CAMBRIDGE
UNIVERSITY PRESS

## Regular Article

# Reactive and control processes in the development of internalizing and externalizing problems across early childhood to adolescence

Jordan L. Harris[1] ⓘ, Brandon LeBeau[2] ⓘ and Isaac T. Petersen[1] ⓘ

[1]Department of Psychological and Brain Sciences, University of Iowa, IA, USA and [2]Department of Psychological and Quantitative Foundations, University of Iowa, IA, USA

## Abstract

Reactive and control processes – e.g., negative emotionality and immediacy preference – may predict distinct psychopathology trajectories. However, externalizing and internalizing problems change in behavioral manifestation across development and across contexts, thus necessitating the use of different measures and informants across ages. This is the first study that created developmental scales for both internalizing and externalizing problems by putting scores from different informants and measures onto the same scale to examine temperament facets as risk factors. Multidimensional linking allowed us to examine trajectories of internalizing and externalizing problems from ages 2 to 15 years ($N = 1,364$) using near-annual ratings by mothers, fathers, teachers, other caregivers, and self report. We examined reactive and control processes in early childhood as predictors of the trajectories and as predictors of general versus specific psychopathology in adolescence. Negative emotionality at age 4 predicted general psychopathology and unique externalizing problems at age 15. Wait times on an immediacy preference task at age 4 were negatively associated with age 15 general psychopathology, and positively associated with unique internalizing problems. Findings demonstrate the value of developmental scaling for examining development of psychopathology across a lengthy developmental span and the importance of considering reactive and control processes in development of psychopathology.

**Keywords:** Bifactor; delay of gratification; heterotypic continuity; longitudinal; negative emotionality

## Introduction

Internalizing and externalizing problems are among the most common, costly, and burdensome issues facing children and adults across the lifespan (Forbes et al., 2016). Externalizing disorders, typically encompassing symptoms of aggression, conduct problems, and oppositionality, have a global prevalence of over 5% (Polanczyk et al., 2015). Moreover, untreated externalizing problems lead to future problems such as academic difficulties (Shi & Ettekal, 2021), social delinquency, incarceration, and substance abuse (Krueger et al., 2021; Loeber et al., 1998). Internalizing psychopathology, which encompasses anxiety and depression, has an even higher prevalence ranging from 20 to 25% in children and adolescents, and has greatly increased since the COVID-19 pandemic (Racine et al., 2021). Thus, it is crucial to identify processes that prevent the development of internalizing and externalizing psychopathology.

### General psychopathology

Diagnoses from the internalizing and the externalizing spectra show a high rate of co-occurrence (Caspi et al., 2014; Clark et al., 2021; Murray et al., 2016). Internalizing and externalizing problems are

most accurately modeled dimensionally rather than as categorical phenomena (Markon et al., 2011), and numerous factor analytic modeling studies have identified strong covariation among internalizing, externalizing, and thought-disordered dimensions of psychopathology (Caspi et al., 2014; Cervin et al., 2021; Choate et al., 2022; Gluschkoff et al., 2019; Lahey et al., 2012). The strong covariation between these problems has led to the hypothesis that internalizing, externalizing, and thought-disordered problems share a common cause – a higher-order factor called "*p* factor" – that accounts for their covariation (Caspi et al., 2014; Forbes et al., 2016; Krueger & Eaton, 2015). The *p* factor can be modeled in various ways, including a higher-order factor model or a bifactor model. The general factor of psychopathology reflects what is common among dimensions of psychopathology. The general factor is operationalized as the common variance among internalizing and externalizing psychopathology indicators (Ree et al., 2015; Watts, Meyer, et al., 2021; Zinbarg et al., 2005). Specific psychopathology (e.g., externalizing) in a bifactor model is represented by the residual correlations among the specific psychopathology facets that are not accounted for by the general factor.

Given the high cost and burden of internalizing and externalizing psychopathology, it is important to identify early processes that lead to development of specific internalizing or externalizing problems versus co-occurring internalizing and externalizing problems. The present study considers reactive and control processes of temperament that may help explain differing developmental pathways to psychopathology.

**Corresponding author:** Jordan L. Harris; Email: jordan-l-harris@uiowa.edu

Check for updates

## Temperamental reactive and control processes in the development of general and specific psychopathology

Temperament is defined as constitutionally based ways in which individuals regulate and react to their environment (Rothbart & Bates, 2006). Individual differences in temperament are early appearing, biologically based, and relatively stable across development (Rothbart & Bates, 2006). The structure of temperament is broadly categorized by three relatively orthogonal dimensions – two reactivity dimensions: positive and negative emotionality; and one control dimension: self-regulation (Rothbart & Bates, 2006; Rothbart, 2011). Positive emotionality reflects the tendency to experience positive emotions, e.g., enthusiasm and joy, often reflecting extraversion (Watson et al., 1988). Negative emotionality reflects a propensity toward anger/frustration, fear, and sadness (Rothbart & Bates, 2006). Self-regulation reflects a child's ability to regulate behavior, cognition and emotions, and often includes subordinate constructs of executive functioning, attentional control, and effortful control. In the present study, we focus on facets of the negative emotionality and self-regulation dimensions of temperament given their robust associations with psychopathology (e.g., Eisenberg et al., 2005; Muris & Ollendick, 2005). One of the most widely studied predictors of children's adjustment is immediacy preference, which is a facet of self-regulation and is a control process (Krueger et al., 1996).

### Immediacy preference

Immediacy preference is the selection of a smaller, immediate reward over a larger, distal one (Mischel & Ebbesen, 1970; Stephens & Anderson, 2001). The inverse of immediacy preference is delay preference (Rachlin & Jones, 2008), which is frequently called delay of gratification. Immediacy preference is often assessed with a self-imposed waiting task (Metcalfe & Mischel, 1999) that is designed to assess a person's ability or preference to resist the temptation of the immediate reward (i.e., gratification) in favor of a more motivationally salient, distal reward. In a temperament framework, delay of gratification is considered a control process because it describes a self-regulatory process by which an individual might suppress a dominant response in favor of a subordinate one (Moran et al., 2013). This conceptualization of delay of gratification as a control process of temperament is supported by prior research (Bjorklund & Kipp, 1996; Moran et al., 2013; Murray & Kochanska, 2002). We use the term "delay of gratification" when referring to the duration of waiting in a self-imposed waiting task, whereas we use the term "immediacy preference" when referring to the process as a risk factor. Immediacy preference has been shown to be associated with deficits in self-control, a common impairment in individuals with psychopathology (Kidd et al., 2013; Michaelson & Munakata, 2020).

Immediacy preference (i.e., a stronger relative preference for an immediate reward) has been associated with many facets of psychopathology, including impulsive decision making and externalizing psychopathology. Several studies have found that immediacy preference was associated with both parent- and teacher-reported aggressive and delinquent behaviors in children ranging from ages 3 to 13 years (Campbell & von Stauffenberg, 2009; Krueger et al., 1996). A 40-year follow-up study found that adults with greater immediacy preference in childhood continued to show more externalizing problems compared to their peers who had a delay preference in childhood (Casey et al., 2011). Numerous studies have identified a strong association of immediacy preference with externalizing problems and delinquent outcomes (Campbell & von Stauffenberg, 2009; Krueger et al., 1996).

Previous studies have implicated immediacy preference as a risk factor that is specific to externalizing problems, i.e., not internalizing problems (Krueger et al., 1996). However, little work has examined whether immediacy preference is associated with general psychopathology. Watts et al. (2018) used the Eunice Kennedy Shriver National Institute of Child Health and Development (NICHD) Study of Early Child Care and Youth Development (SECCYD) sample, the same sample as the present study. Importantly, Watts and colleagues found that when accounting for numerous covariates, including family background, home environment, and early cognitive ability, immediacy preference was no longer significantly associated with behavioral outcomes (Watts et al., 2018). However, the behavioral outcomes variable was a total behavior problems score computed as an average of internalizing and externalizing problems. Thus, the study did not allow for determining whether immediacy preference was associated with general versus specific psychopathology. Partitioning internalizing and externalizing problems is especially important because there might be reason to expect that immediacy preference is more strongly associated with externalizing problems compared to internalizing problems (Campbell & von Stauffenberg, 2009; Krueger et al., 1996).

Another study using the SECCYD sample (Deutz et al., 2020) examined the association between antecedent factors, including immediacy preference, and specific and general psychopathology at later ages (8 and 14 years old). Results indicated that immediacy preference was not associated with later general or specific psychopathology. However, the study did not examine immediacy preference in relation to the *development* (i.e., change over time) of behavior problems. Moreover, the study examined only mother- and self report, and did not include perspectives from other informants such as fathers.

In sum, much remains unknown regarding the association between immediacy preference and psychopathology. For instance, it is not known the degree to which immediacy preference in early childhood predicts general versus specific psychopathology in adolescence (Deutz et al., 2020). In addition to immediacy preference, the present study considers temperamental negative emotionality, a reactive process, which has been proposed as a transdiagnostic risk factor for psychopathology (Mikolajewski et al., 2013; Weissman et al., 2019). Transdiagnostic risk factors contribute to the etiology and maintenance of a broad range of emotional and behavioral difficulties (Egan et al., 2011). Transdiagnostic risk factors can occur within a given specific factor – e.g., perfectionism is a risk factor for eating disorders, anxiety, and depression (Egan et al., 2011) – or can reflect risk for general psychopathology, meaning they contribute to a broader range of behaviors (Lynch et al., 2021).

### Temperamental negative emotionality

Temperamental negative emotionality is an individual's characteristic reaction to stimuli, and includes the tendency to display overt sadness, irritability, fear/withdrawal, distress, and/or anger, and may be characterized by somatic or autonomic physiological reactivity in response to stimuli (Fox, 1989; Rothbart & Derryberry, 1981). As such, negative emotionality is considered a reactive process within a temperament framework (Eisenberg et al., 1996; Moran et al., 2013). Numerous studies have found that negative emotionality is highly correlated with internalizing and externalizing psychopathology (Leaberry et al., 2019; McLaughlin & Nolen-Hoeksema, 2011; Steinberg & Drabick, 2015), and evidence suggests that negative emotionality plays a causal role in

the development of internalizing and externalizing psychopathology (Lilienfeld, 2003). There is some evidence that negative emotionality is associated with general psychopathology (i.e., *p* factor; Hankin et al., 2017). Negative emotionality is thought to evolve and change in its manifestation, much like symptoms of psychopathology more broadly (e.g., Aldao et al., 2016; De Los Reyes et al., 2009, 2013; Mischel & Shoda, 1995; Pettersson et al., 2018; Rutter & Sroufe, 2000).

A previous study found that negative emotionality, as observed in a frustration task, was associated with both internalizing and externalizing problems, and partially accounted for their covariation (Mikolajewski et al., 2013). Negative emotionality predicts psychopathology (Brandes et al., 2019; Briggs-Gowan et al., 2006; Hawes et al., 2020) and they overlap in behavioral indicators of distress (e.g., dysthymia and depression; Greene & Eaton, 2017), suggesting that they conceptually overlap to a degree. However, prior literature has indicated that temperament/personality constructs differ from psychopathology. Temperament is primarily concerned with the how of a behavior (e.g., how intensely a child cries), whereas psychopathology focuses on the content of the behavior (e.g., what does the child cry about; Bates et al., 2014; De Pauw & Mervielde, 2010; Lemery et al., 2002; Thomas & Chess, 1977). However, to our knowledge, no previous study has examined control and reactive temperamental processes simultaneously in predicting general and specific psychopathology in the same model.

Effortful control, an aspect of self-regulation which comprises immediacy preference, and negative emotionality are related but separate aspects of temperament (Eisenberg et al., 2005; Rothbart et al., 2001). Children with less effortful control tend to experience greater adjustment difficulties in the face of stress, thus leading them to show more negative emotionality (Moran et al., 2013; Muris & Ollendick, 2005). A prior study using the same sample as the present study found that anger was associated with higher mother-reported externalizing problems. This same study found that higher levels of effortful control at 54 months indirectly predicted lower levels of externalizing problems at age 15 years (Crockett et al., 2018). Furthermore, when examined together, negative emotionality and effortful control have shown additive effects on psychopathology (Eisenberg et al., 1996, 2000, 2005). Relatedly, one study indicated that an imbalance in approach behavior – a reactive process – and control processes was associated with externalizing problems, indicating an interaction of control and reactive processes in externalizing problems (Jonas & Kochanska, 2018). Surprisingly, few studies have simultaneously examined reactive and control systems, such as immediacy preference and negative emotionality, despite evidence that they likely influence each other (Moran et al., 2013; Rothbart & Bates, 2006). To our knowledge, no previous studies have examined these two processes simultaneously as transdiagnostic risk factors for dimensions of psychopathology across childhood to adolescence. However, one barrier to identifying early mechanisms in the development of later psychopathology is that the behavioral manifestations of psychopathology change across development.

## Heterotypic continuity

A salient issue in developmental psychopathology is that the behavioral manifestations of psychopathology change across development, a phenomenon called heterotypic continuity (Cicchetti & Rogosch, 2002). Heterotypic continuity occurs when the same psychological phenomenon manifests as different behaviors across development (Petersen et al., 2018, 2020). For example, externalizing problems in children often appear as tantruming and overt oppositionality, whereas in adolescents and adults, externalizing behaviors become more covert and tend to take the form of indirect behaviors (e.g., substance use; Mikolajewski et al., 2013; Miller et al., 2009; Petersen & LeBeau, 2022). Patterson (1993) described externalizing problems using the analogy of a "chimera," a mythological creature with a goat's body that, with development, grows the head of a lion and the tail of a serpent. This was meant to highlight that although individual differences in externalizing behavior are relatively stable across time, externalizing behavior manifests in different ways across the lifespan (Patterson, 1993). This notion of changing manifestations of behavior across development has also been identified in internalizing problems (e.g., Avenevoli & Steinberg, 2001; Petersen et al., 2018; Weems, 2008). For example, separation anxiety and fear of animals is common in younger children, whereas social anxiety is more common in adolescence (Weems, 2008).

Consistent with the developmental issues framework (Sroufe, 2016), the changing behavioral manifestation of psychopathology reflects a combination of time-varying genetic and environmental factors, such as school entry transition, in combination with varying developmental tasks and greater experience-dependent capacity. For example, developmental tasks in preschool (e.g., self-regulation) differ from those in adolescence (e.g., peer acceptance), which changes how behavior problems tend to manifest. Thus, for externalizing and internalizing problems, the underlying construct persists across development, but their behavioral manifestations change. However, there are key challenges in identifying early mechanisms in the development of psychopathology.

A key challenge of identifying early mechanisms in the development of psychopathology deals with longitudinal assessment. It is difficult to examine internalizing and externalizing psychopathology across a lengthy span of development in meaningful ways because behavioral manifestations of psychopathology change across development (McElroy et al., 2018). Assessment and analysis become even more challenging when considering that internalizing and externalizing psychopathology often co-occur (Pettersson et al., 2018). Heterotypic continuity poses challenges for measurement because different measures from different informants across ages are needed to capture developmental changes (Petersen & LeBeau, 2022). If the measures do not align with the changes in the construct's manifestation, studies will yield faulty conclusions (Chen & Jaffee, 2015; Petersen et al., 2018, 2021). In addition to using different measures across ages, it is also important to consider using different informants.

## Different informants across development

Before children enter schooling, the most accurate informants on the children's behavior tend to be their parents and caregivers, as reflected in the proliferation of parent- and caregiver-report measures for early childhood (Achenbach & Rescorla, 2000). Many have argued that using multiple informants is the best approach for assessing child psychopathology (De Los Reyes & Makol, 2021, 2022; Makol et al., 2020; Watts, Makol, et al., 2021). In early-to-mid-childhood, when children attend school and preschool, teachers are important informants on children's behavior, because teachers help account for children's behavior across multiple contexts (De Los Reyes & Kazdin, 2005). Context of measurement is important. Prior research has shown that parent and teacher reports of disruptive behaviors that occur in both home and school

tend to result in stronger convergence, whereas context-specific behaviors (i.e., disruptive only at home, not school) result in weaker convergence (De Los Reyes & Makol, 2021; Hartley et al., 2011; Kwon et al., 2012). A review on correlations between mother, father, and teacher ratings of ADHD symptoms in children and adolescents indicated that correlations between mother and father ratings of inattention and hyperactivity–impulsivity were high ($r = .67$–$.70$). Correlations between parent and teacher ratings of inattention and hyperactivity–impulsivity were somewhat lower ($r = .28$–$.47$). Taken together, evidence indicates that context matters in ratings of observed psychopathology symptoms (Martel et al., 2017).

