020). Depression and cognitive control across the lifespan: A systematic review and meta-analysis. *Neuropsychology Review*, *30*(4), 461–476. https://doi.org/10.1007/s11065-020-09436-6

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

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*(2), 326–340. https://doi.org/10.1037/dev0000075

García, L. F., Merchán, A., Phillips-Silver, J., & González, M. T. D. (2021). Neuropsychological development of cool and hot executive functions between 6 and 12 years of age: A systematic review. *Frontiers in Psychology*, *12*, 687337. https://doi.org/10.3389/fpsyg.2021.687337

Gioia, G., Isquith, P., Guy, S. C., & Kenworthy, L. (2000). Behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235–238. https://doi.org/10.1076/chin.6.3.235.3152

Goldsmith, H. H., Hilton, E. C., Phan, J., Sarkisian, K. L., Carroll, I. C., Lemery-Chalfant, K., & Planalp, E. M. (2022). Childhood inhibition predicts adolescent social anxiety: Findings from a longitudinal twin study. *Development and Psychopathology*, *34*(5), 1666–1685. https://doi.org/10.1017/s0954579422000864

Goodall, J., Fisher, C., Hetrick, S., Phillips, L., Parrish, E. M., & Allott, K. (2018). Neurocognitive functioning in depressed young people: A systematic review and meta-analysis. *Neuropsychology Review*, *28*(2), 216–231. https://doi.org/10.1007/s11065-018-9373-9

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015

Grant, K. E., Compas, B. E., Thurm, A., McMahon, S. D., Gipson, P. Y., Campbell, A., Krochock, K., & Westerholm, R. I. (2006). Stressors and child and adolescent psychopathology: Evidence of moderating and mediating effects. *Clinical Psychology Review*, *26*(3), 257–283. https://doi.org/10.1016/j.cpr.2005.06.011

Halse, M., Steinsbekk, S., Hammar, A., Belsky, J., & Wichstrøm, L. (2019). Parental predictors of children's executive functioning from ages 6 to 10. *British Journal of Developmental Psychology*, *37*(3), 410–426. https://doi.org/10.1111/bjdp.12282

Halse, M., Steinsbekk, S., Hammar, A., & Wichstrom, L. (2022). Longitudinal relations between impaired executive function and symptoms of psychiatric disorders in childhood. *Journal of Child Psychology and Psychiatry*, *63*(12), 1574–1582. https://doi.org/10.1111/jcpp.13622

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Han, G., Klimes-Dougan, B., Jepsen, S., Ballard, K., Nelson, M., Houri, A., Kumra, S., & Cullen, K. (2012). Selective neurocognitive impairments in adolescents with major depressive disorder. *Journal of Adolescence*, *35*(1), 11–20. https://doi.org/10.1016/j.adolescence.2011.06.009

Hayward, C., & Sanborn, K. (2002). Puberty and the emergence of gender differences in psychopathology. *The Journal of Adolescent Health*, *30*(4), 49–58. https://doi.org/10.1016/s1054-139x(02)00336-1

Kavanaugh, B. C., Dupont-Frechette, J. A., Jerskey, B. A., & Holler, K. A. (2017). Neurocognitive deficits in children and adolescents following maltreatment: Neurodevelopmental consequences and neuropsychological implications of traumatic stress. *Applied Neuropsychology: Child*, *6*(1), 64–78. https://doi.org/10.1080/21622965.2015.1079712

Kertz, S., Belden, A., Tillman, R., & Luby, J. (2015). Cognitive control deficits in shifting and inhibition in preschool age children are associated with increased depression and anxiety over 7.5 years of development. *Journal of Abnormal Child Psychology*, *44*(6), 1185–1196. https://doi.org/10.1007/s10802-015-0101-0

Kim, Y., Richards, J. S., & Oldehinkel, A. J. (2022). Self-control, mental health problems, and family functioning in adolescence and young adulthood: Between-person differences and within-person effects. *Journal of Youth and Adolescence*, *51*(6), 1181–1195. https://doi.org/10.1007/s10964-021-01564-3

Klasen, F., Petermann, F., Meyrose, A., Barkmann, C., Otto, C., Haller, A. C., Schlack, R., Schulte-Markwort, M., & Ravens-Sieberer, U. (2016). Trajectory of mental disorders in children and adolescents [Verlauf psychischer Auffälligkeiten von Kindern und Jugendlichen]. *Kindheit und Entwicklung*, *25*(1), 10–20. https://doi.org/10.1026/0942-5403/a000184

