



Published in final edited form as:

Clin Psychol Sci. 2022 July ; 10(4): 640–661. doi:10.1177/21677026211055170.

How Robust Is the p Factor? Using Multitrait-Multimethod Modeling to Inform the Meaning of General Factors of Youth Psychopathology

Ashley L. Watts¹, Bridget A. Makol², Isabella M. Palumbo³, Andres De Los Reyes², Thomas M. Olino⁴, Robert D. Latzman³, Colin G. DeYoung⁵, Phillip K. Wood¹, Kenneth J. Sher¹

¹Department of Psychological Sciences, University of Missouri, Columbia, MO, 65211.

²Department of Psychology, University of Maryland, College Park, MD, 20742.

³Department of Psychology, Georgia State University, Atlanta, GA, 30303.

⁴Department of Psychology, Temple University, Philadelphia, PA, 19122.

⁵Department of Psychology, University of Minnesota, Minneapolis, MN 55455.

Abstract

We used multitrait-multimethod (MTMM) modeling to examine general factors of psychopathology in three samples of youth (N s = 2119, 303, 592) for whom three informants reported on the youth's psychopathology (e.g., child, parent, teacher). Empirical support for the p -factor diminished in multi-informant models compared with mono-informant models: the correlation between externalizing and internalizing factors decreased and the general factor in bifactor models essentially reflected externalizing. Widely used MTMM-informed approaches for modeling multi-informant data cannot distinguish between competing interpretations of the patterns of effects we observed, including that the p -factor reflects, in part, evaluative consistency bias or that psychopathology manifests differently across contexts (e.g., home vs. school).

Author correspondence: Correspondence should be addressed to Ashley L. Watts, 200 South, Seventh Street, Columbia, MO, 65211. ashlelywatts@gmail.com.

Author Contribution Statement

We define author contribution using the Contributor Roles Taxonomy (CRediT; <http://credit.niso.org/>). Watts was responsible for conceptualization, data curation, formal analysis, methodology, software, visualization, and writing (original draft preparation, review and editing). Makol was responsible for writing (review and editing). Palumbo was responsible for data curation and writing (review and editing). De Los Reyes, Wood, and Sher were responsible for supervision and writing (review and editing). Olino, Latzman, and DeYoung were responsible for writing (review and editing).

Data Use Statement

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (<https://abcdstudy.org>), which is held in the NIMH Data Archive (NDA), and from a limited access dataset provided by the Child Mind Institute Biobank, Healthy Brain Network (<http://www.healthybrainnetwork.org>). The ABCD Study is a multisite, longitudinal study designed to recruit more than 10,000 children age 9–10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at <https://abcdstudy.org/federal-partners.html>. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/Consortium_Members.pdf. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from DOI 10.15154/1503209. This manuscript reflects the views of the authors and does not necessarily reflect the opinions or views of the Child Mind Institute.

Ultimately, support for the p -factor may be stronger in mono-informant designs, although it is does not entirely vanish in multi-informant models. Instead, the general factor of psychopathology in any given mono-informant model likely reflects a complex mix of variances, some substantive and some methodological.

Keywords

p -factor; general factor of psychopathology; multitrait-multimethod modeling; multi-informant psychopathology structures

One source of contention in psychopathology research is the notion that there is a single dimension reflecting risk for all forms of psychopathology, often referred to as the p -factor or general factor of psychopathology (Caspi et al., 2014; Lahey et al., 2012).¹ Various research teams have proposed mechanisms that could cause the theoretical p -factor (e.g., disordered thought, emotion dysregulation, negative emotionality; (Caspi & Moffitt, 2018; Smith et al., 2020), but what general factors of psychopathology actually capture in any given sample, measure, and model remains debated (Caspi & Moffitt, 2018; Lahey et al., 2021; Smith et al., 2020; Watts et al., 2019, 2020). What might general factors of psychopathology reflect, if not a *real* tendency for all forms of psychopathology to covary?

Few studies have grappled with this question, in part due to the tendency to presuppose that general factors of psychopathology reflect something substantive (cf., Smith et al., 2020; Watts et al., 2020). Research thus far has largely overlooked the possibility that general factors of psychopathology arise due to other factors, including a variety of method artifacts (see Leising et al., 2020, for a related review). Informed by other literatures, this paper uses multitrait-multimethod modeling to test the degree to which general factors of psychopathology are robust when using multiple informants to rate a target's psychopathology.

One reason the p -factor might not be robust is that a particular method artifact, *evaluative consistency bias*, is likely to be a source of systematic variance in general factors of psychopathology. Evaluative consistency (or halo) bias refers to the tendency to rate oneself or someone else more positively or negatively than is warranted across a wide range of characteristics. It has been demonstrated in ratings of personality and other characteristics (Anusic et al., 2009; Feeley, 2002; Forgas & Laham, 2017). Socially desirable responding is related to evaluative consistency bias but is narrower in scope because socially desirable responding entails only the tendency to report overly positive evaluations.

When a person exhibits evaluative consistency, they do not necessarily respond to *all* items in the extreme, but their ratings are systematically biased away from the true response in the desirable or undesirable direction. For instance, a teacher who is prone to view a student in an overly negative light may report that the student exhibits some level of a broad swath of psychopathology features that span higher-order externalizing and internalizing

¹We use " p -factor" to refer to the theoretical construct and "general factor of psychopathology" to refer to general factors extracted in a statistical model to avoid conflating theoretical and statistical constructs (Watts et al., 2020).

dimensions – sadness, emotional outbursts, anxiety, conduct problems, aggression, and so on – even though in reality the child displays conduct problems and aggression but not mood or anxiety problems. The presence of raters who exhibit evaluative consistency, regardless of whether the bias is toward identifying desirable or undesirable characteristics, tends to produce a positive manifold and less differentiated higher-order dimensions (e.g., externalizing, internalizing; see Forgas & Laham, 2017, for a recent review of empirical findings).

The hypothetical example of the teacher bears clear relevance to the p -factor, which is often claimed both to cause a positive manifold in psychopathology data and to reflect the source of the correlation between higher-order psychopathology dimensions (Caspi et al., 2014; Lahey et al., 2012). Given that a substantive p -factor and evaluative consistency bias would tend to generate very similar statistical phenomena, determining whether general factors of psychopathology are probably substantive requires a measurement design that can remove evaluative consistency bias. Here, we use multitrait-multimethod modeling (MTMM; Campbell & Fiske, 1959) to incorporate multiple perspectives on the same target participant. Using multiple informants and modeling each construct as the shared variance of multiple informants' ratings reduces the influence of evaluative consistency bias from any given informant on structural models (Chang et al., 2012; Eid et al., 2008). In this respect, one value of using MTMM modeling approaches lies in the ability to compare them to mono-informant models. If empirical support for the p -factor is more robust in mono-informant models relative to MTMM models, it suggests that general factors of psychopathology arise, at least in part, from evaluative consistency bias.