Furthermore, when children enter their adolescent years, adolescents become more reliable reporters on their internal experience, which is particularly meaningful for internalizing symptoms such as anxiety and depression that may be less overtly visible to outside observers (Damme et al., 2022). Taken together, it is important to consider multiple informants on children's externalizing and internalizing symptoms across development to help account for (a) differing manifestations of behavior in multiple contexts (e.g., school and home) and (b) rater-specific bias.

### Developmental scaling

The challenge is in how to meaningfully combine the scores from the different measures as rated by the various informants so that the scores are on a comparable metric for assessing children's change over time. The combination of heterotypic continuity and having different informants across ages poses important challenges. Heterotypic continuity requires age-differing measures to account for the changing manifestation of the construct. Similarly, informants are differentially capable of rating various aspects of the child's behavior (e.g., relational aggression versus social anxiety) in different contexts (e.g., home versus school), thus requiring different measures for different types of informants (e.g., parents, teachers, self report) and at different ages. For example, the Child Behavior Checklist 1.5–5 (Achenbach & Rescorla, 2000) and the Caregiver–Teacher Report Form (Achenbach & Rescorla, 2001) both assess children at the same age range, but the Caregiver–Teacher Report Form includes slightly different question content aimed at examining school rather than home context for behaviors. Furthermore, the Child Behavior Checklist 6–18 (Achenbach & Rescorla, 2000) includes age-differing items (compared to the ages 1.5–5 form) to maintain developmentally appropriate content, e.g., substance use, that assess problems specific to mid-childhood to adolescence. The Teacher's Report Form (Achenbach & Rescorla, 2001) also adjusts its question content to account for development. Therefore, a single informant from one context (e.g., mother in the home context) might only capture one potentially biased view of a given child's behavior (De Los Reyes & Kazdin, 2005). By contrast, including multiple informants across multiple contexts reduces the impact of informant and context-specific biases.

Traditionally, studies have largely ignored heterotypic continuity when examining children's development of psychopathology (Chen & Jaffee, 2015; Petersen et al., 2018, 2021). Many studies have addressed the challenge of differing measures and informants by using (a) only those ages where the same measure or items are assessed and (b) only those informants who provide ratings across the full age span. However, it is problematic to exclude ages because of a developmental change in the manifestation of the construct. Excluding ages due to a change in measurement would exclude important developmental periods and transitions associated with meaningful developmental change, such as the transition from preschool to school entry or the transition to adolescence. Moreover, it is problematic to exclude informants merely because an informant did not provide ratings across all ages of the study. For instance, in a study from early childhood to adolescence, this would result in the exclusion of self report, which is important for assessing internalizing problems in adolescence. It is thus crucial to leverage approaches that use all available information from all possible informants at all possible ages to get the best estimate of people's development on a comparable scale.

Developmental scaling is a recommended approach to ensure statistical equivalence of scores across different measures (Kolen & Brennan, 2014). Developmental scaling approaches have been used successfully to place scores from different measures and informants onto the same scale (Petersen & LeBeau, 2022). Developmental scaling has been widely used in educational psychology to link children's academic achievement scores across ages (e.g., Kenyon et al., 2011; McArdle, 2009; Murayama et al., 2013). However, relatively few studies have used developmental scaling approaches to study social development. For instance, few studies have used developmental scaling to study development of internalizing psychopathology (Petersen et al., 2018) and externalizing psychopathology (Petersen & LeBeau, 2022).

One approach to developmental scaling uses item response theory (IRT). A two-parameter IRT model estimates item difficulty and discrimination. Item difficulty, also called severity, is the point of median probability at which 50% of respondents endorse a given response. Item discrimination is how well the item distinguishes between the high and low levels of a given construct. Based on the items' difficulty and discrimination parameters, one can generate an item characteristic curve, which represents the expected score on the item as a function of the person's level on the latent psychopathology factor. Combining the individual items, one can generate a test characteristic curve, which represents the expected score on the measure as a function of the person's level on the latent psychopathology factor. To link any given pair of measures and raters, IRT uses scaling parameters to minimize the differences between the two test characteristic curves of the common items across the two measures. The scaling parameters are determined as the linear transformation (i.e., intercept and slope parameter) of the test characteristic curves of the common items between the two measures, that, when applied to the second measure, minimize differences between the test characteristic curves of the common items. Essentially, the scaling parameters minimize the differences in the probability of a rater endorsing the age- and rater-common items across the two measures. That is, IRT links measures' scales based on the severity and discrimination of the age- and rater-common items. IRT uses the age- and rater-common items to set the common scale. However, *all* items for a given rater at a given age are used to estimate a person's score on the common scale.

We are aware of only one prior study that has linked scores from different measures and raters across ages (Petersen & LeBeau, 2022). And, to our knowledge, no prior studies have used developmental scaling to link scores from different measures and raters for multiple dimensions of psychopathology. Performing developmental scaling of multiple dimensions of psychopathology leverages the strong covariation between internalizing and externalizing problems to obtain more accurate estimates of each. Researchers have called for studies that implement developmental scaling for multiple dimensions of psychopathology simultaneously (Tackett & Hallquist, 2022).

This approach of having multiple informants and measures and placing their scores on the same scale is essential to accurately estimate changes in externalizing and internalizing symptoms, given considerable differences in informant reports and measure scales (Petersen & LeBeau, 2022). Developmental scaling allows for individuals to have their informant and self-reported psychopathology mapped onto growth curves, to show people's change in psychopathology across a long development span. To our knowledge, only one study has used developmental scaling from multiple informants to account for heterotypic continuity in order to study individuals' development of psychopathology (Petersen & LeBeau, 2022).

### The present study

The present study examines whether immediacy preference and negative emotionality, control and reactive processes, respectively, predict the development of externalizing and internalizing problems across early childhood to adolescence, using a large sample. We leverage multi-informant ratings from mothers, fathers, teachers, afterschool caregivers, other caregivers, and self report. This is the first study to link scores from multiple measures, raters, and psychopathology dimensions onto the same scale. Prior literature has emphasized the importance of accounting for internalizing and externalizing symptoms concurrently to be a more ecologically valid representation of psychopathology (Ruggero et al., 2019). This is the first study to account for heterotypic continuity of multiple dimensions of psychopathology simultaneously to borrow information from each in the estimation of the other, for more accurate estimates given considerable covariation between internalizing and externalizing problems.

Our study has three primary aims: 1) describe the trajectories of internalizing and externalizing problems across a lengthy developmental span. We use an IRT approach to developmental scaling that places the scores from age-differing measures and raters of internalizing and externalizing problems onto the same scale to account for heterotypic continuity and effects of informant type. 2) We aim to evaluate whether measures of reactive and control processes of temperament predict trajectories of internalizing and externalizing problems. The developmental scaling approach allows us to chart children's trajectories of internalizing and externalizing problems across ages 2–15 years, and to examine immediacy preference and negative emotionality as predictors of children's trajectories. 3) Using a bifactor model, we aim to determine whether immediacy preference and negative emotionality factors predict general versus specific psychopathology.

We hypothesize that immediacy preference will predict general psychopathology and specific externalizing problems, but not specific internalizing problems, because immediacy preference has been more strongly associated with externalizing problems compared to internalizing problems in prior work (Campbell & von Stauffenberg, 2009; Krueger et al., 1996). Little work has examined the association between immediacy preference and general psychopathology, but the strong association between immediacy preference and externalizing problems may drive this association. In addition, we hypothesize that temperamental negative emotionality will predict general psychopathology and specific internalizing problems, but not specific externalizing problems. Negative emotionality, like general psychopathology, is often thought to reflect a general liability (Forbes et al., 2019; Phillips et al., 2022). Additionally, the link between negative emotionality and general psychopathology along with internalizing problems has

been established in prior literature (Castellanos-Ryan et al., 2016; Lahey et al., 2021; Lahey, 2009; Olino et al., 2014; Tackett et al., 2013).

## Method

Children ($N = 1,364$) and their families were recruited for the NICHD SECCYD study in 1991 from 31 hospitals near one of 10 locations in the United States: Little Rock, AR; Irvine, CA; Lawrence and Topeka, KS; Boston, MA; Morganton and Hickory, NC; Charlottesville, VA; Seattle, WA; and Madison, WI. Children were recruited at birth and were followed for data collection in four total phases with multiple timepoints at each phase (Phase I, ages 0–3; Phase II, though 1st grade; Phase III, through 6th grade; Phase IV, through 9th grade) until they were 15 years old. The present study involves behavior problem ratings that were assessed near-annually from ages 2–15 years (except ages 13 and 14). The sample was 48% female, 80.4% White, 12.9% Black, 6.1% Hispanic, 1.6% Asian American, 0.4% American Indian, and 4.7% of "other" ethnicity. At intake, the mother's age ranged from 18 to 46 years of age ($M = 28.11$, $SD = 5.63$), 77% of households had fathers living in the home, and there was an average of 4.27 people living in the household ($SD = 1.17$). For more information about the study methods and participants, see NICHD Early Child Care Research Network (2005).

Exclusion criteria included: (1) The mother was younger than 18 years of age at the time of the child's birth; (2) the family did not anticipate remaining in data collection for at least three years; (3) the child had obvious disabilities at birth and/or remained in the hospital for more than seven days after birth; or (4) the mother was not able to speak conversational English. At enrollment, trained researchers visited the family homes, and families were scheduled for periodic data collection. During a given phase, research assistants visited family homes, childcare, and invited families to the laboratory playroom to collect observations and administer study measures.

### Measures

Analysis scripts and a data dictionary of study variables were published at https://osf.io/yz4we/. Descriptive statistics and correlations of study variables are in Table 1.

### Behavior problems

Children's externalizing and internalizing behavior problems were rated by mothers, fathers, teachers, afterschool caregivers, other caregivers (e.g., daycare workers and babysitters), and self report. Ratings from different informants and measures were used, depending on the child's age. Behavior problem ratings were completed on the following Achenbach measures: Child Behavior Checklist 2–3 (CBCL 2–3; Achenbach, 1992), Child Behavior Checklist 4–18 (CBCL 4–18; Achenbach, 1991a, 1991b), Caregiver–Teacher Report Form (C–TRF; Achenbach & Rescorla, 2000), Teacher's Report Form (TRF; Achenbach, 1991a), and Youth Self-Report (YSR; Achenbach, 1991b). The ages at which each rater provided ratings is provided in Table 2. The measures that were completed by each rater at each age are depicted in Figure 1.

Items were rated as 0, 1, or 2 corresponding to "not true," "somewhat or sometimes true," or "very true or often true," respectively. The Externalizing scale of the CBCL 2–3 includes the Aggressive Behavior and Destructive Behavior subscales. The Externalizing scale of the CBCL 4–18, TRF, and YSR includes

**Table 1.** Correlation matrix of model variables

| Variables | Age | Female | African American | Hispanic | INR | Time Waited | Negative Affect | Externalizing | Internalizing |
|---|---|---|---|---|---|---|---|---|---|
| Age | — | | | | | | | | |
| Female | .00 | — | | | | | | | |
| African American | .00 | .00 | — | | | | | | |
| Hispanic | .00 | .00 | −.07*** | — | | | | | |
| INR | .00 | .01*** | −.22*** | −.06*** | — | | | | |
| Time Waited | .00 | .06*** | −.25*** | −.04*** | .20*** | — | | | |
| Negative Affect | .00 | −.04*** | .08*** | .02*** | −.11*** | −.11*** | — | | |
| Externalizing | −.21*** | −.12*** | .11*** | .01 | −.11*** | −.12*** | .21*** | — | |
| Internalizing | −.07*** | .01† | .02*** | .00 | −.06*** | −.05*** | .15*** | .49*** | — |
| Data points | 42,284 | 42,284 | 42,284 | 42,284 | 39,463 | 29,791 | 33,542 | 25,455 | 25,455 |
| Missingness | 0.00 | 0.00 | 0.00 | 0.00 | 6.67 | 29.55 | 20.67 | 39.80 | 39.80 |
| M | 7.90 | 0.48 | 0.13 | 0.06 | 2.86 | 4.48 | 3.97 | −0.09 | −0.10 |
| SD | 3.51 | 0.50 | 0.34 | 0.24 | 2.61 | 3.01 | 0.66 | 1.07 | 1.16 |
| Min | 2.00 | 0.00 | 0.00 | 0.00 | 0.08 | 0.00 | 1.69 | −1.95 | −2.14 |
| Max | 15.00 | 1.00 | 1.00 | 1.00 | 25.08 | 7.00 | 6.25 | 4.93 | 5.98 |
| Skewness | 0.27 | 0.07 | 2.21 | 3.67 | 2.53 | −0.49 | −0.03 | 0.42 | 0.74 |
| Kurtosis | −0.42 | −2.00 | 2.90 | 11.50 | 10.22 | −1.60 | −0.11 | −0.26 | 0.45 |

*Note.* The correlations and descriptive statistics are presented from data are in long format, where each participant has multiple rows: i.e., one row for each informant-by-timepoint combination. "Age" in years; "INR" = income-to-needs ratio; "Min" = lowest score in the sample; "Max" = highest score in the sample. †$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$; all $p$s two-tailed.

**Table 2.** The child's age when each rater provided ratings of the child's behavior problems

|  | Age (years) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rater | 2 | 3 | 4 | 5 | 6* | 7 | 8 | 9 | 10 | 11 | ... | 15 |
| Mother* | x | x | x | x | x | | x | x | x | x | | x |
| Father | | | | | x | | x | x | x | x | | x |
| Teacher | | | | x | x | x | x | x | x | x | | |
| Afterschool Caregiver | | | | | x | | x | x | x | | | |
| Other Caregiver | x | x | x | | | | | | | | | |
| Self-Report | | | | | | | | | | | | x |

*Note.* "x" indicates the measure was collected at the specified age. "*" indicates the referent age and rater.

the Aggressive Behavior and Delinquent Behavior subscales. The Externalizing scale of the C–TRF includes the Aggressive Behavior and Attention Problems subscales. The scales include symptoms of breaking rules, cruelty, aggression, and destruction of property.

The Internalizing scale of the CBCL 2–3 includes the Anxious/Depressed and Withdrawn subscales. The Internalizing scale of the CBCL 4–18, TRF, and YSR includes the Anxious/Depressed, Somatic Complaints, and Withdrawn subscales. The Internalizing scale of the C–TRF includes the Anxious/Obsessive, Fears, and Depressed/Withdrawn subscales.

The Achenbach scales are widely used, and the scores show strong reliability (internal consistency, test–retest reliability, and interrater reliability) and validity (content, construct, and criterion-related validity; Sattler, 2022). Internal consistency estimates by age and rater of the present study are in Supplementary Table S1.

Due to the wide age range of children and adolescents included in the study, we took steps to ensure that the same construct was assessed on the same scale, across time. The number of common items among different measures are in Table 3. To account for developmental changes in both internalizing and externalizing problems, we used an IRT approach to developmental scaling, consistent with previous research (described later; Kolen & Brennan, 2014; Petersen et al., 2018; Petersen & LeBeau, 2022). Developmental scaling linked behavior problem scores across ages and informants onto the same scale. One-year cross-time stability estimates by rater are in Supplementary Table S2. Full descriptive statistics of externalizing and internalizing problems by age and rater are in Table 4, and correlations among internalizing and externalizing problems by rater are in Table 5. Furthermore, the percentage of participants with behavior ratings, by rater type, are in Supplementary Table S3.

## Predictors

### Delay of gratification

Delay of gratification was assessed in the present study with a self-imposed waiting task when the participants were 54 months old. An experimenter elicited the child's preference in treats. Then, the child was told that they would engage in a game where the experimenter would leave the child in the room with the preferred treat. Further, the child was told that if they waited until the experimenter returned, the child could eat the treat and receive an additional portion of treat as reward for waiting the full length of the task. The child was also instructed that if they

ring a bell, that will signal to the experimenter that the child does not want to wait, and that the child would only receive the portion of treats presented and no extra portions. The recorded experimental trial had a 7-minute ceiling, and the child's score was the total number of seconds the child waited until ringing the bell, eating the treat, or 7 minutes (whichever came first). The self-imposed waiting task is one of the most widely used performance-based tasks of inhibitory control and motivational self-regulation in psychological research.