Klein, A. M., Schlesier-Michel, A., Otto, Y., White, L. O., Andreas, A., Sierau, S., Bergmann, S., Perren, S., & Von Klitzing, K. (2019). Latent trajectories of internalizing symptoms from preschool to school age: A multi-informant study in a high-risk sample. *Development and Psychopathology*, *31*(02), 657–681. https://doi.org/10.1017/s0954579418000214

Klinge, J. L., Warschburger, P., Busching, R., & Klein, A. M. (2023). Self-regulation facets differentially predict internalizing symptom trajectories from middle childhood to early adolescence: A longitudinal multimethod study. *Child and Adolescent Psychiatry and Mental Health*, *17*(1), 1–14. https://doi.org/10.1186/s13034-023-00670-3

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

Leadbeater, B., Thompson, K., & Gruppuso, V. (2012). Co-occurring trajectories of symptoms of anxiety, depression, and oppositional defiance from adolescence to young adulthood. *Journal of Clinical Child and Adolescence Psychology*, *41*(6), 719–730. https://doi.org/10.1080/15374416.2012.694608

Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Lynch, S. J., Sunderland, M., Newton, N. C., & Chapman, C. (2021). A systematic review of transdiagnostic risk and protective factors for general and specific psychopathology in young people. *Clinical Psychology Review*, *87*, 102036. https://doi.org/10.1016/j.cpr.2021.102036

Maasalo, K., Lindblom, J., Kiviruusu, O., Santalahti, P., & Aronen, E. T. (2020). Longitudinal associations between inhibitory control and externalizing and internalizing symptoms in school-aged children. *Development and Psychopathology*, *33*(3), 843–855. https://doi.org/10.1017/S0954579420000176

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 638–648. https://doi.org/10.1080/10705511.2020.1784738

Musci, R. J., Masyn, K. E., Benke, K., Maher, B., Uhl, G., & Ialongo, N. S. (2016). The effects of the interplay of genetics and early environmental risk on the course of internalizing symptoms from late childhood through adolescence. *Development and Psychopathology*, *28*(1), 225–237. https://doi.org/10.1017/s0954579415000401

Musliner, K. L., Munk-Olsen, T., Eaton, W. W., & Zandi, P. P. (2016). Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. *Journal of Affective Disorders*, *192*, 199–211. https://doi.org/10.1016/j.jad.2015.12.030

Nelson, T. D., Kidwell, K. M., Nelson, J. M., Tomaso, C. C., Hankey, M., & Espy, K. A. (2018). Preschool executive control and internalizing symptoms in elementary school. *Journal of Abnormal Child Psychology*, *46*(7), 1509–1520. https://doi.org/10.1007/s10802-017-0395-1

Nigg, J. T. (2017). Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

Orm, S., Pollak, Y., Fossum, I. N., Andersen, P., Øie, M. G., & Skogli, E. W. (2022). Decision-making and risky behavior in individuals with attention-

https://doi.org/10.1017/S0954579424001937 Published online by Cambridge University Press

deficit/hyperactivity disorder: A 10-year longitudinal study. *Developmental Neuropsychology*, *47*(4), 193–209. https://doi.org/10.1080/87565641.2022.2082430

Petermann, F., & Petermann, U. (2008). HAWIK-IV. *Childhood and Development [Kindheit und Entwicklung]*, *17*(2), 71–75. https://doi.org/10.1026/0942-5403.17.2.71

Piovesana, A. M., Ross, S., Whittingham, K., Ware, R. S., & Boyd, R. N. (2015). Stability of executive functioning measures in 8-17-year-old children with unilateral cerebral palsy. *The Clinical Neuropsychologist*, *29*(1), 133–149. https://doi.org/10.1080/13854046.2014.999125

Qu, L., Low, J. J. W., Chong, P. S., Lim, D. W. L., Keren-Happuch, F. F. E (2013). Task switching in English-Chinese bilinguals: A life span approach. In Kashima, Y.Kashima, E. S. and Beatson, R. (Eds.), *Selected papers from the 2010 Congress of the International Association for Cross-Cultural Psychology* (pp. 76–84). IACCP.