Disentangling Substance and Style in Structural Models

The foundations of MTMM modeling were strongly influenced by Garner and colleagues' (1956) concept of “converging operations,” whereby convergence across informants more closely approximates reality, or reflects a more accurate (bias-free) judgement of the construct of interest. Campbell and Fiske's (1959) extension of Garner and colleagues' concept further assumes that a score on any measure reflects the construct of interest plus systematic method-specific influences (plus random error), and that two or more methods are necessary to differentiate trait and method. Multi-informant models fall under the umbrella of MTMM models, which include substantive (trait) factors that include variance shared across all informants, as well as informant-specific sources of variance either in the form of factors (Figures 1a-b) or correlated errors (Figure 1c). Informant-specific factors are thought to capture bias, or variations in one's report of the target construct that is completely independent from one's true levels of the construct.

The present study was inspired by the general personality literature and its application of MTMM modeling to explore the role of evaluative consistency bias in structural models of personality. In mono-informant designs (e.g., self-report), higher-order factors of the Big Five personality dimensions (labeled Alpha and Beta or Stability and Plasticity) are moderately correlated (r s generally between .4 and .6; e.g., (Biesanz & West, 2004; DeYoung, 2006; Digman, 1997), which means that a general factor of personality can be modeled above them to account for that correlation. The seminal paper on the general factor

of personality (Musek, 2007) inspired a slew of papers arguing that the general factor of personality reflects a substantive mechanism that is associated with increased self-esteem, popularity, likeability, and well-being, among others. Multiple groups of researchers have converged on the theory that the general factor of personality is a heritable dimension of evolutionarily adaptive social effectiveness (Rushton & Irwing, 2011; van der Linden et al., 2017).

Along with arguments that the general factor of personality reflects a robust, meaningful mechanism integral to human functioning came skepticism (Chang et al., 2012; Hopwood et al., 2011; Revelle & Wilt, 2013). A major debate in the literature surrounded whether correlations among the Big Five and their higher-order dimensions are artefactual and driven by correlated measurement error arising from evaluative consistency bias (Biesanz & West, 2004). Perhaps the death knell for the general factor of personality was three complementary sets of findings. First, in MTMM models, the two higher-order factors of the Big Five are essentially uncorrelated ($r_s \approx .00 - .10$), whereas the Big Five dimensions within each higher-order factor remain moderately intercorrelated (Chang et al., 2012; DeYoung, 2006).² Second, Revelle and Wilt (2013) concluded that a general factor that explains only around 50 percent of the variance in personality is inadequate for a supposedly “general” dimension. Third, in a comprehensive analysis of multiple personality inventories, Hopwood and colleagues (2011) found substantively different general factors depending on the instrument. They concluded that covariation among various personality traits cannot be represented by a single construct (i.e., a general factor), and that any general factor from a given study is probably sample-, method-, and measure-specific. At this point, the general factor of personality is widely regarded as an artifact of evaluative consistency bias, among other method artifacts (Bäckström & Björklund, 2021; Hopwood et al., 2011; Leising et al., 2020).

Sources of Variance in Multi-informant Models of Psychopathology

As with personality, there are at least two reasons to suspect that general factors of psychopathology may be caused by evaluative consistency bias. First, psychopathology, in part, reflects variation in general personality traits (e.g., Suzuki et al., 2015), so relatively stable tendencies toward personality and psychopathology can be described by the same basic dimensions (Krueger & Tackett, 2003; Widiger et al., 2019). In fact, because psychopathology reflects undesirable and dysfunctional behavior and experience, psychopathology ratings may be even more prone to evaluative consistency bias than personality ratings. Second, researchers have simultaneously modeled general factors of personality and psychopathology in adults and found that they are highly correlated (Littlefield et al., 2021; Oltmanns et al., 2018).

²Studies of the general factor of psychopathology typically rely on bifactor models, whereas studies of the general factor of personality often rely on higher-order models. Bifactor and higher-order models differ conceptually. Higher-order models contain second-order factors that explain the covariation among first-order factors that explain covariation among their constituent indicators. Bifactor models explain covariation among all indicators, and additional, narrower (specific factors) explain additional covariation among subsets of indicators. Though these models are conceptually different, they appear to produce general factors that are highly congruent (Clark et al., 2021; Forbes et al., 2021).

That general factors of psychopathology may be prone to capturing evaluative consistency bias is especially concerning given that most studies that model a general factor of psychopathology incorporate data from only one informant. Typically, mono-informant models incorporate self-reports or clinical interviews in studies of adults, and parent reports in studies of children (Achenbach, 2020). Although each of these sources are probably useful, studies that rely on reports of psychopathology from multiple informants reveal that informants rarely agree in terms of their perception of the target's psychopathology (De Los Reyes et al., 2013, 2019). In fact, no informant's report is fully redundant with another's. Meta-analytically, agreement for child behavior problems across informants (e.g., parent-teacher, parent-clinician) tends to be moderate ($r_s \approx .20-.30$), though agreement among pairs of informants who observe the child in the same context (e.g., two parents at home, two teachers at school) tends to be stronger ($r = .60$; for reviews, see: Achenbach et al., 1987; De Los Reyes et al., 2015). Further, cross-informant agreement appears to vary as a function of observability, with agreement being slightly more pronounced for externalizing compared with internalizing ($r_s = .41$ and $.32$, respectively).

Even though the general personality literature tends to interpret informant disagreement as reflecting mostly bias, there is often unique valid variance in different informants' personality ratings as well (Connelly & Ones, 2010; Jackson et al., 2015), and this is also probably true in psychopathology ratings. Further, in developmental psychopathology research, informants are often recruited systematically from different contexts (e.g., parents, teachers). This is different from typical multi-informant personality research, in which informants often include a mix of people with a variety of relationships to the target. Thus, context is typically more important in developmental psychopathology, and it is well established that informant discrepancies to some extent reflect contextual variation in domain-relevant behaviors, or "situational specificity" (Achenbach et al., 1987). Children often behave differently across their social contexts, and informants interact differently with children, so the informants embedded in these contexts have unique opportunities to observe their behavior (Achenbach et al., 1987). For example, certain behaviors (e.g., disruptive behaviors, inattention, social anxiety) may manifest most strongly in more structured environments like classrooms, so they are more likely to get flagged by teachers than parents (e.g., Barkley, 2003).

If children exhibit certain forms of psychopathology more in some settings than in others, it will give rise to a substantively meaningful discrepancy between parent- and teacher-reported psychopathology. De Los Reyes and colleagues (2009) found that parent reports of preschoolers' disruptive behavior often disagreed with reports completed by teachers, such that each endorsed disruptive behavior that the other did not endorse. In this study, children's disruptive behavior was also indexed using an independent, contextually sensitive observational paradigm, and within this paradigm children tended not to display disruptive behavior in contexts where informants did not report elevations. Also, other studies have shown that adolescents are at increased risk for developing substance use problems when parents claim to have more knowledge of their children's activities than children say they have (Lippold et al., 2013), and youth who self-report lower levels of anxiety than their parents report tend to experience poorer treatment outcomes (Becker-Haimes et al., 2018).

Importantly, the unique variance attributable to each informant is consistently associated with theoretically relevant external criteria, above and beyond what is shared across informants and what is unique to other informants. Youth, parent, and confederate reports of adolescent social anxiety have demonstrated incremental validity above and beyond one another in the statistical prediction of behavioral indicators during social interaction tasks, whether the adolescent received mental health services, clinical severity, and the like (Cannon et al., 2020; Glenn et al., 2019; Makol et al., 2020, 2021). Also, informant divergence is associated with poorer outcomes for certain domains. In sum, if children's psychopathology manifests differently in different contexts (e.g., school vs. peer interactions), informant-specific factors in a MTMM probably reflect genuine variance in psychopathology as well as bias when informants systematically have access to different contexts. Thus, finding that general factors remain in MTMM models of psychopathology provides fairly clear evidence for the existence of those factors, but finding that they do not remain or are weakened comes with some ambiguity about whether the variance removed is bias or substance.