### Negative emotionality

Temperamental negative emotionality was assessed from mother- and other caregiver report at 54 months using a modified version of the Children's Behavior Questionnaire (CBQ; Rothbart et al., 2001). Raters on the CBQ were asked to rate how well each item described the child in the past 6 months using a 7-point scale where 1 = *extremely untrue* and 7 = *extremely true*. Of the 196 items in the original CBQ, mothers completed 80 items and other caregivers completed 48 items. We used the Negative Affect scale on the CBQ, which encompasses the following subscales: Anger/Frustration (10 items), Sadness (10 items), and, among mother report, Fear (10 items). Other caregivers did not complete the Fear subscale. Internal consistency estimates were $\alpha = .82$ for mothers' reports and $\alpha = .93$ for other caregivers' reports. Mothers' ratings were modestly correlated with other caregivers' ratings ($r[761] = .09$, $p = .017$). To incorporate a multi-informant perspective into the estimation of the child's negative emotionality, we averaged ratings across raters. Negative emotionality as assessed by the CBQ has been widely used and has shown strong internal consistency, rank-order stability, and construct validity (Rothbart et al., 2001).

Due to potential inflation of prediction of psychopathology outcomes due to overlap in item content between the CBQ and the numerous measures of child psychopathology, we dropped items from the mother report and other-caregiver-report CBQ that we judged to conceptually overlap with items from the Internalizing or Externalizing scales of the CBCL. We dropped two items from the mother-report CBQ: one item that assessed frustration and one item that assessed sadness. We dropped one item from the other-caregiver-report CBQ that assessed temper tantrums. Two items from the mother report and two items from the other caregiver report had similar item content but were distinct enough to retain, e.g., an item pertaining to getting irritated when making a mistake versus being afraid to make a mistake. In this case, both items conceptually assess behaviors pertaining to making a mistake, but one behavior is anticipatory whereas the other is reactive; thus, we considered them conceptually distinct.

## Covariates

Several demographic characteristics were examined as covariates in the growth curve and bifactor models, including the child's sex (1 = girl, 0 = boy), race, and ethnicity, and the income-to-needs ratio of the child's family. These covariates were selected because they have been shown to robustly impact associations between predictors and child behavior problems (e.g., Petersen et al., 2021; Shi et al., 2020). Race was a dummy coded variable where African American children were compared to other races. Ethnicity was dummy coded such that Hispanic children were compared to non-Hispanic children.
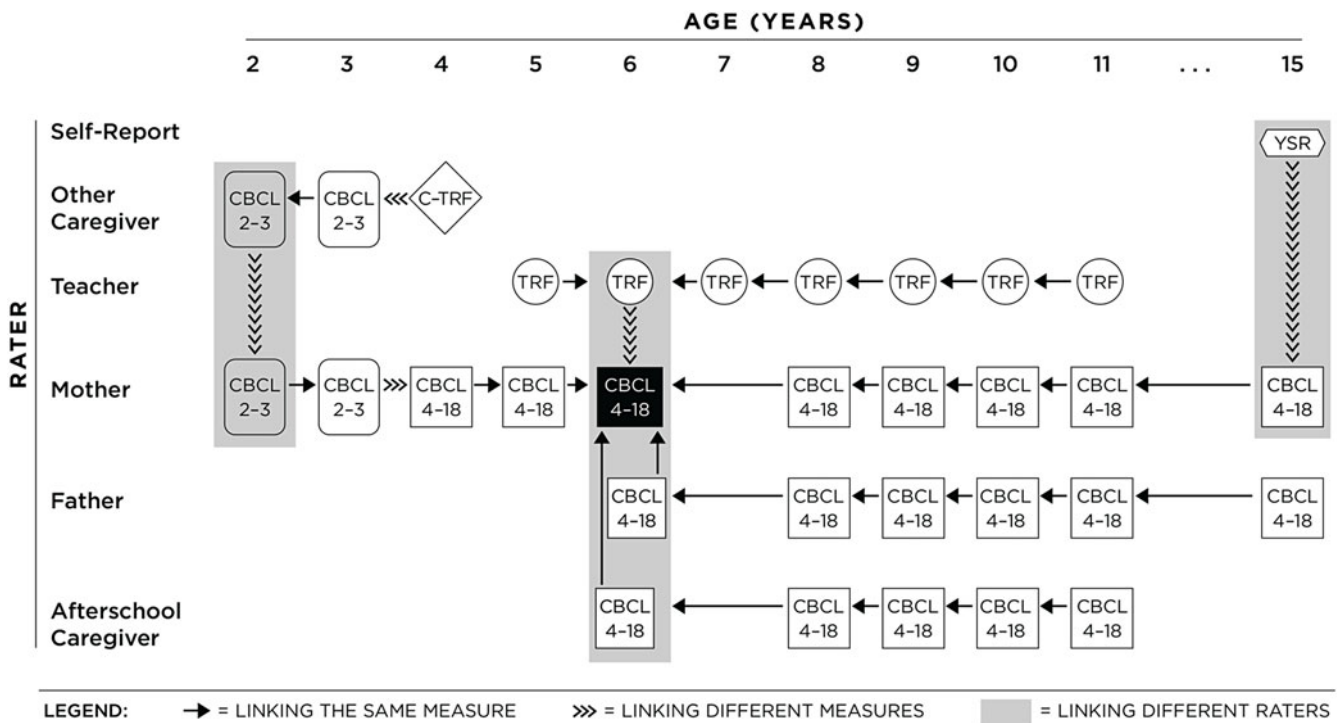
**Figure 1.** Depiction of how the scores from various raters and measures were linked at different ages. *Note*. Raters are depicted in the rows, and the child's age (in years) is depicted in the columns. Different shapes indicate different measures (square = Child Behavior Checklist 4–18; rounded square = Child Behavior Checklist 2–3; circle = Teacher's Report Form; diamond = Caregiver–Teacher Report Form; hexagon = Youth Self-Report). A solid arrow indicates that scores from the same measure were linked using all items (i.e., all items were common items; e.g., mothers' ratings at ages 6 and 8). A broken arrow indicates that scores from different measures were linked using the common items (e.g., mothers' ratings at ages 3 and 4). The direction of the arrow indicates the measure to which the other was linked (e.g., mothers' ratings at age 8 were linked to mothers' ratings at age 6). The solid black box indicates the referent measure (mothers' ratings at age 6) to which every other measure was linked either directly or indirectly. The gray bounding boxes indicate that scores from different raters were linked using the common items (e.g., self-report ratings at age 15 were linked to mothers' ratings at age 15).

## Statistical analysis

### Developmental scale of externalizing and internalizing problems

We used multidimensional IRT (M-IRT) and linking to create a single uniform developmental scale (i.e., developmental scaling) for externalizing and internalizing problems that spans multiple years of development. We conducted this linking in five steps: (1) Fit M-IRT models at each age and for each rater type separately. (2) Link the measures' scores over time within each rater type. (3) Link scores across raters. (4) Calculate latent factor scores on the linked scale. (5) Use linked factor scores in growth curve and bifactor models. We describe this procedure in detail below. Full description of linking details are in Supplementary Appendix S1.

### Step 1. Fit M-IRT models at each age and for each rater type separately

We used the multidimensional graded response IRT model using the mirt package (Chalmers, 2012) in R 3.6.1 (R Core Team, 2022) to estimate item parameters. The mirt package uses a maximum likelihood expectation-maximization algorithm to estimate item parameters. The maximum likelihood estimation procedure uses all available data for each item and provides valid inferences if the data are missing at random or completely at random. The graded response model is a generalized version of the two-parameter logistic model for dichotomous outcomes, accommodating polytomous items that are ordinal in nature. The multidimensional graded response model adds the ability to include multiple latent factors (i.e., externalizing and internalizing problems) – and their covariance – in the same model. This multidimensional graded response IRT model is conceptually like a two-factor categorical confirmatory factor analysis approach (fit to ordinal data) with the internalizing and externalizing latent factors allowed to covary, and with no cross loadings. That is, internalizing and externalizing problem items were included in the same model, but they were allowed to load onto distinct latent factors. The externalizing and internalizing problem items in the current study were questionnaire items rated from 0 to 2. We used the externalizing problems latent factor as the reference group and allowed the mean and variance for internalizing problems latent factor to be estimated freely. Setting the externalizing factor as the reference group, along with linking both internalizing and externalizing items in the same model, placed the internalizing and externalizing problem scores onto the same mathematical scale across ages and raters.

There may be shifts in the externalizing or internalizing problem constructs over time due to natural developmental changes (Petersen et al., 2018). The present study spans a wide age range (ages 2–15 years). When spanning a wide age range, it is considered safer to fit a separate model at each age rather than a single model that spans all ages because a model that spans across a wide age range is more likely to violate IRT dimensionality assumptions (Kolen & Brennan, 2014). We fit two latent factors corresponding to the constructs of interest: i.e., externalizing and internalizing problems. IRT assumes that each latent factor (e.g., externalizing problems) is unidimensional, which is more likely at a single time point than across all time points in the same model.

**Table 3.** The number of common items for each pair of measures

| Measure | CBCL 2–3 | CBCL 4–18 | C–TRF | TRF | YSR |
|---|---|---|---|---|---|
| **Externalizing Problems** | | | | | |
| CBCL 2–3 | 26 | | | | |
| CBCL 4–18 | 9 | 33 | | | |
| C–TRF | 18 | 14 | 40 | | |
| TRF | 10 | 27 | 16 | 34 | |
| YSR | 8 | 30 | 14 | 27 | 30 |
| **Internalizing Problems** | | | | | |
| CBCL 2–3 | 25 | | | | |
| CBCL 4–18 | 8 | 31 | | | |
| C–TRF | 17 | 9 | 34 | | |
| TRF | 8 | 31 | 10 | 35 | |
| YSR | 8 | 29 | 8 | 29 | 31 |

*Note.* "CBCL" = Child Behavior Checklist, "C–TRF" = Caregiver–Teacher Report Form, "TRF" = Teacher's Report Form, "YSR" = Youth Self-Report. The top table presents the number of common items on the Externalizing scale. The bottom table presents the number of common items on the Internalizing scale. Numbers on the diagonal represent the total number of items in the Externalizing scale (top table) or Internalizing scale (bottom table) for that measure (e.g., the CBCL 4–18 has 33 items on the Externalizing scale and 31 items on the Internalizing scale). Numbers below the diagonal represent, for that pair of measures, the number of items that are common to both of the measures. The number of unique items can be calculated by subtracting the number of common items from the total number of items. For instance, the CBCL 4–18 has 6 unique externalizing items when compared with the TRF (i.e., 33 total items minus 27 common items). Conversely, the TRF has 7 unique externalizing items when compared with the CBCL 4–18 (i.e., 34 total items minus 27 common items).

Thus, we fit a separate IRT model at each age and for each rater type in the present study. This approach was also applied by Petersen et al. (2018) and by Petersen & LeBeau (2022) in their creation of a developmental scale for internalizing and externalizing problems, respectively, across a wide age range.

### Step 2. Link the measures' scores over time within each rater type

After successful estimation of the individual IRT models, we used multidimensional linking methodology to create the developmental scale for externalizing and internalizing problems. Developmental scaling is a form of data harmonization that aims to place two measures that assess the same construct but differ based on severity and discrimination onto the same scale. One way to create a developmental scale is to link the two measures. The strength of the linking is enhanced if there are items that overlap across the two measures, often referred to as common items. Developmental scaling based on item parameter invariance theory assumes that any difference in item parameter estimates can be rescaled onto a single unified metric with a linear transformation across adjacent ages. Based on this assumption, the item parameters, and the resulting latent factor scores of externalizing and internalizing problems can be linked across ages by comparing and linearly transforming differences in discrimination and severity of the common items across adjacent ages.

We used multidimensional developmental scaling techniques to link the measures' scores over time within each rater type. We used the plink package (Weeks, 2010) in R to perform the linking by using the multidimensional test characteristic function procedure with an oblique Procrustes rotation (Oshima et al., 2000). The oblique rotation method allowed the latent factors – externalizing and internalizing problems – to be correlated. For

linking, we used a multidimensional Stocking-Lord procedure (Stocking & Lord, 1983). The Stocking-Lord linking procedure iteratively estimates linking constants by minimizing differences in the aggregate scores across common items.

To estimate the Stocking-Lord parameters, we set the reference age at 6 years for each rater because age 6 was the first age when most rater types (except other caregivers and self report) provided ratings of the child's externalizing and internalizing problems. We set the reference rater to be the mother because the mother typically provided the most ratings across the developmental age span. The reference age and rater pair set the scale to which the item parameters at subsequent ages and for other raters were transformed. In other words, we transformed the estimated item parameters at all ages and for all raters to be on the same scale as the item parameters estimated for mothers' ratings at 6 years of age. To achieve this, we first linked the item parameters across ages within rater type. We performed the process of linking iteratively by chaining together multiple linking constants across the age span. First, for a given rater type, we estimated Stocking-Lord linking constants that linked the item parameters at age 7 to be on the same scale as that rater type's item parameters at age 6. We estimated additional linking constants between adjacent age spans, for example between 5 and 6 years of age, 7 and 8 years of age, and so on. We used two estimated scaling constants including an intercept parameter, B, and a slope parameter, A, to link the item parameters onto the reference scale.

After successfully estimating the linking constants, we then transformed all item parameters to be on the age 6 scale for the given rater. Min (2007) provides further technical details on the multivariate linking terms. To shift all item parameters to a common age 6 scale, we applied all previous adjacent scaling constants to the item parameters. For example, when shifting the item parameter estimates for 7-year-olds to the age 6 scale, we used a single set of scaling constants. However, when shifting the item parameters for 8-year-olds, we used two sets of scaling constants: first, we transformed the item parameter estimates for 8-year-olds to the scale of the 7-year-olds, and then we transformed them a second time to be on the age 6 scale. See Figure 1 for a visualization of the linking process. We performed this step of the linking process separately for each row in the figure (i.e., within rater types; horizontal arrows).

### Step 3. Link scores across raters

After creating developmental scales across ages within rater types, we linked scores across raters at age 6 (except for the other caregivers' reports collected at age 2 and self report collected at age 15). As described above, we set the mother as the reference rater. We used a similar process as in step 2; we estimated Stocking-Lord linking constants to link the item parameters across raters within a single age. For example, we estimated a set of linking constants to link the item parameters of the fathers' ratings to the item parameters of mothers' ratings at age 6 to ensure that their factor scores were on the same scale. This step moved the developmental scales for fathers, teachers, and afterschool caregivers to the mothers' scale, anchored at age 6, while preserving the developmental scale created within rater types in step 2. The process of linking scores across raters is depicted in Figure 1 with the gray bounding boxes (vertical arrows).

