

A Developmentally Informed Systematic Review and Meta-Analysis of the Strength of General Psychopathology in Childhood and Adolescence

Jordan L. Harris¹ · Benjamin Swanson² · Isaac T. Petersen¹

Accepted: 6 November 2023 / Published online: 19 December 2023 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Considerable support exists for higher-order dimensional conceptualizations of psychopathology in adults. A growing body of work has focused on understanding the structure of general and specific psychopathology in children and adolescents. No prior meta-analysis has examined whether the strength of the general psychopathology factor (*p* factor)—measured by explained common variance (ECV)—changes from childhood to adolescence. The primary objective of this multilevel meta-analysis was to determine whether general psychopathology strength changes across development (i.e. across ages) in childhood and adolescence. Several databases were searched in November 2021; 65 studies, with 110 effect sizes (ECV), nested within shared data sources, were identified. Included empirical studies used a factor analytic modeling approach that estimated latent factors for child/adolescent internalizing, externalizing, and optionally thought-disordered psychopathology, and a general factor. Studies spanned ages 2–17 years. Across ages, general psychopathology explained over half (~56%) of the reliable variance in symptoms of psychopathology. Age-moderation analyses revealed that general factor strength remained stable across ages, suggesting that general psychopathology strength does not significantly change across childhood to adolescence. Even if the structure of psychopathology changes with development, the prominence of general psychopathology across development has important implications for future research and intervention.

Keywords General psychopathology · p factor · Explained common variance · Meta-analysis · Childhood · Adolescence

The *Diagnostic and Statistical Manual of Mental Disorders (DSM;* American Psychiatric Association, 2022) is in its fifth edition with text revisions, and traditionally distinguishes psychopathology into categorical diagnoses. However, there is growing evidence that a more reliable nosology would reflect dimensionality of psychopathology (i.e. spectrum syndromes) and would more appropriately account for co-occurrence and covariation among disorders (Caspi et al., 2014; Clark et al., 2021; Lahey et al., 2012; Murray et al., 2016). Different aspects of psychopathology, including internalizing (e.g. anxiety and depression), externalizing (e.g. aggression and rule-breaking behaviors), and

thought-disordered psychopathology (e.g. obsessions, compulsions, and mania), tend to covary. One review found that 10-20% of preschoolers with oppositional defiant disorder (ODD) diagnoses, an externalizing disorder, also present with internalizing disorders (Boylan et al., 2007). Estimates of the degree of covariation between internalizing and externalizing problems have shown to range from r = 0.38 - 0.62in a sample of adolescents ages 12-18 (Cosgrove et al., 2011). The strong covariation between internalizing and externalizing psychopathology suggests that higher-order processes may account for their commonality and co-occurrence (Carragher et al., 2015; Caspi et al., 2014; Gluschkoff et al., 2019; Kessler et al., 1999; Kotov et al., 2017). The higher-order factor that accounts for the strong covariation of specific psychopathologies (e.g. internalizing, externalizing, and thought-disorder psychopathology) is referred to as general psychopathology or p factor (Caspi et al., 2014).

Some have suggested that including a general psychopathology or p factor is necessary to fully conceptualize the structure of psychopathology (Avinun et al., 2021; Lahey

Jordan L. Harris Jordan-l-harris@uiowa.edu

¹ Department of Psychological and Brain Sciences, University of Iowa, 340 Iowa Avenue G60, Iowa City, IA 52242, USA

² Department of Psychological Sciences, University of Arkansas, Fayetteville, AR, USA

et al., 2012). A general psychopathology model posits that a single factor influences all symptoms across a range of known categorical and dimensional diagnoses, while specific psychopathology accounts for what is unique to a given set of symptoms above and beyond the general factor (Caspi et al., 2014; Lahey et al., 2012). General psychopathology, as measured by factor analytic modeling (e.g. bifactor model), is meant to reflect the common variance that influences responses on measures of multiple psychopathology dimensions. This common variance indicates that individual differences in people's levels on some symptoms are concurrently associated (Lahey et al., 2021).

General Psychopathology Modeling Approaches

Co-occurrence and correlation among dimensions of psychopathology provide a reasonable justification for a hierarchical structure of psychopathology, but there is no single acceptable method to structuring a general psychopathology factor model (Lahey et al., 2021). Using factor analysis, there are several accepted methods of describing dimensionality among specific and general psychopathology, including higher-order (also called second-order) models, bifactor models, and modified bifactor models (Carragher et al., 2016; Lahey et al., 2021). A bifactor model of general psychopathology is the most common approach to modeling general psychopathology (e.g. Aitken et al., 2020; Clark et al., 2021; Hankin et al., 2017; Huang-Pollock et al., 2017; Neumann et al., 2016; Sheldrick et al., 2012; Vine et al., 2020; Wade et al., 2019; Waldman et al., 2016). In a bifactor model, individual indicators load onto one specific psychopathology factor. The symptoms additionally load directly onto an orthogonal general factor (see Fig. 1, Model A). A traditional bifactor model relies on the general factor to account for commonality among the specific factors, rendering the specific factors to represent the residual correlations among common items in each specific factor, after accounting for what is shared due to the general factor (Lahey et al., 2021). Another modeling approach is the modified bifactor model. Modified bifactor models have an orthogonal general or p factor, like bifactor models, but the specific factors (e.g. internalizing, externalizing, and thought disorder) are allowed to correlate (see Fig. 1, Model B). Some have argued that allowing specific factors to correlate provides a more ecologically valid representation of psychopathology (Afzali et al., 2018; Carragher et al., 2016). Psychopathology may also be modeled using a higher-order model. In a higher-order model, symptom or diagnosis indicators load onto only one first-order specific factor. These first-order factors then load onto a higher-order factor, which represents the general factor (see Fig. 1, Model C). Despite the lack of consensus on the best modeling approach, there is support for modeling the covariation among psychopathology using factor analytic approaches (Canivez, 2016; Lahey et al., 2015, 2021; Patalay et al., 2015).

The Conceptualization of General Psychopathology

The proliferation and wide acceptance of general psychopathology models is due in part to a growing adoption of hierarchical nosologies, such as Hierarchical Taxonomy of Psychopathology or HiTOP (Kotov et al., 2017). HiTOP and other general psychopathological models have inspired a growing body of research aimed at disentangling the structure and manifestation of psychopathology and how it changes across the lifespan (e.g. Forbes et al., 2019; Gomez et al., 2019; Martel et al., 2017; Murray et al., 2016; Waldman et al., 2016). However, HiTOP and other conceptualizations of general psychopathology were developed largely based on adult samples primarily composed of 15-65 year olds, and there is uncertainty as to whether this conceptualization generalizes to younger ages because children rarely exhibit symptoms of end-stage psychopathology (Forbes et al., 2019; Kotov et al., 2021).

Studies have replicated findings that the *p* factor and specific factors exist in both adults and in children (e.g. Laceulle et al., 2015; McElroy et al., 2018a; Olino et al., 2014). Some have argued that HiTOP does not adequately capture developmental changes in behavioral manifestations of psychopathology such as (a) sex-related differences in depression and antisocial behavior during puberty (Hamlat et al., 2019; Van Hulle et al., 2009); (b) restricted access to elicit substances in childhood (Kotov et al., 2021); or (c) the decreasing base rate of aggressive and rule-breaking behavior in adulthood (Achenbach, 2020). Thus, a hierarchical taxonomy of psychopathology in children might not simply be a translation of the adult model, and instead would require more attention to developmentally informed changes in the presentation and covariation of psychopathology (Kotov et al., 2021).

Criticisms of General Psychopathology Modeling

Hierarchical approaches to modeling psychopathology, such as bifactor modeling, are often criticized for numerous reasons. Some have argued that latent factor modeling approaches make assumptions about the data that may be questionable, such as imprecise explanations and predictions on data supporting weak theories (Fried, 2020; van Bork et al., 2017; Watts et al., 2019). For example, bifactor models of general psychopathology are often favored over unidimensional and other correlated factor models because bifactor models yield the best model fit. However, studies



Fig. 1 Representation of Higher-Order, Bifactor, and Modified Bifactor Models. Panel A bifactor model. Panel B modified bifactor model. Panel Chigher-order factor model

have shown that bifactor models fit well even when there are spurious reasons for it, e.g. random patterns and not valid responses, suggesting that both signal and noise are overfitting the data (Fried, 2020; Haeffel et al., 2022; Reise et al., 2016; Snyder et al., 2017b). In light of overreliance on goodness of fit and other flaws in bifactor modeling approaches, Bonifay et al. (2017) suggested that other bifactor statistics, such as explained common variance, may prove more useful than goodness of fit in evaluating indices of general factor modeling.

Others have argued that the general factor in a bifactor model is merely a methodological artifact. Watts et al. (2019) found that general psychopathology factors differ greatly as a function of which indicators are included in the bifactor model. They estimated 15 separate bifactor models that each had a distinct single indicator dropped, leaving a unique combination of 14 indicators in each model. The results indicated that some indicators' loading strength and sign were highly dependent on the presence of the other indicators in the model (Watts et al., 2019). For example, the conduct disorder indicator had a loading strength on the general factor of 0.63 when tics were excluded, but a strength of -0.45 when obsessions were excluded. Evidence from these studies raises questions about whether the general factor of psychopathology might be a methodological artifact and might not be meaningful, when estimated using a bifactor model (Fried et al., 2021; Watts et al., 2021a). However, there is growing support that the general factor represents severity or comorbidity among items of psychopathology (Fried et al., 2021). To better understand the development of the general factor of psychopathology, it is important to examine the general factor strength across ages.

Measurement of General Psychopathology Strength

Many studies have tested higher-order psychopathology models in children that index the amount of variance accounted for by a general psychopathology factor at a given age or a span of ages across development (e.g. Gomez et al., 2019; Haltigan et al., 2018; Hankin et al., 2017; Levin-Aspenson et al., 2019; Waldman et al., 2016). The proportion of variance in ratings of psychopathology accounted for by the general factor is often called explained common variance (ECV). ECV is considered a reliable estimate of the strength of the general factor, when estimating the relative contributions of both the general and specific factors (Martel et al., 2017; Rodriguez et al., 2016). In this regard, explained common variance is a useful metric from which stability of the strength of the general factor can be estimated when examined longitudinally (Murray et al., 2016; Rodriguez et al., 2016). There is a growing need to examine the strength of the general factor meta-analytically and how it changes across development to better understand the contributions of higher-order psychopathology across childhood and adolescence to inform the development of interventions (Forbes et al., 2019; Hopwood et al., 2020; Ruggero et al., 2019).

Developmentally Informed Conceptualizations of Psychopathology

Developmental psychopathology is a framework in which practitioners and researchers study how individuals may or may not develop pathology (e.g. externalizing or internalizing disorders), given social, biological, and psychological risks (Cicchetti, 2020). To this end, it is important to consider the timing of development, new challenges that may arise, and the degree to which the individuals are able to navigate these challenges (Cicchetti, 2020). The earlier and longer that an individual continues along a maladaptive pathway, the more difficult it becomes for them to return toward a normal developmental trajectory (Nigg, 2006). However, individuals may move between states of pathology and non-pathology functioning across development (Cicchetti, 2020; Cicchetti & Sroufe, 2000).

Behavioral manifestations of psychopathology vary at different ages (i.e. heterotypic continuity; Cicchetti & Rogosch, 2002). For example, externalizing problems in a 3 year old may take the form of overt behavior (e.g. temper tantrum), whereas in a 16-year-old adolescent, the manifestation may take a more covert form (e.g. substance use; Miller et al., 2009). Changes in the manifestation of psychopathology at roughly predictable ages may reflect changes in the tasks that children face across development. A review by Hankin and colleagues highlighted that within the internalizing spectrum, specific disorders and syndromes follow developmentally informed patterns (Hankin et al., 2016). Separation anxiety and specific phobias are highest at early ages, then decrease in adolescence. Social anxiety and generalized anxiety are most prevalent in middle childhood, but panicrelated symptoms become most prevalent in adolescence (Beesdo et al., 2009; Costello et al., 2011; Hankin et al., 2016). In summary, there is mounting evidence that there are developmentally informed changes in lower-order spectra of psychopathology that span across diagnoses.

The culmination of the order and consequences of how a child copes with these developmental tasks are called developmental cascades, and they are thought to map onto one's course of pathology development (Cicchetti, 2020). From this perspective, some children have successfully surpassed a given milestone, while others have not (Sroufe, 2009). Because less attention has been paid to general psychopathology in childhood from a developmentally informed perspective, little is known about whether changes in manifestation of psychopathology correspond to periods when children typically encounter developmental tasks. One example is that 5-6-year-old children spend more time away from parents at primary school compared to their youngeraged selves and peers, which occurs concurrently with a developmental task where children start demonstrating a desire for more autonomy-seeking behaviors (Cicchetti & Rogosch, 2002; Lahey et al., 2021; Sroufe, 2009, 2016). Therefore, developmental tasks, and when they occur, may provide a useful metric for accounting for individual differences in behavior at given ages.

This shift in the presentation of psychopathology across development has led to many questions about whether there are higher-order factors, such as *p* factor that might account

for why some individuals are more likely to develop pathology than others (Smith et al., 2020). There are two theoretical frameworks that suggest that the general factor changes in strength over development. Dynamic mutualism suggests that the general factor represents local interactions of symptoms that directly influence and reinforce one another, resulting in the increase in the strength of the general factor over time, due to an increased number of symptoms and correlations (Caspi et al., 2014; McElroy et al., 2018a; Murray et al., 2016). Another theory is that the general factor represents a general liability for psychopathology and is strongest at a young age. This theory, *p-differentiation*, posits that as a child ages, the symptoms of psychopathology differentiate, more specific symptoms emerge, and the strength of general psychopathology decreases (McElroy et al., 2018a; Murray et al., 2016; Patalay et al., 2015).

Prior Studies Examining Stability Versus Change in the General Factor

Prior work has examined stability versus change in multiple aspects of general psychopathology, including stability and change of individual differences, structure, and strength. Studies have shown relative stability in individual differences in the *p* factor, even when different measurements of psychopathology are used to assess symptoms and diagnoses, suggesting that *p* factor is relevant and meaningful across development (Smith et al., 2020). Studies have found stability of individual differences in the *p* factor across ages 2–14 (β =0.52–0.76; McElroy et al., 2018a), and ages 13–15 (β =0.86; Snyder et al., 2017b).

Studies have also examined the stability in the structure of the p factor, and conflicting evidence has emerged. Castellanos-Ryan et al. (2016) found that substance use and internalizing indicator loadings on the p factor were stronger at age 16 compared to age 14 years, consistent with changes, and therefore, instability, in the structure of the p factor across development (heterotypic continuity). By contrast, Snyder et al., (2017b) found that loadings on the p factor were largely invariant from ages 13 to 15 years, suggesting relatively stable structure in the general factor during this adolescent period.

As evidenced by dynamic mutualism and *p*-differentiation theories, there is a lack of consensus of whether explained common variance (ECV) or an equivalent metric of factor strength, is stable, increases, or decreases with age. To our knowledge, only a few longitudinal studies (Castellanos-Ryan et al., 2016; Choate et al., 2022; Constantinou, 2019; McElroy et al., 2018a, 2018b; Murray et al., 2016) have explored the changes in strength of a general psychopathology factor, i.e. ECV, throughout childhood and adolescence. Findings in these studies have varied. Some studies have shown fluctuations of increases and decreases (Choate et al., 2022)—others have shown no change (McElroy et al., 2018a)—in general factor strength across childhood and adolescence. Due to inconsistent findings, there is a need to explore changes, i.e. increases, decreases, or stability, in general psychopathology factor strength in childhood and adolescence through a developmental psychopathology lens to better conceptualize and prevent development of psychopathology. Developmental psychopathologists would be interested in understanding the amount of explained common variance and the timing of fluctuations, because these changes may map onto known developmental tasks and circumstances.

One longitudinal study examined explained common variance in an expanded age range of 14–21 year olds and found that explained common variance in the general factor appeared to increase stepwise (see Fig. 2; Choate et al., 2022). Explained common variance from ages 14 to 16 years slightly decreased (from 0.60 to 0.57), then increased from ages 16 to 18 (from 0.57 to 0.71) where it remained until hitting a peak at age 21 (ECV = 0.75), but overall stayed relatively stable across this period (Choate et al., 2022).

A longitudinal study of children ages 2 to 14 years also found fluctuations, but relative stability, in the strength of the general factor across ages; (ECV = 0.60-0.71; McElroy et al., 2018a). Taken together, findings from two longitudinal studies suggest that the general factor accounts for approximately 60–75% of the reliable variance, and that there are modest fluctuations that occur at different developmental periods (see Fig. 2; Choate et al., 2022; McElroy et al., 2018a).

One review of cross-sectional and longitudinal studies found that, when fit to a nonlinear trajectory, explained common variance in childhood (ages 2–12) showed that the general factor accounted for 56% of explained common variance, whereas in adolescence (ages 13–17), this value declined subtly to 54%, and then increased in adulthood (ages 18–40) to 60%, following a *u-shape* trajectory (Constantinou, 2019). Interestingly, this finding suggests that explained common variance might decrease across childhood and adolescence before increasing into adulthood. Taken together, the studies suggest that there may be small fluctuations in the strength of the general factor in explaining individual differences across the lifespan.

To our knowledge, only one study has conducted a comprehensive systematic review on the changes in explained common variance across early childhood, middle childhood, and adolescence (Constantinou, 2019). However, this review calculated an average explained common variance to estimate general factor strength, rather than a meta-analysis which would have provided confidence intervals in estimations. This review also evaluated general factor strength across development by plotting a study's



Fig. 2 Two Longitudinal Evaluations of ECV Change. *ECV* explained common variance. Data from McElroy et al., 2018a included a longitudinal sample of 2–14 year olds. Data from Choate et al., 2022 included a separate sample of 14–21 year olds

explained common variance against the study's mean age, rather than a meta-regression. Meta-regression is needed to evaluate whether explained common variance changes across time and as a function of other factors. To our knowledge, no previous meta-analysis has aggregated the relevant literature and used robust multilevel metaregression to test whether the proportion of variance in ratings of psychopathology differs across early childhood, middle childhood, and adolescence. Nor has such a review been conducted in adults. Fortunately, a growing number of cross-sectional and a few longitudinal child and adolescent studies have generated general psychopathology models using factor analysis, which provide the information needed to conduct a meta-analysis to address this gap in the literature. In summary, there is little consensus from prior work as to whether the strength of the general factor of psychopathology increases, decreases, or is stable across development. Therefore, there is a need for studies that examine general psychopathology to account for developmentally informed changes across childhood and adolescence. However, prior research has shown strong support for differentiation of specific symptoms of psychopathology across development, which would result in a decrease in general factor strength, supporting the p-differentiation hypothesis (Choate et al., 2022; McElroy et al., 2018a; Murray et al., 2016; Patalay et al., 2015). When paired with evidence that only 10% of mental disorders begin to manifest as observable behaviors at or before age 5, it might be the case that psychopathology is more general at younger ages and then becomes more specific throughout development, leading to a decrease in general factor strength throughout development (Kessler et al., 2005).

The Present Review

The aim of the present meta-analysis is to examine factor analytic models of general psychopathology in children and adolescents (e.g. bifactor, modified bifactor, and higherorder factor models) to determine whether explained common variance, a measure of general factor strength, changes across childhood and adolescence. We hypothesize that general psychopathology will account for more variance in childhood than adolescence, functionally taking the form of a negative age moderation from a meta-regression analysis, supporting the *p*-differentiation theory. We expect change in general psychopathology strength because previous research has indicated that specific symptom expression and presentations are likely not at their "end-stage" earlier in development and that psychopathology becomes more specified throughout the lifespan (Forbes et al., 2019). In addition, informants are likely to observe a heterogeneous expression of psychopathology and may be unable to differentiate specific forms of psychopathology at younger ages, resulting in more broad representation of psychopathology. If we find differences in general factor strength across development, a secondary aim of the present review would be to map these differences onto expected developmental tasks (e.g. Sroufe, 2016). If, for example, we find that general factor strength decreases during preschool age, we might investigate whether development of, or challenges to the development of, self-regulation plays a role in this change in psychopathology strength (Sroufe, 2016).

If the hypothesis is not supported, and general factor strength either increases across development or does not change, these results would still have implications for future interventions and research. For example, if general factor strength does not change across development, it would suggest that general psychopathology strength is stable across development and is interpretable as an overall impairment or liability, identifiable (and potentially treatable) from a young age. We anticipate that several factors may interact with general factor strength to alter the slope of factor strength over development. Several exploratory moderators were evaluated in the present review.

Subgroup Sensitivity Analyses

We examine the strength of the general factor and how it differs between subgroups. These analyses include subgrouping by: model type (i.e. bifactor or higher-order factor models), study wave (i.e. timepoint or measurement occasion), studies that have a small or large variability in age at a given wave, developmental period (i.e. preschool, school-age, adolescence), longitudinal studies, studies that established at least partial longitudinal metric invariance, and explained common variance of specific factors (ECVs). The subgroup analyses are motivated by prior research indicating that general factor strength may differ as a function of differences in sample or modeling characteristics. For example, hierarchical models and bifactor models tend to differ in their factor loading strengths (Lahey et al., 2021), and data from longitudinal studies are thought to provide a stronger test of change in factor strength compared to cross-sectional studies (Ringwald et al., 2021, 2022). In addition, developmental stages-preschool age, school age, and adolescence-and the associated developmental tasks with these key developmental stages have impact on development of psychopathology (Cicchetti, 2020; Cicchetti & Rogosch, 2002). We also examined moderation sensitivity analyses.