Present Study

Typical MTMM approaches with multiple informants cannot readily adjudicate between these competing interpretations of informant-specific factors. Nevertheless, without the use of MTMM modeling, we cannot readily determine whether general factors of psychopathology are *potentially* due to evaluative consistency bias. To our knowledge, at least thirteen studies have extracted general factors of psychopathology using data from multiple informants. Eight studies claimed to find a strong general factor indicated by psychopathology across informants but none modeled informant-specific factors (Allegrini et al., 2020; Brandes et al., 2019; Castellanos-Ryan et al., 2016; Hamlat et al., 2019; Jenness et al., 2020; Lahey et al., 2018; Neumann et al., 2016; Riem et al., 2019; Weissman et al., 2019). Omitting these method factors can artificially inflate the strength of the substantive factors (Geiser & Lockhart, 2012). Nevertheless, three studies modeled informant effects and again found a reasonably strong general factor (Oltmanns et al., 2018; Sallis et al., 2019; Snyder et al., 2017), so it is probably unrealistic to anticipate that empirical support for the p -factor will completely vanish in a MTMM model.

What previous studies lack is systematic examination of whether empirical support for the p -factor diminishes when comparing mono- to multimethod structures. We compared mono- and multi-informant structures for two models relevant to the p -factor, (1) a correlated factors model with externalizing and internalizing factors and (2) a bifactor model with a general factor and externalizing and internalizing specific factors. We conceptualized empirical support for the p -factor as follows:

1. The correlation between externalizing and internalizing factors (e.g., Lahey et al., 2012).
2. The proportion of variance explained in psychopathology indicators by a general factor of psychopathology (Watts et al., 2020).

3. The range in loadings on a general factor of psychopathology (i.e., how well it is represented by psychopathology indicators; Watts et al., 2019, 2020).

If general factors of psychopathology *do not* contain variance that arises from evaluative consistency bias, all of the following should occur when we compare multi-informant models to mono-informant models:

1. The externalizing-internalizing correlation should not become attenuated.
2. The general factor should explain an equivalent amount of variance in its indicators.
3. The range in loadings on the general factor should not widen.

We gathered existing data where broadband psychopathology assessment was available for three or more informants, with the aim of identifying samples with identical assessments of psychopathology to rule out variation in instruments as a contributor to different findings across samples. To examine replication across independent samples, we used three samples of youth – the Adolescent Brain and Cognitive Development Study, the Healthy Brain Network, and the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development – for which 3 informants completed the widely-used Achenbach System of Empirically Based Assessment (ASEBA) School-Age Instruments (e.g., Child Behavior Checklist [CBCL]; (Achenbach, 2009). We relied on youth samples for a practical reason: Reports from multiple informants are often incorporated in structural models of child psychopathology, whereas it is rare to find samples of adults for whom psychopathology was assessed with three or more informants.

Methods

Participants and Procedure

We used data from 3 community-based samples of children and adolescents, the (1) Adolescent Brain and Cognitive Development Study (ABCD, $n = 2119$), (2) Healthy Brain Network (HBN, $n = 303$), and (3) two waves of the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD, $ns = 592, 672$).³ Written informed consent was obtained from participants ages 18 and older, and written consent was obtained from their legal guardian and written assent from the participant when participants were younger than 18. Table S2 displays an array of demographic information for each sample.

ABCD.—Participants were drawn from the 6-month follow-up wave of the ABCD Study (Data Release 2.0), a sample of 11,872 9- and 10-year-olds from the U.S. The ABCD study is a collaboration between 21 sites across the U.S. (Jernigan et al., 2018). We used the subset of the 6-month follow-up data with child-, parent-, and teacher-reports were available

³Because we do not own these data, we cannot make them publicly available. Researchers interested in accessing these data can apply at the following links (ABCD: <https://nda.nih.gov/abcd>, HBN: http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/Pheno_Access.html, NICHD SECCYD: <https://www.icpsr.umich.edu/web/ICPSR/series/00233>). To facilitate reproducibility, our OSF page (<https://osf.io/hvskz/>) includes a data dictionary of all indicators used (Table S1), correlation matrices of all indicators used here (Table S3), and model scripts and outputs.

($n = 2119$). Children were aged 9 through 12 ($M = 10.98$, $SD = .63$); 52% identified as female, and 66% identified as White, 8% as Black, 2% as Asian, 10% as Other, and 14% as Hispanic. See Table S2 for a breakdown of household income.

HBN.—The Child Mind Institute’s HBN consortium (Alexander et al., 2017) is an ongoing initiative that aims to create and share a biobank of data from 10,000 New York area children. We used the subset of HBN ($n = 303$) where self-, parent-, and teacher-report data were available at the time of data analysis. Children included in this subset were aged 10 to 17 ($M = 13.6$, $SD = 1.92$); 38% identified as female, 49% as White, 14% as Black, 18.2% as Hispanic, 3% as Asian, and 23.5% as Biracial or Other. See Table S2 for a breakdown of household income.

NICHD SECCYD.—The NICHD SECCYD began recruitment of 1,364 families in 1991, from hospitals located in 10 cities across the US (NICHD Early Child Care Research Network, 2005). Children completed assessments when they were 1, 6, 15, 24, 36, and 54 months old, when they were in kindergarten and grades 1 to 6, and at age 15 years. We used assessments from grade 6 and age 15 to maintain compatibility in terms of average sample age with those of HBN and ABCD. We used the subsets of data where ratings from both parents (or primary caregivers) and a teacher were available at grade 6 ($n = 592$), and from both parents and the child at age 15 ($n = 672$). At the Grade 6 wave, children were aged 11 through 13 ($M = 11.34$, $SD = .48$); 50% identified as female, 86% as White, 8% as Black, 1% as Asian, 5% as Other, and 6% as Hispanic. The mean household income at this wave was \$94,586 ($SD = \$74,730$). At the Age 15 wave, children were aged between 14 and 15 ($M = 14.42$, $SD = .49$); 48% identified as female, 87% as White, 8% as Black, 1% as Asian, 4% as Other, and 6% as Hispanic. The mean household income at this wave was \$116,013 ($SD = \$111,235$).

Psychopathology Assessment and Modeling

Each of the samples we used included the ASEBA School-Age Instruments, which are widely used, standardized assessments designed to assess emotional, social, and behavioral problems in school-age children (Achenbach, 2009). These instruments include CBCL completed by parents (113 items), Youth Self Report (YSR) completed by youth (112 items), and Teacher Report Form (TRF) completed by teachers (113 items). There is also the abbreviated Brief Problem Monitor (BPM), which includes a version for each type of informant (19 items for children and parents, 18 items for teachers). All BPM items are drawn from their longer-form ASEBA counterparts. Across all instruments, items are rated using a 3-point Likert-type response format (*not true, sometimes true, very/often true*).