### Step 4. Calculate latent factor scores on the linked scale

After successfully placing item parameter estimates onto a single developmental scale (for all raters and ages), we calculated

**Table 4.** Descriptive statistics of externalizing and internalizing problems by age and rater

| | Age (Years) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *M* | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 15 |
| Mother | 0.73 | 0.61 | 0.62 | 0.07 | 0.00 | – | −0.14 | −0.30 | −0.33 | −0.38 | −0.49 |
| | −0.09 | 0.03 | 0.42 | −0.06 | 0.00 | | 0.06 | 0.07 | 0.11 | 0.08 | 0.03 |
| Father | – | – | – | – | −0.11 | – | −0.33 | −0.37 | −0.50 | −0.42 | −0.47 |
| | | | | | −0.10 | | −0.08 | −0.10 | −0.13 | −0.19 | −0.22 |
| Teacher | – | – | – | −0.28 | −0.21 | −0.20 | −0.10 | −0.20 | −0.09 | −0.30 | – |
| | | | | −0.92 | −0.74 | −0.75 | −0.54 | −0.57 | −0.62 | −0.65 | |
| Afterschool Caregiver | – | – | – | – | −0.09 | – | −0.27 | −0.32 | −0.44 | – | – |
| | | | | | −1.22 | | −1.30 | −1.39 | −1.40 | | |
| Other Caregiver | 0.41 | 0.30 | −0.55 | – | – | – | – | – | – | – | – |
| | 1.21 | 1.38 | 1.07 | | | | | | | | |
| Self-Report | – | – | – | – | – | – | – | – | – | – | 0.19 |
| | | | | | | | | | | | 0.77 |

| | Age (Years) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *SD* | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 15 |
| Mother | 0.79 | 0.80 | 0.79 | 0.89 | 0.93 | – | 0.93 | 0.92 | 0.96 | 0.95 | 0.98 |
| | 1.00 | 1.01 | 1.03 | 0.96 | 0.94 | | 1.03 | 1.03 | 1.04 | 1.06 | 1.04 |
| Father | – | – | – | – | 0.98 | – | 0.94 | 1.01 | 1.04 | 1.04 | 1.04 |
| | | | | | 0.96 | | 0.98 | 0.98 | 1.04 | 1.03 | 1.09 |
| Teacher | – | – | – | 1.13 | 1.15 | 1.15 | 1.22 | 1.17 | 1.18 | 1.20 | – |
| | | | | 0.82 | 0.86 | 0.90 | 0.93 | 0.90 | 0.89 | 0.91 | |
| Afterschool Caregiver | – | – | – | – | 0.98 | – | 1.00 | 0.97 | 1.01 | – | – |
| | | | | | 0.65 | | 0.59 | 0.60 | 0.67 | | |
| Other Caregiver | 0.94 | 1.00 | 1.08 | – | – | – | – | – | – | – | – |
| | 1.34 | 1.29 | 1.36 | | | | | | | | |
| Self-Report | – | – | – | – | – | – | – | – | – | – | 0.92 |
| | | | | | | | | | | | 1.18 |

*Note.* "–" indicates not applicable because the particular rater did not provide ratings at the given time point. Means and standard deviations (*SD*s) for externalizing problems are the top number in each box, whereas means and *SD*s for internalizing problems are the bottom number.

**Table 5.** Correlation matrix of externalizing problem scores (below diagonal) and internalizing problem scores (above diagonal) by rater

| Rater | Mother | Father | Teacher | Afterschool Caregiver | Other Caregiver | Self-Report |
|---|---|---|---|---|---|---|
| Mother | .56*** | .42*** | .23*** | .23*** | .13*** | .29*** |
| Father | .57*** | .66*** | .19*** | .22*** | n/a | .29*** |
| Teacher | .33*** | .33*** | .35*** | .20*** | n/a | n/a |
| Afterschool Caregiver | .39*** | .41*** | .44*** | .55*** | n/a | n/a |
| Other Caregiver | .21*** | n/a | n/a | n/a | .57*** | n/a |
| Self-Report | .32*** | .33*** | n/a | n/a | n/a | .48*** |

*Note.* "n/a" indicates not applicable because the two raters did not provide ratings at the same time point(s). Diagonal represents within-informant correlations between externalizing and internalizing problems scores.
†$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$; all *p*s two-tailed.

children's latent externalizing and internalizing problem scores with expected a posteriori factor scores. The linking in the previous two steps scaled the factor scores to be on the single developmental scale while retaining changes in means and variances over time and across raters. The linking constants by measure and age are in Supplementary Table S4.

In sum, the linking of scores within a rater type created a developmental scale for scores from that rater type, so each rater type had their own trajectory (see Figure 2). We then, ultimately, linked each rater type's developmental scale (directly or indirectly) to the mothers' ratings at age 6, so that each rater type's trajectory was on the same developmental scale. Examples of linked scores
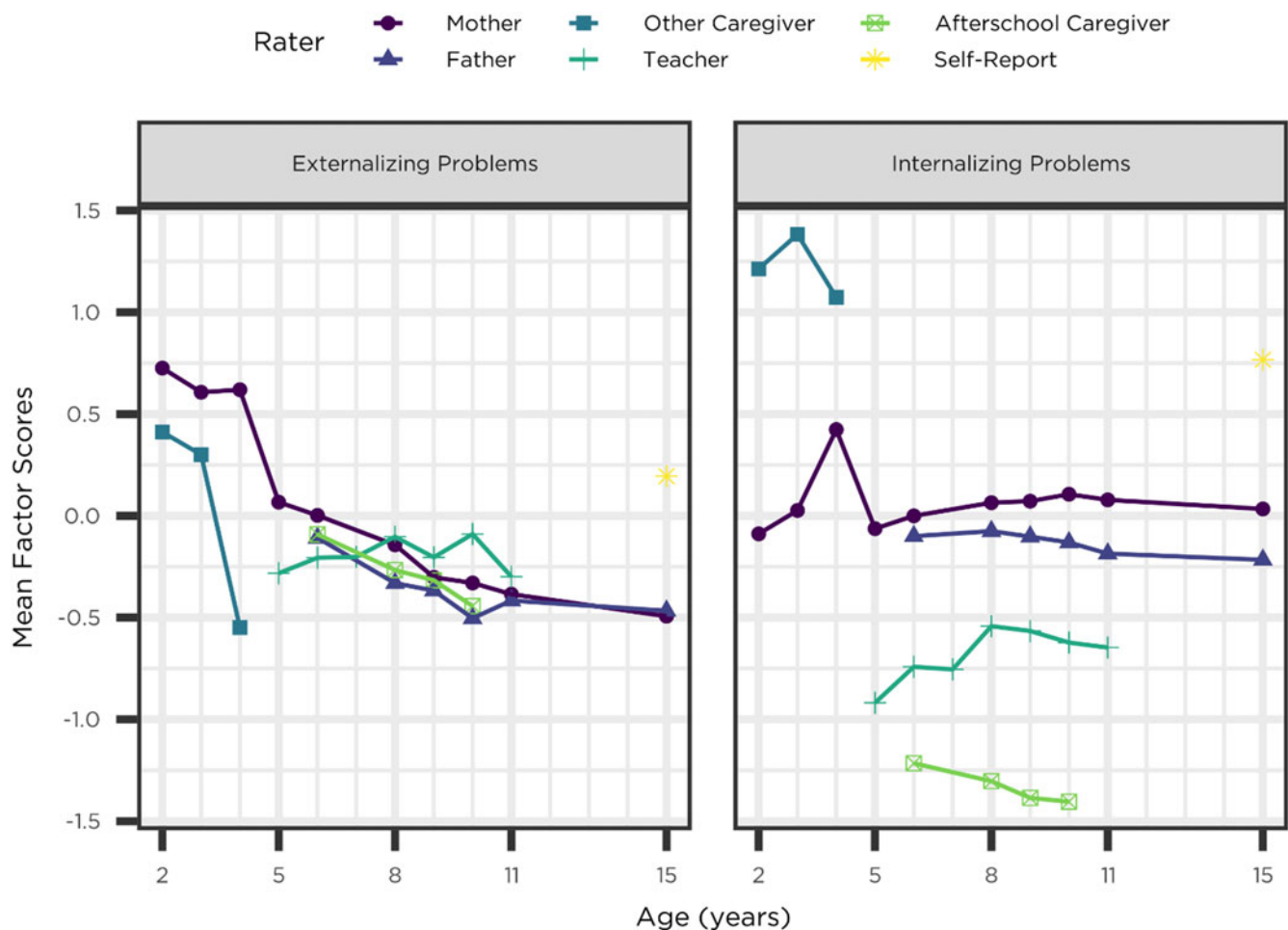
**Figure 2.** Mean factor scores by age, rater, and type of behavior problem.

across raters and years are depicted with test characteristic curves in Supplementary Figures S1 through S4. The test characteristic curves of the linked scores across raters and years were highly similar (and more similar than the test characteristic curves of the pre-linked scores), indicating that we successfully linked scores across raters and years to be on the same scale.

*Differential item functioning.* Post-linking estimates of scale-level DIF between measures used to link scores across different raters and ages are in Supplementary Table S5. Tests of differential item functioning (DIF) by age showed no major concerns at the scale level after linking (see Supplementary Appendix S2). The distribution of DIF effect size statistics between ages by rater type are in Supplementary Figure S5. We also conducted analyses to examine potential differential item functioning by sex and race (see Supplementary Appendix S3). There was some differential item functioning between males and females, and between Whites and Non-Whites. However, most instances of DIF were differences in severity (i.e., uniform DIF) rather than differences in discrimination (non-uniform DIF). Thus, covariate adjustment within the growth models should adjust for group-level differences in the factor score differences by sex and race.

*Secondary analysis of aggression and delinquent behavior.* As a secondary analysis, we also examined aggression and delinquent

subdimensions of externalizing problems given their differing associations with risk factors (J. Murray & Farrington, 2010; Wall & Barth, 2005). Thus, we also conducted developmental scaling with aggression and delinquent behavior (see Supplementary Appendix S4).

*Step 5. Use linked factor scores in growth curve and bifactor models.* After linking factor scores from all raters and at all ages to be on the scale of mothers' ratings at age 6, we used the linked factor scores as the child's estimated level of behavior problems for a given rater and age in subsequent growth curve and bifactor models.

### Growth curve modeling

We modeled children's trajectories of behavior problems using mixed models. We estimated mixed models using the lmer function of the lme4 package (D. Bates et al., 2015) in R. We modeled externalizing problems and internalizing problems separately given their different developmental courses. Because our goal was to predict behavior problems in adolescence, we set the intercepts to be at the last time point (age 15), consistent with prior research. First, we established the form of change of children's behavior problems over time by comparing linear, quadratic, and cubic polynomials. To reduce the potential for multicollinearity among the polynomial terms, we used orthogonal

polynomials calculated using the poly function in R. We used chi-square difference tests to compare nested models.

Upon establishing the form of the trajectory, we added dummy-coded fixed effects for the rater type: mother, father, teacher, afterschool caregiver, other caregiver, or self report. We set mother report to be the reference group because mothers provided the most ratings on average. Then, we added fixed effects for the covariates. After adding covariates, we then added the focal predictors of interest, the child's negative emotionality and delay of gratification. For parsimony and interpretability, we examined the predictors in relation to the intercepts and linear slopes. Model formulas are in Supplementary Appendix S5.

We determined the importance of focal predictors using $R^2$ statistics to evaluate how much variation in behavior problems was explained by the rater predictors, demographic predictors, and focal predictors of interest. We computed $R^2$ statistics defined by Nakagawa and Schielzeth (2013). In Supplementary Appendix S6, we describe tests of systematic missingness and how missing data were handled.

### Bifactor model

We estimated a bifactor model at the last timepoint in the study (i.e., age 15) to determine whether negative emotionality and delay of gratification predicted general versus specific psychopathology. First, we fit a bifactor model at age 15 years with only externalizing and internalizing items and no predictors. The bifactor model included a latent factor for the general factor of psychopathology, in addition to latent factors for externalizing problems and internalizing problems. The latent factors for internalizing and externalizing problems for each participant were estimated using the developmentally scaled factor scores from all informant types who provided ratings at age 15 (mother-, father-, and self report). The latent factors were set to be uncorrelated, so the general factor represented the covariation among all externalizing and internalizing items. By contrast, the specific psychopathology factors – i.e., externalizing problems and internalizing problems – represented the covariation among the items within that dimension after extracting the variance accounted for by the general factor. The externalizing latent factor represented unique externalizing variance, and the internalizing latent factor represented unique internalizing variance. Upon establishing a well-fitting bifactor model, we examined the focal predictors of interest. Predictors were allowed to predict the three latent factors. Then, we added covariates.

Bifactor models were fit in lavaan (Rosseel, 2012) in R. Models with diagonally weighted least squares (WLSMV) were unable to be estimated due to sparse cells in some response categories for some items. Therefore, models used maximum likelihood estimation and a probit link with robust standard errors (MLR-probit) to account for the nonnormally distributed data, which has shown comparable power to WLSMV and better control for Type I error when using ordinal data (Cuhadar & Kalkan, 2023; Suh, 2015). Missing data were handled with full information maximum likelihood estimation, which uses all available data and is the gold standard approach for handling missingness when data are missing at random or completely at random. We evaluated model fit with the root mean square error of approximation (RMSEA), Robust estimate of comparative fit index (CFI), and standardized root mean square residual (SRMR). Model fit was considered good if RMSEA ≤ .05, CFI ≥ .95, and SRMR ≤ .08; model fit was considered acceptable if RMSEA ≤ .08, CFI ≥ .90, and SRMR ≤ .10

(Bentler & Bonett, 1980; Hu & Bentler, 1999; Schermelleh-Engel et al., 2003; Schreiber et al., 2006).

Due to the sparse cells in some response categories for some items, we structured the multi-informant ratings at age 15 in long form by rater, to leverage ratings from all raters at age 15: mothers, fathers, and self-report. That is, each participant had up to three rows (one for each rater). To account for the non-independence of multiple observations per participant, we used the participant as a cluster variable, which calculates robust standard errors using a Huber-White sandwich estimator (Huber, 1967; White, 1980).

The scale of each latent factor was set using the effects coding method (Little et al., 2006) so that the average of the items' factor loadings was equal to one. Composite reliability, indexed by coefficient omega, was estimated using the semTools package (Jorgensen et al., 2022) in R.

### Sensitivity analyses

We conducted several sensitivity analyses. Methods of the sensitivity analyses are in Supplementary Appendices S7 and S8. In a secondary analysis, we also included the child's early cognitive ability as an additional covariate (see Supplementary Appendix S7).

## Results

The present study sought to achieve three aims: First, we sought to describe people's trajectories of externalizing and internalizing problems from ages 2–15 years after performing developmental scaling to put the scores from the different measures, informants, and constructs onto the same scale. Second, we examined whether negative emotionality and immediacy preference predict the slopes from ages 2–15 and intercepts (i.e., ending levels) of people's trajectories of externalizing and internalizing problems at age 15. Third, we examined whether negative emotionality and immediacy preference predict general psychopathology versus specific psychopathology – i.e., unique externalizing problems or unique internalizing problems – at age 15.

### Aim 1: Describe the trajectories
#### Unconditional means model
An unconditional means model (i.e., random intercepts) demonstrated that individual differences in intercepts accounted for 34% of the variance in children's trajectories of externalizing problems and 18% of the variance in children's trajectories of internalizing problems. A model with random intercepts and random linear slopes fit better than a model with only random intercepts and accounted for 40% of the variance in externalizing problems and 22% of the variance in internalizing problems (externalizing: $\Delta\chi^2[3] = 1,809.63$; internalizing: $\Delta\chi^2[3] = 408.11$; $ps < .001$).

#### Functional form
To determine the best-fitting functional form, we compared models with random linear, quadratic, and cubic slopes. A model with random linear and quadratic slopes fit better than a model with only random linear slopes (externalizing: $\Delta\chi^2[4] = 1,543.94$; internalizing: $\Delta\chi^2[4] = 1,937.47$; $ps < .001$). A model with random cubic slopes did not converge due to small variance of the random cubic term. Thus, we selected the quadratic model as the best-fitting functional form of growth.

### Rater type

Then, we added the rater type (e.g., mother, father, teacher) as a dummy-coded predictor of the trajectories to account for systematic differences as a function of rater type. Mothers served as the reference rater group. The model with rater type as a predictor fit better than a model without rater type (externalizing: $\Delta\chi2[13] = 2,927.75$; internalizing: $\Delta\chi2[13] = 8,780.80$; $ps < .001$). The model with rater type predicting linear and quadratic slopes fit better than models with rater type predicting only linear slopes (externalizing: $\Delta\chi2[4] = 417.34$; internalizing: $\Delta\chi2[4] = 69.95$; $ps < .001$). Thus, for the baseline growth model, we selected the model in which the rater type predicted the intercepts, linear slopes, and quadratic slopes to allow different curvature by rater type.

### Baseline growth model

Individuals' growth curves from the baseline growth model are depicted in Figure 3. Model results are in Supplementary Table S6.