Moderation Sensitivity Analyses

As a contrast to the sum-based estimate of ECV, we calculate a mean-based estimate of general factor strength (see Eq. 4 in Supplemental Appendix 3) to determine if the method by which general factor strength is estimated alters the results of the study. This analysis is motivated by concerns that a larger number of indicators on the general factor likely inflates sum-based ECV estimates (Watts et al., 2021a). Similarly, we conduct additional analyses on modeling approaches, including general factor indicator count, and factor count, to further explore method and modeling-related variables (Watts et al., 2019). According to Rodriguez et al. (2016), PUC is a metric of how a measurement of a general factor is 'uncontaminated' by multidimensionality due to specific factors, and represents the suitability of the model to assess a general factor of psychopathology. PUC is calculated as the number of correlations explained by the general factor compared to the number of within-specific factor item correlations. PUC was found to moderated the association between ECV across age, where higher PUC resulted in a strong positive slope, where lower PUC resulted in stronger negative slope (Constantinou, 2019). Therefore, we calculated PUC and included it as a moderator to reexamine these prior analyses.

Due to well-established sex-related differences in the development of specific psychopathology; e.g. boys experience more externalizing symptoms whereas girls experience more internalizing symptoms (Hinnant & El-Sheikh, 2013; Mayes et al., 2020; Mesman et al., 2001); and boys show higher levels of general psychopathology than girls (Lynch et al., 2021), we also examine whether general factor strength across development differs by sex. Ratings of psychopathology were collected from parent, self, and teacher reports from either a questionnaire or a structured clinical interview. Previous research suggests that measure type and informant-related biases influence estimates of general factor strength (Conway et al., 2019; Laceulle et al., 2015; Lahey et al., 2012; Martel et al., 2017). Some have suggested that method effects, such as informants and measure, may account for about 25% of variance in general factor strength (Constantinou, 2019; Cote & Buckley, 1987). Thus, we also examine the strength of the general factor as a function of the measure and informant type. In addition, we evaluated whether ECV changes across development when including sample size as a moderator and setting the sampling variance to constants.

These moderation analyses are informed by prior research, and they aim to elucidate areas for future study. Understanding the degree to which developmental trajectory is associated with differing degrees of general psychopathology risk from a developmental psychopathology perspective, is a novel and important gap in the literature that can be used to inform research and clinical evaluation of child and adolescent psychopathology.

Method

Procedure

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for conducting this meta-analysis and reporting findings (Page et al., 2021). See https://osf.io/pyc9r for our PRIMA 2020 checklist.

Search Strategy and Eligibility Criteria

Studies were compiled by the first author with the assistance of a psychology librarian in November 2021 through a systematic search using the following electronic databases: PsycINFO, PubMED, Embase, CINAHL Plus, Scopus, Web of Science (Core Collection), ProQuest Dissertations, and Theses Global. See https://osf.io/d3a8n for full list of search terms from all databases. In total, 3200 articles were screened for inclusion. Additional description of deduplication and screening procedures are in Supplemental Appendix 1.

Inclusion and Exclusion Criteria

The exclusion criteria were (1) the study did not report empirical findings (e.g. reviews or meta-analyses); (2) mean participant age in the study was > 18.00 years; (3) the study did not use factor analysis to model psychopathology; (4) internalizing and/or externalizing (or their sub-factors) were not evaluated in the study; (5) the study separately evaluated psychopathology factors (i.e. the study did not examine the covariation of internalizing and externalizing); and (6) the psychopathology model included extraneous latent factors that cannot be categorized as sub-factors of internalizing, externalizing, and/or thought-disordered problems (e.g. model includes latent factors of personality, stress, wellbeing, etc.). However, models with latent factors that represented a sub-factor of internalizing or externalizing problems were retained. An example of a sub-factor is the use of both fear and distress in substitution of a single internalizing disorders factor (Martel et al., 2017). Moreover, attention problems and conduct problems may be represented as distinct but related factors that comprise total externalizing problems (Clark et al., 2021; Haltigan et al., 2018; Harden et al., 2020; Neumann et al., 2020; Sheldrick et al., 2012). The sixth exclusion criterion was intended to retain a uniform definition of general psychopathology. Models were limited to what might fall under the scope of Caspi and colleagues' p factor, including only factors of psychopathology conceptualized as thought-disordered, internalizing, and externalizing dimensions Capsi et al. (2014). Models including these three factors have the most empirical support (Bates et al., 2014; Forbes et al., 2021b; Kotov et al., 2017, 2021). Therefore, studies with extraneous non-psychopathology factors (e.g. well-being, stress, and personality dimensions) would introduce heterogenous conceptualizations of general psychopathology, changing the meaning of p factor, and were, thus, excluded.

Title and Abstract Screening

The first and second authors both independently screened all titles and abstracts using Rayyan (Ouzzani et al., 2016). Studies that clearly met eligibility criteria or were inconclusive were passed to the data extraction phase where the full text was reviewed.

Data Extraction Criteria and Study Selection

Prior to undergoing the full data extraction process, inclusion and exclusion criteria were assessed using information from the full text. Information on how exclusion criteria were reported are in Supplemental Appendix 1. A subset of approximately 20% of the studies were independently coded by the first two authors to determine their reliability. Intraclass correlation coefficient (ICC) reliability between the coders ranged from 0.92 to 1 on five key variables (e.g. exclusion criteria, general factor strength, sum of squared factor loadings of externalizing, internalizing, and thought-disordered specific factors). Following reliability check, all discrepancies were discussed and resolved. Remaining studies were divided among the two coders. The meta-analysis coding manual is available on the Open Science Framework (https:// osf.io/fvbsu). If an included study did not contain a table or figure in the full-text or supplemental material, we requested relevant information from the study authors by email. If the authors did not respond, and there was no other factor loading information available, the study was not included in the present meta-analysis.

Statistical Analysis

Effect Size

The study effect size was taken to be the explained common variance in ratings of psychopathology. Standardized (β) loadings for the specific and general/*p* factor were extracted from a table or figure in the manuscript or supplementary materials. The explained common variance was calculated by dividing the variance explained by the general factor (i.e. sum of squared general factor loadings) by the total reliable variance (i.e. sum of squared general and specific factor loadings) using Eq. 1 (Constantinou, 2019; Rodriguez et al., 2016). Reliable variance is similar to a total variance estimation, but reliable variance, as measured in the present meta-analysis, does not include an error estimate. The sum of squared loadings from a given sub-factor are summed to represent the variance explained by the given specific factor.

$$\frac{\left(\sum \beta^2_{Gen}\right)}{\left(\sum \beta^2_{Gen}\right) + \left(\sum \beta^2_{Ext}\right) + \left(\sum \beta^2_{Int}\right) + \left(\beta^2_{TD}\right)} \tag{1}$$

To calculate ECV for higher-order models, an additional step was taken. For higher-order models, we follow path tracing rules (Loehlin, 2003), where the β s between the general factor and the indicator, typically passing through at least a specific factor, are multiplied to derive the value that represents the regression coefficient from indicator to general factor. The specific factor may also require subordinate sub-factors that require similar path tracing multiplication. The derived β s are then calculated into the explained common variance using Eq. 1.

Given that the effect size was taken to be a calculated proportion score (p), there was no provided sampling variance; therefore, one needed to be estimated to provide a metric of standard error to the effect size due to sample size. Based on prior literature (e.g. Moeyaert et al., 2017), we derived the sampling variance for each proportion (p) using the proportion score and the sample size (n). As seen in Eq. 2, the sample size (n) appears in the denominator of the square root, indicating that larger samples have less sampling variability compared to smaller samples. Effect sizes are normally distributed and the sample sizes are large enough to justify this method of sampling variance calculation (Moeyaert et al., 2017).

$$\sqrt{\frac{p(1-p)}{n}} \tag{2}$$

Accounting for Non-independence of Nested Data

Given the prevalence of large cohort study datasets and independent samples used across multiple studies, we accounted for nonindependence of observations by nesting within a given study, within wave (i.e. prospective measurement occasion or timepoint from a given data source), and within sources of participant data (i.e. large cohort studies or independently collected samples shared by multiple studies). Each unique data source was given a categorical code, e.g. ABCD was given the corresponding code of 1. If the study used, for example, wave 2 of the data source (e.g. wave 2 of ABCD = 11 years of age), then all studies that used the same data source at a given wave were assigned the same corresponding wave code (e.g. 2), in addition to the same data source code (e.g. 1). This method also allowed us to retain as many effect sizes at as many ages as possible, which was essential for testing our primary hypotheses. These methods of nesting data are known to be robust for accounting for nonindependence of data within meta-analyses (e.g. Konstantopoulos, 2011; McCurdy et al., 2020). A multilevel meta-analysis approach in the R package metafor was used to derive the pooled statistical effect size, explained common variance, while accounting for the nested structure of the data (i.e. study within waves(s) within data sources(s); Viechtbauer & Viechtbauer, 2021).

Evidence of heterogeneity from effect sizes was examined using the Q statistic found in *metafor*, which outputs a chisquare distribution based on k-1 degrees of freedom in which k is the number of effect sizes derived from studies to account for between-study variance (Cochran, 1954; Huedo-Medina et al., 2006; Ringwald et al., 2021). Total amount of heterogeneity was calculated by I^2 statistic, a robust estimate of the amount of heterogeneity present in a given dataset (Higgins & Thompson, 2002). Multilevel meta-analyses with nonindependence of data (i.e. nested data) have used the I^2 to determine the percentage of heterogeneity due to between-cluster and within-cluster levels, given nested data (Konstantopoulos, 2011). To calculate I^2 , we used analysis script templates provided by experts in multivariate metaanalyses (e.g. Konstantopoulos, 2011; Viechtbauer & Viechtbauer, 2021).

Age-Moderated Changes

The primary goal of the present meta-analysis was to determine whether strength of the general factor of psychopathology changes across childhood to adolescence. Metaregression moderation analysis was used to calculate the degree to which this strength changes across the mean ages of included studies. Data were nested within a given study, within wave, and within sources of participant data. Sample mean age was examined as a potential moderator in metafor to conduct meta-regression analysis (Viechtbauer & Viechtbauer, 2021). A significant age moderation would indicate that the slope of ECV as a function of sample mean age is different from 0. We hypothesized a decrease, or negative slope, of ECV across sample mean ages. If the moderation is not significantly different from 0, that would indicate that general factor strength does not significantly change across ages.

Publication Bias

Currently, multilevel meta-analyses are unable to use graphical and quantitative methods of publication bias (e.g. funnel plot, trim-and-fill plot) using nested data because these approaches do not adequately account for multiple effect sizes that come from a single study or sample (Assink & Wibbelink, 2016; Rodgers & Pustejovsky, 2021). Nevertheless, we generated a contour-enhanced funnel plot and a trim-and-fill plot with analyses using the effect sizes without nesting the data. A traditional funnel plot visualizes the effect size of each study plotted against the standard error, a function of the study's sample size, to determine whether effects from smaller studies are more variable than effects from larger studies (Peters et al., 2008). A contour-enhanced funnel plot includes colored areas of significance thresholds, one showing effects between p = 0.05-0.1, and another showing p = 0.01-0.05, representing significant deviations from the pooled meta-analyzed effect size (Peters et al., 2008).

A trim-and-fill plot attempts to correct for asymmetry by estimating the number of studies needed to be imputed on one side of the figure to balance the asymmetry (Duval & Tweedie, 2000; Rodgers & Pustejovsky, 2021). A commonly used method for assessing publication bias, even in multilevel meta-analyses, is Egger's regression test (Egger et al., **1997**). The metric of interest is the intercept, β_0 , because a regression line through symmetrical data in a funnel plot would have a β_0 not significantly different from 0, whereas asymmetrical data due to small sample size influences will have an intercept significantly different from 0 (Egger et al., 1997). Egger's test tends to overestimate bias and may lead to false positives and, thus, should be interpreted carefully (Pustejovsky & Rodgers, 2019). Egger's test is calculated by performing a meta-regression with the standard error of the effect size (i.e. sampling variance) as a moderator.

Study Quality

To assess the quality of the included studies, we modified the Downs and Black (1998) checklist for assessing methodological quality. See Supplemental Appendix 2 for the modified study quality checklist and scoring of study quality.

Subgroup Analyses

Several subgroup analyses were conducted to determine if general factor strength differed based on groups. These analyses included: model type (i.e. bifactor or higherorder model), longitudinal studies, studies that established at least partial measurement invariance, study wave (i.e. measurement occasion or timepoint), studies that have a small or large variability in age at a given wave, developmental period (i.e. preschool age, school age, adolescence), explained common variance of specific factors (ECVs), and a mean-based estimate of general factor strength. For each subgroup analysis, we first generated the meta-analytic estimate of ECV without age as a moderator. Second, within each subgroup, we examined age as a moderator to determine whether ECV differed by age. For details on subgroup analyses, see Supplemental Appendix 3.

Moderation Analyses

Several moderation analyses were conducted to determine if the strength of the general factor, and whether the age moderation, differs when including (separately) each of the following factors as a potential moderator: mean-based estimate of general factor strength, general factor indicator count, factor count, sex composition of the sample, informant type, mono- versus multi-informant, whether ratings of psychopathology came from questionnaire versus interview, sample size, and percent uncontaminated correlations. Among the combinations of informants and measures in the meta-analysis, 5 effect sizes-4.5%-derived from models that included multiple informants and both interviews and questionnaires. Nine effect sizes-8.2%-were derived from models that included only questionnaires and had multiple informants, and 10 effect sizes-9.1%-were derived from models that only included interviews and had multiple informants. For each moderator examined, we first evaluated whether the moderator was associated with the estimate of ECV. Second, we added age as a moderator, to determine whether age was associated with ECV when accounting for a given moderator. For details on the moderator analyses, see Supplemental Appendix 3.

Nonlinear Trajectory of Age

We examined whether the change in general factor strength followed a nonlinear trajectory. We centered age such that the intercept was set at the youngest age, 2 years of age, by subtracting 2 from each age. Next, we squared these centered ages to derive a quadratic term (i.e. centered age²) that allowed for a test of nonlinear moderation of age. To evaluate whether there was nonlinear age moderation, the quadratic-centered age term was added as a second moderator along with the linear-centered age term. To further evaluate the possibility of nonlinear age moderation, a cubiccentered age term (i.e. centered age³) was added as a third moderator along with the linear and quadratic-centered age terms in a separate analysis.

Results

Inclusion of Studies

As shown in Fig. 3 3200 deduplicated studies were identified. After screening abstracts and titles, 233 studies were sought for retrieval. 63 of these studies were conference poster or symposium abstracts, and the authors were contacted for additional details, 10 articles were not in English and their abstracts were translated to determine inclusion, and 6 articles were found to be duplicates of other retrieved studies. Authors were contacted, yielding an additional 14 studies to be assessed for eligibility. Including additional studies from contacted authors, a total of 168 articles were assessed for eligibility. Of these 168 articles, a total of k=65 articles were included in this review. Given the nested data structure in which multiple effect sizes might be found in a given study, a total of 110 distinct effect



Fig. 3 PRISMA Diagram. Rayyan was used for reviewing study abstracts. REDCap was used for full-text eligibility assessment and study data extraction. Diagram template from Page et al., (2021)

sizes were derived from the 65 studies. The included studies and a snapshot of study characteristics are shown in Table 1.

Overall Explained Common Variance and Heterogeneity

The 110 distinct effect sizes derived an aggregate effect size representing the proportion of variance in

psychopathology ratings accounted for by the general factor of 0.56, SE = 0.02, p < 0.001. These results indicate that general psychopathology accounted for approximately 56% of the reliable variance across the included studies. The forest plot is shown in Fig. 4. The homogeneity Q statistic (Q = 250.24, p < 0.001) indicated significant variability in the 110 individual effect sizes nested within the 65 studies. A total I^2 value of 54.83% indicated that approximately

Table 1 Include	ed study characte	eristics											
Study	Data Source	Sample Size (N)	Mean Age (SD)	% Female	Country	Model Type	Informant	Measure	# EXT	# INI #	I QL #	3CV Sampling Variance)	Study Quality (MeanlSum)
Afzali et al., 2018	Co-Venture	3826	12.8 (0.4)	49.2	CA	Mod. Bifactor	Self	SDQ, BSI, APSS	7	20	6	.42 (0.008)	0.85 11
Aitken et al., 2020	IMPACT	465	15 (NA)	75	GB	Bifactor	Self, Parent(s)	MFQ, RCMAS, LOI, BC	Γ	30	NA ().58 (0.023)	1 16
Blanco et al., 2015	NCS-A	6483	15 (NA)	51.1	SU	Hierarch	Parent(s), Teacher	CIDI	9	10	NA ().38 (0.006)	0.73 11
Bloemen et al., 2018	TRAILS	2230	11.1 (0.6)	51	NL	Bifactor	Parent(s)	CBCL, CSBQ	9	7	9).68 (0.01)	0.94 15
Brandes et al., 2019	From Tackett, 2011	695	9.87 (NA)	52	SU	Bifactor	Parent(s)	CBCL, C-DISC IV	5	6	NA (.42 (0.019)	0.94 15
Calkins et al., 2015	PNC	9421	14.3 (3.67)	51.7	SU	Bifactor	Parent(s)	GOASSESS	4	11	NA ().69 (0.005)	0.94 15
Carragher et al., 2016	CAP	2175	13.3 (0.48)	42.6	AU, NZ	Mod. Bifactor	Self	SDQ, BSI, RAPI	15	20	6).42 (0.011)	0.93 14
Castellanos- Ryan et al., 2016	IMAGEN	2144	14.39 (0.77)	51	GB, IE, FR, DE	Bifactor	Self, Parent(s)	DAWBA, AUDIT	4	6) AN).47 (0.011)	0.94115
Chen et al., 2022	CATSS	4786	15 (NA)	51	SE	Bifactor	Self	SDQ	10	5	NA (.44 (0.007)	0.87 13
Chen et al., 2022	CATSS	4786	15 (NA)	51	SE	Bifactor	Parent(s)	SDQ	10	4	NA ().51 (0.007)	0.87 13
Chen et al., 2022	PGS	2450	14 (NA)	66.66	SU	Bifactor	Self	SCARED, ASI-4, SCAS	4	7	NA ().59 (0.01)	1 16
Choate et al., 2022	PGS	2450	15 (NA)	66.66	SU	Bifactor	Self	SCARED, ASI-4, SCAS	4	0) NA).49 (0.01)	1 16
Choate et al., 2022	PGS	2450	16 (NA)	66.66	SU	Bifactor	Self	SCARED, ASI-4, SCAS	4	7	NA ().5 (0.01)	1 16
Choate et al., 2022	PGS	2450	17 (NA)	66.66	SU	Bifactor	Self	SCARED, ASI-4, SCAS	4	7) AN).57 (0.01)	1 16
Clark et al., 2021	ABCD	11,875	9.92 (0.62)	48.4	SU	Bifactor	Parent(s)	CBCL	29	31	NA ().85 (0.003)	0.93 13
Clark et al., 2021	ABCD	11,875	9.92 (0.62)	48.4	SU	Hierarch	Parent(s)	CBCL	35	31	NA ().42 (0.005)	0.93 13
Clark et al., 2021	ABCD	11,875	9.92 (0.62)	48.4	SU	Bifactor	Parent(s)	CBCL	5	6	NA ().81 (0.004)	0.93 13

Table 1 (contir	(pənı												
Study	Data Source	Sample Size (N)	Mean Age (SD)	% Female	Country	Model Type	Informant	Measure	# EXT	# INI #	I QL#	3CV Sampling Variance)	Study Quality (MeanlSum)
Clark et al., 2021	ABCD	11,875	9.92 (0.62)	48.4	US	Hierarch	Parent(s)	CBCL	2	ς Γ	NA (.41 (0.005)	0.93 13
Class et al., 2019	SLL	3990	13.6 (2.5)	52.1	NS	Bifactor	Parent(s)	CADS	9	1) VA	.47 (0.008)	1116
Class et al., 2019	SLL	3990	13.6 (2.5)	52.1	NS	Bifactor	Self	CADS	9	1) VA	.46 (0.008)	1116
Constantinou et al., 2019	START	684	13.8 (1.4)	18	GB	Bifactor	Self	SDQ, MFQ	11	13]	NA (.53 (0.019)	0.86 12
Deutz et al., 2020	SECCYD	1073	11 (NA)	49.8	SU	Bifactor	Mother	CBCL	26	13]	NA ().67 (0.014)	0.8112
Deutz et al., 2020	SECCYD	1073	11 (NA)	49.8	SU	Bifactor	Mother	CBCL	26	31]	NA ().66 (0.014)	0.8 12
Etkin et al., 2020	Ind. Sample	835	14.35 (1.58)	49	ES, RO, MA	Bifactor	Self	SENA	33	ς Γ	NA ().62 (0.017)	0.6919
Etkin et al., 2021	From Etkin et al., 2020	831	14.35 (1.58)	50.6	ES	Bifactor	Self	SENA	33	5	NA ().6 (0.017)	0.81113
Etkin et al., 2021	From Etkin et al., 2020	619	14.74 (1.22)	50.8	ES	Bifactor	Self	SENA	33	5	NA ().65 (0.019)	0.81113
Etkin et al., 2021	From Etkin et al., 2020	465	15.22 (1)	49.9	ES	Bifactor	Self	SENA	33	5	NA ().56 (0.023)	0.81113
Faber & Chaplin, 2020	Ind. Sample	294	10.35 (3.6)	47.85	SU	Bifactor	Parent(s)	у-оо	11	12) VA).6 (0.029)	0.83110
Forbes et al., 2020	LSAC	3335	14.41 (0.49)	49.1	AU	Hierarch	Self	SDQ, SMFQ	б	6	NA ().39 (0.008)	0.94 15
Gomez et al., 2019	ABCD	1233	11.22 (3.1)	28.2	AU	Bifactor	Mother, Father	CBCL, ADISC	б	10	NA ().42 (0.014)	0.93 14
Gomez et al., 2019	ABCD	866	15 (NA)	28.2	AU	Bifactor	Mother, Father	CBCL, ADISC	33	10	NA ().52 (0.017)	0.93 14
Haltigan et al., 2018	Ind. Sample	1552	12.52 (3.91)	23.2	CA	Bifactor	Parent(s)	CBCL	36	30	12 ().63 (0.012)	1113
Hamlat et al., 2019	GEM	567	13.58 (2.37)	55.5	US	Bifactor	Self, Parent(s)	CBCL, YSR, MASC, CDI, EATQ-R, SNAP-IV	ŝ	4	NA ().52 (0.021)	0.92112
Harden et al., 2020	Texas Twin Project	1913	13.1 (NA)	49	SU	Bifactor	Self	CBCL, Con- ner's 3, BFI	8	ε	NA ().55 (0.011)	1115
Harden et al., 2020	Texas Twin Project	1913	13.1 (NA)	49	SU	Bifactor	Parent(s)	CBCL, Con- ner's 3, BFI	∞	с,	NA ().62 (0.011)	1115