Samples differed in terms of specific assessments available (see Table S4 for an outline of the assessments available for each sample). For instance, in ABCD, parents completed the CBCL, whereas children and teachers completed the abbreviated BPM. All other samples included the long form measures for informants (YSR, CBCL, and TRF). We elected to use the 19 items (18 for teachers) from the BPM for the following reasons. First, because of the length of the long form variants of the ASEBA scales (e.g., YSR, CBCL, TRF), modeling their item-level MTMM structure was not possible (i.e., the models do not converge).

Second and relatedly, we could not model ASEBA subscales across samples because the BPM does not yield subscale scores.⁴ Third, modeling BPM items allowed us to maintain compatibility across samples in terms of identical assessment.

The BPM contains items assessing internalizing (6 items), externalizing (7 items for children and parents, 6 for teachers), and attention problems (6 items), but not thought problems. The omission of thought problems items precluded our ability to model a higher-order thought problems factor. Also, although the ASEBA instruments do not include Attention Problems items in the Externalizing composite, we elected to model attention problems items with externalizing given that attention-deficit/hyperactivity disorder is widely regarded as a robust marker of child externalizing (e.g., Tackett et al., 2014). Other researchers have similarly used ASEBA Attention Problems as markers of externalizing, and they loaded strongly onto an externalizing latent factor (e.g., Lacey et al., 2015).⁵

Linear Mixed Effects Models

We conducted linear mixed models to examine mean-level differences in externalizing and internalizing across informant type using R version 4.0.2 and the `gamm4` (Wood & Scheipl, 2020) and `emmeans` (Lenth, 2021) packages. We included composite scores of externalizing and internalizing (reflected as the mean of the score for each item within the composite) as dependent variables. We began with an omnibus test with psychopathology composite (collapsed across externalizing and internalizing) as the dependent variable; informant type, psychopathology, and sample as fixed effects; and site as a random effect to account for nonindependence among observations collected at the same location. There were significant main effects for sample ($F = 242.26$, $p < .001$, $dfs = 3$, 22085), informant type ($F = 149.83$, $p < .001$, $dfs = 4$, 22085) and psychopathology type ($F = 336.32$, $p < .001$, $dfs = 1$, 22085; Table S6), so we then conducted linear mixed effects models (1) within sample and across informant types to test for mean differences across informant type, and (2) across samples and within informant type to test for mean differences across samples and within informant type. We conducted these tests for externalizing and internalizing separately.

Multitrait-Multimethod Models

A wide array of MTMM models have been developed for use in confirmatory factor analysis (CFA). As we noted earlier, this study treats each informant as a different method. Most all of the widely used MTMM models have been criticized for one or more reasons (Eid et al., 2003, 2008; Lance et al., 2002), so we fit a number of them for the sake of comprehensiveness, to minimize the problems associated with any given model, and to examine the robustness of our conclusions across models:

1. *Correlated traits, correlated methods* (CTCM, Figure 1a). The CTCM model decomposes an observed trait into three components: trait, method, and a residual component that is independent of the trait and method components.

⁴For the sake of comprehensiveness, we fit subscale-level MTMM models in HBN and NICHD SECCYD (see Table S5). Major findings were consistent across item- and subscale-level models, but the subscale-level models fit poorly, further justifying our use of the BPM items.

⁵We also conducted all analyses with Attention Problems items excluded. The latent correlation between externalizing and internalizing did not vary as a function of whether attention problems items were included in the model (Figure S1, Table S8).

It also includes one method factor per informant, with all psychopathology indicators within-informant loading onto them. Method factors are allowed to correlate. An advantage of the CTCM model is that it accommodates the possibility that two people (e.g., parents) provide correlated ratings above and beyond what is captured in the common trait factors, but it is often associated with estimation issues. Also, even when the CTCM model produces a solution that converges, method factors tend to absorb trait variance, which weakens the trait factors (Lance et al., 2002).

2. *2. Correlated traits, orthogonal methods* (CTOM, Figure 1b). The CTOM model is nearly identical to the CTCM model, but the method factors are uncorrelated. Major advantages of the CTOM include that it has fewer estimation problems than the CTCM model, yields unique informant factors, and better differentiates trait and method effects. The CTOM model is conceptually similar to Kraemer's "satellite method" (2003), which extracts principal components for the trait of interest and each informant (setting aside differences between factor analysis and principal components analysis), and Bauer and colleagues' (2013) trifactor model, which specifies orthogonal informant factors.
3. *3. Correlated traits, correlated uniquenesses* (CTCU, Figure 1c). In lieu of method factors, the CTCU model allows for correlated residuals (uniquenesses) among indicators within but not across informants. The CTCU model therefore is more agnostic with respect to the dimensionality of method effects given that the CTCM and CTOM models assume they are unidimensional. The CTCU model has been criticized because it produces upwardly biased estimates of trait factor loadings and does not directly estimate method variance (e.g., Lance et al., 2002).
4. *4. M-1 models*. Extensions of the CT-CM and CT-OM models, the CT-CM minus one (M-1) and CT-OM minus one models (Eid et al., 2003), claim to resolve identification issues associated with CTCM models because they contain one less informant factor than informants. One method (e.g., mother-report) is chosen as the reference method and its respective method factor is dropped from the model. Functionally, indicators of the dropped method factor define the substantive (trait) factors and serve as the comparison standard by which all other methods (e.g., father-report, teacher-report) are compared. For example, if the mother's method factor is dropped, the externalizing and internalizing factors are defined by the mother's externalizing and internalizing ratings. Because there is no obvious *a priori* referent informant (De Los Reyes, 2011), we tested all possible CT-CU/OM minus one models.

In addition to correlated traits models, we tested bifactor models to explicitly model a "general" psychopathology dimension, given recent interest in bifactor models and despite criticisms of them (Bonifay et al., 2017; Watts et al., 2019). These models included general factors, two specific factors for externalizing and internalizing, and either method factors or correlated uniquenesses between items within informants (Figures 1d-f). Across all MTMM

models tested, trait and method factors were orthogonal to disentangle trait and method variance.⁶

Confirmatory Factor Analyses

We conducted all CFAs using Mplus version 8.4 (Muthén & Muthén, 1998-2017). Because each of the datasets we used are drawn from large, multi-site efforts, we accounted for the complex survey design. In HBN and NICHD SECCYD, we clustered by site. In ABCD, we treated site as a stratum variable and further clustered by family to account for nonindependence of siblings within families.

We estimated CFAs with WLSMV, a robust estimator that provides the best option for modeling ordinal data (Muthén & Muthén, 1998-2017), and with latent factor means fixed to 0 and variances to 1. We considered the following goodness-of-fit statistics to evaluate model fit: the Comparative Fit Index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). Although we present χ^2 values, their p values, and their degrees of freedom, we did not use them to determine adequacy of model fit because χ^2 significance tests are highly sensitive to sample size and would be virtually certain to be rejected in our larger samples. Instead, the fit of a single model was evaluated using the combination of CFI, TLI and RMSEA. To characterize factor similarity across MTMM models, we report Tucker's phi (Φ). Like a correlation coefficient, Φ ranges from -1 to $+1$.