For externalizing problems, the model explained 47% of the variance (fixed effects: 10%; random effects: 37%). Ratings by fathers, teachers, and afterschool caregivers had lower intercepts compared to ratings by mothers. Self-report had higher intercepts than ratings by mothers. When setting the intercepts to the first timepoint when the target informant type provided ratings, ratings by other caregivers had higher intercepts than ratings by mothers (see Supplementary Appendix S9). In summary, mothers tended to rate their child as showing more externalizing problems than did fathers, afterschool caregivers, and teachers (except for teachers' ratings after age 8; see Figure 2); however, mothers tended to rate their child as showing fewer externalizing problems than did other caregivers or self-report. On average, externalizing problems decreased across ages 2–15, but trajectories differed by rater. Mothers' ratings showed average decreases across ages 2–15, whereas teachers' ratings showed modest average increases from ages 4–8 and then stayed relatively stable with slight declines from ages 8–11.

For internalizing problems, the model explained 47% of the variance (fixed effects: 24%; random effects: 23%). Ratings by fathers, teachers, and afterschool caregivers had lower intercepts compared to ratings by mothers. Self-report had higher intercepts than ratings by mothers. When setting the intercepts to the first timepoint when the target informant type provided ratings, ratings by other caregivers had higher intercepts than ratings by mothers (see Supplementary Appendix S9). In summary, mothers tended to rate their child as showing more internalizing problems than did fathers, afterschool caregivers, and teachers; however, mothers tended to rate their child as showing fewer internalizing problems than did other caregivers or self-report. On average, internalizing problems showed decreases from ages 2–10 and increases from ages 10–15, but trajectories differed by rater. Mothers' ratings were relatively stable from ages 2–15, whereas teachers' ratings showed modest average increases from ages 4–8 and then stayed relatively stable from ages 8–11.

### Demographic covariates

Then, we added demographic characteristics to the model as covariates. Model results are in Supplementary Table S7. For externalizing problems, the fixed effects explained 12% of the variance (Δ2%). Compared to girls, boys had higher intercepts but did not differ in slopes. Relative to other races, African American children had higher intercepts but did not differ in slopes. Relative to non-Hispanic children, Hispanic children did not differ in intercepts but showed marginally significant differences in slopes

in terms of shallower decreases with age. A lower income-to-needs ratio was associated with higher intercepts but not different slopes. In other words, boys, African Americans, and children from families with a lower income-to-needs ratio had higher ratings of externalizing problems. Compared to non-Hispanic children, Hispanic children showed lesser decreases in ratings of externalizing problems across development.

For internalizing problems, the fixed effects in the baseline growth model explained 24% of the variance (Δ <1%). Compared to boys, girls had higher intercepts and had steeper increases over time. Relative to other races, African American children had steeper decreases over time but did not differ in intercepts. Relative to non-Hispanic children, Hispanic children did not differ in intercepts or slopes. A lower income-to-needs ratio was associated with higher intercepts but not different slopes. In other words, girls and children from families with a lower income-to-needs ratio had higher ratings of internalizing problems. Girls showed steeper increases in internalizing problems compared to boys; African Americans showed steeper decreases in internalizing problems compared to non-African Americans.

### Aim 2: Predicting the trajectories

Then, we added negative emotionality and delay of gratification as predictors in the model. Model results are in Supplementary Table S8.

For externalizing problems, the fixed effects in the baseline growth model explained 17% of the variance. Thus, the predictors of interest collectively accounted for ~5% of additional variance. Higher negative emotionality was associated with higher intercepts but steeper declines in externalizing problems over time. Greater delay of gratification was associated with lower intercepts but not with differences in slopes. Prototypical growth curves of externalizing problems as a function of delay of gratification are in Figure 4.

For internalizing problems, the fixed effects in the baseline growth model explained 26% of the variance. Thus, the predictors of interest collectively accounted for ~3% of additional variance. Higher negative emotionality was associated with higher intercepts but steeper declines in internalizing problems over time. Greater delay of gratification was associated with lower intercepts but not with differences in slopes.

### Sensitivity analysis results: growth curve models

For results of sensitivity analyses of growth curves, see Supplementary Appendix S9. Early cognitive ability was associated with lower intercepts of both internalizing and externalizing problems. When accounting for early cognitive ability as a covariate, its presence attenuated the previously significant associations between delay of gratification and intercepts of internalizing and externalizing problems. When examining mother-only ratings of internalizing and externalizing problems, negative emotionality and delay of gratification did not predict differences in slopes in internalizing and externalizing problems that were present when examining ratings from all informants. Results regarding intercepts did not differ significantly.

When excluding ratings prior to 54 months of age, results did not substantively change. When examining anger/frustration and fear, separately, instead of a negative emotionality composite, anger/frustration was strongly associated with higher externalizing and internalizing problem intercepts, and steeper declines in externalizing problems and, at a trend-level, internalizing problems. By contrast, fear was not associated with intercepts or
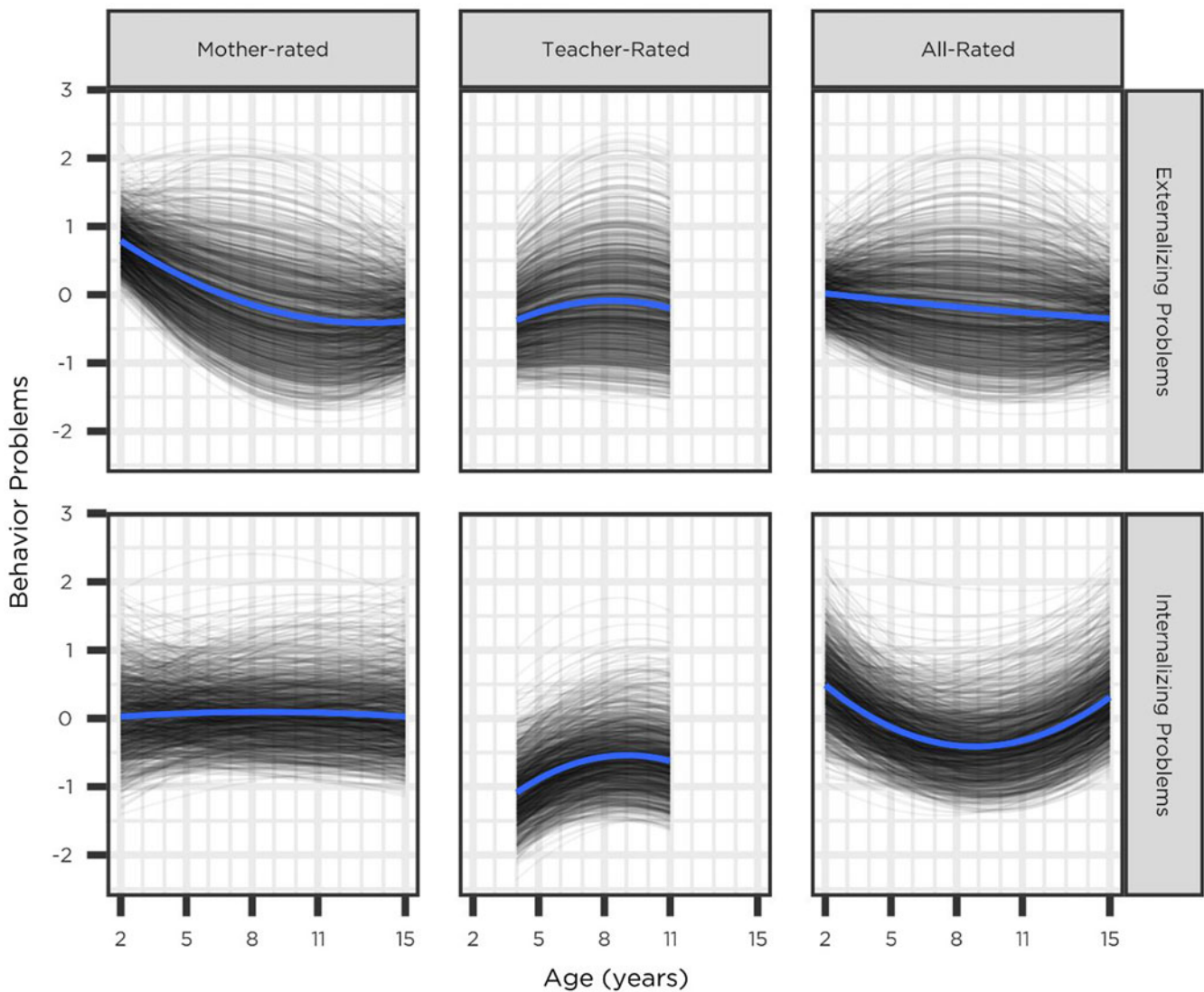
**Figure 3.** Children's model-implied growth curves of internalizing and externalizing problems by rater. *Note.* "All-Rated" refers to the model-implied ratings for the "average" rater.

slopes of externalizing problems but was strongly associated with higher intercepts and steeper declines in internalizing problems. When examining mother and caregiver report of negative emotionality separately, mother-reported negative emotionality was associated with higher intercepts of internalizing and externalizing problems, and steeper decreases in internalizing but not externalizing problems. Caregiver report had similar results as mother report, except that caregiver-reported negative emotionality predicted steeper declines in internalizing problems at a trend-level. Finally, when examining aggressive and delinquent behavior separately, associations with predictors and aggressive behavior factor did not substantively change from the primary analyses. Associations between predictors and delinquent behaviors also did not substantively change from the primary analyses.

### Aim 3: Predicting general versus specific psychopathology at age 15

Given associations between the risk factors and ending levels of externalizing and internalizing problems at age 15, we examined

the risk factors in relation to general versus specific psychopathology at age 15. The bifactor model with all items (externalizing: 33 items; internalizing: 33 items) did not fit. Thus, we modified the model by dropping items with low endorsement rates (externalizing: 7 items; internalizing: 0 items), leaving 26 externalizing items and 33 internalizing items. Then, we removed loadings onto the specific factors that were not significant and positive (externalizing: 8 items; internalizing: 1 item), leaving 18 externalizing items and 32 internalizing items that loaded onto the respective specific factor. The model fit well according to RMSEA (.051) and SRMR (.050) but did not fit well according to CFI (.796). Thus, we made model modifications.

We allowed item residuals to be correlated for which the modification index was large ($\Delta\chi2 > 20$), indicating local non-independence of items, if the modification was also consistent with theory (i.e., both items were within the same domain). This led to 104 correlated residuals out of 649 possible within-domain correlated residuals. The model fit well according to RMSEA (.034) and SRMR (.047) and showed acceptable fit according to CFI (.907).
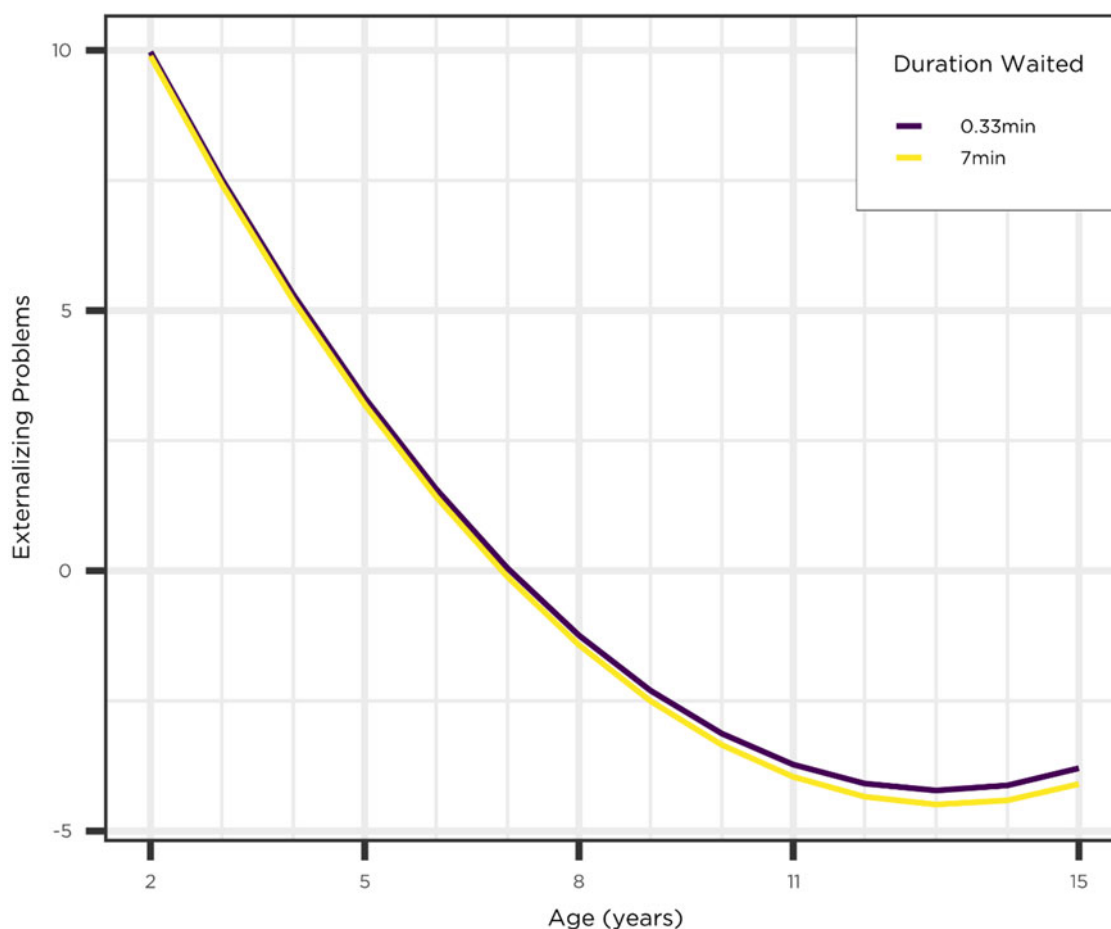
**Figure 4.** Children's model-implied externalizing problems trajectory as a function of duration of time waited in a self-imposed waiting task.

Items' factor loadings on the general factor and specific externalizing and internalizing factors are in Supplementary Table S9. The proportion of variance in ratings that was accounted for by the general factor (i.e., the explained common variance, ECV) was .69. The ECV of specific factors (ECVs) for the specific externalizing factor and specific internalizing factor was .09 and .22, respectively. Coefficient omega was .73, .21, and .42 for the general factor, specific externalizing factor, and specific internalizing factor, respectively. Thus, the reliability was acceptable for the general factor but was low for the specific factors, as evidenced by items' relatively weak factor loadings on the specific factors. Findings in relation to unique externalizing and unique internalizing problems should therefore be interpreted carefully.

Then, we added the predictors: negative emotionality and delay of gratification. Regression coefficients are in Supplementary Table S10. Negative emotionality was positively associated with the general factor and unique externalizing problems; it was not significantly associated with unique internalizing problems. Delay of gratification was negatively associated with the general factor but was positively associated with unique internalizing problems; it was not significantly associated with unique externalizing problems.

Then we added covariates. Regression coefficients are in Supplementary Table S11. The association between negative emotionality and the general factor remained significantly associated even when controlling for covariates. However, the association between negative emotionality and unique externalizing was attenuated to trend-level significance after controlling for covariates. The association between negative emotionality and unique internalizing problems remained nonsignificant. The association between delay of gratification and externalizing problems remained nonsignificant. However, delay of gratification was no longer significantly associated with the general factor or unique internalizing problems after controlling for covariates When compared to boys, girls showed lower ratings of general psychopathology, and higher ratings of unique externalizing and internalizing problems. When compared to non-African Americans, African Americans showed lower ratings of unique internalizing and externalizing problems and, to a trend-level, higher ratings of general psychopathology. When compared to non-Hispanics, Hispanics showed higher ratings of unique externalizing problems and, to a trend-level, general psychopathology. A higher income-to-needs ratio was associated with lower ratings of general psychopathology and higher ratings of unique internalizing problems, at a trend-level. Compared to mother report, fathers rated their child as showing greater general psychopathology, but lower unique externalizing problems unique internalizing problems, all at a trend-level. Compared to mother report, self-report showed higher ratings of general psychopathology and unique internalizing and externalizing problems.