Table 1 (conti	nued)												
Study	Data Source	Sample Size (N)	Mean Age (SD)	% Female	Country	Model Type	Informant	Measure	# EXT	# INT	#TD	ECV (Sampling Variance)	Study Quality (MeanlSum)
He & Li, 2021	PNC	3088	13.86 (0.07)	54.05	SU	Bifactor	Self, Parent(s)	GOASSESS	б	6	б	0.73 (0.008)	0.85 11
He & Li, 2021	PNC	5147	13.73 (0.05)	49.64	SU	Bifactor	Self, Parent(s)	GOASSESS	ю	6	ю	0.72 (0.006)	0.85 11
Herzhoff & Tackett, 2018	From Tackett, 2011	349	9.82 (0.66)	53	SU	Bifactor	Mother	CBCL, C-DISC IV	5	Ś	NA	0.5 (0.027)	1114
Huang-Pol- lock et al., 2017	Ind. Sample	415	10 (NA)	40.96	SU	Bifactor	Mother, Father	BASC-2	4	б	NA	0.59 (0.024)	1115
Huang et al., 2020	TTC	3171	10.18 (0.28)	46.9	ď	Bifactor	Self, Mother, Father	CBCL, SDQ, APSS, SMFQ	33	×	ε	0.38 (0.009)	0.81 13
Huang et al., 2020	TTC	3007	12.17 (0.31)	47.2	dſ	Bifactor	Self, Mother, Father	CBCL, SDQ, APSS, SMFQ	3	×	б	0.34 (0.009)	0.81 13
Jenness et al., 2021	Ind. Sample	160	12.63 (2.68)	52	SU	Bifactor	Self, Parent(s)	K-SADS	ŝ	б	NA	0.45 (0.039)	1 14
Laceulle et al., 2020	TRAILS	2230	13.6 (0.59)	50.8	NL	Mod. Bifactor	Parent(s)	CBCL	ŝ	5	NA	0.76 (0.009)	1116
Laceulle et al., 2020	TRAILS	2230	13.6 (0.59)	50.8	NL	Mod. Bifactor	Self	YSR, CADS, CAPE	б	9	NA	0.78 (0.009)	1116
Laceulle et al., 2015	TRAILS	2230	13.6 (0.59)	NA	N	Mod. Bifactor	Self, Parent(s)	CBCL, YSR, CADS, CAPE	б	9	NA	0.76 (0.009)	0.75 12
Lahey et al., 2015	PGS	2450	8 (NA)	66'66	SU	Bifactor	Parent(s)	CSI-4, SCARED	5	9	NA	0.51 (0.01)	1116
Lambert et al., 2018	Ind. Sample	626	8.62 (1.74)	51	SU	Bifactor	Teacher	EBS	9	б	NA	0.74 (0.018)	0.77 10
Lees et al., 2021	ABCD	11,875	(9.0) 6.6	47.9	SU	Hierarch	Parent(s)	K-SADS	ŝ	٢	4	0.43 (0.005)	0.62 10
Martel et al., 2017	HRC	2512	9.65 (1.93)	46.2	BR	Mod. Bifactor	Parent(s)	DAWBA	ŝ	11	NA	0.55 (0.01)	1115
McElroy et al., 2018a	SECCYD	1253	2 (NA)	51	SU	Bifactor	Mother	CBCL	24	36	NA	0.59 (0.014)	0.69 11
McElroy et al., 2018a	SECCYD	1253	3 (NA)	51	SU	Bifactor	Mother	CBCL	24	36	NA	0.61 (0.014)	0.69 11
McElroy et al., 2018a	SECCYD	1253	5 (NA)	51	SU	Bifactor	Mother	CBCL	37	30	NA	0.71 (0.013)	0.69 11
McElroy et al., 2018a	SECCYD	1253	6 (NA)	51	SU	Bifactor	Mother	CBCL	35	31	NA	0.66 (0.013)	0.69111

Table 1 (contir	(pənu												
Study	Data Source	Sample Size (N)	Mean Age (SD)	% Female	Country	Model Type	Informant	Measure	# EXT	# INT	#TD	ECV (Sampling Variance)	Study Quality (MeanlSum)
McElroy et al., 2018a	SECCYD	1253	6 (NA)	51	NS	Bifactor	Mother	CBCL	35	31	NA	0.64 (0.014)	0.69111
McElroy et al., 2018a	SECCYD	1253	9 (NA)	51	SU	Bifactor	Mother	CBCL	34	31	NA	0.68 (0.013)	0.69 11
McElroy et al., 2018a	SECCYD	1253	10 (NA)	51	SU	Bifactor	Mother	CBCL	35	31	NA	0.69 (0.013)	0.69 11
McElroy et al., 2018a	SECCYD	1253	11 (NA)	51	SU	Bifactor	Mother	CBCL	37	31	NA	0.65 (0.013)	0.69 11
McElroy et al., 2018a	SECCYD	1253	14 (NA)	51	SU	Bifactor	Mother	CBCL	39	31	NA	0.71 (0.013)	0.69 11
Mewton et al., 2021	ABCD	11,721	9.91 (0.62)	47.8	SU	Hierarch	Parent(s)	K-SADS	б	٢	4	0.43 (0.005)	1116
Mollon et al., 2021	PNC	9421	13.8 (3.6)	49.7	SU	Bifactor	Self, Parent(s)	GOASSESS	25	09	27	0.61 (0.005)	0.92 12
Mollon et al., 2021	PNC	9421	13.8 (3.6)	49.7	SU	Hierarch	Self, Parent(s)	GOASSESS	25	09	27	0.42 (0.005)	0.92 12
Moore et al., 2019	PNC	5563	15.7 (2.6)	51.7	SU	Bifactor	Self	GOASSESS	24	62	24	0.61 (0.007)	0.86 12
Moore et al., 2020	ABCD	5934	9.5 (NA)	47.9	SU	Bifactor	Parent(s)	CBCL	41	26	NA	0.73 (0.006)	0.93 13
Moore et al., 2020	ABCD	5934	9.5 (NA)	47.9	SU	Bifactor	Parent(s)	CBCL	47	19	NA	0.7 (0.006)	0.93 13
Moroney & Lee, 2021	From Hin- shaw, 2002; Shemmas- sian & Lee, 2012	460	8.72 (1.67)	65.65	SU	Bifactor	Parent(s)	CBCL, C-TRF, C-DISC IV	4	٢	NA	0.37 (0.022)	1114
Neumann et al., 2016	Generation R	2115	6.8 (1.3)	NA	NL, TR, SR, MA	Bifactor	Self, Parent(s), Teacher	CBCL, TRF, SSRS/ SRS, BPI, CPRS-R	11	10	NA	0.76 (0.009)	1 15
Neumann et al., 2020	Generation R	3030	5.9 (0.3)	50.4	NL	Bifactor	Mother	CBCL	7	4	NA	0.86 (0.006)	1115
Neumann et al., 2020	Generation R	3030	10 (0.3)	50.4	NL	Bifactor	Mother	CBCL	Э	ŝ	NA	0.8 (0.007)	1115
Neumann et al., 2020	Generation R	3030	10 (0.3)	50.4	NL	Bifactor	Father	CBCL	ю	ŝ	NA	0.79 (0.007)	1115
Neumann et al., 2020	Generation R	3030	6.5 (1.1)	50.4	NL	Bifactor	Teacher	TRF	Э	ŝ	NA	0.64 (0.009)	1115
Neumann et al., 2020	Generation R	3030	6 (0.4)	50.4	NL	Bifactor	Self	BPI	б	e	ΝA	0.5 (0.009)	1115

Table 1 (contin	ued)												
Study	Data Source	Sample Size (N)	Mean Age (SD)	% Female	Country	Model Type	Informant	Measure	# EXT	# INI #	TD #	ECV (Sampling Variance)	Study Quality (MeanlSum)
Niarchou et al., 2017	From Gur et al., 2014; Tang et al., 2014a; Tang et al., 2014b	331	16.9 (8.7)	49	US	Bifactor	Self, Parent(s)	K-SADS, SIPS, SCID	19	26	15 (0.42 (0.027)	0.93 13
O'Reilly et al., 2020	CATSS	60,888	10.5 (NA)	NA	SE	Bifactor	Mother, Father	A-TAC	31	12	NA	0.73 (0.002)	0.94 15
Olino et al., 2014	From Olino et al., 2010	559	3.56 (0.27)	46.1	SN	Bifactor	Parent(s)	PAPA, CBQ, Unknown/ Other	4	5	ΥN	0.45 (0.021)	0.86 12
Olino et al., 2018	From Olino et al., 2010	541	3.56 (0.27)	46.1	SU	Bifactor	Parent(s)	PAPA	4	5	NA NA	0.4 (0.021)	0.75 12
Olino et al., 2018	From Olino et al., 2010	466	6.09 (0.44)	46.1	SU	Bifactor	Parent(s)	PAPA	4	5	NA	0.4 (0.023)	0.75 12
Oro et al., 2021	WTP	502	13.24 (1.52)	NA	SU	Bifactor	Self, Parent(s)	CIDI, C-DISC IV	ŝ	1	NA NA	0.68 (0.021)	0.81113
Oro et al., 2019	WTP	1004	13.24 (1.52)	52.5	SU	Bifactor	Self	C-DISC IV	б	[NA	0.68 (0.015)	0.93 13
Patalay et al., 2015	Me and My School	23,477	12.05 (0.56)	50.4	GB	Bifactor	Self	SDQ, Me & My School	11	14	NA NA	0.42 (0.003)	0.8 12
Pettersson et al., 2018	CATSS	16,806	16.7 (NA)	48.25	SE	Mod. Bifactor	Parent(s)	A-TAC	31	12	ΝA	0.64 (0.004)	0.81113
Riglin et al., 2019	ALSPAC	8161	7 (NA)	48	GB	Bifactor	Parent(s)	DAWBA, SCDC	7	5	NA	0.64 (0.005)	0.94 15
Riglin et al., 2019	ALSPAC	7017	13 (NA)	48	GB	Bifactor	Parent(s)	DAWBA, SCDC	7	5	NA	0.63 (0.006)	0.94 15
Rijlaarsdam et al., 2021a	Generation R	440	10 (NA)	50.7	NL	Mod. Bifactor	Mother	CBCL	ŝ	ŝ	NA	0.87 (0.016)	1 14
Rijlaarsdam et al., 2021a	ALSPAC	7814	7 (NA)	50	GB	Mod. Bifactor	Parent(s)	DAWBA	ŝ	4	NA	0.43 (0.006)	1114
Rijlaarsdam et al., 2021b	ALSPAC	7814	7 (NA)	51	GB	Bifactor	Parent(s)	DAWBA	ŝ	4	NA	0.43 (0.006)	0.94 15
Rijlaarsdam et al., 2021b	ALSPAC	7145	10 (NA)	51	GB	Bifactor	Parent(s)	DAWBA	6	S L	NA	0.48 (0.006)	0.94 15
Rijlaarsdam et al., 2021b	ALSPAC	6210	13 (NA)	51	GB	Bifactor	Parent(s)	DAWBA	ŝ	5	NA	0.44 (0.006)	0.94 15
Sheldrick et al., 2012	Ind. Sample	817	3.5 (NA)	45	NS	Bifactor	Parent(s)	PPSC	7	6]	NA	0.8 (0.014)	0.94 15

Sample Size Mean Age % Female Cou (N) (SD) 895 11 54 (NA) 57 11S	Mean Age % Female Cou (SD) 11 54 (NA) 57 11S	% Female Cou	Cou	ntry	Model Type Bifactor	Informant Parent(s)	Measure	# EXT	# LNI # 9	TD ECV (Sampling Variance)	Study Quality (MeanlSum)
	895	11.54 (NA)	52	US, CA	Bifactor	Parent(s)	C-DISC IV C-DISC IV		۲ o	A 0.47 (0.01 /	0.93113
254		7.92 (1.38)	46	SU	Bifactor	Parent(s)	CBCL, CBQ	~	4 Z	A 0.81 (0.024	0.91110
571		13.58 (2.37)	55.5	NS	Bifactor	Self, Parent(s)	CBCL, YSR, MASC, CDI, EATQ-R, SNAP-IV	ŝ	4 N	A 0.52 (0.021	0.81113
510	0	15.07 (2.36)	55.5	US	Bifactor	Self, Parent(s)	CBCL, YSR, MASC, CDI, EATQ-R, SNAP-IV	60	4 Z	A 0.53 (0.022	0.81113
20	03	15.5 (0.03)	48.6	AU	Hierarch	Parent(s)	C-DISC IV	5	5 5	0.25 (0.01)	0.71 10
1 554	4	7.7 (1.35)	49.8	SU	Bifactor	Mother	CBCL, CBQ	60	4 N	A 0.82 (0.016	1114
15	69	13 (NA)	51	SU	Bifactor	Self, Parent(s)	CAPS, CADS	4	5 N	A 0.69 (0.012	0.93 13
16	6,091	16 (NA)	50.7	GB	Bifactor	Parent(s)	Rutter Scales	11	5 N	A 0.68 (0.004	0.88 14
	528	16 (NA)	52.2	GB	Bifactor	Parent(s)	Rutter Scales	11	5 N	A 0.69 (0.012	0.88 14
L	31	2 (NA)	49	SU	Bifactor	Parent(s)	CBCL	24	22 N	A 0.58 (0.018	1 15
	31	3 (NA)	49	SU	Bifactor	Parent(s)	CBCL	24	22 N	A 0.66 (0.018	1 15
(-	731	4 (NA)	49	SU	Bifactor	Parent(s)	CBCL	24	22 N	A 0.67 (0.017	1 15
	731	7.5 (NA)	49	SU	Bifactor	Parent(s)	CBCL	38	24 N	A 0.6 (0.018)	1 15

Table 1 (conti	nued)												
Study	Data Source	Sample Size (N)	Mean Age (SD)	% Female	e Country	Model Type	Informant	Measure	# EXT	# INI #	I QL #	ECV (Sampling Variance)	Study Quality (MeanlSum)
Tein et al., 2023	Early Steps	731	8.5 (NA)	49	NS	Bifactor	Parent(s)	CBCL	38	24	NA (0.65 (0.018)	1 15
Tein et al., 2023	Early Steps	731	9.5 (NA)	49	SU	Bifactor	Parent(s)	CBCL	38	24	NA (0.62 (0.018)	1115
Tein et al., 2023	Early Steps	731	10.5 (NA)	49	SU	Bifactor	Parent(s)	CBCL	38	24	NA (0.63 (0.018)	1115
Vine et al., 2020	Ind. Sample	162	12.03 (0.92)	47	SU	Bifactor	Self, Parent(s)	K-SADS	S	5	NA (0.29 (0.036)	0.79 11
Wade et al., 2019	BEIP	220	8 (NA)	50	RO	Bifactor	Parent(s)	MacArthur HBQ	5	33	NA (0.61 (0.033)	0.75 9
Wade et al., 2019	BEIP	220	12 (NA)	50	RO	Bifactor	Parent(s)	MacArthur HBQ	5	e G	NA (0.52 (0.034)	0.75 9
Wade et al., 2019	BEIP	220	16 (NA)	50	RO	Bifactor	Parent(s)	MacArthur HBQ	5	e G	NA (0.68 (0.032)	0.75 9
Wade et al., 2021	iKFP	501	3.15 (0.27)	49.3	CA	Bifactor	Mother, Father	BITSEA, OCHS	9	4	NA (0.56 (0.022)	0.94 15
Waldman et al., 2016	STT	1568	11.7 (3.3)	51	SU	Bifactor	Parent(s)	CAPS	4	[NA (0.56 (0.013)	0.86 12
NA Data are n Project, CAP C ronment, Moo Consortium St Supplement, F and Adolescen vidual Lives S Measures. AD ADHD and ot Inventory, BIT CAPE Commu dren's Depress Conners' Parer and Well-Bein	at available in-te Study, General d Study, General d Study, General udy, IMPACT II GS Pittsburgh C t Survey of Met urvey, TTS Tenr (SC Anxiety Dis ner Comorbiditis SEA Brief Infan nity Assessment ion Inventory, C tt Rating Scale- tt Rating Scale-	xt. Data Source and Preventue, ttion R Generatio MPACT Trial, L Jirls Study, PVC tal Health and V tessee Twin Stud sorders Interview s, AUDIT Alcoh te-Toddler Social t of Psychic Exp t of Psychic Exp . <i>DISC IV</i> Diagn . <i>Revised, CSBQ</i>	ABCD Adolesc CATSS Child an R Study, HRC SAC Longitudin SAC Longitudin Philadelphia N Vellbeing (Youn ly, WTP Wiscom ly, WTP Wiscom ly, WTP Wiscom ly, UTP W	cent Brain (High Risk High Risk all Study o eurodevelo g Minds M Sin Twin P Sin Twin P Sin Twin A Sin Twin A Sin A A Sessemer Child and A Schedule fou il Behavior rament Qu	Cognitive Deve ent Twin Study (Cohort Study f Australian Cr papenental Cohort latter), START 6 roject, 1958 MC PSS Adolescen tion Test, BAS(nt, BPI Berkele Adolescent Psy r Children, vers ouestionnaire, estionnaire Rev	Iopment, ALSPAC in Sweden, Co-W for Psychiatric Di inidren, Me and M t, SECCYD NICH Systemic Therapy Systemic Therapy DS 1958 Nationa it Psychotic-Like { C-2 Behavior Asse y Puppet Interviey copathology Scal sion IV, CIDI Com CSI-4 Child Sym ised, EBS Emotio ised, EBS Emotio	"Avon Longitudi muture Co-Ventur sorders, <i>iKFP</i> in by <i>School</i> Me and ID Study of Earl for At Risk Teer for	anal Study of Pa e Project, <i>Earl</i> y tensive sample d My School Si y Child Care at s Trial, <i>TTC</i> To nent Study, <i>197</i> et, <i>ASI-4</i> The <i>i</i> for Children, 2n ptom Inventor tehavior Checkl anal Diagnostic at Screener, <i>G</i>	 Steps and Steps Ead Steps Ead of the Kic udy, NCS udy, NCS udy, NCS udy, NCS udy, NCS ad Youth Adolescen d Ed., BC y, CADS ist, CBQ interview, OASSESSS 	Children, chysteps J s, Familie 'A Nation Developm Cohort, 7 70 British t Symptoi t Symptoi evised C Children's Conner's Teacher J Compute	, <i>BEIP</i> 1, Multisit is Place and Comment <i>TRAILS</i> is Cohorn in Loohorn in Loohorn in Liver al Behav is 3 Conr Report 1 Report 1 rrized, s	Bucharest Earl Bucharest Earl as study, <i>IMAG</i> se study, <i>IMAG</i> orbidity Surv orbidity Surv Add Ad <i>ITracking</i> Add <i>ITracking</i>	y Intervention A Gene, Envi- EN IMAGEN sy Adolescent Istralian Child lescents' Indi- lescents' Indi- ression Scale, ression Scale, rire, CDI Chil- cales, CPRS-R Development view, K-SADS

Country: Country: country codes were the two-letter ISO 3166 Alpha-2 codes from the International Organization of Standardization (ISO) found at https://www.iso.org/obp/ui/#home. Studies derived and Pelham scale, 4th ed., SRS Social Responsiveness Scale, TRF Teacher's Rating Form, Y-OQ Youth Outcome Questionnaire, YSR Youth Self Report from multiple samples were comma separated

Kiddie Schedule for Affective Disorders and Schizophrenia, LOI Short Leyton Obsessional Inventory, MacArthur HBQ The MacArthur Health and Behavior Questionnaire, MASC Manifest Anxiety Scale for Children, M&MS Me and My School questionnaire, MFQ Mood and Feelings Questionnaire, OCHS Ontario Child Health Study Scales, PAPA Preschool Age Psychiatric Assessment, PPSC Preschool Pediatric Symptom Checklist, RAPI Rutgers Alcohol Problem Index, RCMAS Revised Children's Manifest Anxiety Scale, SCAS The Spence Children's Anxiety Scale, SCAS Description of the second sec Screen for Child Anxiety Related Disorders, SCDC Social and Communication Disorders Checklist, SCID Structured Clinical Interview for DSM Disorders, SDQ Strengths and Difficulties Questionnaire, SENA Assessment System for Children and Adolescents, SIPS Structured Interview for Prodromal Syndromes, SMFQ Short Mood and Feeling Questionnaire, SNAP-IV Swanson, Nolan, half of the heterogeneity is attributable to the included nested components of data source, wave, and study. By proxy, just under half (45.17%) of the heterogeneity is due to sampling variance. The 54.83% of heterogeneity due to nested components is broken down into 38.11% of heterogeneity accounted for by data source, with the remaining 16.72% accounted for by the study. Study wave did not account for any additional heterogeneity.



Fig. 4 Forest plot. Some lower- and upper-bound confidence intervals fell outside of 0–1 range and were cut off due to being improbable values for a proportion (shown with arrows). Every effect size is in the figure, including multiple effect sizes from a given study

Age-Moderation Analyses

We hypothesized that general factor strength would decrease across development. Study mean ages ranged from 2 to 17 years of age, M_{age} (*SD*) = 10.95 (3.72). Figure 5 describes the distribution of effect sizes by sample mean age included in the review. Screened studies that had 17- or 18-year-old participants tended to be included in adult samples with a mean sample size of > 18.00; thus, there were no studies that included a mean sample size of ≥ 18.0 years of age. Agemoderation analysis results yielded a Test of Moderator (*QM*) statistic, *QM*(1) of 0.63, *p*=0.43. The slope of ECV as a function of age was β = - 0.003, SE=0.004, *p*=0.43. Findings suggest that general psychopathology strength did not change significantly across childhood and adolescence.