Reiß, F., Kaman, A., Napp, A., Devine, J., Li, L., Strelow, L., Erhart, M., Hölling, H., Schlack, R., & Ravens-Sieberer, U. (2023). Epidemiology of mental well-being of children and adolescents in Germany. Results from 3 studies before and during the COVID-19 pandemic [Epidemiologie seelischen Wohlbefindens von Kindern und Jugendlichen in Deutschland. Ergebnisse aus 3 Studien vor und während der COVID-19-Pandemie], Bundesgesundheitsblatt - Gesundheitsforschung. *Gesundheitsschutz*, *66*(7), 727–735. https://doi.org/10.1007/s00103-023-03720-5

Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*, *146*(4), 324–354. https://doi.org/10.1037/bul0000227

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–2040. https://doi.org/10.1017/s0033291713002535

Roebers, C. M., & Kauer, M. (2009). Motor and cognitive control in a normative sample of 7-year-olds. *Developmental Science*, *12*(1), 175–181. https://doi.org/10.1111/j.1467-7687.2008.00755.x

Roebers, C. M., Röthlisberger, M., Cimeli, P., Michel, E., & Neuenschwander, R. (2011). School enrolment and executive functioning: A longitudinal perspective on developmental changes, the influence of learning context, and the prediction of pre-academic skills. *European Journal of Developmental Psychology*, *8*(5), 526–540. https://doi.org/10.1080/17405629.2011.571841

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.

Röthlisberger, M., Neuenschwander, R., Cimeli, P., & Roebers, C. (2013). Executive functions in 5- to 8-year olds: Developmental changes and relationship to academic achievement. *Journal of Educational and Developmental Psychology*, *3*(2), 153–167. https://doi.org/10.5539/jedp.v3n2p153

Rudolph, K. D., Davis, M. M., & Monti, J. D. (2017). Cognition-emotion interaction as a predictor of adolescent depressive symptoms. *Developmental Psychology*, *53*(12), 2377–2383.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Schwartz, C. E., Snidman, N., & Kagan, J. (1999). Adolescent social anxiety as an outcome of inhibited temperament in childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*(8), 1008–1015. https://doi.org/10.1097/00004583-199908000-00017

Simonds, J., Kieras, J. E., Rueda, M. R., & Rothbart, M. K. (2007). Effortful control, executive attention, and emotional regulation in 7-10-year-old children. *Cognitive Development*, *22*(4), 474–488. https://doi.org/10.1016/j.cogdev.2007.08.009

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, *139*(1), 81–132. https://doi.org/10.1037/a0028727

Stifter, C., & Dollar, J. (2016). Temperament and developmental psychopathology. In D. Cichetti (Eds.), *Developmental psychopathology: Risk, resilience, and intervention*. vol. *4*, 3rd ed. pp. 546–607). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119125556.devpsy411

Tackett, J. L. (2006). Evaluating models of the personality-psychopathology relationship in children and adolescents. *Clinical Psychology Review*, *26*(5), 584–599. https://doi.org/10.1016/j.cpr.2006.04.003

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*(5), 637–657. https://doi.org/10.1037/met0000210

Warschburger, P., Gmeiner, M. S., Bondü, R., Klein, A. M., Busching, R., & Elsner, B. (2023). Self-regulation as a resource for coping with developmental challenges during middle childhood and adolescence: The prospective longitudinal PIERYOUTH-study. *BMC Psychology*, *11*(1), Article–97. https://doi.org/10.1186/s40359-023-01140-3

Wesarg-Menzel, C., Ebbes, R., Hensums, M., Wagemaker, E., Zaharieva, M. S., Staaks, J. P., Van den Akker, A. L., Visser, I., Hoeve, M., Brummelman, E., Dekkers, T. J., Schuitema, J. A., Larsen, H., Colonnesi, C., Jansen, B. R., Overbeek, G., Huizenga, H. M., & Wiers, R. W. (2023). Development and socialization of self-regulation from infancy to adolescence: A meta-review differentiating between self-regulatory abilities, goals, and motivation. *Developmental Review*, *69*, 101090. https://doi.org/10.1016/j.dr.2023.101090

Wille, N., Bettge, S., Ravens-Sieberer, U., & BELLA Study Group (2008). Risk and protective factors for children's and adolescents' mental health: Results of the BELLA study. *European Child and Adolescence Psychiatry*, *17*(S1), 133–147. https://doi.org/10.1007/s00787-008-1015-y

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

Zelazo, P. D. (2020). Executive function and psychopathology: A neurodevelopmental perspective. *Annual Review of Clinical Psychology*, *16*(1), 431–454. https://doi.org/10.1146/annurev-clinpsy-072319-024242

https://doi.org/10.1017/S0954579424001937 Published online by Cambridge University Press