## Sensitivity analysis results: Bifactor models

Full results from the bifactor model sensitivity analyses are in Supplementary Appendix S10. When including early cognitive ability as an additional covariate in a bifactor framework, the association between delay of gratification and general and specific psychopathology remained nonsignificant. The association between negative emotionality and specific externalizing problems became statistically significant, after previously being associated at a trend-level. Regression coefficients, including early cognitive ability, are in Supplementary Table S12. When examining anger/frustration and fear, separately, as predictors of general and specific psychopathology, anger/frustration was not associated with unique internalizing problems, but predicted general psychopathology and unique externalizing problems more strongly than did fear. Fear was not significantly associated with general or specific psychopathology. Mother report of negative emotionality significantly predicted general and unique externalizing problems, and, at a trend-level, unique internalizing problems. Caregiver report of negative emotionality did not predict general or unique behavior problems. When examining predictors of unique aggressive and delinquent behavior, negative emotionality predicted unique delinquent behavior, but not unique aggressive behavior. Delay of gratification was not significantly associated with unique aggressive or delinquent behavior. Finally, when separately estimating bifactor models derived from mother report and self report, model fit according to CFI was poor for self-report but was acceptable for mother report. Predictors and covariates were added to these analyses and the full results are in Supplementary Appendix S10. However, we caution the interpretation of the findings from the self-report model due to poor model fit. For mother report, negative emotionality was positively associated with general psychopathology and unique externalizing problems, but was not associated with unique internalizing problems. Delay of gratification was negatively associated with general psychopathology and was positively associated with unique internalizing problems, but was not associated with unique externalizing problems.

## Discussion

The present study had three goals: First, we aimed to describe people's trajectories of externalizing and internalizing problems from ages 2–15 years while accounting for heterotypic continuity via developmental scaling. Second, we examined whether negative emotionality and immediacy preference predicted the slopes from ages 2–15 and intercepts (i.e., ending levels) of people's trajectories of externalizing and internalizing problems at age 15. Third, we examined whether negative emotionality and immediacy preference predicted general psychopathology versus specific psychopathology at age 15.

## Developmental trajectories

The present study is the first to chart the trajectories of externalizing and internalizing psychopathology concurrently across a lengthy developmental span while linking ratings from multiple raters. Similar to previous findings, ratings of externalizing problems decreased across development (e.g., Keiley et al., 2000; Leve et al., 2005; Petersen et al., 2015). Boys showed higher levels of externalizing problems than girls at age 15, but they did not differ in slopes.

For internalizing problems, results indicated an average decrease from ages 2 to 10 and an average increase from ages 10 to 15 years. An increase in internalizing symptoms at around 10–15 years of age maps onto pubertal development, and is expected given previous research (e.g., Hamlat et al., 2019). Mothers tended to endorse higher levels of both internalizing and externalizing problems (for the child/adolescent) compared to ratings by fathers, teachers, and afterschool caregivers; mothers tended to endorse lower rates of internalizing and externalizing problems compared to other caregivers and self-report. Mothers' ratings of internalizing problems tended to be relatively stable (on average) from 2 to 15 years of age, whereas teachers' ratings showed average increases from 4 to 8 years, and then were relatively stable (on average) from 8 to 11 years of age. Thus, developmentally scaled trajectories of internalizing and externalizing problems from ages 2 to 15 years were consistent with the expected age-related decrease in externalizing problems, and with the age-related increase in internalizing problems that aligns with pubertal development. Girls showed steeper increases in internalizing problems than boys and showed higher levels at age 15.

## Reactive and control processes as predictors of trajectories

We examined only one dimension of reactive (negative emotionality) and one dimension of control (immediacy preference) processes as predictors of psychopathology trajectories. Higher informant-reported negative emotionality assessed as the average of mother and other caregiver report at 54 months was associated with steeper declines in externalizing and internalizing problems across development but remained associated with higher ending levels of externalizing and internalizing problems at age 15, controlling for covariates – though effect sizes were small. These results indicate that negative emotionality assessed in early childhood is associated with relatively high levels of internalizing and externalizing problems that endure across childhood to adolescence. In a sensitivity analysis examining the subscales of negative emotionality – fear and anger/frustration – in place of negative emotionality, they showed differential associations with trajectories of internalizing and externalizing problems. As would be expected based on prior studies (e.g., Crockett et al., 2018), anger/frustration was associated with externalizing problems but also internalizing problems, whereas fear was strongly associated with only internalizing problems. These results highlight the importance of considering multiple aspects of negative emotionality.

Immediacy preference assessed at 54 months was associated with higher ending levels of externalizing and internalizing problems at age 15, controlling for covariates. That is, a delay preference, i.e., longer wait times in a self-imposed waiting task, was associated with lower ratings of externalizing and internalizing problems at age 15 – though effect sizes were small. The association between immediacy preference and externalizing problems is consistent with previous research (e.g., Krueger et al., 1996; Mischel et al., 1989; Peake, 2017). However, immediacy preference was not associated with changes in externalizing or internalizing problems across development. These results point to the small, but potentially meaningful, association of specific facets of reactive and control processes with levels of internalizing and externalizing psychopathology across childhood and adolescence.

Predictors of mother-reported trajectories differed from predictors of trajectories derived from multiple raters, such that negative emotionality did not predict slopes of internalizing or externalizing problems, and delay of gratification was not significantly associated with lower intercepts of internalizing

problems. All other results were consistent with primary analyses. These results indicate that analyses including only mother ratings were just as well suited as when using multiple raters to detect that negative emotionality was associated with higher levels but not slopes of behavior problems, and that delay of gratification did not predict slopes across the developmental span. Of note, mother-reported negative emotionality was not able to detect differences in predictions of slopes, which provides further evidence for the utility of examining multiple perspectives of the child's behavior.

When accounting for early cognitive ability in a sensitivity analysis, the associations between delay of gratification and the intercepts of internalizing and externalizing problems were attenuated to non-significance and trend-level significance, respectively. Early cognitive ability such as verbal skills are thought to support executive function – whose deficits underlie externalizing problems – and delay of gratification. Given these findings, future studies should examine the role of early cognitive ability and executive functioning when examining reactive and control processes as predictors of psychopathology. Prior studies have shown that executive functioning may play a key role in the relations of reactive and control processes on behavior problems (Ursache et al., 2013; Watts et al., 2018).

### General versus specific risk factors for psychopathology at age 15

In a bifactor model, negative emotionality at 54 months was significantly associated with general psychopathology and unique externalizing problems, but not unique internalizing problems at age 15 years. When accounting for covariates, the association between negative emotionality and unique externalizing was attenuated to trend-level significance. Sensitivity analysis elucidated that among symptoms of externalizing problems, delinquent behavior, but not aggressive behavior, appears to be most strongly related to negative emotionality. Furthermore, anger/frustration, not fearfulness, was strongly associated with general psychopathology and unique externalizing problems. A possible interpretation is that negative emotionality as assessed by the CBQ is overt and observable in nature, and thus may reflect more externalizing problems (Eisenberg et al., 2009). Furthermore, when accounting for covariates, the association between negative emotionality and general psychopathology remained significant. Interestingly, negative emotionality is often considered associated with internalizing psychopathology (Greene & Eaton, 2017), but our findings indicate that when accounting for general psychopathology, there was no association between childhood negative emotionality and later internalizing problems. The exception was that when examining mother- vs. caregiver-reported negative emotionality as predictors, only mother-reported negative emotionality was significantly associated with general psychopathology, unique externalizing problems, and, at a trend-level, unique internalizing problems.

One potential explanation for the consistently strong relation between negative emotionality and general psychopathology is that general psychopathology might be functionally interpreted as negative emotionality, such that temperamental negative emotionality may partially reflect a trait-like version of what is shared between internalizing and externalizing problems. An implication of these findings is that negative emotionality assessed at young ages may be a clinically relevant factor to consider when assessing psychopathology, as previous work has indicated (Forbes et al., 2019).

In a sensitivity analysis in which early cognitive ability was included as an additional covariate along with demographic characteristics, the association between negative emotionality and general psychopathology was attenuated to trend-level significance and the association with unique externalizing problems returned to statistical significance with the inclusion of early cognitive ability as a covariate. These findings suggest that other variables, potentially early cognitive ability, or demographic or environmental features, might partially account for the association between negative emotionality and psychopathology, which otherwise show strong associations.

Immediacy preference, i.e., shorter wait time, assessed at 54 months was associated with higher general psychopathology and with *lower* ratings of unique internalizing problems at age 15 years. Surprisingly, there was no association between immediacy preference and unique externalizing problems. Although the association between shorter wait times and lower ratings of internalizing problems was not hypothesized, previous research has observed such an association (Ho et al., 2022). The authors interpreted the association as possibly reflecting that choosing an immediate reward over a distal one is adaptive in reducing feelings of anxiety or depression. Alternatively, the finding could reflect that internalizing problems may be characterized by over-regulation (Murray & Kochanska, 2002). The finding that there was no association between immediacy preference and unique externalizing problems was contrary to what many studies have found (e.g., Ip et al., 2019; Krueger et al., 1996). However, these studies have not examined the association between immediacy preference and *unique* externalizing problems by controlling for the general factor of psychopathology. Nevertheless, previous studies have shown that after accounting for the general factor, self-regulation and executive functioning are not strongly associated with unique psychopathology (Bloemen et al., 2018). However, consistent with Watts et al. (2018), when controlling for covariates, the associations between immediacy preference and psychopathology outcomes were attenuated. This may reflect that individual differences in delay of gratification may be partially accounted for by demographic and background factors, including sex, culture, and socioeconomic status.

Watts et al. (2018) examined the association between immediacy preference and internalizing and externalizing psychopathology. However, Watts et al., assessed a summation of internalizing and externalizing symptoms, whereas we derived general psychopathology using bifactor models, which allowed us to distinguish shared and unique psychopathology. This difference is important because we found that the signs of the association of delay of gratification with general psychopathology (positive) versus unique internalizing problems (negative) were in the opposite direction. The sign difference in the association with immediacy preference would not have been observable with a total behavior problems score, as was used in Watts et al. (2018), or by examining total externalizing problems or total internalizing problems, as we separately examined in the present study. Others have argued against the use of behavior composites because they do not allow for partitioning variance between levels (Michaelson & Munakata, 2020). Therefore, accounting for the shared variance of internalizing and externalizing psychopathology may help identify potentially divergent effects of immediacy preference on various forms of psychopathology.

In a separate model including predictors and covariates together, immediacy preference was no longer associated with unique internalizing problems or the general factor. These results

indicate that much of the variance in the association of immediacy preference with internalizing and general psychopathology may be accounted for demographic and socioeconomic factors. This result was similar to those found by Watts et al. (2018), suggesting that immediacy preference may partially reflect environmental and dispositional processes. Although Watts et al. (2018) focused primarily on children of mothers without college degrees, we leveraged a larger sample while linking scores from multiple informants to obtain a more robust estimate.

The numerous sensitivity analyses (growth curve models: Supplementary Appendix S9; bifactor models: Supplementary Appendix S10) indicated utility in examining subdimensions of constructs. Negative emotionality, in particular anger/frustration, was a robust predictor of externalizing problems across many sensitivity analyses. It was also useful to examine subdimensions of externalizing problems, including delinquent versus aggressive behavior; negative emotionality was more strongly associated with delinquent than aggressive behavior. Furthermore, when examining the association between temperament and psychopathology, accounting for covariates such as early cognitive ability provided potential alternative explanations for future work to consider. The effect sizes of findings in the sensitivity analyses were generally small in magnitude ($\beta = -.01$–$.23$); therefore, we caution that smaller samples might be underpowered to detect these associations.

### Explained common variance

Partitioning unique and general psychopathology also allows estimating the proportion of variance in psychopathology ratings that was accounted for by general and specific psychopathology factors in the sample. The general factor accounted for ∼69% of the reliable variance in ratings of psychopathology at age 15 while specific externalizing problems accounted for ∼9% and specific internalizing problems accounted for ∼22%. This indicates that the general factor was substantially stronger than the specific factors at age 15. Because internalizing problems increased whereas externalizing problems decreased from childhood to adolescence, it is unsurprising that the specific internalizing factor had a larger ECVs compared to the specific externalizing factor. These results are important to consider, because predictors of unique externalizing psychopathology at age 15 may not be as robust given that there is little strength in ratings of unique externalizing problems at this age. By contrast, the unique internalizing and general factors were much stronger at age 15.

### Comparison to prior studies using SECCYD dataset

Prior researchers have studied internalizing and externalizing problems using the same dataset as the present study. Fanti & Heinrich (2010) examined pure and co-occurring mother-reported internalizing and externalizing problems from age 2 to 12. In their study, "pure" referred to a latent class where individuals endorsed high levels in externalizing or internalizing problems but low levels in the other. "Co-occurring" referred to a latent class where individuals endorsed high levels of both. This study found that children with a difficult temperament were more likely to be classified in a co-occurring group compared to "pure" externalizing or internalizing problems, which is consistent with our findings with general psychopathology. They also found similar results where early cognitive development difficulties were associated most strongly with pure chronic externalizing problems compared to other groups, and were less likely to be associated with pure internalizing problems (Fanti & Henrich, 2010).

Cao et al. (2021) examined various inhibitory control tasks, including immediacy preference, as mediators between early tobacco smoke exposure and an average composite of mother- and father-reported internalizing and externalizing problems at the earlier time point of 6th grade. They found that delay of gratification did not predict internalizing or externalizing problems in 6th grade, which differed from our results. Wang & Liu (2021) examined internalizing and externalizing problems at multiple timepoints of 1st, 3rd, 4th, 5th, and 6th grade from teacher ratings. Unlike our linking approach, they estimated a latent factor score for intercepts and slopes of internalizing and externalizing problems composed of scores from each grade as manifest variables. In this study, executive functioning and social competence measured at 1st grade were the predictors of interest (Wang & Liu, 2021). The researchers found that poor executive functioning and poor social competence both predicted higher intercepts of internalizing and externalizing problems (Wang & Liu, 2021).

In comparison to prior studies, the present study uses multidimensional linking to chart trajectories across multiple raters and uses robust methods of bifactor modeling to evaluate predictors of general and specific psychopathology. Furthermore, like Cao et al. (2021) who examined "hot" and "cool" inhibitory control as risk factors, our study examines different facets of a given construct, i.e., reactive and control processes.

To our knowledge, there are two prior studies, Deutz et al. (2020) and McElroy et al. (2018), that examined factor analytic models of general psychopathology using the SECCYD dataset. Deutz et al. (2020) found that immediacy preference did not predict separately analyzed mother- or self-reported general psychopathology after controlling for covariates, similar to our study; however, we also included father report, and we linked scores across raters. Furthermore, we examined these questions for general and specific psychopathology with bifactor modeling. McElroy et al. (2018) examined phenotypic stability in the general and specific factors across all measurements of mother-reported psychopathology from ages 2 to 14, but did not examine predictors of general versus specific psychopathology.

### Strengths

The present study had key strengths. First, we examined children's internalizing and externalizing problems from multiple informants. Second, we examined a lengthy span of development in a large and diverse sample. Third, we used a robust IRT approach to developmental scaling to estimate trajectories of internalizing and externalizing psychopathology on the same scale using ratings at different ages, from numerous raters, and from different measures. Fourth, we examined aspects of reactive and control processes – negative emotionality and immediacy preference – along with demographic characteristics in association with general versus specific psychopathology. The present study provides novel and important contributions to the theory of reactive and control processes of temperament as predictors and the measurement and analysis of psychopathology across a lengthy developmental span.

### Limitations

The present study also had weaknesses. First, we are unable to make causal inferences due to the observational design of the study. Although we examined negative emotionality and immediacy

preference at 54 months in relation to psychopathology at age 15, we are unable to rule out the possibility of the reverse direction of effect, or the possibility of unmeasured third variables. Second, the reliability of the specific psychopathology factors was relatively low, which may have hindered our ability to detect associations with unique externalizing or internalizing problems. It will be important for future work to identify ways of assessing specific psychopathology that yield greater internal consistency.

Third, the linking approach used in the study assumes that item parameters and factor scores are linearly related across measures, raters, and periods of measurements. However, the results indicated that the approach to linking was successful across measures, raters, and timepoints (see Supplement Figures S2–S5). The linking assumption of linearity is between item parameters at *adjacent* ages; the linking does not assume that the change is linear across the entire age span – the changes can be larger or smaller at various ages as needed to adjust for differences. Our developmental scaling approach linked scores across the age- and rater-common items at the aggregate construct level rather than at the item level. Thus, our approach would not be well-suited for interpreting scores for any individual behavior/item across time or raters. Moreover, our approach would also not be appropriate if there is not a common factor (e.g., externalizing problems) that influences scores across the home and school contexts. Our approach may also potentially overlook meaningful qualitative differences that can occur from one year to the next (e.g., transition from elementary to middle school; Shi & Ettekal, 2021).