Results

Cross-informant Agreement on Psychopathology

Mean-level agreement.—Informants reported higher levels of externalizing compared with internalizing, but there was significant variability in reported mean levels of externalizing across informants (Tables S6, S7). Children tended to report higher levels of externalizing compared with parents (including when mothers and fathers served as separate informants; Cohen's d s ranged from .08 to .95). Also, children and parents reported higher levels of externalizing than teachers (Cohen's d s for children vs. teachers ranged from .15 to .19; parents vs. teachers ranged from .04 to .23), though all mean differences were relatively small in magnitude with one exception: differences in externalizing ratings across children and parents were especially pronounced in NICHD SECCYD Age 15 (d s ranged from .91 to .95; see for a review of cross-informant agreement). Mothers and fathers reported comparable levels of externalizing (d s ranged from .01 to .04). In contrast with externalizing, there were no differences in mean levels of internalizing across informants (d s ranged from .01 to .07, excluding NICHD SECCYD Age 15), with the exception that

⁶We also tested models with symptom-specific latent factors where each symptom was modeled as a latent factor using each informant's rating of that item as indicators. These latent item variables then served as indicators of higher-order externalizing or internalizing factors. That is, the substantive (trait) portions of the MTMM models reflected a higher-order structure with observed items, intermediate symptom-specific latent factors, and higher-order factors indicated by the symptom-specific factors, as opposed to a structure where items loaded directly onto the substantive factors. These models often failed to converge. Many of the item-specific latent factors' variances were negative, and models would not converge unless they were constrained to be positive, so we elected not to present these models here. Generally, they showed similar results to models presented here, with significant correlations between internalizing and externalizing that were weaker than the corresponding mono-method correlations. See the OSF page for model outputs.

children reported higher levels than mothers and fathers in NICHD SECCYD Age 15 (d_s ranged from .37 to .39).

Correlational agreement.—To examine convergent and discriminant associations among latent psychopathology factors and informants (*convergent*: same factor, different informants; *discriminant*: different factors, different informants), as well as the presence of method variance, we estimated confirmatory factor models with two factors (externalizing and internalizing) and for each informant (Table S7). We allowed all factors within and across informants to correlate and did not model method factors (see Figure S2). Homotrait-heteromethod correlations (i.e., interinformant agreement) were generally strong (externalizing range: from .40 to .76, $mdn = .49$; internalizing range: from .30 to .74, $mdn = .46$); homotrait-heteromethod correlations generally exceeded heterotrait-heteromethod correlations, which were small to moderate on average (range: $-.10$ to .39, $mdn r = .27$); and heterotrait-homomethod correlations ($mdn r = .62$) generally well exceeded heterotrait-heteromethod correlations ($mdn r = .27$).

Adjudicating Multi-informant Models

The CTCU, CTCM, and CTOM models fit the data well, whereas the M-1 models did not (Table S8). Because we tested several multi-informant models, we compared them in terms of whether they converged on the same general patterns of effects with respect to: (a) the latent externalizing-internalizing correlation and (b) the extent to which the general factor in a bifactor model was well-represented by its indicators. Doing so led us to rule out the CTCM models because they tended to yield relatively anomalous results compared with the other multi-informant models (Figure 2; see also Lance et al., 2002).

We also ruled out the M-1 models for two major reasons. First, they fit the data poorly in both a relative and absolute sense (based on CFI, TLI, RMSEA, and WRMR), particularly in comparison with their counterpart CTCM and CTOM models. Second, the contents of the general factors from the bifactor models shifted based on the informant.⁷ As is intended with M-1 models, substantive factors were defined by whichever informant was chosen as the referent informant and were generally underrepresented by the other informants. Considering that there is no clear *a priori* referent informant in child psychopathology data (De Los Reyes, 2011), we viewed such shifting in the contents of latent psychopathology dimensions as concerning.

This left us to consider the CTCU and CTOM models. These models are conceptually similar insofar as they do not allow for cross-method covariance. Both models yielded similar externalizing and internalizing factors (Φ_s ranged from .92 to 1.00), as well as general factors (Φ_s were 1.00 for all samples but HBN, which was .70). We elected to move forward with the CTOM model for three reasons. First, whether there is a clear reason to think the CTCU model is preferable to the CTOM depends on one's priors regarding

⁷The correlations between latent externalizing and internalizing factors in the M-1 models were comparable to those of the other MTMM models, with the exception of the CTCM model (see Table S8). Also, general factors from the M-1 models were not necessarily general, and instead were essentially defined by the items from the rater who was chosen as the referent (i.e., the rater whose method factor was dropped; Table S10).

the dimensionality of the rater bias of interest. CTOM models are most appropriate in this context given that evaluative consistency bias is conceptualized as unidimensional. Second, the CTOM model is conceptually most compatible with other widely used methods to model multi-informant youth psychopathology data (e.g., Bauer et al., 2013; Kraemer et al., 2003; Makol et al., 2020).⁸ Results for all multi-informant models are reported in Tables S8-11.

Correlated Externalizing and Internalizing Factors across Mono- and Multi-informant Structures

Figure 2 displays the latent externalizing and internalizing correlations across mono- and multi-informant models. In mono-informant models, externalizing and internalizing factors were strongly represented by their items (Table S9; median loadings on externalizing factors = .75, internalizing factors = .76). Externalizing and internalizing were moderately to highly correlated, but their correlation varied across informants and datasets (r s ranged from .37 [HBN teacher report] to .90 [NICHD SECCYD age 15 teacher report], median = .62). In CTOM models, externalizing and internalizing were again moderately to highly correlated, although the magnitude of the correlation was generally attenuated compared with those of the mono-informant models (r s ranged from .44 to .60, median = .50).

Whether the CTOM externalizing-internalizing correlation was attenuated relative to mono-informant correlations again depended on the dataset and informant. For instance, in ABCD and HBN, the CTOM externalizing-internalizing correlation (r s were .44 and .36, respectively) was significantly lower than the mono-method correlations of all three informants (ABCD r s ranged from .51 to .71; HBN r s ranged from .41 to .59). In NICHD SECCYD grade 6 and age 15, however, the CTOM externalizing-internalizing correlation (r s were .57 and .60, respectively) was only significantly lower than the correlation for one informant, the father (r s were .68 and .79). In these two datasets, the CTOM correlation was actually higher than those for other informants, such as teachers at grade 6 ($r = .48$) and children at age 15 ($r = .45$).

General Factors of Psychopathology across Mono- and Multi-informant Structures

Figure 3 displays density plots of general factor loadings across mono- and multi-informant models. Mono- and multi-informant bifactor models fit the data well (Table S8). In mono-informant models, general factors were strongly represented by their items (median λ s ranged from .50 [HBN parent] to .79 [ABCD teacher]), with no appreciable differences in loading magnitudes across externalizing and internalizing. Explained common variance (ECV) in each general factor was high, ranging from .53 (HBN, teacher) to .80 (father, NICHD SECCYD Age 15) across informants and samples (Table S11). In the CTOM models, loadings on the general factor were lower on average relative to the mono-informant models (median λ s ranged from .20 [HBN] to .54 [NICHD SECCYD grade 6]). In contrast with the mono-informant models, ECV for general factors in the CTOM models was low, ranging from .20 (HBN) to .46 (NICHD SECCYD Grade 6).

⁸The CTCU model fit slightly better than the CTOM model but this is not all that surprising because it is inherently more flexible. As we mentioned, there is no clear conceptual reason to think the CTCU model is conceptually superior to the CTOM model (Lance et al., 2002).