Mean-Based Estimate of General Factor Strength

We also examined general factor strength using the mean (rather than sum) of squared factor loadings, to reduce the impact of the number of indicators on estimates of general factor strength (see Supplemental Appendices 3 and 4). Using the mean of squared standardized factor loadings, the estimate of general factor strength was somewhat smaller (0.34, SE = 0.02, p < 0.001). Age-moderation analysis results yielded a Test of Moderator (*QM*) statistic, *QM*(1) of 0.96, p = 0.33. The slope of the moderation was - 0.004, SE = 0.004, p = 0.33. These results increase confidence in the finding that, even when accounting for potential ECV inflation due to indicator count, general psychopathology strength does not change significantly across childhood and adolescence.

Publication Bias

The results of the Egger's test of the present meta-analyses indicated an intercept of $\beta_0 = 0.63$, SE = 0.0795%CI = 0.50–0.76. *t*-test results were t(108) = 9.40, p < 0.001. That is, the intercept of the sampling variance as a moderator was significantly different from 0, indicating the possibility of publication bias due to fewer studies with small sample sizes being published compared to larger samples.

A contour-enhanced funnel plot is in Fig. 6. Results show that there is likely a bias toward publishing results that indicated larger explained common variance estimates, albeit only a small bias. There was significant variability in explained common variance at different sample sizes, and there was some indication that studies with smaller sample sizes (i.e. larger standard errors) tended to have smaller explained common variance values, depicted on the left side of the plot.

Results from the trim-and-fill analysis indicate that 8 effect sizes, depicted as white dots in Fig. 7, would need to be imputed to render the current findings symmetrical. All imputed points were placed to the right of the plot, resulting in an increased estimate where the general factor accounts for ~ 60% of total reliable variance; ECV = 0.60, SE = 0.01, 95% CI = 0.57–0.63. Because these analyses do not account for nesting, they should be interpreted carefully. The results likely indicate that the meta-analysis result of an explained common variance of 0.56 is likely a slight underestimate, falling just short of the confidence interval of the trim-and-fill estimate (0.57–0.63). These values are close enough to one another to suggest that publication bias may exist but does not greatly affect the findings of the present study.

Taken together, the trim-and-fill and funnel plot provided some evidence that there may be a slight bias such that studies with larger samples published results that indicated larger explained common variance estimates. However, only 8 studies were needed on the right side of the plot to balance the symmetry. Furthermore, the values implied by the trim-and-fill plot were close to 0.85, which would indicate unidimensionality of the general factor (Forbes et al., 2021a; Reise & Revicki, 2014; Stucky & Edelen, 2014). Given the low number of studies that have found unidimensionality in the general factor, and that the publication bias analyses did not account for nonindependent effect sizes, the trim-and-fill and funnel plot results should be interpreted with caution.



Fig. 5 Distribution of effect sizes across included ages. The present sample of papers did not have any studies with a mean age of 1 or 18 years of age

Fig. 6 Contour-enhanced funnel plot with standard error and sampling variance predictors. Plot generated using non-nested data. Areas with light gray show effects between p = 0.05 - 0.1, and dark gray showing p = 0.01 -0.05. Solid line represents standard error as a predictor in the association between standard error and explained common variance. Dashed line represents sampling variance as a predictor in the association between standard error and explained common variance



Explained Common Variance



Explained Common Variance

Fig. 7 Trim and fill plot. Plot generated using non-nested data. Areas with light gray show effects between p = 0.05-0.1, and dark gray showing p = 0.01-0.05. White dots are imputed effect sizes. A total of 8, *SE* = 6.61 imputed values would need to render the current findings symmetrical

Study Quality

We assessed the study quality of included studies using the modified Downs and Black (1998) checklist. Mean study quality had a mean score of 0.88, SD = 0.11. Scores ranged from 0.62 to 1.00 and had a median score of 0.93. See Supplemental Appendix 2 for more information.

A total of k = 65 effect sizes were derived from high quality studies, i.e. greater than or equal to a mean study quality score of 88%. Results indicated that the general factor strength accounted for 59% of the reliable variance. The results yielded a QM(1) = 1.29, p = 0.256. The slope of moderation was $\beta - 0.007$, SE = 0.006, p = 0.256. A total of k = 43 effect sizes were derived from lower quality studies, i.e. less than a mean quality score of 87%. Results indicated that the general factor strength accounted for 52% of the reliable variance. The results yielded a QM(1) = 0.30, p = 0.586. The slope of the moderation was: $\beta = 0.004$, SE = 0.007, p = 0.586. Taken together, results indicate that higher-quality studies had stronger general factor strength, but the association between general factor strength and age was not moderated by study quality.

Nonlinear Trajectory of Age

We examined potential nonlinearity in the ECV estimates as a function of age. Neither the quadratic [QM(2) of 0.85, p = 0.65] nor the cubic [QM(3) = 1.42, p = 0.70] terms showed evidence of moderation. Figure 8 depicts a bubble plot of model-implied estimates of ECV as a function of sample mean age, with the size of bubbles corresponding to the sample size.

Moderation and Subgroup Analysis Results

All moderation and subgroup analyses' results, regardless of statistical significance, are reported in Supplemental Appendix 4. Below we highlight results that yielded at least trend-level statistical significance ($p \le 0.10$).

Developmental Period Subgroups

When analyses separately analyzed whether ECV changes as a function of developmental period, school age (mean age of $\geq 6.00 \& < 13.00$ years; k = 54), and adolescents (mean age of > 13.00 years; k = 45) derived a general factor strength of 0.57 and 0.56, respectively, and ECV did not significantly change as a function of age.



Fig.8 Bubble plot of effect sizes as a function of age overlaid with model-implied nonlinear ECV curve. Each bubble represents a single effect size. The bubble size corresponds to the sample size (larger bubbles representing larger samples). 44 uniquely colored data

sources are represented numerically. The nonlinear trendline of ECV as a function of age was calculated using the model-implied quadratic trajectory of ECV

For preschool age (mean age of < 6.00 years; k=11), general factor strength was 0.66 and results indicated a significant increase across age within the age range of 2–5.9 years; M(SD)=3.52 (1.16). The slope of moderation was $\beta=0.065$, SE=0.032, p=0.043.

To expand upon these results, we examined whether the mean factor loadings for general, internalizing, and externalizing factors changed across preschool ages, age moderation indicated that the mean internalizing factor loadings decreased across this age range: QM(1) of 4.29, p = 0.038; slope_{int}: $\beta = -0.057$, SE = 0.027, p = 0.038. Neither the mean externalizing factor loadings nor mean general factor loadings significantly changed across this age range.

Percent Uncontaminated Correlations (PUC)

When PUC was included as a moderator, greater PUC values were associated with higher ECV at a trend level (β =0.304, SE=0.160, *p*=0.057). However, ECV was not associated with age when including PUC as a moderator.

General Factor Indicator Count

The number of indicators loading onto the general factor ranged from 5 to 116, $M_{\text{indicator}}$ (*SD*) = 27.81 (28.31). A greater number of indicators on the general factor was associated with greater ECV (β =0.001, *SE*=0.001, *p*=0.020). ECV was not associated with age when including general factor indicator count as a moderator.

Factor Count

Among the separate analyses on factor count—total, externalizing, internalizing, and thought disorder—only the count of externalizing factors and the presence of a thought-disorder factor emerged as at least trend-level significant moderators in general factor strength. Having more than one externalizing factor, or sub-factors, was associated with greater ECV at a trend level (β =0.050, *SE*=0.029, *p*=0.084). The presence of a thought-disorder factor was associated with weaker ECV (β =-0.137, *SE*=0.041, *p*=0.001). ECV was not associated with age when controlling for factor count.

Measure Type Moderation

Among the included studies, 89 effect sizes were derived using results from questionnaires, while 38 were derived using results from structured clinical interview. Several factor analytic models included both questionnaires and interviews. Results from moderation analysis of whether a questionnaire or interview was used indicated that a moderation was present, QM(2) of 6.09, p = 0.05; slope_{questionnaire}: $\beta = 0.09$, SE = 0.04, p = 0.02. Studies with questionnaire ratings tended to yield a stronger ECV estimate than studies with interviews. ECV was not associated with age when controlling for measure type.

Discussion

In the present meta-analysis, we aimed to determine whether general psychopathology strength changes across childhood to adolescence. The present meta-analysis expands the scope of what is understood about general psychopathology from evidence of aggregated studies to include children as young as 2 and includes information from almost every developmental period in childhood and adolescence (except 1 and 18 years). Included studies (k=65) examined internalizing and externalizing psychopathology factors (or sub-factors) at a minimum, and included a thought-disorder factor at a maximum (Caspi et al., 2014). Standardized factor loadings were used to estimate explained common variance that represents the general factor's strength in relation to the total reliable psychopathology variance, while accounting for interdependencies due to shared data source, waves, and studies. When meta-analyzed, these results showed that general psychopathology accounted for approximately 56% of the reliable variance in ratings of child and adolescent psychopathology across the included studies, and this factor strength did not significantly change across development.

General Factor Strength

While there are no cutoffs for explained common variance values (McElroy et al., 2018a), the suggested range to denote that the general factor is the main source of shared variance ranges from 0.6 or 0.7, which would indicate high importance of general factor relative to specific factors (Forbes et al., 2021a; Reise et al., 2013; Stucky & Edelen, 2014), to 0.85, which would indicate unidimensionality (Forbes et al., 2021a; Reise & Revicki, 2014; Stucky & Edelen, 2014). At a value of 0.56, our findings suggest that the general factor is not the main source of the shared variance, but does account for a nontrivial amount of variance of psychopathology symptoms across childhood and adolescence (Rodriguez et al., 2016). One interpretation for this finding is that general psychopathology (i.e. covariation of internalizing, externalizing, and thought disorder) meaningfully represents a considerable proportion of the total symptoms as reported by parents, teachers, secondary caregivers, and self-report across childhood and adolescence. These results support prior literature that has noted that symptom- and syndromespecific diagnoses and treatments do not adequately cover the entirety of psychopathology, and that a general psychopathology representation would account for the considerable overlap in symptoms (Conway et al., 2019; Forbes et al.,

2019; Hopwood et al., 2020; Kotov et al., 2017; Ruggero et al., 2019).

Role of Development in General Factor Strength

Inconsistent with hypotheses, general psychopathology strength did not differ as a function of sample mean age. When allowed to fit a nonlinear trajectory, the modelimplied change in general factor strength was near-identical to the linear trajectory, suggesting that the general factor did not fluctuate in its strength at specific ages, and it did not significantly increase or decrease from early childhood to late adolescence when assessed meta-analytically. It is possible that prior studies that found random fluctuations in general factor strength may have captured sampling and measurement error (Watts et al., 2021a). In the present meta-analysis, general factor strength showed modest but non-significant decreases across school age and adolescent ages. By contrast, ECV significantly increased in the preschool developmental period encompassing ages 2-6 years. However, additional analyses indicated that these changes in ECV were driven by a decrease across this age range in the average factor loadings on the specific internalizing factor; mean general factor loadings did not change. One potential hypothesis for why psychopathology during the preschool ages may differ from other developmental periods is that many preschool-age children enter daycare or preschool settings and begin to spend more time away from parents. Preschool ages also coincide with a developmental task of more autonomy-seeking behaviors (Cicchetti & Rogosch, 2002; Lahey et al., 2021; Sroufe, 2009, 2016). The number of effect sizes in the preschool-age range was small at only k = 11; therefore, it is important for future studies to estimate factor analytic models of general psychopathology in preschool-age children to replicate these findings longitudinally.

The results suggest that—even with a potentially slight increase in preschool age—general psychopathology is as meaningful in young children as it is in adolescents nearing adulthood. These findings align with previous longitudinal studies (e.g. Choate et al., 2022; McElroy et al., 2018a) that found that the general factor strength did not change with age. The findings in this meta-analysis cannot address changes in the general factor within individuals over development. Therefore, we are unable to make any claims about whether there are developmental changes in the level, structure, or strength of the general factor for an individual. Future longitudinal studies will be needed to examine these questions. However, our analysis of only longitudinal studies did not show changes in general factor strength across development.

When conducting a subgroup analysis on longitudinal studies that established at least partial metric invariance, only 4 of the 12 longitudinal studies met this criterion (e.g. Choate et al., 2022; Etkin et al., 2021; Snyder et al., 2017b; Wade et al., 2019). The remaining 8 (e.g. Chen et al., 2022; Huang et al., 2020; McElroy et al., 2018a; Neumann et al., 2020; Olino et al., 2018; Riglin et al., 2019; Rijlaarsdam et al., 2021b; Tein et al., 2023) either did not evaluate or attempted and determined that factor loadings of measures of general psychopathology were non-invariant across age. General factor strength also did not show changes across ages among studies that established longitudinal measurement invariance. In addition, within studies that established longitudinal measurement invariance, there were very little fluctuations in ECV across the ages. Given the scarcity of studies that evaluated, let alone established, measurement invariance, the ability to detect an effect of age among the longitudinal studies in the present meta-analysis is limited. Future studies should evaluate longitudinal measurement invariance to estimate changes in general factor strength from longitudinal designs.

The present findings suggest that the general factor likely represents a variable that is both transdiagnostic and present at all stages of childhood and adolescent development, but we are unable to generalize these findings to the withinperson level.

Potential Interpretations of General Factor

The finding that the strength of the general factor did not vary across ages has implications for how researchers and clinicians may conceptualize and interpret the meaning of the general factor. Stability in this factor's strength suggests that the general factor is not differentially strong at specific ages but potentially: overall impairment (Smith et al., 2020); a risk factor for developing symptoms (Ringwald et al., 2021); or a dimensional alternative to categorical diagnoses conceptualization (Forbes et al., 2019; Kotov et al., 2017; Ringwald et al., 2021). Simply put, the evidence is consistent with the idea that the general factor represents something that influences the presentation of symptoms relatively evenly across development. Therefore, an interpretation that the general p factor represents overall impairment would suggest that experiencing general and transdiagnostic difficulties equally affects children and adolescents, even though specific symptoms—and their frequency or severity—may change at different ages, i.e. heterotypic continuity (Smith et al., 2020). However, the general factor may instead be a statistical artifact, and therefore, would not influence presentation of symptoms (Watts et al., 2021a, 2021b).

There are many potential candidates for how we might interpret the general factor of psychopathology given the findings from the present meta-analysis. In a recent systematic review, Lynch et al. (2021) found that general psychopathology in young people aged 10–24 years of age was associated with a number of risk factors. Among biological risk processes for general psychopathology, they identified genetic risk for ADHD and schizophrenia (Brikell et al., 2020; Riglin et al., 2020); being male (Riglin et al., 2020; Wade et al., 2018; Wang et al., 2020). They also identified early pubertal timing (Hamlat et al., 2019); and executive functioning deficits (Hatoum et al., 2018; Shields et al., 2019; Wade et al., 2019). Among psychological risk processes for general psychopathology, they identified high negative affectivity (Deutz et al., 2020; Hankin et al., 2017; Mann et al., 2020; Wang et al., 2020); difficult temperament (Deutz et al., 2020; Levin-Aspenson et al., 2019); and low effortful control (Hankin et al., 2017; Shields et al., 2019). Among social risk processes for general psychopathology, they identified stressful life events (Hamlat et al., 2019); and maternal depression (Deutz et al., 2020; McCutcheon et al., 2013).

Although the Lynch et al. (2021) review did not include children under 10 years old, these findings highlight the diversity of potential candidates that influence the general factor of psychopathology. We contribute to this literature by describing developmentally informed conceptualizations of the general factor.

Risk of Developing Symptoms of Co-occurring Psychopathology

Plausibly, the findings in the meta-analysis could support the hypothesis that the general factor might represent the likelihood of experiencing co-occurring symptoms of psychopathology remains stable across development provided that an individual shows psychopathology symptoms (Ringwald et al., 2021). Although genetic and environmental risks for developing psychopathology would be higher in some individuals (e.g. Brikell et al., 2020; Chen et al., 2022; Grotzinger et al., 2019; Neumann et al., 2016), the present findings might indicate that, on a population level, the risk of developing symptoms of co-occurring psychopathology would be approximately equally likely across childhood and adolescence, rather than at certain age ranges. However, findings from analyses examining developmental periods suggest that there might be a slightly higher risk at younger ages.

Temperamental Negative Emotionality

Another possibility, given that the general factor is proposed to represent what is common among symptoms of psychopathology, is that the general factor and its relatively equal contributions across development might represent temperamental affective behavior. Temperamental affective behavior is easily observed by an informant and found across the lifespan. One aspect of temperamental affective behavior that is present across development, and thus, a potential interpretation for general psychopathology, is dysregulated emotionality, also called difficulty. Dysregulated emotionality changes in its behavioral manifestations throughout development and has been implicated as a transdiagnostic mechanism of psychopathology (Damme et al., 2020; Weissman et al., 2019). Particularly at younger ages, dysregulated emotionality has been labeled irritability, a dispositional tendency to respond with anger when faced with slowed or blocked goal attainment (Damme et al., 2020; Wakschlag et al., 2018; Wiggins et al., 2018, 2021). Furthermore, irritability is present throughout development, even through adolescence (Hawes et al., 2020). Negative affect/irritability present at younger ages predicts future psychopathology, even when accounting for the introduction to novel contexts and challenges, e.g. child going to school, seeking more autonomy (Briggs-Gowan et al., 2006). Due to its presence throughout development during the ages assessed in the present meta-analysis, 2–17, irritability might be a candidate interpretation for the general factor.

Measurement and/or Informant Effects

Other considerations might include factors exterior to the symptoms themselves, such as method- or informant-related effects. The current review found a few instances where there was meaningful difference in ECV as a function of differences in method or analytic choices. The general factor was stronger in models that included questionnaires compared to models that included interviews. Including an interview weakens the general factor strength across development. An interview is typically administered and interpreted by a trained professional who may be more likely to assess psychopathology objectively. Nevertheless, Constantinou (2019) found that separating questionnaires from interviews and evaluating them in two separate models resulted in a non-significant ECV variability difference.

The weaker general factor strength from interviews may indicate that the use of interviews reduces reporter bias, compared to questionnaires completed by parents, teachers, or self-report. Alternatively, the difference between factor strength from interviews and questionnaires might be due to questionnaires containing more items that load onto the general factor, which would potentially inflate ECV estimations. Alternatively, due to more and diverse questionnaires compared to fewer interviews present in the study, the finding that questionnaires yielded a strong ECV might reflect a more reliable estimate of the general factor (Constantinou, 2019).

Another important consideration is informant-related biases, such as method biases specific to an informant type (e.g. child, parent, teacher). It is plausible that the general factor, what is common among ratings of psychopathology, might represent reporter bias to a degree (Constantinou, 2019; Martel et al., 2017). When Watts et al. (2021a) compared ECV estimates from mono-informant models to ECV estimates from models that included multiple informants and accounted for method factors of informant type, ECV estimates decreased on average from 0.68 (range: 0.53–0.80) to 0.37 (range: 0.20-0.46). The proportion of variance in ECV estimates that were attributable to method factors ranged from 0.29 to 0.67 (M = 0.46). Thus, around half of the variance in ECV estimates may be attributable to method variance. ECV estimates from models that do not control for variance attributable to informant may overestimate the true strength of the general factor. Applying this adjustment to account for method factors to the present study, one might expect that the true strength of the general factor in the present meta-analysis may be closer to $0.30 ([1 - 0.46] \times 0.56)$. However, our moderation analyses on informants yielded non-significant results. One potential explanation is that while instances of self and parent report were evenly split, there were only four instances of teacher report, which limited the variability in objective measurement across settings. Prior research has asserted that it is important to assess behavior problems from multiple raters in a given setting (e.g. mothers and fathers in the home), and across different settings (e.g. teachers in schools) to capture context-specific variability in reporting to reduce bias (De Los Reves & Makol, 2021; Kraemer et al., 2003; Makol et al., 2020). It is important for future studies to examine measurementrelated biases to determine the extent to which the general factor represents something other than the covariation of psychopathology.

Measuring General Psychopathology

A previous review had noted that measures differ in their validity and reliability to detect general psychopathology (Constantinou, 2019). Along with explained common variance to detect the general factor strength, another method that is often paired with ECV is percent uncontaminated correlations (PUC). Higher PUC values reflect a higher number of subscales with fewer items in each subscale, making them more suitable for estimating the general factor (Constantinou, 2019). Constantinou (2019) examined the interaction between age and PUC in predicting explained common variance values. As noted above, their results suggest that PUC might strongly influence interpretation of strength of the general factor over development (Constantinou, 2019).

In the present meta-analysis, we found that among bifactor models, PUC was associated with higher levels of ECV, which replicated the findings from Constantinou (2019). However, when assessed meta-analytically, including PUC as a moderator did not result in a change in ECV across development. Furthermore, when examining the role of PUC as a moderator of the mean-based estimate of general factor strength, PUC was negatively associated with this composite at a trend level, and there was no significant change in the mean-based estimate of general factor strength across development. One potential culprit that is responsible for these seemingly contradictory results might be due to differences in influence from the number of indicators loading onto the general factor. It is important to consider that PUC is calculated using the number of indicators and number of factors. In fact, the number of indicators loading onto the general factor was positively associated with higher levels of ECV, but not with the mean-based estimate of general factor strength. Therefore, deciding to calculate ECV using the sum, or the average of squared standardized factor loadings may influence the degree to which the number of indicators on the general factor indicator impacts the strength of the general factor. The number of specific factors, specifically of externalizing problems, and whether to include a thoughtdisorder factor will also be important decision points for future research.