Fourth, negative emotionality and psychopathology were both derived from the same method (i.e., informant report), whereas immediacy preference was assessed using a laboratory task. Shared method variance might inflate associations between ratings of negative emotionality and psychopathology, when compared to more context-dependent performance on an immediacy preference task (Brock et al., 2014; Makol et al., 2020). Finally, there are potential limitations to the sample. Data collection for the NICHD SECCYD study began approximately 30 years ago, which may impact the relevance of the findings to current understanding of temperament and psychopathology. Moreover, the sample is not nationally representative in terms of race or ethnicity, thus limiting the potential generalizability of the findings to the U.S. population. However, the sample is diverse economically and geographically.

## Conclusion

The findings from the present study extend previous work (Petersen & LeBeau, 2022) that has linked ratings of psychopathology from multiple informants across a wide age range. Very few studies have used developmental scaling to study the development of psychopathology across a lengthy span (Petersen et al., 2018; Petersen & LeBeau, 2022). The present study demonstrates a useful approach to link ratings of internalizing and externalizing psychopathology across time and across raters. The approach was valuable because it allowed us to chart children's development in internalizing and externalizing problems across a lengthy developmental span and to examine early risk factors of their trajectories. Additionally, findings demonstrate the importance of partitioning general versus specific psychopathology.

**Supplementary material.** The supplementary material for this article can be found at https://doi.org/10.1017/S0954579424000713.

## References

Achenbach, T. M. (1991a). *Manual for the Teacher's Report Form and 1991 profile*. University of Vermont, Department of Psychiatry. https://aseba.org/.

Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 profile*, University of Vermont, Department of Psychiatry. https://aseba.org/.

Achenbach, T. M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 profile: Profile for boys and girls*. University of Vermont, Department of Psychiatry. https://aseba.org/.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles (Vol. 30)*. University of Vermont, University of Vermont, Department of Psychiatry. https://aseba.org/.

Achenbach, T. M., & Rescorla, L. A. (2001, *Manual for the ASEBA school-age forms & profiles*. University of Vermont, Department of Psychiatry. https://aseba.org/.

Aldao, A., Gee, D. G., De Los Reyes, A., & Seager, I. (2016). Emotion regulation as a transdiagnostic factor in the development of internalizing and externalizing psychopathology: Current and future directions. *Development and Psychopathology*, *28*(4pt1), 927–946. https://doi.org/10.1017/S0954579416000638

Avenevoli, S., & Steinberg, L. (2001). The continuity of depression across the adolescent transition. *Advances in Child Development and Behavior*, *28*, 139–173. https://doi.org/10.1016/s0065-2407(02)80064-7

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, J. E., Schermerhorn, A. C., & Petersen, I. T. (2014). Temperament concepts in developmental psychopathology. In M. Lewis, & K. D. Rudolph (Eds.), *Handbook of developmental psychopathology* (pp. 311–329). Springer US, https://doi.org/10.1007/978-1-4614-9608-3_16

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bjorklund, D. F., & Kipp, K. (1996). Parental investment theory and gender differences in the evolution of inhibition mechanisms. *Psychological Bulletin*, *120*(2), 163–188.

Bloemen, A., Oldehinkel, A. J., Laceulle, O. M., Ormel, J., Rommelse, N., & Hartman, C. A. (2018). The association between executive functioning and psychopathology: General or specific? *Psychological Medicine*, *48*(11), 1787–1794. https://doi.org/10.1017/S0033291717003269

Brandes, C. M., Herzhoff, K., Smack, A. J., & Tackett, J. L. (2019). The p factor and the n factor: Associations between the general factors of psychopathology and neuroticism in children. *Clinical Psychological Science*, *7*(6), 1266–1284. https://doi.org/10.1177/2167702619859332

Briggs-Gowan, M. J., Carter, A. S., Bosson-Heenan, J., Guyer, A. E., & Horwitz, S. M. (2006). Are infant-toddler social-emotional and behavioral problems transient? *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(7), 849–858. https://doi.org/10.1097/01.chi.0000220849.48650.59

Brock, L. L., Rimm-Kaufman, S. E., & Wanless, S. B. (2014). Delay of gratification in first grade: The role of instructional context. *Learning and Individual Differences*, *29*, 81–88. https://doi.org/10.1016/j.lindif.2013.10.012

Campbell, S. B., & von Stauffenberg, C. (2009). Delay and inhibition as early predictors of ADHD symptoms in third grade. *Journal of Abnormal Child Psychology*, *37*(1), 1–15. https://doi.org/10.1007/s10802-008-9270-4

Cao, H., Liang, Y., & Zhou, N. (2021). Early tobacco smoke exposure, preschool cool/hot inhibitory control, and young adolescents' externalizing/internalizing problems. *Journal of Family Psychology*, *35*(3), 311–323.

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., Jonides, J., Berman, M. G., Wilson, N. L., Teslovich, T., Glover, G., Zayas, V., Mischel, W., & Shoda, Y. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences of The United States of America*, *108*(36), 14998–15003. https://doi.org/10.1073/pnas.1108561108

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure

of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137. https://10.1177/2167702613497473

Castellanos-Ryan, N., Brière, F. N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., Bromberg, U., Büchel, C., Flor, H., Frouin, V., Gallinat, J., Garavan, H., Martinot, J.-L., Nees, F., Paus, T., Pausova, Z., Rietschel, M., Smolka, M. N., Robbins, T. W., Whelan, R., Schumann, G., Conrod, P., The IMAGEN Consortium (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology*, *125*(8), 1039–1052.

Cervin, M., Norris, L. A., Ginsburg, G., Gosch, E. A., Compton, S. N., Piacentini, J., Albano, A. M., Sakolsky, D., Birmaher, B., Keeton, C., Storch, E. A., & Kendall, P. C. (2021). The p factor consistently predicts long-term psychiatric and functional outcomes in anxiety-disordered youth. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(7), 902–912.e5. https://doi.org/10.1016/j.jaac.2020.08.440

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Chen, F. R., & Jaffee, S. R. (2015). The heterogeneity in the development of homotypic and heterotypic antisocial behavior. *Journal of Developmental and Life-Course Criminology*, *1*(3), 269–288. https://doi.org/10.1007/s40865-015-0012-3

Choate, A. M., Bornovalova, M. A., Hipwell, A. E., Chung, T., & Stepp, S. D. (2022). The general psychopathology factor (p) from adolescence to adulthood: Exploring the developmental trajectories of p using a multi-method approach. *Development and Psychopathology*, *1-19*(4), 1775–1793. https://doi.org/10.1017/S0954579422000463

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, *70*(1), 6–20. https://doi.org/10.1037/0022-006X.70.1.6

Clark, D. A., Hicks, B. M., Angstadt, M., Rutherford, S., Taxali, A., Hyde, L., Weigard, A. S., Heitzeg, M. M., & Sripada, C. (2021). The general factor of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study: A comparison of alternative modeling approaches. *Clinical Psychological Science*, *9*(2), 169–182. https://doi.org/10.1177/21677026 20959317

Crockett, L. J., Wasserman, A. M., Rudasill, K. M., Hoffman, L., & Kalutskaya, I. (2018). Temperamental anger and effortful control, teacher-child conflict, and externalizing behavior across the elementary school years. *Child Development*, *89*(6), 2176–2195.

Cuhadar, I., & Kalkan, Ö.K. (2023). Performance of estimation methods in bifactor models with ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(2), 329–339. https://doi.org/10.1080/10705 511.2023.2247567

Damme, K. S. F., Norton, E. S., Briggs-Gowan, M. J., Wakschlag, L. S., & Mittal, V. A. (2022). Developmental patterning of irritability enhances prediction of psychopathology in preadolescence: Improving RDoC with developmental science. *Journal of Psychopathology and Clinical Science*, *131*(6), 556–566. https://doi.org/10.1037/abn0000655

De Los Reyes, A., Bunnell, B. E., & Beidel, D. C. (2013). Informant discrepancies in adult social anxiety disorder assessments: Links with contextual variations in observed behavior. *Journal of Abnormal Psychology*, *122*(2), 376–386. https://doi.org/10.1037/a0031150

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*(5), 637–652. https://doi.org/10.1007/s10802-009-9307-3

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*(4), 483–509. https://doi.org/10.1037/0033-2909.131.4.483

De Los Reyes, A., & Makol, B. A. (2021). Interpreting convergences and divergences in multi-informant, multimethod assessment. In Joni L. Mihura (Eds.), *The oxford handbook of personality and psychopathology assessment* (2nd. ed.). Oxford Academic,

De Los Reyes, A., & Makol, B. A. (2022). Informant reports in clinical assessment. In G. J. G. Asmundson (Eds.), *Comprehensive clinical psychology*

(2nd ed. pp. 105–122). Elsevier, https://doi.org/10.1016/B978-0-12-818697-8.00113-8

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

Deutz, M. H. F., Geeraerts, S. B., Belsky, J., Deković, M., van Baar, A. L., Prinzie, P., & Patalay, P. (2020). General psychopathology and dysregulation profile in a longitudinal community sample: Stability, antecedents and outcomes. *Child Psychiatry and Human Development*, *51*(1), 114–126. https://doi.org/10.1007/s10578-019-00916-2

Egan, S. J., Wade, T. D., & Shafran, R. (2011). Perfectionism as a transdiagnostic process: A clinical review. *Clinical Psychology Review*, *31*(2), 203–212.

Eisenberg, N., Fabes, R. A., Guthrie, I. K., Murphy, B. C., Maszk, P., Holmgren, R., & Suh, K. (1996). The relations of regulation and emotionality to problem behavior in elementary school children. *Development and Psychopathology*, *8*(1), 141–162. https://doi.org/10.1017/S095457940000701X

Eisenberg, N., Guthrie, I. K., Fabes, R. A., Shepard, S., Losoya, S., Murphy, B., Jones, S., Poulin, R., & Reiser, M. (2000). Prediction of elementary school children's externalizing problem behaviors from attentional and behavioral regulation and negative emotionality. *Child Development*, *71*(5), 1367–1382. https://doi.org/10.1111/1467-8624.00233

Eisenberg, N., Sadovsky, A., Spinrad, T. L., Fabes, R. A., Losoya, S. H., Valiente, C., Reiser, M., Cumberland, A., & Shepard, S. A. (2005). The relations of problem behavior status to children's negative emotionality, effortful control, and impulsivity: Concurrent relations and prediction of change. *Developmental Psychology*, *41*(1), 193–211. https://doi.org/10.1037/0012-1649.41.1.193

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

Fanti, K. A., & Henrich, C. C. (2010). Trajectories of pure and co-occurring internalizing and externalizing problems from age 2 to age 12: Findings from the National Institute of Child Health and Human Development Study of Early Child Care. *Developmental Psychology*, *46*(5), 1159–1175. https://doi.org/10.1037/a0020659.

Forbes, M. K., Rapee, R. M., & Krueger, R. F. (2019). Opportunities for the prevention of mental disorders by reducing general psychopathology in early childhood. *Behaviour Research and Therapy*, *119*, 103411–103411. https://doi.org/10.1016/j.brat.2019.103411

Forbes, M. K., Tackett, J. L., Markon, K. E., & Krueger, R. F. (2016). Beyond comorbidity: Toward a dimensional and hierarchical approach to understanding psychopathology across the life span. *Development and Psychopathology*, *28*(4pt1), 971–986. https://doi.org/10.1017/S09545794 16000651

Fox, N. A. (1989). Psychophysiological correlates of emotional reactivity during the first year of life. *Developmental Psychology*, *25*(3), 364–372. https://doi.org/10.1037/0012-1649.25.3.364

Gluschkoff, K., Jokela, M., & Rosenström, T. (2019). The general psychopathology factor: Structural stability and generalizability to within-individual changes. *Frontiers in Psychiatry*, *10*, 594. https://doi.org/10.3389/fpsyt.2019.00594

Greene, A. L., & Eaton, N. R. (2017). The temporal stability of the bifactor model of comorbidity: An examination of moderated continuity pathways. *Comprehensive Psychiatry*, *72*, 74–82. https://doi.org/10.1016/j.comppsych.2016.09.010

Hamlat, E. J., Snyder, H. R., Young, J. F., & Hankin, B. L. (2019). Pubertal timing as a transdiagnostic risk for psychopathology in youth. *Clinical Psychological Science*, *7*(3), 411–429. https://doi.org/10.1177/2167702618810518

Hankin, B. L., Davis, E. P., Snyder, H., Young, J. F., Glynn, L. M., & Sandman, C. A. (2017). Temperament factors and dimensional, latent bifactor models of child psychopathology: Transdiagnostic and specific

associations in two youth samples. *Psychiatry Research*, *252*, 139–146. https://doi.org/10.1016/j.psychres.2017.02.061

Hartley, A. G., Zakriski, A. L., & Wright, J. C. (2011). Probing the depths of informant discrepancies: Contextual influences on divergence and convergence. *Journal of Clinical Child & Adolescent Psychology*, *40*(1), 54–66. https://doi.org/10.1080/15374416.2011.533404

Hawes, M. T., Carlson, G. A., Finsaas, M. C., Olino, T. M., Seely, J. R., & Klein, D. N. (2020). Dimensions of irritability in adolescents: Longitudinal associations with psychopathology in adulthood. *Psychological Medicine*, *50*(16), 2759–2767. https://doi.org/10.1017/S0033291719002903.

Ho, H., Dang, H. M., Odum, A. L., DeHart, W. B., & Weiss, B. (2022). Sooner is better: Longitudinal relations between delay discounting, and depression and anxiety symptoms among Vietnamese adolescents. *Research on Child and Adolescent Psychopathology*, *51*(1), 133–147. https://doi.org/10.1007/s10802-022-00959-5.https://doi.org/10.1007/s10802-022-00959-5

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

Huber, P. J. The behavior of maximum likelihood estimates under nonstandard conditions. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, University of California Press, 1967

Ip, K. I., Jester, J. M., Sameroff, A., & Olson, S. L. (2019). Linking Research Domain Criteria (RDoC) constructs to developmental psychopathology: The role of self-regulation and emotion knowledge in the development of internalizing and externalizing growth trajectories from ages 3 to 10. *Development and Psychopathology*, *31*(4), 1557–1574. https://doi.org/10.1017/S0954579418001323

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

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. *semTools: Useful tools for structural equation modeling*. [Computer software], 2022. https://CRAN.R-project.org/package=semTools

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: Externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology*, *28*(2), 161–179. https://doi.org/10.1023/a:1005122814723

Kenyon, D. M., MacGregor, D., Li, D., & Cook, H. G. (2011). Issues in vertical scaling of a K-12 English language proficiency test. *Language Testing*, *28*(3), 383–400. https://doi.org/10.1177/0265532211404190

Kidd, C., Palmeri, H., & Aslin, R. N. (2013). Rational snacking: Young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition*, *126*(1), 109–114. https://doi.org/10.1016/j.cognition.2012.08.004

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed., pp. xxvi, 566). https://doi.org/10.1007/978-1-4939-0317-7

Krueger, R. F., Caspi, A., Moffitt, T. E., White, J., & Stouthamer-Loeber, M. (1996). Delay of gratification, psychopathology, and personality: Is low self-control specific to externalizing problems? *Journal of Personality*, *64*(1), 107–129. https://doi.org/10.1111/j.1467-6494.1996.tb00816.x

Krueger, R. F., & Eaton, N. R. (2015). Transdiagnostic factors of mental disorders. *World Psychiatry*, *14*(1), 27–29. https://doi.org/10.1002/wps.20175

Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Keyes, K. M., Latzman, R. D., Michelini, G., Patrick, C. J., Sellbom, M., Slade, T., South, S C., Sunderland, M., Tackett, J., Waldman, I., Waszczuk, M. A., Wright, A. G. C., Zald, D. H., Watson, D., Kotov, R., HiTOP Utility Workgroup (2021). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry*, *20*(2), 171–193. https://doi.org/10.1002/wps.20844

Kwon, K., Kim, E. M., & Sheridan, S. M. (2012). A contextual approach to social skills assessment in the peer group: Who is the best judge? *School Psychology Quarterly*, *27*(3), 121–133. https://doi.org/10.1037/a0028696

Lahey, B. B. (2009). Public health significance of neuroticism. *American Psychologist*, *64*(4), 241–256.