General factor loadings also varied as a function of psychopathology dimension such that they were well-represented by externalizing items (median λ s ranged from .25 [HBN] to .59 [NICHD SECCYD grade 6]) and underrepresented by internalizing items (median λ s ranged from .01 [HBN] to .34 [NICHD grade 6]). Thus, to summarize general factors as accounting for up to 46% of the variance in psychopathology is somewhat misleading. On average, 7 to 15% of the variance in internalizing items was explained by the general factors across samples, whereas 14 to 36% of the variance in externalizing items was explained (HBN was the outlier; the other samples hovered around 30%). There were two exceptions worth noting: The general factor in HBN tended was not represented by attention-problems items, and two internalizing items probing depression and feelings of worthlessness tended to load fairly strongly on the general factor (median λ across samples was .45).⁹

Discussion

On balance, we found mixed empirical support for a substantive p -factor when using a multi-informant design. The correlation between latent externalizing and internalizing factors remained significant but was generally lower in multi-informant models compared with mono-informant models (see also Clark et al., 2021). The differences in these correlations between mono- and multi-informant models were nevertheless modest. Also, though general factors were strong in mono-informant models, they were not necessarily “general” in the multi-informant models (Revelle & Wilt, 2013): they explained less than 47 percent of the variance in psychopathology indicators, most of which was attributable to externalizing as opposed to internalizing.

That our multi-informant general factors of psychopathology were not all that “general” and largely reflected externalizing is noteworthy because they are incompatible with most other published general factors of psychopathology. Across studies, general factors of psychopathology have tended to be more strongly indicated by thought problems¹⁰ or internalizing than by externalizing (see Watts et al., 2020, for a review). Nonetheless, general factors of psychopathology do not replicate very strongly across studies (ICC across 15 studies = .24; Watts et al., 2020), so the relative contribution of externalizing and internalizing to general factors probably also depends on the population examined and the measure being modeled (see also Hopwood et al., 2011). Given that different assessments can produce different general factors, it is especially noteworthy that our multi-informant general factors even differ from mono-informant models derived from the CBCL and its variants (e.g., Noordhof et al., 2015). Mono-informant general factors of psychopathology derived from the CBCL and its variants tend to be fairly similarly indicated by externalizing and internalizing, with some shifts in indicator representation

⁹Attention problems items across all informants did not strongly indicate the CTOM general factors in HBN, and child-reported attention problems did not strongly indicate the general factors in NICHD SECCYD Age 15. When we dropped attention problems items from all multi-informant models, our conclusion remained the same: general factors were strongly represented by externalizing but not internalizing (Table S10).

¹⁰A reasonable criticism of our study is that we did not model thought problems indicators, which tend to be strong markers of general factors of psychopathology (e.g., Caspi et al., 2014), though many other studies do not include them (e.g., Lahey et al., 2013). In subsidiary analyses, we modeled CTOM models using CBCL subscales in HBN and NICHD SECCYD. As Table S11 displays, Thought Problems tended not to strongly mark our general factors, again reinforcing that our multi-informant general factors are discrepant from most of the published literature.

across studies. Thus, the differences we observed between our mono- and multi-informant general factors are probably not exclusively caused by measurement or sampling differences.

Interpreting Multi-informant Structures of Psychopathology

Our multi-informant general factors of psychopathology are open to several interpretations, among which we cannot entirely adjudicate. As in the general personality literature, and like the modal researcher who uses MTMM modeling, we could interpret our findings strictly through the lens of MTMM assumptions (Campbell & Fiske, 1959; Garner, 1956) and conclude that only cross-informant factors contain the substantive variance of interest. In this interpretation, informant-specific factors are not considered to be relevant to the construct of interest, because they reflect largely error and various rater biases, such as evaluative consistency bias. From the perspective of these assumptions, the fact that we did not consistently find a truly general factor of psychopathology but did find that externalizing and internalizing factors were substantially correlated across informants somewhat indicates that general factors in mono-informant models are inflated by evaluative consistency and other biases.

But there are compelling challenges to that line of reasoning. As we noted earlier, method factors may represent valid incremental variance as well as bias (Connelly & Ones, 2010; Jackson et al., 2015). By definition, bias consists of systematic variation that is *irrelevant* to the construct assessed (Millsap, 2012). Nevertheless, an informant's report that disagrees with other informants' reports may nonetheless provide clinically useful information (Meehl, 1945). For instance, if a client perceives themselves as having a broad array of psychopathology, even if that is not substantiated by clinical interview or by their peers, it is probably an important indicator of the client's experience, perhaps reflective of real distress or impairment. We cannot necessarily assume that observed disagreements among informants should be dismissed as meaningless artifacts, given that they may harbor important clinical meaning. Individuals know things about themselves that others do not, and vice versa (Connelly & Ones, 2010).

In developmental psychopathology, the fact that different informants provide incremental validity appears to reflect systematic contextual variation in the expression of psychopathology (Achenbach et al., 2017; De Los Reyes, 2011; De Los Reyes & Makol, 2021). Therefore, our cross-informant factors in MTMM models of psychopathology may miss some important valid variance (c.f., Bauer et al., 2013; Campbell & Fiske, 1959) because they capture sources of variance *shared across contexts* and omit those specific to individual contexts. The fact that we did not find strong cross-informant general factors of psychopathology may indicate that there may be certain situations that evoke both externalizing and internalizing psychopathology, but that they do not jointly manifest across contexts in a consistent and systematic manner.

In our data, there are findings that bear directly on the question of whether contextual variation is likely to be present, namely the pattern of mono-informant correlations between internalizing and externalizing (Figure 2). Contextual variation is due to the fact that different informants see the target in different contexts. Nevertheless, one informant (the self) sees the target in all contexts. This implies that, to the extent that the nonshared

variance among the informants reflects real psychopathology variance, the correlation between internalizing and externalizing should be stronger in self-ratings than in parent and teacher ratings. This is the case in ABCD and HBN, but it is not the case in the NICHD SECCYD Age 15 data, where the self-ratings show a considerably lower correlation than the other informants' ratings. Thus, evidence regarding the contextual variance interpretation is ambiguous.

We are hesitant to chalk up all informant differences in psychopathology to contextual variation, not only because of the inconsistent pattern of mono-method correlations just described, but also for another reason. Heterotrait-homomethod correlations exceeded heterotrait-heteromethod correlations, in some cases quite dramatically. These effects were consistent across all pairs of informants, even including pairs who witness the child in largely overlapping contexts (i.e., mothers and fathers), and this seems to provide some evidence against the contextual specificity interpretation of unique informant variance. If it were the case that some aspects of psychopathology do not manifest in a given context for a systematic reason, there should be greater variation in terms of the magnitudes of heterotrait-homomethod correlations across pairs of raters. More generally, the heterotrait-homomethod correlations should in general be nearly as strong, as opposed to much stronger, than heterotrait-heteromethod correlations, which was not the case (median heterotrait-homomethod correlation: .62, heterotrait-heteromethod correlation: .27).