Clinical Implications

The stability of the general factor's strength in ratings of psychopathology across childhood and adolescence suggests that what is common among symptoms of psychopathology, p factor, may be detectable from a young age. If the general factor represents a general liability, risk, or negative emotionality present at all stages of life course, then there is a need for early detection of general psychopathology in childhood and the use of domain-general treatment approaches. Examples of domain-general approaches might include teaching emotion regulation skills and parenting training, which might potentially target and forestall development of end-stage specific psychopathology symptoms (Forbes et al., 2019; Martel et al., 2017).

The findings from the present meta-analysis suggest that emphasizing single symptoms or syndromes will not be sufficient to conceptualize the entirety of presenting concerns for youths seeking psychological treatment (Forbes et al., 2019). Evidence of a robust general factor of psychopathology across development has the potential to motivate the shift away from single-disorder treatment protocols toward transdiagnostic approaches that also better account for heterotypic continuity of problem behaviors, such as the Unified Protocol for Transdiagnostic Treatment for Emotional Disorders for Children and Adolescents (Ehrenreich-May et al., 2017). That is not to say that focusing on lower-order or specific symptoms or their treatment protocols should not also be a primary concern; in reality, we, along with others, argue that it is increasingly important to study homogeneous specific psychopathology concerns (McGrath, 2005; Smith et al., 2003, 2020; Strauss & Smith, 2009).

Higher general factor scores are associated with more functional impairment and an increased risk for suicidal behavior and non-suicidal self-injurious behavior (Haltigan et al., 2018; Hoertel et al., 2015; Lahey et al., 2015, 2021; Pettersson et al., 2018; Sallis et al., 2019). We urge clinicians and researchers, regardless of the presenting concern of the child, to assess for broad ranges of psychopathology (e.g. internalizing, externalizing, thought-disordered, and other dimensions) in all clients or research participants. This perspective is in line with suggestions made from supporters of the HiTOP model (Conway et al., 2019; DeYoung et al., 2021; Forbes et al., 2019; Hopwood et al., 2020; Ruggero et al., 2019). We feel that assessing general psychopathology will better capture the full range of covarying symptoms to account for overlaps in symptoms often dismissed as a specific syndrome. Future work needs to develop measures that better account for heterotypic continuity to assess a wide scope of covarying symptoms that suitably estimate a general factor (Harris et al., 2023; Petersen & LeBeau, 2022). Ideally, measures might take the form of a computer adaptive test (CAT) such as the Overall Mental Illness (OMI) screener (Moore et al., 2019) that provide more rapid and accurate assessments to be used in research and clinical settings to develop more transdiagnostic treatment approaches.

One approach for treatment of general psychopathology is a transdiagnostic stepped-care approach to prevention is proposed by Forbes et al. (2019). This approach provides a framework for more universal interventions for broad and limited-modifiable risk factors (e.g. harsh parenting, emotional reactivity) at ages 3-6, and increases slightly in specificity at ages 7-10 to incorporate more targeted treatment, with a focus on emergent symptoms in adolescence and through adulthood (Forbes et al., 2019). Evidence from the present meta-analysis shows that general psychopathology strength is stable across ages at the population level and provides support for the need for prevention and early intervention of dimensional psychopathology problems. However, the present review did not account for functional impairment that would be relevant to consider in clinical treatment (Ruggero et al., 2019).

Limitations

Several limitations should be noted when interpreting the results of the present meta-analysis. First, we opted to include only those studies that assessed, at a maximum, internalizing, externalizing, and thought-disordered specific factors (or multiple sub-factors representing these specific factors). Studies that included additional factors (e.g. stress, personality, well-being, prosocial behavior) were excluded (e.g. Black et al., 2019), because the conceptual interpretation of general psychopathology would differ greatly as a function of the specific factors

from which it was composed. Additional factors, such as maladaptive personality traits, are important to consider because these traits informed the HiTOP structure (Kotov et al., 2017). Our defining characteristics of general psychopathology most closely align with the extant literature in children and adults (e.g. Caspi et al., 2014; Haywood et al., 2021; Lahey et al., 2017; Ringwald et al., 2021). Future studies should also include personality factors in the conceptualization of general psychopathology.

A second limitation is the method by which effect sizes were calculated. There are no single universal methods of factor analytic modeling. We calculated our effect sizes by using the information we had available, the standardized factor loadings to estimate the reliable variance. We ultimately chose the present method because it was a robust approach to estimating explained common variance, a metric of factor strength. The method chosen to calculate the effect size, explained common variance, inflates estimates of factor strength for factors that have more loadings (i.e. the general factor; Reise et al., 2013). However, averaging squared factor loadings, rather than summing them resulted in a reduced general factor strength of 34% in the present review, and these findings also did not significantly change across development. Published studies likely were biased to include only the best fitting version of models, which may have had many indicators. In addition, variability in factor loadings across studies challenges the comparability of the latent factor itself. Therefore, estimation of ECV may be an overestimate and poses questions about possible interpretation of the general factor. However, because the averaging squared factor loadings also resulted in no change across development, we have further confidence that general factor strength does not change across development. We did not estimate "unreliable" error/residual variance of standardized factor loadings because the correlated nature of higher-order and modified bifactor models pose challenges to interpreting residual variance. Future studies should examine whether residual variance in the indicators or factors affects ECV interpretation, specifically for traditional bifactor models.

A third limitation is that there are potential limits to comparing ECV across models due to significant heterogeneity in methods of model estimation, including measures, indicator count, factor count, and type/number of informants across studies. Several sensitivity analyses elucidated that to some degree all of these were associated with differences in ECV. A higher percent of uncontaminated correlations (PUC) was also positively associated with higher ECV. However, in this analysis, ECV did not substantially change across age. In sum, it is likely that differences in what goes into a model slightly change the meaning of ECV. The findings did not substantively change when examining only longitudinal studies that established longitudinal measurement invariance. Findings from this subset of studies provide greater confidence that the strength of the general factor does not substantively change over the developmental span.

Fourth, we did not include studies with the mean age of over 18 years of age. This constraint limited the scope of the meta-analysis. Because we found that general psychopathology strength may change in preschool ages, it is possible that general factor strength might also change from adolescence to adulthood. Despite limitations, this was the first systematic review and meta-analysis to examine the change in general factor strength across childhood and adolescence and included robust multilevel meta-analysis methods that allowed estimating explained common variance at the population level, and its changes across development, while accounting for interdependence of data from shared data sources.

Future Directions and Reporting Guidelines

To better evaluate changes in general factor strength, longitudinal studies should test whether the findings from the present meta-analysis replicate after establishing longitudinal measurement invariance. Future studies should evaluate the degree to which questionnaires or interviews differ in their ability to detect general psychopathology. For ease of transparency for future meta-analyses, empirical studies should publish standardized factor loadings, variance-covariance matrices of included factor analysis indicators, and specific modeling methods in-text, in supplementary materials, or on an open-source pre-registration database (e.g. Open Science Framework; OSF). Future studies should also clearly state the source of the participant sample pool; note whether sample has been included in prior studies; and provide details about data collection, methods, and demographic information about the data source. Furthermore, future research should be dedicated to developing approaches to assessing publication bias using nested data. Future intervention research should target transdiagnostic symptoms, such as difficulty or negative emotionality, starting in young children with the aim of preventing onset of more pronounced specific psychopathology in later years (Damme et al., 2020; Weissman et al., 2019). Future studies should also examine how best to interpret the general factor across developmental periods and the extent to which it involves method biases.

Conclusion

In conclusion, our meta-analytic findings suggest that general psychopathology makes up over half of the total reliable variance in ratings of psychopathology in children and adolescents. The strength of the general factor did not change across childhood to adolescence, suggesting that the strength of higher-order dimensional psychopathology is stable across childhood to adolescence at the population level, with a possible modest increase during preschool age. Research on the strength and stability of the general factor of psychopathology across childhood and adolescence will continue to accumulate, but our meta-analysis shows that the general psychopathology factor is meaningful and represents something that is robustly prominent at all stages of childhood and adolescent development.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10567-023-00464-1.

Funding The authors and their research were supported by Grants HD098235 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), T32GM108540 from the National Institute of General Medical Sciences (NIGMS), and UL1TR002537 from the National Center for Advancing Translational Sciences (NCATS). We acknowledge Drs. Molly Nikolas and Grazyna Kochanska for their invaluable feedback on meta-analysis design. We thank Dr. Kristian Markon for his help with analysis interpretation. We also thank Kelly Hangauer for his assistance and guidance with developing the keywords and search terms. The research questions, hypotheses, search terms, and coding manual were pre-registered with the Open Science Framework (OSF) at https://osf.io/jkupe. The data (https://osf.io/k47cs) and a computational notebook (https://osf.io/5hzka) are also available in the OSF repository.

Declarations

Conflict of interest We have no conflicts of interest to disclose.

Ethical approval Not applicable.

References

References marked with an asterisk indicate studies included in the meta-analysis,

- Achenbach, T. M. (2020). Bottom-up and top-down paradigms for psychopathology: A half-century odyssey. *Annual Review of Clinical Psychology*, 16(1), 1–24. https://doi.org/10.1146/annurev-clinp sy-071119-115831
- *Afzali, M. H., Sunderland, M., Carragher, N., & Conrod, P. (2018). The structure of psychopathology in early adolescence: Study of a Canadian sample: La structure de la psychopathologie au début de l'adolescence: Étude d'un échantillon canadien. *Canadian Journal of Psychiatry*, 63(4), 223–230. https://doi.org/10.1177/ 070674371773703
- *Aitken, M., Haltigan, J. D., Szatmari, P., Dubicka, B., Fonagy, P., Kelvin, R., Midgley, N., Reynolds, S., Wilkinson, P. O., & Goodyer, I. M. (2020). Toward precision therapeutics: General and specific factors differentiate symptom change in depressed adolescents. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 61(9), 998–1008. https://doi.org/10.1111/jcpp.13194
- American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). https://doi.org/10. 1176/appi.books.9780890425787

- Assink, M., & Wibbelink, C. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12, 154.
- Avinun, R., Knafo-Noam, A., & Israel, S. (2021). The general psychopathology factor from early to middle childhood: Longitudinal genetic and risk analyses. *MedRxiv*. https://doi.org/10. 1101/2021.03.20.21253838
- 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). US, Boston: Springer.
- Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: Developmental issues and implications for DSM-V. *Psychiatric Clinics*, 32(3), 483–524.
- Black, L., Panayiotou, M., & Humphrey, N. (2019). The dimensionality and latent structure of mental health difficulties and wellbeing in early adolescence. *PLoS ONE*, *14*(2), e0213018. https://doi.org/ 10.1371/journal.pone.0213018
- *Blanco, C., Wall, M. M., He, J.-P., Krueger, R. F., Olfson, M., Jin, C. J., Burstein, M., & Merikangas, K. R. (2015). The space of common psychiatric disorders in adolescents: Comorbidity structure and individual latent liabilities. *Journal of the American Academy of Child & Adolescent Psychiatry*, 54(1), 45–52. https://doi.org/10.1016/j.jaac.2014.10.007
- *Bloemen, A. J. P., Oldehinkel, A. J., Laceulle, O. M., Ormel, J., Rommelse, N. N. J., & 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
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184–186. https://doi.org/ 10.1177/2167702616657069
- Boylan, K., Vaillancourt, T., Boyle, M., & Szatmari, P. (2007). Comorbidity of internalizing disorders in children with oppositional defiant disorder. *European Child & Adolescent Psychiatry*, 16(8), 484–494. https://doi.org/10.1007/s00787-007-0624-1
- *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/ 216770261985933
- 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
- Brikell, I., Larsson, H., Lu, Y., Pettersson, E., Chen, Q., Kuja-Halkola, R., Karlsson, R., Lahey, B. B., Lichtenstein, P., & Martin, J. (2020). The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. *Molecular Psychiatry*, 25(8), 1809–1821. https://doi.org/10.1038/ s41380-018-0109-2
- *Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., Ruparel, K., Wolf, D. H., Roalf, D. R., Mentch, F. D., Qiu, H., Chiavacci, R., Connolly, J. J., Sleiman, P. M. A., Gur, R. C., Hakonarson, H., & Gur, R. E. (2015). The Philadelphia Neurodevelopmental Cohort: Constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(12), 1356–1369. https://doi.org/10.1111/jcpp.12416
- Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction:*

🖄 Springer

Standards and recent advances (pp. 247–271). Germany: Hogrefe Publishing.

- Carragher, N., Krueger, R. F., Eaton, N. R., & Slade, T. (2015). Disorders without borders: Current and future directions in the meta-structure of mental disorders. *Social Psychiatry and Psychiatric Epidemiology*, 50(3), 339–350. https://doi.org/10.1007/s00127-014-1004-z
- *Carragher, N., Teesson, M., Sunderland, M., Newton, N. C., Krueger, R. F., Conrod, P. J., Barrett, E. L., Champion, K. E., Nair, N. K., & Slade, T. (2016). The structure of adolescent psychopathology: A symptom-level analysis. *Psychological Medicine*, 46(5), 981–994. https://doi.org/10.1017/S0033291715002470
- 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://doi.org/ 10.1177/2167702613497473
- *Castellanos-Ryan, N., Briere, F. N., O'Leary-Barrett, M., Banaschewski, T., Bokde, A., Bromberg, U., Biichel, 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., & The, I. C. (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology*, *125*(8), 1039–1052.
- *Chen, C., Lu, Y., Lundström, S., Larsson, H., Lichtenstein, P., & Pettersson, E. (2022). Associations between psychiatric polygenic risk scores and general and specific psychopathology symptoms in childhood and adolescence between and within dizygotic twin pairs. *Journal of Child Psychology and Psychiatry*, 63(12), 1513–1522. https://doi.org/10.1111/jcpp.13605
- *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*. https://doi.org/10.1017/S0954579422000463
- Cicchetti, D. (2020). Studying psychopathology in early life: Foundations of developmental psychopathology. In A. G. C. Wright & M. N. Hallquist (Eds.), *The cambridge handbook of research methods in clinical psychology* (pp. 193–204). Cambridge University Press; Cambridge Core.
- Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, 70(1), 6–20.
- Cicchetti, D., & Sroufe, L. A. (2000). The past as prologue to the future: The times, they've been a-changin'. *Development and Psychopathology*, 12(3), 255–264.
- *Clark, D. A., Hicks, B. M., Angstadt, M., Rutherford, S., Taxali, A., Hyde, L., Weigard, A., 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/10.1177/2167702620959317
- *Class, Q. A., Van Hulle, C. A., Rathouz, P. J., Applegate, B., Zald, D. H., & Lahey, B. B. (2019). Socioemotional dispositions of children and adolescents predict general and specific secondorder factors of psychopathology in early adulthood: A 12-year prospective study. *Journal of Abnormal Psychology*, *128*(6), 574–584. https://doi.org/10.1037/abn0000433
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*(1), 101–129.
- Constantinou, M. (2019). The bifactor model of psychopathology: Methodological issues and clinical applications (Doctoral dissertation, UCL (University College London)).
- *Constantinou, M. P., Goodyer, I. M., Eisler, I., Butler, S., Kraam, A., Scott, S., Pilling, S., Simes, E., Ellison, R., Allison, E., &

Fonagy, P. (2019). Changes in general and specific psychopathology factors over a psychosocial intervention. *Journal of the American Academy of Child and Adolescent Psychiatry*, 58(8), 776–786. https://doi.org/10.1016/j.jaac.2018.11.011

- Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., Mullins-Sweatt, S. N., Shackman, A. J., Skodol, A. E., South, S. C., Sunderland, M., Waszczuk, M. A., Zald, D. H., Afzali, M. H., Bornovalova, M. A., Carragher, N., Docherty, A. R., Jonas, K. G., Krueger, R. F., & Eaton, N. R. (2019). A hierarchical taxonomy of psychopathology can transform mental health research. *Perspectives on Psychological Science*, 14(3), 419–436.
- Cosgrove, V. E., Rhee, S. H., Gelhorn, H. L., Boeldt, D., Corley, R. C., Ehringer, M. A., Young, S. E., & Hewitt, J. K. (2011). Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. *Journal of Abnormal Child Psychology*, 39(1), 109–123. https://doi.org/10.1007/s10802-010-9444-8
- Costello, E. J., Copeland, W., & Angold, A. (2011). Trends in psychopathology across the adolescent years: What changes when children become adolescents, and when adolescents become adults? *Journal of Child Psychology and Psychiatry*, 52(10), 1015–1025. https://doi.org/10.1111/j.1469-7610.2011.02446.x
- Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research*, 24(3), 315–318. https://doi. org/10.1177/002224378702400308
- Damme, K. S. F., Wakschlag, L. S., Briggs-Gowan, M. J., Norton, E. S., & Mittal, V. A. (2020). Developmental patterning of irritability enhances prediction of psychopathology in pre-adolescence: Improving RDoC with developmental science. *BioRxiv*. https:// doi.org/10.1101/2020.04.30.070714
- De Los Reyes, A., & Makol, B. A. (2021). Interpreting convergences and divergences in multi-informant, multimethod assessment. In J. L. Mihura (Ed.),) *The Oxford Handbook of Personality and Psychopathology Assessment*. Oxford: Oxford University Press.
- *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
- DeYoung, C. G., Kotov, R., Krueger, R. F., Cicero, D. C., Conway, C. C., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Jonas, K. G., Latzman, R. D., Rodriguez-Seijas, C., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., Widiger, T. A., & Wright, A. G. C. (2021). Answering questions about the Hierarchical Taxonomy of Psychopathology (HiTOP): Analogies to whales and sharks miss the boat. *Clinical Psychological Science*. https://doi.org/10.1177/21677026211049390
- Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, 52(6), 377. https://doi.org/10.1136/jech.52.6.377
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plotbased method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. https://doi.org/10. 1111/j.0006-341x.2000.00455.x
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629. https://doi.org/10.1136/bmj.315. 7109.629
- Ehrenreich-May, J., Kennedy, S. M., Sherman, J. A., Bilek, E. L., Buzzella, B. A., Bennett, S. M., & Barlow, D. H. (2017). Unified protocols for transdiagnostic treatment of emotional disorders in children and adolescents: Therapist guide. Oxford: Oxford University Press.

- *Etkin, P., Ibáñez, M. I., Ortet, G., & Mezquita, L. (2021). Longitudinal associations between the five-factor model of personality and the bi-factor model of psychopathology: Continuity, pathoplasty and complication effects in adolescents. *Journal of Psychopathology and Behavioral Assessment*. https://doi.org/10.1007/ s10862-021-09903-1
- *Etkin, P., Mezquita, L., López-Fernández, F. J., Ortet, G., & Ibáñez, M. I. (2020). Five factor model of personality and structure of psychopathological symptoms in adolescents. *Personality and Individual Differences*. https://doi.org/10.1016/j.paid.2020. 110063
- *Faber, A., & Chaplin, W. (2020). The confusing problem of overlapping internalizing and externalizing dimensions: What should we do? [St. John's University (New York)]. ProQuest Dissertations & Theses Global.
- Forbes, M. K., Greene, A. L., Levin-Aspenson, H. F., Watts, A. L., Hallquist, M., Lahey, B. B., Markon, K. E., Patrick, C. J., Tackett, J. L., Waldman, I. D., Wright, A. G. C., Caspi, A., Ivanova, M., Kotov, R., Samuel, D. B., Eaton, N. R., & Krueger, R. F. (2021a). Three recommendations based on a comparison of the reliability and validity of the predominant models used in research on the empirical structure of psychopathology. *Journal of Abnormal Psychology*, *130*(3), 297–317. https://doi.org/10.1037/abn00 00533
- *Forbes, M. K., Magson, N. R., & Rapee, R. M. (2020). Evidence that different types of peer victimization have equivalent associations with transdiagnostic psychopathology in adolescence. *Journal of Youth and Adolescence*, 49(3), 590–604. https://doi.org/10.1007/ s10964-020-01202-4
- 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., Sunderland, M., Rapee, R. M., Batterham, P. J., Calear, A. L., Carragher, N., Ruggero, C., Zimmerman, M., Baillie, A. J., Lynch, S. J., Mewton, L., Slade, T., & Krueger, R. F. (2021b). A detailed hierarchical model of psychopathology: From individual symptoms up to the general factor of psychopathology. *Clinical Psychological Science*, 9(2), 139–168. https://doi.org/10.1177/ 2167702620954799
- Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.18534 61
- Fried, E. I., Greene, A. L., & Eaton, N. R. (2021). The p factor is the sum of its parts, for now. *World Psychiatry*, 20(1), 69–70. https:// doi.org/10.1002/wps.20814
- 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
- *Gomez, R., Stavropoulos, V., Vance, A., & Griffiths, M. D. (2019). Re-evaluation of the latent structure of common childhood disorders Is there a general psychopathology factor (p-factor)? *International Journal of Mental Health and Addiction*, 17(2), 258–278. https://doi.org/10.1007/s11469-018-0017-3
- Grotzinger, A. D., Cheung, A. K., Patterson, M. W., Harden, K. P., & Tucker-Drob, E. M. (2019). Genetic and environmental links between general factors of psychopathology and cognitive ability in early childhood. *Clinical Psychological Science*, 7(3), 430– 444. https://doi.org/10.1177/2167702618820018
- Gur, R. E., Yi, J. J., McDonald-McGinn, D. M., Tang, S. X., Calkins, M. E., Whinna, D., Souders, M. C., Savitt, A., Zackai, E. H., Moberg, P. J., Emanuel, B. S., & Gur, R. C. (2014). Neurocognitive development in 22q11.2 deletion syndrome: Comparison

with youth having developmental delay and medical comorbidities. *Molecular Psychiatry*, *19*(11), 1205–1211. https://doi.org/ 10.1038/mp.2013.189