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology*, *121*(4), 971–977. https://doi.org/10.1037/a0028355

Lahey, B. B., Moore, T. M., Kaczkurkin, A. N., & Zald, D. H. (2021). Hierarchical models of psychopathology: Empirical support, implications, and remaining issues. *World Psychiatry*, *20*(1), 57–63. https://doi.org/10.1002/wps.20824

Leaberry, K. D., Rosen, P. J., Slaughter, K. E., Reese, J., & Fogleman, N. D. (2019). Temperamental negative affect, emotion-specific regulation, and concurrent internalizing and externalizing pathology among children with ADHD. *ADHD Attention Deficit and Hyperactivity Disorders*, *11*(3), 311–324. https://doi.org/10.1007/s12402-019-00294-8

Lemery, K. S., Essex, M. J., & Smider, N. A. (2002). Revealing the relation between temperament and behavior problem symptoms by eliminating measurement confounding: Expert ratings and factor analyses. *Child Development*, *73*(3), 867–882. https://doi.org/10.1111/1467-8624.00444

Leve, L. D., Kim, H. K., & Pears, K. C. (2005). Childhood temperament and family environment as predictors of internalizing and externalizing trajectories from ages 5 to 17. *Journal of Abnormal Child Psychology*, *33*(5), 505–520. https://doi.org/10.1007/s10802-005-6734-7

Lilienfeld, S. O. (2003). Comorbidity between and within childhood externalizing and internalizing disorders: Reflections and directions. *Journal of Abnormal Child Psychology*, *31*(3), 285–291. https://doi.org/10.1023/a:1023229529866

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & Van Kammen, W. B. (1998). Multiple risk factors for multiproblem boys: Co-occurrence of delinquency, substance use, attention deficit, conduct problems, physical aggression, covert behavior, depressed mood, and shy/withdrawn behavior. In R. Jessor (Eds.), *New perspectives on adolescent risk behavior* (pp. 90–149). Cambridge University Press. https://doi.org/10.1017/CBO9780511571138.005

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

Makol, B. A., Youngstrom, E. A., Racz, S. J., Qasmieh, N., Glenn, L. E., & De Los Reyes, A. (2020). Integrating multiple informants' reports: How conceptual and measurement models may address long-standing problems in clinical decision-making. *Clinical Psychological Science*, *8*(6), 953–970. https://doi.org/10.1177/2167702620924439

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin*, *137*(5), 856–879. https://doi.org/10.1037/a0023678

Martel, M. M., Markon, K., & Smith, G. T. (2017). Research review: Multi-informant integration in child and adolescent psychopathology diagnosis. *Journal of Child Psychology and Psychiatry*, *58*(2), 116–128. https://doi.org/10.1111/jcpp.12611

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

McLaughlin, K. A., & Nolen-Hoeksema, S. (2011). Rumination as a transdiagnostic factor in depression and anxiety. *Behaviour Research and Therapy*, *49*(3), 186–193. https://doi.org/10.1016/j.brat.2010.12.006

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*, *106*(1), 3–19. https://doi.org/10.1037/0033-295X.106.1.3

Michaelson, L. E., & Munakata, Y. (2020). Same data set, different conclusions: Preschool delay of gratification predicts later behavioral outcomes in a

preregistered study. *Psychological Science*, *31*(2), 193–201. https://doi.org/10.1177/0956797619896270

Mikolajewski, A. J., Allan, N. P., Hart, S. A., Lonigan, C. J., & Taylor, J. (2013). Negative affect shares genetic and environmental influences with symptoms of childhood internalizing and externalizing disorders. *Journal of Abnormal Child Psychology*, *41*(3), 411–423. https://doi.org/10.1007/s10802-012-9681-0

Miller, J. L., Vaillancourt, T., & Boyle, M. H. (2009). Examining the heterotypic continuity of aggression using teacher reports: Results from a national Canadian study. *Social Development*, *18*(1), 164–180. https://doi.org/10.1111/j.1467-9507.2008.00480.x

Min, K.-S. (2007). Evaluation of linking methods for multidimensional irt calibrations. *Asia Pacific Education Review*, *8*(1), 41–55. https://doi.org/10.1007/BF03025832

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, *16*(2), 329–337. https://doi.org/10.1037/h0029815

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246–268. https://doi.org/10.1037/0033-295x.102.2.246

Mischel, W., Shoda, Y., & Rodriguez, M. (1989). Delay of gratification in children. *Science*, *244*(4907), 933–938. https://doi.org/10.1126/science.2658056

Moran, L. R., Lengua, L. J., & Zalewski, M. (2013). The interaction between negative emotionality and effortful control in early social-emotional development. *Social Development*, *22*(2), 340–362. https://doi.org/10.1111/sode.12025

Murayama, K., Pekrun, R., Lichtenfeld, S., & Vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development*, *84*(4), 1475–1490. https://doi.org/10.1111/cdev.12036

Muris, P., & Ollendick, T. H. (2005). The role of temperament in the etiology of child psychopathology. *Clinical Child and Family Psychology Review*, *8*(4), 271–289. https://doi.org/10.1007/s10567-005-8809-y

Murray, A. L., Eisner, M., & Ribeaud, D. (2016). The development of the general factor of psychopathology 'p factor' through childhood and adolescence. *Journal of Abnormal Child Psychology*, *44*(8), 1573–1586. https://doi.org/10.1007/s10802-016-0132-1

Murray, J., & Farrington, D. P. (2010). Risk factors for conduct disorder and delinquency: Key findings from longitudinal studies. *The Canadian Journal of Psychiatry*, *55*(10), 633–642.

Murray, K. T., & Kochanska, G. (2002). Effortful control: Factor structure and relation to externalizing and internalizing behaviors. *Journal of Abnormal Child Psychology*, *30*(5), 503–514. https://doi.org/10.1023/a:1019821031523

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

NICHD Early Child Care Research Network 2005). *Child care and child development: Results from the NICHD study of early child care and youth development*. Guilford Press.

Olino, T. M., Dougherty, L. R., Bufferd, S. J., Carlson, G. A., & Klein, D. N. (2014). Testing models of psychopathology in preschool-aged children using a structured interview-based assessment. *Journal of Abnormal Child Psychology*, *42*(7), 1201–1211. https://doi.org/10.1007/s10802-014-9865-x

Oshima, T. C., Davey, T. C., & Lee, K. (2000). Multidimensional linking: Four practical approaches. *Journal of Educational Measurement*, *37*(4), 357–373. http://www.jstor.org/stable/1435246

Patterson, G. R. (1993). Orderly change in a stable world: The antisocial trait as a chimera. *Journal of Consulting and Clinical Psychology*, *61*(6), 911–919. https://doi.org/10.1037/0022-006X.61.6.911

Peake, P. K. (2017). Delay of gratification: Explorations of how and why children wait and its linkages to outcomes over the life course. *Nebraska Symposium on Motivation*, *64*, 7–60. https://doi.org/10.1007/978-3-319-51721-6_2

Petersen, I. T., Bates, J. E., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood. *Development and Psychopathology*, *27*(3), 791–818. https://doi.org/10.1017/S0954579414000789

Petersen, I. T., Choe, D. E., & LeBeau, B. (2020). Studying a moving target in development: The challenge and opportunity of heterotypic continuity. *Developmental Review*, *58*, 100935. https://doi.org/10.1016/j.dr.2020.100935

Petersen, I. T., & LeBeau, B. (2022). Creating a developmental scale to chart the development of psychopathology with different informants and measures across time. *Journal of Psychopathology and Clinical Science*, *131*(6), 611–625. https://doi.org/10.1037/abn0000649

Petersen, I. T., LeBeau, B., & Choe, D. E. (2021). Creating a developmental scale to account for heterotypic continuity in development: A simulation study. *Child Development*, *92*(1), e1–e19. https://doi.org/10.1111/cdev.13433

Petersen, I. T., Lindhiem, O., LeBeau, B., Bates, J. E., Pettit, G. S., Lansford, J. E., & Dodge, K. A. (2018). Development of internalizing problems from adolescence to emerging adulthood: Accounting for heterotypic continuity with vertical scaling. *Developmental Psychology*, *54*(3), 586–599. https://doi.org/10.1037/dev0000449

Pettersson, E., Lahey, B. B., Larsson, H., & Lichtenstein, P. (2018). Criterion validity and utility of the general factor of psychopathology in childhood: Predictive associations with independently measured severe adverse mental health outcomes in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, *57*(6), 372–383. https://doi.org/10.1016/j.jaac.2017.12.016

Phillips, E. M., Brock, R. L., James, T. D., Nelson, J. M., Espy, K. A., & Nelson, T. D. (2022). Empirical support for a dual process model of the p-factor: Interaction effects between preschool executive control and preschool negative emotionality on general psychopathology. *Journal of Psychopathology and Clinical Science*, *131*(8), 817–829. https://doi.org/10.1037/abn0000777

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, *56*(3), 345–365. https://doi.org/10.1111/jcpp.12381

R Core Team (2022). *R: A language and environment for statistical computing*.

Rachlin, H., & Jones, B. A. (2008). Social discounting and delay discounting. *Journal of Behavioral Decision Making*, *21*(1), 29–43. https://doi.org/10.1002/bdm.567

Racine, N., McArthur, B. A., Cooke, J. E., Eirich, R., Zhu, J., & Madigan, S. (2021). Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: A meta-analysis. *JAMA Pediatrics*, *175*(11), 1142–1150. https://doi.org/10.1001/jamapediatrics.2021.2482

Ree, M. J., Carretta, T. R., & Teachout, M. S. (2015). Pervasiveness of dominant general factors in organizational measurement. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *8*(3), 409–427. https://doi.org/10.1017/iop.2015.16

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rothbart, M., & Derryberry, D. (1981). Development of individual differences in temperament. In *Advances in developmental psychology erlbaum*. vol. *1*, pp. 33–86.

Rothbart, M. K. (2011). *Becoming who we are: Temperament and personality in development*. Guilford Press.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, *72*(5), 1394–1408. https://doi.org/10.1111/1467-8624.00355

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In *Handbook of child psychology: Social, emotional, and personality development*. vol. *3*, 6th ed. pp. 99–166). John Wiley & Sons, Inc.

Ruggero, C. J., Kotov, R., Hopwood, C. J., First, M., Clark, L. A., Skodol, A. E., Mullins-Sweatt, S. N., Patrick, C. J., Bach, B., Cicero, D. C., Docherty, A., Simms, L. J., Bagby, R. M., Krueger, R. F., Callahan, J. L., Chmielewski, M., Conway, C. C., De Clercq, B., Dornbach-Bender, A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Haltigan, J. D., Miller, J. D., Morey, L. C., Patalay, P., Regier, D. A., Reininghaus, U., Shackman, A. J., Waszczuk, M. A., Watson, D., Wright, A. G. C., Zimmermann, J. (2019). Integrating the Hierarchical Taxonomy of Psychopathology (HiTOP) into clinical

practice. *Journal of Consulting and Clinical Psychology*, *87*(12), 1069–1084. https://doi.org/10.1037/ccp0000452.

Rutter, M., & Sroufe, L. A. (2000). Developmental psychopathology: Concepts and challenges. *Development and Psychopathology*, *12*(3), 265–296. https://doi.org/10.1017/s0954579400003023

Sattler, J. M. (2022). *Foundations of behavioral, social, and clinical assessment of children*. Jerome M. Sattler, Publisher, Inc.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2), 23–74.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Shi, Q., & Ettekal, I. (2021). Co-occurring trajectories of internalizing and externalizing problems from grades 1 to 12: Longitudinal associations with teacher-child relationship quality and academic performance. *Journal of Educational Psychology*, *113*(4), 808–829. https://doi.org/10.1037/edu0000525

Shi, Q., Ettekal, I., Deutz, M. H. F., & Woltering, S. (2020). Trajectories of pure and co-occurring internalizing and externalizing problems from early childhood to adolescence: Associations with early childhood individual and contextual antecedents. *Developmental Psychology*, *56*(10), 1906–1918. https://doi.org/10.1037/dev0001095

Sroufe, L. A. (2016). The place of attachment in development. In J. Cassidy, & P. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications*. vol. *3*, p. 997–1011). Guilford Press.

Steinberg, E. A., & Drabick, D. A. G. (2015). A developmental psychopathology perspective on ADHD and comorbid conditions: The role of emotion regulation. *Child Psychiatry & Human Development*, *46*(6), 951–966. https://doi.org/10.1007/s10578-015-0534-2

Stephens, D. W., & Anderson, D. (2001). The adaptive value of preference for immediacy: When shortsighted rules have farsighted consequences. *Behavioral Ecology*, *12*(3), 330–339. https://doi.org/10.1093/beheco/12.3.330

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–210. https://doi.org/10.1177/014662168300700208

Suh, Y. (2015). The performance of maximum likelihood and weighted least square mean and variance adjusted estimators in testing differential item functioning with nonnormal trait distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(4), 568–580. https://doi.org/10.1080/10705511.2014.937669

Tackett, J. L., & Hallquist, M. (2022). The need to grow: Developmental considerations and challenges for modern psychiatric taxonomies. *Journal of Psychopathology and Clinical Science*, *131*(6), 660–663. https://doi.org/10.1037/abn0000751

Tackett, J. L., Lahey, B. B., van Hulle, C., Waldman, I., Krueger, R. F., & Rathouz, P. J. (2013). Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *Journal of Abnormal Psychology*, *122*(4), 1142–1153.

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

Ursache, A., Blair, C., Stifter, C., & Voegtline, K. (2013). Emotional reactivity and regulation in infancy interact to predict executive functioning in early childhood. *Developmental Psychology*, *49*(1), 127–137. https://doi.org/10.1037/a0027728

Wall, A. E., & Barth, R. P. (2005). Aggressive and delinquent behavior of maltreated adolescents: Risk factors and gender differences. *Stress, Trauma, and Crisis*, *8*(1), 1–24. https://doi.org/10.1080/15434610490888081

Wang, Y., & Liu, Y. (2021). The development of internalizing and externalizing problems in primary school: Contributions of executive function and social competence. *Child Development*, *92*(3), 889–903.

Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology*, *97*(3), 346–53.

Watts, A. L., Makol, B. A., Palumbo, I. M., De Los Reyes, A., Olino, T. M., Latzman, R. D., DeYoung, C. G., Wood, P. K., & Sher, K. J. (2021). How robust is the p factor? Using multitrait-multimethod modeling to inform the meaning of general factors of youth psychopathology. *Clinical Psychological Science*, *10*(4), 640–661. https://doi.org/10.1177/21677026211055170

Watts, A. L., Meyer, F. A. C., Greene, A. L., Wood, P., Trull, T., Steinley, D., & Sher, K. (2021). Spurious empirical support for the p-factor arises with the inclusion of undiagnosed cases. *PsyArXiv*. Preprint: https://doi.org/10.31234/osf.io/4tazx

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the marshmallow test: A conceptual replication investigating links between early delay of gratification and later outcomes. *Psychological Science*, *29*(7), 1159–1177. https://doi.org/10.1177/0956797618761661

Weeks, J. P. (2010). plink: An R package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software*, *35*(12), 1–33. https://doi.org/10.18637/jss.v035.i12

Weems, C. F. (2008). Developmental trajectories of childhood anxiety: Identifying continuity and change in anxious emotion. *Developmental Review*, *28*(4), 488–502. https://doi.org/10.1016/j.dr.2008.01.001

Weissman, D. G., Bitran, D., Miller, A. B., Schaefer, J. D., Sheridan, M. A., & McLaughlin, K. A. (2019). Difficulties with emotion regulation as a transdiagnostic mechanism linking child maltreatment with the emergence of psychopathology. *Development and Psychopathology*, *31*(3), 899–915. https://doi.org/10.1017/S0954579419000348

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838. https://doi.org/10.2307/1912934

Zinbarg, R., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7