All told, it seems most likely that each informant factor includes both bias and unique valid variance. In contrast with common practice in the MTMM literature, interpreting *all* informant disagreement as bias or error may overlook an important part of the often complex clinical picture, especially when informants come from systematically different contexts. Unfortunately, informants and contexts are often confounded given that various informants observe children in different settings, so MTMM models cannot necessarily tease apart variation reflecting domain-relevant information from between-informant variation attributable to a variety of method artifacts. Ultimately, we cannot conclude exactly what amount of the variance in psychopathology ratings reflects evaluative consistency bias or contextual variation, but we suspect the true magnitudes of the correlation between externalizing and internalizing probably falls somewhere between those seen in mono-informant and multi-informant models.

Such competing interpretations of MTMM models are not fully appreciated in the psychopathology literature. As we mentioned earlier, only three (Oltmanns et al., 2018; Sallis et al., 2019; Snyder et al., 2017) of the thirteen studies that have extracted a multi-informant general factor of psychopathology have modeled informant-specific factors. This practice is associated with two related assumptions (that are either explicit or implicit), (1) that only cross-informant factors capture construct-relevant variance, and (2) that there is no systematic method variance in the psychopathology ratings, whether error or substance. Moreover, failing to include informant-specific factors in a MTMM model precludes our ability to examine their external validity and longitudinal stability, and it also precludes our ability to examine their incremental validity above and beyond cross-informant factors. Even when informant specific factors were modeled, only the cross-informant factors' correlates were examined such that the external validity of the informant-specific factors is overlooked

altogether (Oltmanns et al., 2018; Sallis et al., 2019; Snyder et al., 2019). Again, this assumes that only the cross-informant psychopathology factors capture meaningful variance.

Limitations and Future Directions

For the purposes of examining replication across various samples, this study benefitted from the use of several widely used datasets in the developmental psychopathology literature. We specifically selected datasets with identical measurement to rule out measurement as an explanation of any differences across samples that might arise. In terms of examining cross-informant structures, we further benefitted from the use of identical, dimensional assessments (though there may be subtle differences in assessment, such as the fact that items were framed in the first- versus third person for children and parents, respectively). Cross-informant correspondence is stronger when informants complete the same as opposed to different measures (Achenbach et al., 2005) and stronger when informants complete dimensional as opposed to categorical assessments (De Los Reyes et al., 2015). Finally, we tested an array of MTMM models, which revealed how they can yield different findings across studies.

Many researchers report only a single type of MTMM model, but which one we choose can lead us to dramatically different conclusions. For instance, we settled on the CTOM model, which yielded mixed evidence in terms of empirical support for the p -factor. In contrast, the CTCM model tended to yield results that were less favorable in terms of support of the p -factor, with latent correlations between externalizing and internalizing hovering around zero in most samples (in one sample, it was even slightly negative). Finally, various versions of CTCM-1 and CTOM-1 models yielded general factors that were inconsistent with each other, suggesting that contents of the general factor were strongly dependent on which informant was chosen as the referent informant. Although Eid and colleagues (2008) encouraged researchers to choose the referent informant on an a priori basis, there is no clearly preferred informant in the developmental psychopathology literature.

One major conclusion we drew from our replication effort is that our findings were not always consistent across sets of informants and samples. Notably, externalizing and internalizing were especially highly correlated for certain informants, but effects for informant type (e.g., child, parent) were not always consistent across samples. There are probably several explanations for the observed differences across samples, with one being that we relied on different sets of informants across samples. ABCD and HBN relied on ratings from children, parents, and teachers, whereas NICHD SECCYD relied on ratings from mothers, fathers, and teachers at grade 6 and children, mothers, and fathers at age 15. Findings for ABCD and HBN were generally consistent with each other, and findings across NICHD SECCYD waves were consistent with each other, so this seems like a reasonable explanation. In particular, there was stronger evidence for a multi-informant general factor of psychopathology in NICHD SECCYD compared with the other samples.

As to what is causing differences across these pairs of samples, we suspect the inclusion of two parents' ratings in NICHD may have an undue influence on the psychopathology structure for a couple of reasons. Informant convergence tends to be highest for mothers and fathers, as well as when informants observe the target in largely overlapping contexts

(De Los Reyes & Kazdin, 2005; De Los Reyes et al., 2015). Because parents observe their child in largely overlapping contexts and are likely to discuss their child's problems together, thereby aligning their perspectives even further, they are likely to share more variance in their ratings and therefore to induce a stronger general factor. When multiple informants, such as parents, are not only biased, but also share their bias, this will increase the general factor of psychopathology, especially if the correlation between the parents' method factors is not modeled.¹¹ If the goal of multi-informant assessment is to capture individuals across a wide range of contexts, it may be especially advantageous not to include both parents, or at least it would be wise to collect teacher reports, in addition to reports by parents and children. Advantages of teacher reports include that teachers probably have a better sense of what are normative youth behaviors given that they interact with a relatively wide array of youth. There are also advantages to including children's reports of their own psychopathology, including they may have greater access to their subjective experiences, such as the negative emotions associated with internalizing.

Age differences across the samples are probably unlikely to cause the differences we observed. For instance, there were differences between ABCD and NICHD SECCYD grade 6, but these samples were comparable in terms of mean age and its variance. Similarly, HBN and NICHD SECCYD age 15 were reasonably comparable in terms of mean age and its variance. Should age differences explain sample differences, it would imply substantive differences in psychopathology across development such that its dimensionality varies as a function of age. A handful of studies have examined longitudinal stability of psychopathology structures across time and have not supported this possibility (Greene & Eaton, 2017; Mann et al., 2020; Murray et al., 2016; Olino et al., 2018). Although small differences in psychopathology structure may occur across development, these differences do not appear to reflect weakening of a general factor of psychopathology across time (e.g., Olino et al., 2018).

Another lingering question is whether our findings would generalize to adult samples. Cross-informant correlations among adult psychopathology assessments are not substantially higher than those observed for youth (e.g., Achenbach et al., 2005; Klonsky et al., 2002). In fact, parents' correlations tend to be higher than any typical pair of informants on adults. Nonetheless, there is a relative paucity of data with three or more psychopathology ratings on the same target adult (e.g., Achenbach et al., 2005). Further, if one wanted to study contextual variation in adults, it would be challenging to find good analogs for parents and teachers. Perhaps one could use friends and co-workers, but this may not provide a particularly strong parallel given the variety of adult contexts. There is also a relative paucity of validation of multi-informant structures of psychopathology with reasonably objective external criteria (e.g., job performance, romantic partner interactions, physical health), which is necessary to adjudicate the degree to which informant method factors reflect evaluative consistency bias versus substantive contextual variation (see also

¹¹In post hoc exploratory analyses, we tested a model in NICHD SECCYD that allowed for correlated parent method factors but no other correlated method factors (e.g., mother-child; Table S8). The externalizing-internalizing correlation did not decrease significantly compared with that of the CTOM model, but the variance in general factor loadings in the bifactor model did increase slightly, suggesting that the general factor tended to weaken. All told, allowing for correlated method factors among parents did not change our conclusions.

De Los Reyes et al., 2015). Because informants are strategically selected to observe children across multiple contexts, typical assessment methods used in developmental psychopathology harbor the potential to produce disagreements among sources, perhaps more so than in the adult personality and psychopathology literatures (see also Kraemer et al., 2003).