- Haeffel, G. J., Jeronimus, B. F., Fisher, A. J., Kaiser, B. N., Weaver, L. J., Vargas, I., Goodson, J. T., Soyster, P. D., & Lu, W. (2022). HiTOP Is not an improvement over the DSM. *Clinical Psychological Science*. https://doi.org/10.1177/21677026211068873
- *Haltigan, J. D., Aitken, M., Skilling, T., Henderson, J., Hawke, L., Battaglia, M., Strauss, J., Szatmari, P., & Andrade, B. F. (2018). "P" and "DP:" Examining symptom-level bifactor models of psychopathology and dysregulation in clinically referred children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 57(6), 384–396. https://doi.org/10.1016/j. jaac.2018.03.010
- *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
- Hankin, B. L., Snyder, H. R., Gulley, L. D., Schweizer, T. H., Bijttebier, P., Nelis, S., Toh, G., & Vasey, M. W. (2016). Understanding comorbidity among internalizing problems: Integrating latent structural models of psychopathology and risk mechanisms. *Development and Psychopathology*, 28(4), 987–1012. https:// doi.org/10.1017/S0954579416000663
- *Harden, K. P., Engelhardt, L. E., Mann, F. D., Patterson, M. W., Grotzinger, A. D., Savicki, S. L., Thibodeaux, M. L., Freis, S. M., Tackett, J. L., Church, J. A., & Tucker-Drob, E. M. (2020). Genetic associations between executive functions and a general factor of psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, 59(6), 749–758. https://doi. org/10.1016/j.jaac.2019.05.006
- Harris, J. L., LeBeau, B., & Petersen, I. T. (2023). Reactive and control processes in the development of internalizing and externalizing problems across early childhood to adolescence. Preprint: doi: https://doi.org/10.31234/osf.io/js96k
- Hatoum, A. S., Rhee, S. H., Corley, R. P., Hewitt, J. K., & Friedman, N. P. (2018). Do executive functions explain the covariance between internalizing and externalizing behaviors? *Development and Psychopathology*, 30(4), 1371–1387.
- 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
- Haywood, D., Baughman, F. D., Mullan, B. A., & Heslop, K. R. (2021). One p-factor for all? Exploring the applicability of structural models of psychopathology within subgroups of a population. *International Journal of Environmental Research and Public Health.* https://doi.org/10.3390/ijerph18137108
- *He, Q., & Li, J. J. (2021). Factorial invariance in hierarchical factor models of mental disorders in African American and European American youths. *Journal of Child Psychology and Psychiatry* and Allied Disciplines, 62(3), 289–298. https://doi.org/10.1111/ jcpp.13243
- *Herzhoff, K., & Tackett, J. L. (2018). Gender differences in youth externalizing comorbidity [Northwestern University]. ProQuest Dissertations & Theses Global.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. https://doi.org/10.1002/sim.1186

- Hinnant, J. B., & El-Sheikh, M. (2013). Codevelopment of externalizing and internalizing symptoms in middle to late childhood: Sex, baseline respiratory sinus arrhythmia, and respiratory sinus arrhythmia reactivity as predictors. *Development and Psychopathology*, 25(2), 419–436. https://doi.org/10.1017/S095457941 2001150
- Hinshaw, S. P. (2002). Preadolescent girls with attention-deficit/hyperactivity disorder: Background characteristics, comorbidity, cognitive and social functioning, and parenting practices. *Journal of Consulting and Clinical Psychology*, 70(5), 1086–1098. https:// doi.org/10.1037/0022-006X.70.5.1086
- Hoertel, N., Franco, S., Wall, M. M., Oquendo, M. A., Kerridge, B. T., Limosin, F., & Blanco, C. (2015). Mental disorders and risk of suicide attempt: A national prospective study. *Molecular Psychiatry*, 20(6), 718–726. https://doi.org/10.1038/mp.2015.19
- Hopwood, C. J., Bagby, R. M., Gralnick, T., Ro, E., Ruggero, C., Mullins-Sweatt, S., Kotov, R., Bach, B., Cicero, D. C., Krueger, R. F., Patrick, C. J., Chmielewski, M., DeYoung, C. G., Docherty, A. R., Eaton, N. R., Forbush, K. T., Ivanova, M. Y., Latzman, R. D., Pincus, A. L., & Zimmermann, J. (2020). Integrating psychotherapy with the hierarchical taxonomy of psychopathology (HiTOP). *Journal of Psychotherapy Integration*, 30(4), 477–497.
- *Huang, Z., Endo, K., Yamasaki, S., Fujikawa, S., Ando, S., Hiraiwa-Hasegawa, M., Kasai, K., Nishida, A., & Koike, S. (2020). Bidirectional relationships between psychological symptoms and environmental factors in early adolescence. *Frontiers in Psychiatry*. https://doi.org/10.3389/fpsyt.2020.574182
- *Huang-Pollock, C., Shapiro, Z., Galloway-Long, H., & Weigard, A. (2017). Is poor working memory a transdiagnostic risk factor for psychopathology? *Journal of Abnormal Child Psychology*, 45(8), 1477–1490. https://doi.org/10.1007/s10802-016-0219-8
- Huedo-Medina, T. B., Sánchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² index? *Psychological Methods*, 11(2), 193. https:// doi.org/10.1037/1082-989X.11.2.193
- *Jenness, J. L., Peverill, M., Miller, A. B., Heleniak, C., Robertson, M. M., Sambrook, K. A., Sheridan, M. A., & McLaughlin, K. A. (2021). Alterations in neural circuits underlying emotion regulation following child maltreatment: A mechanism underlying trauma-related psychopathology. *Psychological Medicine*, *51*(11), 1880–1889. https://doi.org/10.1017/S00332917200006 41
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry, 62(6), 593–602. https://doi.org/10.1001/archpsyc.62.6.593
- Kessler, R. C., Mickelson, K. D., & Williams, D. R. (1999). The prevalence, distribution, and mental health correlates of perceived discrimination in the United States. *Journal of Health and Social Behavior*, 40(3), 208–230.
- Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, 2(1), 61–76. https://doi.org/10.1002/jrsm.35
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., & Clark, L. A. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454. https://doi. org/10.1037/abn0000258
- Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweatt, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A quantitative nosology based on consensus of evidence. *Annual*

Review of Clinical Psychology, 17(1), 83–108. https://doi.org/ 10.1146/annurev-clinpsy-081219-093304

- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*, 160(9), 1566–1577. https://doi.org/10.1176/ appi.ajp.160.9.1566
- *Laceulle, O. M., Chung, J. M., Vollebergh, W. A. M., & Ormel, J. (2020). The wide-ranging life outcome correlates of a general psychopathology factor in adolescent psychopathology. *Personality & Mental Health*, 14(1), 9–29. https://doi.org/10.1002/pmh. 1465
- *Laceulle, O. M., Vollebergh, W. A. M., & Ormel, J. (2015). The structure of psychopathology in adolescence: Replication of a general psychopathology factor in the TRAILS study. *Clinical Psychological Science*, 3(6), 850–860. https://doi.org/10.1177/ 2167702614560750
- 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. https://doi.org/10.1037/a0028355
- Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). Validity and utility of the general factor of psychopathology. *World Psychiatry*, 16(2), 142–144. https://doi.org/10. 1002/wps.20410
- 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
- *Lahey, B. B., Rathouz, P. J., Keenan, K., Stepp, S. D., Loeber, R., & Hipwell, A. E. (2015). Criterion validity of the general factor of psychopathology in a prospective study of girls. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 56(4), 415–422. https://doi.org/10.1111/jcpp.12300
- *Lambert, M. C., January, S. A. A., & Pierce, C. D. (2018). Latent structure of scores from the emotional and behavioral screener. *Journal of Psychoeducational Assessment*, 36(3), 249–260. https://doi.org/10.1177/0734282916676130
- *Lees, B., Squeglia, L. M., McTeague, L. M., Forbes, M. K., Krueger, R. F., Sunderland, M., Baillie, A. J., Koch, F., Teesson, M., & Mewton, L. (2021). Altered neurocognitive functional connectivity and activation patterns underlie psychopathology in preadolescence. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(4), 387–398. https://doi.org/10.1016/j. bpsc.2020.09.007
- Levin-Aspenson, H. F., Khoo, S., & Kotelnikova, Y. (2019). Hierarchical taxonomy of psychopathology across development: Associations with personality. *Journal of Research in Personality*, 81, 72–78. https://doi.org/10.1016/j.jrp.2019.05.006
- Loehlin, J. C. (2003). Latent variable models: An introduction to factor, path, and structural equation analysis (4th ed.). UK: Psychology Press.
- 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.
- Mann, F. D., Atherton, O. E., DeYoung, C. G., Krueger, R. F., & Robins, R. W. (2020). Big five personality traits and common mental disorders within a hierarchical taxonomy of

psychopathology: A longitudinal study of Mexican-origin youth. Journal of Abnormal Psychology, 129(8), 769-787.

- *Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., Manfro, G. G., Miguel, E. C., Paus, T., Bressan, R. A., Rohde, L. A., & Salum, G. A. (2017). A general psychopathology factor (p Factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *Journal of Abnormal Psychology*, 126(1), 137–148. https://doi.org/10.1037/abn00 00205
- Mayes, S. D., Castagna, P. J., & Waschbusch, D. A. (2020). Sex differences in externalizing and internalizing symptoms in ADHD, Autism, and general population samples. *Journal of Psychopathology and Behavioral Assessment*, 42(3), 519–526. https://doi. org/10.1007/s10862-020-09798-4
- McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review*, 27(6), 1139–1165. https://doi.org/10.3758/s13423-020-01762-3
- McCutcheon, V. V., Scherrer, J. F., Grant, J. D., Xian, H., Haber, J. R., Jacob, T., & Bucholz, K. K. (2013). Parent, sibling and peer associations with subtypes of psychiatric and substance use disorder comorbidity in offspring. *Drug and Alcohol Dependence*, 128(1), 20–29. https://doi.org/10.1016/j.drugalcdep.2012.07.015
- *McElroy, E., Belsky, J., Carragher, N., Fearon, P., & Patalay, P. (2018a). 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
- *McElroy, E., Shevlin, M., Murphy, J., & McBride, O. (2018b). Cooccurring internalizing and externalizing psychopathology in childhood and adolescence: A network approach. *European Child & Adolescent Psychiatry*, 27(11), 1449–1457. https://doi. org/10.1007/s00787-018-1128-x
- McGrath, R. E. (2005). Conceptual complexity and construct validity. Journal of Personality Assessment, 85(2), 112–124. https://doi. org/10.1207/s15327752jpa8502_02
- Mesman, J., Bongers, I. L., & Koot, H. M. (2001). Preschool developmental pathways to preadolescent internalizing and externalizing problems. *The Journal of Child Psychology and Psychiatry* and Allied Disciplines, 42(5), 679–689. https://doi.org/10.1017/ S0021963001007351
- *Mewton, L., Lees, B., Squeglia, L. M., Forbes, M. K., Sunderland, M., Krueger, R., Koch, F. C., Baillie, A., Slade, T., Hoy, N., & Teesson, M. (2021). The relationship between brain structure and general psychopathology in preadolescents. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 63(7), 734–744. https://doi.org/10.1111/jcpp.13513
- 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
- Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel metaanalysis. *International Journal of Social Research Methodology*, 20(6), 559–572. https://doi.org/10.1080/13645579.2016.1252189
- *Mollon, J., Knowles, E. E. M., Mathias, S. R., Rodrigue, A., Moore, T. M., Calkins, M. E., Gur, R. C., Peralta, J. M., Weiner, D. J., Robinson, E. B., Gur, R. E., Blangero, J., Almasy, L., & Glahn, D. C. (2021). Genetic influences on externalizing psychopathology overlap with cognitive functioning and show developmental

variation. European Psychiatry, 64(1), e29. https://doi.org/10. 1192/j.eurpsy.2021.21

- *Moore, T. M., Calkins, M. E., Satterthwaite, T. D., Roalf, D. R., Rosen, A. F. G., Gur, R. C., & Gur, R. E. (2019). Development of a computerized adaptive screening tool for overall psychopathology ("p"). *Journal of Psychiatric Research*, *116*, 26–33. https://doi.org/10.1016/j.jpsychires.2019.05.028
- *Moore, T. M., Kaczkurkin, A. N., Durham, E. L., Jeong, H. J., McDowell, M. G., Dupont, R. M., Applegate, B., Tackett, J. L., Cardenas-Iniguez, C., Kardan, O., Akcelik, G. N., Stier, A. J., Rosenberg, M. D., Hedeker, D., Berman, M. G., & Lahey, B. B. (2020). Criterion validity and relationships between alternative hierarchical dimensional models of general and specific psychopathology. *Journal of Abnormal Psychology*, *129*(7), 677–688. https://doi.org/10.1037/abn0000601
- *Moroney, E. C., & Lee, S. S. (2021). Multi-modal approaches for transdiagnostic risk processes in youth psychopathology [University of California, Los Angeles]. ProQuest Dissertations & Theses Global.
- 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
- *Neumann, A., Muetzel, R. L., Lahey, B. B., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Jaddoe, V. W., Hillegers, M. H. J., White, T., & Tiemeier, H. (2020). White matter microstructure and the general psychopathology factor in children. *Journal* of the American Academy of Child and Adolescent Psychiatry, 59(11), 1285–1296. https://doi.org/10.1016/j.jaac.2019.12.006
- *Neumann, A., Verhulst, F. C., Tiemeier, H., Pappa, I., Lahey, B. B., Medina-Gomez, C., Jaddoe, V. W., Bakermans-Kranenburg, M. J., van Ijzendoorn, M. H., & Moffitt, T. E. (2016). Single Nucleotide Polymorphism heritability of a general psychopathology factor in children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(12), 1038-1045.e4. https://doi.org/10. 1016/j.jaac.2016.09.498
- *Niarchou, M., Moore, T. M., Tang, S. X., Calkins, M. E., McDonald-McGuinn, D. M., Zackai, E. H., Emanuel, B. S., Gur, R. C., & Gur, R. E. (2017). The dimensional structure of psychopathology in 22q11.2 deletion syndrome. *Journal of Psychiatric Research*, 92, 124–131. https://doi.org/10.1016/j.jpsychires.2017.04.006
- Nigg, J. T. (2006). Temperament and developmental psychopathology. Journal of Child Psychology and Psychiatry, 47(3–4), 395–422. https://doi.org/10.1111/j.1469-7610.2006.01612.x
- *Olino, T. M., Bufferd, S. J., Dougherty, L. R., Dyson, M. W., Carlson, G. A., & Klein, D. N. (2018). The development of latent dimensions of psychopathology across early childhood: Stability of dimensions and moderators of change. *Journal of Abnormal Child Psychology*, 46(7), 1373–1383. https://doi.org/10.1007/s10802-018-0398-6
- *Olino, T. M., Dougherty, L. R., Bufferd, S. J., Carlson, G. A., & Klein, D. N. (2014). Testing models of psychopathology in preschoolaged 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
- Olino, T. M., Klein, D. N., Dyson, M. W., Rose, S. A., & Durbin, C. E. (2010). Temperamental emotionality in preschool-aged children and depressive disorders in parents: Associations in a large community sample. *Journal of Abnormal Psychology*, 119(3), 468–478. https://doi.org/10.1037/a0020112
- *O'Reilly, L. M., Pettersson, E., Quinn, P. D., Klonsky, E. D., Lundström, S., Larsson, H., Lichtenstein, P., & D'Onofrio, B. M. (2020). The association between general childhood psychopathology and adolescent suicide attempt and self-harm: A prospective, population-based twin study. *Journal of Abnormal Psychology*, 129(4), 364–375. https://doi.org/10.1037/abn0000512

- *Oro, V., Van Hulle, C., Goldsmith, H., & Lemery-Chalfant, K. (2019). Elucidating mechanisms of familial transmission for general psychopathology: Results from a nuclear twin-family model. *Behavior Genetics*, 49(6), 553. https://doi.org/10.1007/ s10519-021-10072-w
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. https://doi.org/10. 1186/s13643-016-0384-4
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. https://doi.org/10.1136/bmj.n71
- *Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M. (2015). A general psychopathology factor in early adolescence. *British Journal of Psychiatry*, 207(1), 15–22. https://doi.org/10.1192/bjp.bp.114.149591
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991–996. https:// doi.org/10.1016/j.jclinepi.2007.11.010
- 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, 611–625. https://doi.org/ 10.1037/abn0000649
- *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
- Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthe*sis Methods, 10(1), 57–71. https://doi.org/10.1002/jrsm.1332
- Reise, S. P., Kim, D. S., Mansolf, M., & Widaman, K. F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg Self-Esteem Scale. *Multivariate Behavioral Research*, 51(6), 818–838.
- Reise, S. P., & Revicki, D. A. (2014). *Handbook of item response* theory modeling. Taylor & Francis New York.
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73(1), 5–26. https:// doi.org/10.1177/0013164412449831
- *Riglin, L., Collishaw, S., Thapar, A. K., Leppert, B., Martin, J., Richards, A., Anney, R., Smith, G. D., Tilling, K., Stergiakouli, E., O'Donovan, M., & Thapar, A. (2019). SU93 the contribution of psychiatric risk alleles to a general liability to psychopathology in early life: Investigating one explanation for pleiotropy. *European Neuropsychopharmacology*, 29, S1315. https://doi.org/10.1007/s10519-019-09985-4

- Riglin, L., Thapar, A. K., Leppert, B., Martin, J., Richards, A., Anney, R., Davey Smith, G., Tilling, K., Stergiakouli, E., Lahey, B. B., O'Donovan, M. C., Collishaw, S., & Thapar, A. (2020). Using genetics to examine a general liability to childhood psychopathology. *Behavior Genetics*, 50(4), 213–220. https://doi.org/10. 1007/s10519-019-09985-4
- *Rijlaarsdam, J., Barker, E. D., Caserini, C., Koopman-Verhoeff, M. E., Mulder, R. H., Felix, J. F., & Cecil, C. A. M. (2021a). Genome-wide DNA methylation patterns associated with general psychopathology in children. *Journal of Psychiatric Research*, 140, 214–220. https://doi.org/10.1016/j.jpsychires.2021.05.029
- *Rijlaarsdam, J., Cecil, C. A. M., Buil, J. M., van Lier, P. A. C., & Barker, E. D. (2021b). Exposure to bullying and general psychopathology: A prospective, longitudinal study. *Research on Child* and Adolescent Psychopathology, 49(6), 727–736. https://doi. org/10.1007/s10802-020-00760-2
- Ringwald, W. R., Forbes, M. K., & Wright, A. G. C. (2021). Metaanalysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychological Medicine*. https://doi.org/10.1017/S0033291721001902
- Ringwald, W. R., Forbes, M. K., & Wright, A. G. (2022). Meta-analytic tests of measurement invariance of internalizing and externalizing psychopathology across common methodological characteristics. *Journal of Psychopathology and Clinical Science*, 131(8), 847. https://doi.org/10.1037/abn0000785
- Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, 26(2), 141–160. https:// doi.org/10.1037/met0000300
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137. https://doi.org/10.1037/ met0000045
- 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., & Zimmermann, J. (2019). Integrating the Hierarchical Taxonomy of Psychopathology (HiTOP) into clinical practice. *Journal of Consulting and Clinical Psychology*, 87(12), 1069–1084.
- Sallis, H., Szekely, E., Neumann, A., Jolicoeur-Martineau, A., van Ijzendoorn, M., Hillegers, M., Greenwood, C. M. T., Meaney, M. J., Steiner, M., Tiemeier, H., Wazana, A., Pearson, R. M., & Evans, J. (2019). General psychopathology, internalising and externalising in children and functional outcomes in late adolescence. *Journal of Child Psychology and Psychiatry*, 60(11), 1183–1190. https://doi.org/10.1111/jcpp.13067
- *Sheldrick, R. C., Henson, B. S., Merchant, S., Neger, E. N., Murphy, J. M., & Perrin, E. C. (2012). The Preschool Pediatric Symptom Checklist (PPSC): Development and initial validation of a new social/emotional screening instrument. *Academic Pediatrics*, 12(5), 456–467. https://doi.org/10.1016/j.acap.2012.06.008
- Shemmassian, S. K., & Lee, S. S. (2012). Comparing four methods of integrating parent and teacher symptom ratings of attentiondeficit/hyperactivity disorder (ADHD). Journal of Psychopathology and Behavioral Assessment, 34(1), 1–10. https://doi.org/10. 1007/s10862-011-9262-5
- *Shields, A. N., Reardon, K. W., Brandes, C. M., & Tackett, J. L. (2019). The p factor in children: Relationships with executive functions and effortful control. *Journal of Research in Personality*, 82, 103853. https://doi.org/10.1016/j.jrp.2019.103853
- Smith, G. T., Atkinson, E. A., Davis, H. A., Riley, E. N., & Oltmanns, J. R. (2020). The general factor of psychopathology. *Annual Review* of Clinical Psychology, 16(1), 75–98. https://doi.org/10.1146/ annurev-clinpsy-071119-115848

- Smith, G. T., Fischer, S., & Fister, S. M. (2003). Incremental validity principles in test construction. *Psychological Assessment*, 15(4), 467–477. https://doi.org/10.1037/1040-3590.15.4.467
- *Snyder, H. R., Hankin, B. L., Sandman, C. A., Head, K., & Davis, E. P. (2017a). Distinct patterns of reduced prefrontal and limbic grey matter volume in childhood general and internalizing psychopathology. *Clinical Psychological Science*, 5(6), 1001–1013. https://doi.org/10.1177/216770261771456
- *Snyder, H. R., Young, J. F., & Hankin, B. L. (2017b). Strong homotypic continuity in common psychopathology-, internalizing-, and externalizing-specific factors over time in adolescents. *Clinical Psychological Science*, 5(1), 98–110. https://doi.org/10.1177/ 2167702616651076
- Sroufe, L. A. (2016). The place of attachment in development (J. Cassidy & P. R. Shaver, Eds.; 3rd ed.). Guilford Press.
- Sroufe, L. A. (2009). The concept of development in developmental psychopathology. *Child Development Perspectives*, 3(3), 178– 183. https://doi.org/10.1111/j.1750-8606.2009.00103.x
- Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychol*ogy, 5(1), 1–25. https://doi.org/10.1146/annurev.clinpsy.032408. 153639
- Stucky, B. D., & Edelen, M. O. (2014). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response* theory modeling: Applications to typical performance assessment (p. 183206). UK: Taylor & Francis.
- *Sunderland, M., Forbes, M. K., Mewton, L., Baillie, A., Carragher, N., Lynch, S. J., Batterham, P. J., Calear, A. L., Chapman, C., Newton, N. C., Teesson, M., & Slade, T. (2021). The structure of psychopathology and association with poor sleep, self-harm, suicidality, risky sexual behavior, and low self-esteem in a population sample of adolescents. *Development and Psychopathology*, 33(4), 1208–1219. https://doi.org/10.1017/S095457942000043
- *Swales, D. A., Snyder, H. R., Hankin, B. L., Sandman, C. A., Glynn, L. M., & Davis, E. P. (2020). Maternal depressive symptoms predict general liability in child psychopathology. *Journal of Clinical Child and Adolescent Psychology*, 51(1), 85–96. https://doi. org/10.1080/15374416.2020.1723598
- Tackett, J. L. (2011). Parent informants for child personality: Agreement, discrepancies, and clinical utility. *Journal of Personality Assessment*, 93(6), 539–544. https://doi.org/10.1080/00223891. 2011.608763
- *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. https://doi.org/10.1037/a0034151
- Tang, S. X., Yi, J. J., Calkins, M. E., Whinna, D. A., Kohler, C. G., Souders, M. C., McDonald-McGinn, D. M., Zackai, E. H., Emanuel, B. S., Gur, R. C., & Gur, R. E. (2014a). Psychiatric disorders in 22q11.2 deletion syndrome are prevalent but undertreated. *Psychological Medicine*, 44(6), 1267–1277. https://doi. org/10.1017/S0033291713001669
- Tang, S. X., Yi, J. J., Moore, T. M., Calkins, M. E., Kohler, C. G., Whinna, D. A., Souders, M. C., Zackai, E. H., McDonald-McGinn, D. M., Emanuel, B. S., Bilker, W. B., Gur, R. C., & Gur, R. E. (2014b). Subthreshold psychotic symptoms in 22q11.2 deletion syndrome. *Journal of the American Academy of Child & Adolescent Psychiatry*. https://doi.org/10.1016/j.jaac.2014. 05.009
- *Tein, J., Oro, V., Wang, F., Shaw, D., Wilson, M., & Lemery-Chalfant, K. (2023). The role of early intervention for adolescent psychopathology and polydrug use: Mediation through growth in early and middle childhood p factor. Revision invited. *Developmental Psychology*. https://doi.org/10.1037/dev0001543

- *Thompson, E. J., Richards, M., Ploubidis, G. B., Fonagy, P., & Patalay, P. (2021). Changes in the adult consequences of adolescent mental ill-health: Findings from the 1958 and 1970 British birth cohorts. *Psychological Medicine*. https://doi.org/10.1017/S0033 291721002506
- van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. J. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology*, 27(6), 759–773. https://doi.org/10.1177/0959354317737185
- Van Hulle, C. A., Waldman, I. D., D'Onofrio, B. M., Rodgers, J. L., Rathouz, P. J., & Lahey, B. B. (2009). Developmental structure of genetic influences on antisocial behavior across childhood and adolescence. *Journal of Abnormal Psychology*, *118*(4), 711–721. https://doi.org/10.1037/a0016761
- Viechtbauer, W., & Viechtbauer, M. W. (2021). Package 'metafor.' The Comprehensive R archive network. Package 'Metafor'. https:// Cran.r-Project.Org/Web/Packages/Metafor/Metafor.Pdf.
- *Vine, V., Byrd, A. L., Mohr, H., Scott, L. N., Beeney, J. E., & Stepp, S. D. (2020). The structure of psychopathology in a sample of clinically referred, emotionally dysregulated early adolescents. *Journal of Abnormal Child Psychology*, 48(11), 1379–1393. https://doi.org/10.1007/s10802-020-00684-x
- Wade, M., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2018). Effect of foster care intervention on trajectories of general and specific psychopathology among children with histories of institutional rearing: A randomized clinical trial. JAMA Psychiatry, 75(11), 1137–1145. https://doi.org/10.1001/jamapsychiatry.2018.2556
- *Wade, M., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2019). Effect of foster care intervention on trajectories of general and specific psychopathology among children with histories of institutional rearing: A randomized clinical trial. JAMA Psychiatry, 76(4), 447–447. https://doi.org/10.1001/jamapsychiatry.2018.2556
- *Wade, M., Plamondon, A., & Jenkins, J. M. (2021). A family socialization model of transdiagnostic risk for pychopathology in preschool children. *Research on Child and Adolescent Psyhopathology*, 49(8), 975–988. https://doi.org/10.1007/ s10802-021-00789-x
- Wakschlag, L. S., Perlman, S. B., Blair, R. J., Leibenluft, E., Briggs-Gowan, M. J., & Pine, D. S. (2018). The neurodevelopmental basis of early childhood disruptive behavior: Irritable and callous phenotypes as exemplars. *American Journal of Psychiatry*, 175(2), 114–130. https://doi.org/10.1176/appi.ajp.2017.17010 045
- *Waldman, I. D., Poore, H. E., van Hulle, C., Rathouz, P. J., & Lahey, B. B. (2016). External validity of a hierarchical dimensional model of child and adolescent psychopathology: Tests using confirmatory factor analyses and multivariate behavior genetic analyses. *Journal of Abnormal Psychology*, 125(8), 1053–1066. https://doi.org/10.1037/abn0000183

- Wang, F. L., Galán, C. A., Lemery-Chalfant, K., Wilson, M. N., & Shaw, D. S. (2020). Evidence for two genetically distinct pathways to co-occurring internalizing and externalizing problems in adolescence characterized by negative affectivity or behavioral inhibition. *Journal of Abnormal Psychology*, 129(6), 633–645. https://doi.org/10.1037/abn0000525
- 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. Preprint. Doi: https://doi.org/10.31234/osf.io/4tazx
- 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. (2021a). 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., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, 7(6), 1285–1303. https://doi.org/10.1177/ 2167702619855035
- 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/S0954 579419000348
- Wiggins, J. L., Briggs-Gowan, M. J., Brotman, M. A., Leibenluft, E., & Wakschlag, L. S. (2021). Toward a developmental nosology for Disruptive Mood Dysregulation Disorder in early childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, 60(3), 388–397. https://doi.org/10.1016/j.jaac.2020.04. 015
- Wiggins, J. L., Briggs-Gowan, M. J., Estabrook, R., Brotman, M. A., Pine, D. S., Leibenluft, E., & Wakschlag, L. S. (2018). Identifying clinically significant irritability in early childhood. *Journal of* the American Academy of Child & Adolescent Psychiatry, 57(3), 191-199.e2. https://doi.org/10.1016/j.jaac.2017.12.008

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Supplemental Appendix 1. Deduplication, Screening, and Exclusion Procedures.

Before the abstract and title screening stage, duplicates were identified by finding overlaps in information from one or more combinations of the following criteria: author name(s), title, year of publication, journal title, and reference type (e.g., journal article, book chapter) in EndNote X9 (The EndNote Team, 2013). A total of 3,606 duplicated articles were found and removed by the first author. A deduplicated file including abstracts and titles was exported to the screening software, Rayyan (Ouzzani et al., 2016). An additional deduplication screening was conducted in Rayyan, where the software identified instances where two or more articles had a high percentage of common words in their abstract and title. These articles were reviewed by the first author.

The exclusion criteria were: (1) the study did not report empirical findings (e.g., reviews or meta-analyses); (2) mean participant age in the study was > 18.00 years; (3) the study did not use factor analysis to model psychopathology; (4) internalizing and/or externalizing (or their sub-factors) were not evaluated in the study; (5) the study separately evaluated psychopathology factors (i.e., the study did not examine the covariation of internalizing and externalizing); and (6) the psychopathology model included extraneous latent factors that cannot be categorized as sub-factors of internalizing, externalizing, and/or thought disordered problems (e.g., model includes latent factors of personality, stress, wellbeing, etc.). These criteria were ranked in order numerically from broad to specific, and if a given study was excluded for multiple reasons, the exclusion criterion with the broadest rationale was selected for ease of reporting. For example, the exclusion criterion that the study did not use factor analysis took priority over the exclusion criterion that the study did not use factor analysis took priority ore.

Supplemental Appendix 2. Modified Study Quality Measure.

The original Downs and Black checklist (1998) was intended for randomized and nonrandomized clinical trial studies; however the structure of this measure as a numerical coding checklist makes it ideal as a measure of study quality, allowing us to calculate mean and sum quality scores for each study. Items that were not relevant to those in the present review (i.e., items specific to interventions and group comparisons) were eliminated. The wording on several questions was modified to be less specific to interventions and loss to intervention follow up. The modified Downs & Black checklist had 16 items with response options being 1 = yes, 0 =no, as well as a not applicable (NA) option for when we were unable to determine (i.e., there was not enough information present to make conclusion). See Supplemental Table 1 for items. Studies were split evenly among the first two authors to code. The mean score was calculated as the average score, ranging from 0-1, allowing to determine average quality without penalizing for NA responses. The sum score was calculated as the sum of all 16 questions, with lower scores representing worse quality and potentially more NA responses. Both sum and mean study quality scores are in Table 1 in the main text. Sum quality had a mean score of 13.29 (out of a possible 16), SD = 1.91. Scores ranged from 9 to 16 and had a median score of 13. Secondary analyses examined low and high study quality papers separately based on a mean split (M = .88) to determine whether general factor strength differed as a function of study quality.

Supplemental Table 1.

Study Quality Measure

- 11 the main outcomes are first mentioned in the Results section, the question should be 2. answered no.
- *Are the characteristics of the patients/subjects included in the study clearly described?* 3.

Is the hypothesis/aim/objective of the study clearly described? Are the main outcomes to be measured clearly described in the Introduction or Methods section?
 If the main outcomes are first mentioned in the Results section, the question should be

In cohort studies and trials, inclusion and/or exclusion criteria should be given. In case-control studies, a case-definition and the source for controls should be given. *Are the principal confounders and limitations in the study clearly described?*

4. A list of principal confounders is provided and study reports limitations. *Are the main findings of the study clearly described?*

Simple outcome data (including denominators and numerators) should be reported for all major findings so that the reader can check the major analyses and conclusions. (This question

5. does not cover statistical tests which are considered below).

Does the study provide estimates of the random variability in the data for the main outcomes?

In non-normally distributed data the inter-quartile range of results should be reported. In normally distributed data the standard error, standard deviation or confidence intervals should be reported. If the distribution of the data is not described, it must be assumed that the

6. estimates used were appropriate and the question should be answered yes.

In longitudinal data, were attrition and characteristics of participants lost to follow-up described?

This should be answered "unable to determine" for cross sectional or non-longitudinal studies. If longitudinal studies, yes where there were no losses to follow-up or where losses to follow-up were so small that findings would be unaffected by their inclusion. This should be answered nowhere a longitudinal study does not report the number of participants lost to follow-

7. up/attrition.

Have actual probability values been reported (e.g., 0.035 rather than < 0.05) for the main 8. outcomes except where the probability value is less than 0.001?

All the following criteria attempt to address the representativeness of the findings of the study and whether they may be generalized to the population from which the study External Validity subjects were derived.

Were the subjects **asked to participate in the study** representative of the entire population from which they were recruited?

The study must identify the source population for participants and describe how the patients were selected. Participants would be representative if they comprised the entire source population, an unselected sample of consecutive participants, or a random sample. Random sampling is only feasible where a list of all members of the relevant population exists. Where a study does not report the proportion of the source population from which the participants are 9. derived, the question should be answered as unable to determine.

Were those **subjects who were prepared to participate** *representative of the entire population from which they were recruited?*

The proportion of those asked who agreed should be stated. Validation that the sample was representative would include demonstrating that the distribution of the main confounding

10. factors was the same in the study sample and the source population

Were the staff, places, and facilities where the research took place, suitable and approved to conduct the research on this population?

For the question to be answered, yes, the study should demonstrate that the staff and facility was well suited to study the representatives of the source population. The question should be answered no if, for example, the study was not Institutionally affiliated (e.g., IRB) approved, 11. or conducted in a setting unsuited to study the sample.

Internal Validity - bias

	If any of the results of the study were based on "data dredging", was this made clear?
12.	Any analyses that had not been planned at the outset of the study should be clearly indicated. If no retrospective unplanned subgroup analyses were reported, then answer yes.
	were the statistical tests used to assess the main outcomes appropriate?
13.	The statistical techniques used must be appropriate to the data. For example nonparametric methods should be used for small sample sizes. Where little statistical analysis has been undertaken but where there is no evidence of bias, the question should be answered yes. If the distribution of the data (normal or not) is not described, it must be assumed that the estimates used were appropriate and the question should be answered yes.
	Were the main outcome measures used accurate (valid and reliable)?
14.	For studies where the outcome measures are clearly described, the question should be answered yes. For studies which refer to other work or that demonstrates the outcome measures are accurate, the question should be answered as yes.
	Internal Validity - confounding (selection bias)
	Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?
15.	Yes, if the study acknowledges and address confounding variables in the analyses. No, if the study does not include included confounders in the analyses. Unable to determine if no confounders were noted
	Were losses of patients to follow-up taken into account?
	If the numbers of subjects lost to follow-up are not reported (or it is not longitudinal), the question should be answered as unable to determine. If the proportion lost to follow-up was too small to affect the main findings, the question should be answered yes. Should also be

Supplemental Appendix 3. Subgroup and Moderation Analyses Methods.

Subgroup Analyses

Model-Type Analyses

Because hierarchical models and bifactor models tend to differ in their factor loading strengths (Lahey et al., 2021), we conducted a subgroup analysis where we dummy coded the model type such that 0 = hierarchical models, and 1 = bifactor and modified bifactor models. This method would have allowed for an additional test of homogeneity and to determine if the factor analysis approach influenced estimates of general factor strength.

Longitudinal & Measurement Invariance Analysis

Although we accounted for inter-dependency of data given shared data sources and data from the same wave of a study, data from longitudinal studies are thought to provide a stronger test of whether factor strength changes when compared to cross-sectional data (Ringwald et al., 2021, 2022). In an exploratory analysis, we isolated longitudinal studies to determine whether explained common variance changed as a function of age. We also conducted an exploratory analysis on the sample of longitudinal studies that established, at a minimum, partial metric measurement invariance (i.e., invariance of some indicators' factor loadings) across ages. The studies that established partial metric measurement invariance across ages provide greater confidence that the factor strength across ages can be meaningfully compared on a common metric, and thus might provide the most robust tests of stability in general factor strength.

Study Wave Moderation

In this exploratory analysis, we included study wave as a moderator among longitudinal studies and in a separate analysis with studies that established at least partial metric measurement invariance across ages.

Age-Range Analyses

Due to the importance that age plays in the present meta-analysis, and potentially in the development of psychopathology, we assessed whether the strength of the general factor and its change over time differs as a function of models derived from samples with small versus large age-ranges (difference between minimum and maximum age) in their samples. We defined small age-range as < 2.5 years age range, and large age range as ≥ 2.5 years.

Developmental Periods Subgroup Analyses

Developmental periods—preschool age, school age, and adolescence—and the associated developmental tasks with these key developmental stages have impact on development of psychopathology (Cicchetti, 2020; Cicchetti & Rogosch, 2002). To evaluate whether ECV changes as a function of age, within a given developmental period, we subgroup the dataset into the three corresponding developmental periods: preschool age, school age, and adolescence. We defined preschool age as < 6.0 years, school age as \geq 6.00 & < 13.00 years, and adolescence as \geq 13.00 years of age.

Explained Common Variance of Specific Factors

Explained common variance for a specific factor (ECVs) was calculated the same way as general factor but the numerator was the sum of squared specific factor loadings (e.g., calculating ECVs for externalizing; Equation 3; Constantinou, 2019).

$$\frac{(\Sigma\beta_{Ext}^{2})}{(\Sigma\beta_{Gen}^{2}) + (\Sigma\beta_{Ext}^{2}) + (\Sigma\beta_{Int}^{2}) + (\Sigma\beta_{TD}^{2})}$$
(3)

Models that include the presence of the thought disorder specific factor have been shown to load strongly onto the general factor, affecting the distribution of ECVs among the specific factors (see Caspi et al., 2014; Constantinou, 2019). Due to this potential impact, we separately analyzed models that included a thought disorder specific factor from those that included only internalizing, externalizing, and the general factor. The purpose of this exploratory analysis was to determine whether the proportion of variance in the ratings of specific factors of psychopathology change with development. However, some have argued that current methods of modeling general psychopathology are limited in their ability to appropriately capture what is unique above and beyond the general factor, putting the reliability of interpreting ECVs into question (Martel et al., 2017).

Specific Factors Within Developmental Periods

As a follow-up to developmental period analyses, we also examined the ECVs to estimate the proportion of variance in ratings of specific factors of psychopathology within each developmental period. Although reliability is equally called into question in these cases, understanding changes in ECVs across development within developmental periods might elucidate potentially meaningful fluctuations across development. Furthermore, any significant result within a given age group would motivate additional analyses to investigate the result. In this case, we would conduct a meta-analysis and meta-regression on the average of the factor loading strength of the models for the general factor and specific factors in separate analyses. The goal would be to determine if change in one or more factor loading strengths influences a global change in ECV across age, within a given developmental period.

Moderation Analyses Methods

Mean-Based Estimate of General Factor Strength

A different approach to estimating general factor strength, taking the average of squared standardized factor loadings, rather than the sum, is not as influenced by the effect of outlier values and gives a more conservative estimate of general factor strength. Standardized (β)

loadings for the specific and general/*p* factor were extracted in a table or figure in the paper or supplementary materials. The mean-based estimate of general factor strength was calculated by dividing the variance explained by the general factor (i.e., mean of squared general factor loadings) by the total reliable variance (i.e., mean of squared general and specific factor loadings) using Equation 4 (Constantinou, 2019; Rodriguez et al., 2016).

The aforementioned process was straightforward for bifactor and modified bifactor models, but an additional step was needed to calculate the effect sizes for higher-order models. For higher-order models, we follow path tracing rules (Loehlin, 2003), where the β s that are between the general factor and the indicator, typically passing through at least a specific factor, are multiplied to derive the value that represent the regression coefficient from indicator to general factor. The specific factor may also require subordinate sub factors that require similar path tracing multiplication. The mean-based estimate of the strength of the general factor was then calculated using the derived β s using Equation 4. In addition to calculating mean-based estimate of general factor strength, the mean-based estimate of general factor strength was used to derive a sampling variance for each estimate proportion (*p*) using the proportion score and the sample size (*n*), as seen in Equation 2 in main text.

$$\frac{(\bar{x}\beta_{Gen}^2)}{(\bar{x}\beta_{Gen}^2) + (\bar{x}\beta_{Ext}^2) + (\bar{x}\beta_{Int}^2) + (\bar{x}\beta_{TD}^2)}$$
(4)

General Factor Indicator Count Moderation

To determine whether the number of indicators loading onto the general factor influences the explained common variance of the general factor, we included the count of indicators on the general factor in an exploratory moderation analysis. This analysis addresses concerns about whether estimates of explained common variance are potentially inflated as a function of the number of indicators included in the model (Watts et al., 2019). We also analyzed this using the mean-based estimate of general factor strength to determine whether the potential impact of indicator count on ECV influences both sum- and mean-based estimates of general factor strength.

Factor Count Moderation

Similar to the potential influence of indicator count, another factor modeling consideration that has the potential to influence general factor strength is the make-up of the general factor model (van Bork et al., 2017). To evaluate whether the number of specific factors influences general factor strength, we conducted four separate analyses. We included the total number of factors, i.e., one general factor plus the sum of all the specific factors, as a moderator. In separate analyses, we included the count of internalizing, externalizing, and thought disorder factors as moderators in an exploratory analysis. We replicated these analyses using mean-based estimate of general factor strength as a foil to the traditional ECV score. In the k = 15 effect sizes where a thought disorder factor was present, no studies had more than one thought disorder factor in the model. Therefore, the thought disorder factor count was a dummy coded variable where 1 = thought disorder factor present, and 0 = thought disorder factor not present.

Percent Uncontaminated Correlations (PUC)

According to Rodriguez and colleagues (2016), PUC is a metric of how a measurement of a general factor is 'uncontaminated' by multidimensionality due to specific factors. In bifactor models, each item is influenced by both specific factors and the general factor. PUC is the number of correlations explained by the general factor, but not the specific factors. PUC was found to moderate the association between ECV across age, where higher PUC resulted in a strong positive slope, whereas lower PUC resulted in stronger negative slope (Constantinou, 2019). To examine these prior findings meta-analytically in the present study, we examined whether age moderates ECV when including PUC as a moderator. Due to 'contamination' of correlations in higher-order and modified bifactor models due to hierarchical and correlated associations among specific factors, respectively, we only conducted these analyses in traditional bifactor models, due to their orthogonal structure (Reise et al., 2013). As an additional test, we included PUC as a moderator in the association between age and the mean-based estimate of general factor strength (see below).

To estimate PUC for each factor analytic model, we followed the approach by Constantinou & Fonagy (2019) and Reise et al. (2013). We first collected the number of indicators (p), number of specific factors (f), and divided the number of indicators by the number of specific factors to derive an estimate of number of items per specific factor (s). We calculated the number of correlations among the indicators using the formula, p(p-1)/2. We also calculated the number of correlations that arise from the general and specific factors using a similar formula, s(s-1)/2. Dividing the number of correlations that arise from the general and specific factors by the number of correlations among the indicators that can be described among the specific factors in the absence of the general factor. Subtracting this value from 1 derives the PUC, which can be interpreted as the number of correlations described by the general factor, excluding the specific factors. A PUC of \geq .7 indicates that more than 70% of the possible correlations come from a single source, i.e., the general factor of psychopathology (Rodriguez et al., 2016).

$$1 - \frac{\left(\frac{s(s-1)}{2}\right)}{\left(\frac{p(p-1)}{2}\right)} \tag{5}$$

Sex-Related Moderation

. .

To evaluate sex-related differences in the general factor's strength over development, the percent of each sample comprised of females was included as a moderator along with sample mean age. This exploratory test was aimed at assessing whether general factor strength had a sex-related difference.

Informant-Related Moderation

Factor analytic models were generated using information about the child/adolescent's symptoms or syndromes reported by parents, teachers, and self-report informants, or any combination of the three. To analyze whether informant type moderated the effect of explained common variance across ages, binary coded (0,1) variables were added to a new meta-regression: (1) whether parent report contributed to the factor analytic model, (2) whether teacher report contributed to the factor analytic model, or (3) whether self-report by the child or adolescent contributed to the factor analytic model. These informant variables were added as moderators along with sample mean age in an exploratory meta-regression. For an additional test of informant-related moderation, we tested whether having multiple informants was associated with differences in general factor strength. We created a binary coded variable (0,1) with 1 indicating multiple informants and 0 indicating a single informant that was added as a moderator in a separate meta-regression.

Measure-Related Moderation

In a test to determine if age was associated with ECV when including measure type as a moderator, two new variables were binary coded (0,1) and added to the meta-regression: (1) a variable indicating if the model included a questionnaire as a measure, or (2) a variable indicating if the model included a structured or semi-structured clinical interview as a measure. One measure, the Development and Well-Being Assessment (DAWBA), included both a

questionnaire and an interview component, thus receiving a "1" in both variables. Questionnaire and interview variables were added as moderators in a single and separate exploratory analysis.

Measure Type Controlling for General Factor Indicator Count

Additionally, we included the number of indicators loading onto the general factor as a moderator, in addition to the measure type (questionnaire or interview), in the strength of general psychology. These analyses tested whether a moderation effect of number of indicators may instead be better captured by the effects measure type, given the possibility that factor models that include questionnaires may be expected to have more indicators on average compared to factor models that include interviews.

Among the combinations of informants and measures in the present meta-analysis, 5 effect sizes—4.5%—derived from models that included multiple informants and both interviews and questionnaires. 9 effect sizes—8.2%—were derived from models that included only questionnaires and had multiple informants, and 10 effect sizes—9.1%— were derived from models that only included interviews and had multiple informants.

Study Sample Size

To determine whether study sample size influences the strength of the general factor across development, we included sample size as a moderator along with sample mean age. As an additional test of the influence of sample size, we conducted the meta-analysis while setting the sampling variance to constants.

Sampling Variance Constants

In separate analyses, we set the constants to the minimum and maximum sampling variance among the included studies. Adjusting the sampling variance to these constants provided an additional test in the potential conflation between sampling variance and sample size to determine if there is an association between sample size and ECV across development.

Supplemental Appendix 4. Subgroup and Moderation Analysis Results.

Subgroup Analysis Results

Model-Type Analyses

Among bifactor and modified bifactor models (k = 102), the results indicated that the general factor accounted for 58% of the total reliable variance. When examining ECV across development among these models, results yielded a QM(1) of .16, p = .688. The slope of the moderation was: $\beta = -.002$, SE = .004, p = .688.

Among higher-order models (k = 8), the results showed a significantly lower ECV estimate of .40. When examining ECV across development among these models, results yielded a QM(1) of 1.53, p = .216. The slope of the moderation was: $\beta = -.014$, SE = .011, p = .216. The results of these analyses indicate that higher-order factor models are potentially associated with weaker general factor strength compared to bifactor models, but these differences do not translate to any change in ECV across development. However, we caution interpretation of this finding because of limited information from hierarchical models (i.e., only 8 effect sizes derived from this modeling approach).

Longitudinal & Measurement Invariance Analysis

A total of k = 11 longitudinal studies from 8 data sources derived 44 ECV effect sizes. The results indicated that the general factor accounted for 56% of the total reliable variance, which was the same as the estimate when examining all studies, including cross-sectional studies. These results yielded a QM(1) of 0.06, p = .80. The slope of the moderation was: $\beta =$.001, SE = .006, p = .80. These results indicate that general factor strength did not significantly change as a function of age in longitudinal studies. A subset of longitudinal studies established at least partial metric measurement invariance across ages, k = 4, from 4 data sources, yielding 12 ECV effect sizes. Results indicated that the general factor accounted for 56% of the total reliable variance. These results yielded a QM(1) of 0.06, p = .996. The slope of the moderation was: $\beta = .005$, SE = .020 p = .803. Due to small sample size, we caution interpretation of these findings, but there was no significant change in general factor strength when analyzed only in longitudinal samples, and in those that established measurement invariance.

Study Wave Moderation

Among longitudinal studies, results indicated that study wave was not associated with ECV, QM(1) of 0.89, p = .347. The slope of the moderation was: $\beta = .012$, SE = .011, p = .347. Age was not associated with ECV when including study wave as a moderator, QM(2) of 1.20, p = .548. The slope of the moderation was: $\beta = -.004$, SE = .01, p = .581. When study wave was removed from the nesting structure, the results did not significantly differ.

Among studies that established at least partial metric invariance across ages, results indicated that study wave as a moderator was not associated with ECV, QM(1) of 0.18, p = .669. The slope of the moderation was: $\beta = -.012$, SE = .029, p = .669. Age was not associated with ECV when including study wave as a moderator, QM(2) of 0.19, p = .908. The slope of the moderation was: $\beta = .003$, SE = .030, p = .919. When study wave was removed from the nesting structure, the analyses did not converge. These results were only when study wave was included in the nesting structure.

Age-Range Analyses

Among the studies with large age-ranges (>= 2.5 years), k = 50 effect sizes derived a general factor strength of .57, SE = .03. Results indicated that general factor strength in studies with large age-ranges did not differ as a function of age, QM(1) = 2.31, p = .13. Slope of moderation was: $\beta = -.013$, SE = .01, p = .13. Among studies with small age-ranges (< 2.5 years),

k = 56 effect sizes derived a general factor strength of .56, SE = .03. Results indicated that general factor strength in studies with small age-ranges did not differ as a function of age, QM(1) = .16, p = .69. The slope of moderation was: $\beta = .002$, SE = .01, p = .69.

Developmental Period Subgroup Analyses

Preschool Age

There were k = 11 effect sizes that derived a general factor strength of .66 among preschool aged children. Ages ranged from 2 to 5.9 years; M(SD) = 3.52 (1.16). Results indicated that general factor strength significantly increased with age among preschool age children. QM(1) = 4.09, p = .043. The slope of moderation was consistent with a modest effect size, $\beta =$.065, SE = .032, p = .043.

School Age

There were k = 54 effect sizes that derived a general factor strength of .57 among school aged children. Ages ranged from 6.0 to 12.8 years; M(SD) = 9.55 (1.86). Results indicated that general factor strength did not differ as a function of age among school-aged children. QM(1) = 0.026, p = .873. The slope of moderation was: $\beta = -.002$, SE = .011, p = .873.

Adolescence

There were k = 45 effect sizes that derived a general factor strength of .56 among adolescents. Ages ranged from 13.0 to 17.0 years; M(SD) = 14.44 (*1.11*). Results indicated that general factor strength did not differ as a function of age among adolescents. QM(1) = 0.183, p = .669. The slope of moderation was: $\beta = -.008$, SE = .020, p = .669.

Explained Common Variance of Specific Factors (ECVs)

Models without Thought Disorder Factor

In models that did not include a thought disorder factor (k = 95), the general factor accounted for 59% of the reliable variance in ratings of psychopathology, the specific externalizing factor accounted for 21%, and the specific internalizing factor accounted for 19%. Age moderation analysis found that none of the ECVs estimates from specific factors in models without a thought disorder factor were significantly moderated by age; slope_{ext}: $\beta = .005$, SE =.003, p = .16; and slope_{int}: $\beta = -.001$, SE = .003, p = .97.

Models with Thought Disorder Factor

For models that included a thought disorder factor (k = 15), the general factor accounted for 47% of the reliable variance in ratings of psychopathology, the specific externalizing factor accounted for 15%, the specific internalizing factor accounted for 23%, and the specific thought disorder factor accounted for 12%. Age moderation analysis found that none of the ECVs estimates from specific factors in models with a thought disorder factor were significantly moderated by age; slope_{ext}: $\beta = .003$, SE = .011, p = .77; slope_{int}: $\beta = -.018$, SE = .01, p = .21; slope_{thought}: $\beta = -.002$, SE = .01, p = .86.

Specific Factors Within Developmental Periods

Preschool Age

For school age children, there were no models that included a thought disorder factor. Among models of preschool age children (k = 11), the specific externalizing factor accounted for 17% and the specific internalizing factor accounted for 16% of the reliable variance in ratings of psychopathology. Age moderation analysis found that none of the ECVs estimates from specific factors in preschool age children's models were significantly moderated by age; slope_{ext}: $\beta = -.026$, SE = .027, p = .337; and slope_{int}: $\beta = -.044$, SE = .027, p = .106. A significant age moderation within preschool age motivated additional analyses. When examining whether the average of factor loadings for general, internalizing, and externalizing factors changed across preschool ages, age moderation found that the average of factor loadings on the specific internalizing factor decreased across this age range: QM(1) of 4.29, p = .038; slope_{int}: $\beta = -.057$, SE = .027, p = .038. The average of factor loadings on the specific externalizing did not significantly change across this age range: QM(1) of .560, p = .454; slope_{ext}: $\beta = -.023$, SE = .030, p = .454. The average of factor loadings on the general factor also did not significantly change: QM(1) of .177, p = .775; slope_{general}: $\beta = .012$, SE = .027, p = .674.

School Age

For models of school-age children that do not include a thought disorder factor (k = 47), the general factor accounted for 58% of the reliable variance in ratings of psychopathology, the specific externalizing factor accounted for 22%, and the specific internalizing factor accounted for 18%. Age moderation analysis found that none of the ECVs estimates from specific factors in models without a thought disorder factor were significantly moderated by age; slope_{ext}: $\beta = .012$, SE = .009, p = .161; and slope_{int}: $\beta = -.011$, SE = .009, p = .227.

For models of school age children that included a thought disorder factor (k = 7), the general factor accounted for 49% of the reliable variance in ratings of psychopathology, the specific externalizing factor accounted for 14%, the specific internalizing factor accounted for 25%, and the specific thought disorder accounted for 13%. Age moderation analysis found that none of the ECVs estimates from specific factors in models with a thought disorder factor were significantly moderated by age; slope_{ext}: $\beta = -.006$, SE = .027, p = .819; slope_{int}: $\beta = -.004$, SE = .042, p = .933; slope_{thought}: $\beta = .008$, SE = .027, p = .779.

Adolescence

For models of adolescents that do not include a thought disorder factor (k = 37), the general factor accounted for 58% of the reliable variance in ratings of psychopathology, the specific externalizing factor accounted for 22%, and the specific internalizing factor accounted for 20%. Age moderation analysis found that none of the ECVs estimates from specific factors in models without a thought disorder factor were significantly moderated by age; slope_{ext}: $\beta = .022$, SE = .015, p = .137; and slope_{int}: $\beta = -.023$, SE = .016, p = .158.

For models of adolescents that included a thought disorder factor (k = 8), the general factor accounted for 45% of the reliable variance in ratings of psychopathology, the specific externalizing factor accounted for 16%, the specific internalizing factor accounted for 18%, and the specific thought disorder factor accounted for 12%. Age moderation analysis found that none of the ECVs estimates from specific factors in models with a thought disorder factor were significantly moderated by age; slope_{ext}: $\beta = .022$, SE = .032, p = .494; slope_{int}: $\beta = -.013$, SE = .034, p = .711; slope_{thought}: $\beta = .012$, SE = .033, p = .721.

Percent Uncontaminated Correlations

There were k = 93 bifactor models. PUC among bifactor models ranged from .51 to .82; M(SD) = .63 (.09). When PUC was included as a moderator, results yielded a QM(1) of 3.612, p = .057. The slope was consistent with a small effect size, at a trend level, $\beta = .304$, SE = .160, p = .057. Age moderation results yielded a QM(2) of 4.416, p = .110. The slope of the moderation was: $\beta = -.004$, SE = .004, p = .382. These results indicated that higher percent uncontaminated correlations was associated with higher ECV at a trend level. Age was not associated with ECV when including PUC as a moderator.

When PUC was included as a moderator of the mean-based estimate of general factor strength, there was a trend level moderation; QM(1) of 3.07, p = .080. The slope of the

moderation was: $\beta = -.257$, SE = .146, p = .080. Age moderation results yielded a QM(2) of 3.65, p = .161. The slope of the moderation was: $\beta = -.003$, SE = .004, p = .430. Percent uncontaminated correlations was negatively associated with mean-based estimate of general factor strength at a trend level. Age was not associated with the mean-based estimate of general factor strength when including PUC as a moderator.

Moderation Analysis Results

Mean-Based Estimate of General Factor Strength

For the mean-based estimate of general factor strength, the results indicated that the general psychopathology accounted for 33% of the reliable variance. Age moderation results yielded a QM(1) of 0.842, p = .359. The slope of the moderation was: $\beta = -.004$, SE = .004, p = .359. These results indicate that the mean-based estimate of general factor strength did not significantly change as a function of age.

General Factor Indicator Count Moderation

The number of indicators that loaded onto the general factor had a range from 5 to 116, $M_{\text{indicator}}(SD) = 27.81 \ (28.31)$. When general factor indicator count was included as a moderator, it was significantly and positively associated with ECV, QM(1) = 5.40, p = .020. The slope of moderation was consistent with a small effect size: $\beta = .001$, SE = .001, p = .020. Age was not associated with ECV when including general factor indicator count as a moderator, QM(2) =5.65, p = .059. The slope of moderation was: $\beta = -.002$, SE = .004, p = .61.

Results differed when examining the associations with the mean-based estimate of general factor strength. General factor indicator count was not significantly associated with mean-based estimate of general factor strength, QM(1) = 0.187, p = .665; slope: $\beta = .000$, SE = .001, p = .665. Age was not associated with mean-based estimate of general factor strength when

including general factor indicator count as a moderator, QM(2) = .828, p = .661; slope: $\beta = .003$, SE = .004, p = .421.

Factor Count Moderation

Total Factor Count

Included studies had a total factor count (including the general factor and any specific factors) ranging from 3 to 6, $M_{\text{factors}}(SD) = 3.52 \ (0.71)$. When included as a moderator, total factor count was not significantly associated with ECV, QM(1) = 0.679, p = .410. The slope of moderation was $\beta = -.018$, SE = .022, p = .410. Age was not associated with ECV when including total factor count as a moderator, QM(2) of .814, p = .666; slope: $\beta = -.002$, SE = .004, p = .718.

Externalizing Factor Count

Included studies had an externalizing factor count ranging from 1 to 3, $M_{\text{extFactors}}(SD) =$ 1.27 (0.49). When included as a moderator, externalizing factor count was positively associated with ECV at a trend level, QM(1) = 2.99, p = .084. The slope of moderation was $\beta = .050$, SE = .029, p = .084. Age was not associated with ECV when including externalizing factor count as a moderator, QM(2) of 3.27, p = .195; slope: $\beta = -.002$, SE = .004, p = .627.

Internalizing Factor Count

Included studies had an internalizing factor count ranging from 1 to 4, $M_{intFactors}$ (*SD*) = 1.11 (0.39). When included as a moderator, internalizing factor count was not associated with ECV, QM(1) = .858, p = .354. The slope of moderation was $\beta = -.041$, SE = .044, p = .354. Age was not associated with ECV when including internalizing factor count as a moderator, QM(2) of .980, p = .613; slope: $\beta = -.002$, SE = .004, p = .733.

Thought Disorder Factor Presence

When the dummy-coded variable indicating thought disorder presence (1) or absence (0) was a moderator, thought disorder factor presence was negatively associated with ECV, QM(1) = 11.20, p = .001. ECV estimates were approximately .137 units smaller for studies that included a thought disorder factor compared to studies that did not include a thought disorder factor ($\beta = -.137$, SE = .041, p = .001). Age was not associated with ECV when including thought disorder presence as a moderator, QM(2) of 11.34, p = .003; slope: $\beta = -.001$, SE = .004, p = .763.

Sex-Related Moderation

Included samples had a percentage of females ranging from 18% to 99.99%, $M_{\text{female}}(SD)$ = 51.58 (12.46). Age was not associated with ECV when including sex as a moderator, QM(2) of 0.19, p = .91; slope_{female}: $\beta = -.000$, SE = .002, p = .89.

Informant-Related Moderation

Among the included studies, factor analytic models were derived from 88 instances of parent report, 84 instances of self-report, and 4 instances of teacher report. The informant type did not significantly moderate the slope of the ECV across development, QM(4) of 7.29, p = .12; slope_{parent}: $\beta = .038$, SE = .04, p = .38; slope_{teacher}: $\beta = -.034$, SE = .07, p = .61; slope_{self}: $\beta = -.05$, SE = .04, p = .20. A total of 24 effect sizes were derived from multiple informants, leaving 86 from single informants. Age was not associated with ECV when one vs. multiple informants were included as moderators, QM(2) of 0.66, p = .72; slope_{multiple}: $\beta = -.001$, SE = .05, p = .84.

Measure-Related Moderation

Among the included studies, 89 effect sizes were derived using results from questionnaires, while 38 were derived using results from structured clinical interview. 17 factor analytic models included both questionnaires and interviews. Results from moderation analysis of whether a questionnaire or interview was used indicated that a moderation was present, QM(2)

of 6.09, p = .05; slope_{questionnaire}: $\beta = .09$, SE = .04, p = .02. In models derived from questionnaires, there was a .09 standard deviation increase in ECV compared to models derived from interviews.

Measure Type Controlling for General Factor Indicator Count

Among factor analytic models that were derived from questionnaires, when controlling for general factor indicator count, a moderation was present, QM(2) of 10.613, p = .05 where both questionnaire and general factor indicator count were positively associated with general factor strength, which replicates the findings from separate analyses. Age was not associated with ECV when when including both general factor indicator count and presence of questionnaires as moderators, QM(3) of 10.990, p = .012; slope of moderation: $\beta = -.003$, SE = .004, p = .533.

Among factor analytic models that were derived from interviews, when controlling for general factor indicator count, a moderation was present, QM(2) of 12.20, p = .002 where interviews were negatively associated and general factor count was positively associated with general factor strength, which replicates findings from separate analyses. Age was not associated with ECV when including both general factor indicator count and presence of interviews as moderators, QM(3) of 12.97, p = .005; slope of moderation: $\beta = -.004$, SE = .004, p = .381.

Study Sample Size

Included studies had a range of sample size from 160 to 60,888, $M_{\text{sample}}(SD) = 3800.13$ (6852.77). Sample size was not significantly associated with ECV, QM(1) = 1.30, p = .254. The slope of moderation was $\beta = .000$, SE = .000, p = .254. Age was not associated with ECV when including sample size as a moderator, QM(2) of 1.45, p = .485; slope: $\beta = -.002$, SE = .004, p = .702.

Sampling Variance Constants

Among the included studies, the minimum sampling variance was .002 and the maximum sampling variance was .039. When all sampling variances were set to the minimum sampling variance as a constant, ECV was .56 and did not significantly change across development. When set to the maximum sampling variance as a constant, ECV was .58 and also did not significantly change across development.

References

- 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.
- Cicchetti, D. (2020). Studying Psychopathology in Early Life: Foundations of Developmental Psychopathology. In A. G. C. Wright & M. N. Hallquist (Eds.), *The Cambridge Handbook of Research Methods in Clinical Psychology* (pp. 193–204). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/9781316995808.019
- Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, 70(1), 6–20. APA PsycArticles. https://doi.org/10.1037/0022-006X.70.1.6
- Constantinou, M. (2019). *The bifactor model of psychopathology: Methodological issues and clinical applications* (Doctoral dissertation, UCL (University College London)).
- Constantinou, M., & Fonagy, P. (2019). Evaluating bifactor models of psychopathology using model-based reliability indices.
- Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, 52(6), 377. https://doi.org/10.1136/jech.52.6.377
- 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

- Loehlin, J. C. (2003). Latent variable models: An introduction to factor, path, and structural equation analysis (4th ed.). Psychology Press. https://doi.org/10.4324/9781410609823
- Martel, M. M., Pan, P. M., Hoffmann, M. S., Gadelha, A., do Rosário, M. C., Mari, J. J., Manfro, G. G., Miguel, E. C., Paus, T., Bressan, R. A., Rohde, L. A., & Salum, G. A. (2017). A general psychopathology factor (P Factor) in children: Structural model analysis and external validation through familial risk and child global executive function. *Journal of Abnormal Psychology*, *126*(1), 137–148.
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. https://doi.org/10.1186/s13643-016-0384-4
- Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective.
 Educational and Psychological Measurement, 73(1), 5–26.
 https://doi.org/10.1177/0013164412449831
- Ringwald, W. R., Forbes, M. K., & Wright, A. G. C. (2021). Meta-analysis of structural evidence for the Hierarchical Taxonomy of Psychopathology (HiTOP) model. *Psychological Medicine*, 1–14. Cambridge Core. https://doi.org/10.1017/S0033291721001902
- Ringwald, W. R., Wright, A. G., & Forbes, M. K. (2022). Meta-analytic measurement invariance of internalizing and externalizing psychopathology across methodological characteristics. *PsyArXiv. April*, 19.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137.

The EndNote Team. (2013). EndNote (EndNote X9) [Computer software]. Clarivate.

- van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. J. (2017).
 What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory* & *Psychology*, 27(6), 759–773. https://doi.org/10.1177/0959354317737185
- Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, 7(6), 1285–1303. https://doi.org/10.1177/2167702619855035