In our youth samples, the lack of independent criterion variables (e.g., observed behavior, school records) is perhaps the most significant limitation of this study. This limitation prevented us from further investigating the meaning of cross-informant and informant-specific factors in the MTMM models. Earlier, we reviewed several studies from the developmental psychopathology literature that use sophisticated and thorough construct validation of multi-informant assessment across modalities (see De Los Reyes & Makol, 2021, for a review). Generally, these types of studies collect multi-modal data, such as behavior within various tasks (e.g., peer social interaction, classroom observation, parent-child discussion), and then compare task performance (e.g., safety behaviors with peers, avoidance of public speaking in the classroom, and confidence during parent-child conflict) with self- and informant-reported psychopathology (e.g., social anxiety). Reported psychopathology may be further compared with other external criteria, such as school performance, treatment outcomes, and so on.

Critical to these efforts is the identification of external criteria that should vary across situations, as well as those that should not. If an informant-specific factor reflects genuine contextual variation, it should be associated with external criteria especially salient to the context in which a particular observer witnesses the child, and other informant-specific factors should not be associated with this same set of external criteria. Likewise, if a construct is thought *not* to vary across contexts, measures of it should be associated with cross-informant but not single-informant factors. This type of validation work is potentially the only method to adjudicate between competing interpretations of informant-specific factors in MTMM models.

A major downside to MTMM modeling is that it requires considerable sample sizes, which often hinders validation work with indicators from modalities other than informants' reports, like laboratory task performance. To capture the scope of a broad range of psychopathology, such research would require numerous multi-modal indicators of psychopathology symptoms and relevant behaviors. Typically, samples have relatively high depth or resolution in terms of participants or modalities, but not both. The present study is perhaps the most comprehensive evaluation of the MTMM metastructure of youth psychopathology, which is an important first step in evaluating their validity, utility, and continuity across development, but we look forward to future MTMM modeling that incorporates independent indicators as external validators of both cross-informant and informant-specific factors.

Conclusion

We designed this study to contribute to ongoing discussion surrounding the structure of psychopathology, the interpretation of published general factors of psychopathology, and the

nature of the theoretical p -factor. Strong empirical evidence for the p -factor did not arise in multi-informant models that take into account the perspectives of multiple individuals who know the target well. MTMM modeling is ultimately useful to partition sources of variance in psychopathology data, but the conclusions we draw about the nature of each factor require careful consideration. There are at least two competing interpretations of our findings, and both may be in play simultaneously. Our mono-informant general factors of psychopathology may contain at least some variance from evaluative consistency bias. Likewise, the informant-specific factors in our models probably contain some variance due to evaluative consistency bias, and they probably also contain variance reflecting genuine contextual variation in psychopathology. Either way, externalizing and internalizing do not appear to consistently form a coherent, p -factor-like dimension across informants and contexts. Future work should incorporate carefully chosen external validators to adjudicate the competing interpretations of informant-specific factors in MTMM models of psychopathology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

ALW is funded through K99AA028306 (Principal Investigator: Watts). We thank Christopher Hopwood and two other reviewers for their feedback on this paper.

References

- Achenbach TM (2020). Bottom-Up and Top-Down Paradigms for Psychopathology: A Half-Century Odyssey. *Annual Review of Clinical Psychology*, 16(1), 1–24. 10.1146/annurev-clinpsy-071119-115831
- Achenbach TM, Ivanova MY, & Rescorla LA (2017). Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. *Comprehensive Psychiatry*, 79, 4–18. 10.1016/j.comppsy.2017.03.006 [PubMed: 28356192]
- Achenbach TM, Krukowski RA, Dumenci L, & Ivanova MY (2005). Assessment of Adult Psychopathology: Meta-Analyses and Implications of Cross-Informant Correlations. *Psychological Bulletin*, 131(3), 361–382. 10.1037/0033-2909.131.3.361 [PubMed: 15869333]
- Achenbach TM, McConaughy SH, & Howell CT (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101(2), 213–232. 10.1037/0033-2909.101.2.213 [PubMed: 3562706]
- Achenbach TM (2009). *The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory and Applications*.
- Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, Vega-Potler N, ...Milham MP (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, 4(1), 170181. 10.1038/sdata.2017.181
- Allegrini AG, Cheesman R, Rimfeld K, Selzam S, Pingault J-B, Eley TC, & Plomin R (2020). The p factor: Genetic analyses support a general dimension of psychopathology in childhood and adolescence. *Journal of Child Psychology and Psychiatry*, 61(1), 30–39. 10.1111/jcpp.13113 [PubMed: 31541466]
- Anusic I, Schimmack U, Pinkus RT, & Lockwood P (2009). The nature and structure of correlations among Big Five ratings: The halo-alpha-beta model. *Journal of Personality and Social Psychology*, 97(6), 1142–1156. 10.1037/a0017159 [PubMed: 19968424]

- Bäckström M, & Björklund F (2021). Is the general factor of personality really related to frequency of agreeable, conscientious, emotionally stable, extraverted, and open behavior? An experience sampling study. *Journal of Individual Differences*, No Pagination Specified-No Pagination Specified. 10.1027/1614-0001/a000341
- Barkley RA (2003). Issues in the diagnosis of attention-deficit/hyperactivity disorder in children. *Brain and Development*, 25(2), 77–83. 10.1016/S0387-7604(02)00152-3 [PubMed: 12581803]
- Bauer DJ, Howard AL, Baldasaro RE, Curran PJ, Hussong AM, Chassin L, & Zucker RA (2013). A trifactor model for integrating ratings across multiple informants. *Psychological Methods*, 18(4), 475–493. 10.1037/a0032475 [PubMed: 24079932]
- Becker-Haimes EM, Jensen-Doss A, Birmaher B, Kendall PC, & Ginsburg GS (2018). Parent–youth informant disagreement: Implications for youth anxiety treatment. *Clinical Child Psychology and Psychiatry*, 23(1), 42–56. 10.1177/1359104516689586 [PubMed: 28191794]
- Biesanz JC, & West SG (2004). Towards understanding assessments of the Big Five: Multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality*, 72(4), 845–876. 10.1111/j.0022-3506.2004.00282.x [PubMed: 15210019]
- Bonifay W, Lane SP, & Reise SP (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184–186. 10.1177/2167702616657069
- Brandes CM, Herzhoff K, Smack AJ, & Tackett JL (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. 10.1177/2167702619859332
- Campbell DT, & Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. 10.1037/h0046016 [PubMed: 13634291]
- Cannon CJ, Makol BA, Keeley LM, Qasmieh N, Okuno H, Racz SJ, & De Los Reyes A (2020). A paradigm for understanding adolescent social anxiety with unfamiliar peers: Conceptual foundations and directions for future research. *Clinical Child and Family Psychology Review*, 23(3), 338–364. 10.1007/s10567-020-00314-4 [PubMed: 32140896]
- Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, Meier MH, ... & Moffitt TE (2014). The p Factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2), 119–137. 10.1177/2167702613497473 [PubMed: 25360393]
- Caspi A, & Moffitt TE (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, 175(9), 831–844. 10.1176/appi.ajp.2018.17121383 [PubMed: 29621902]
- Castellanos-Ryan N, Brière FN, O’Leary-Barrett M, Banaschewski T, Bokde A, Bromberg U, Büchel C, ... & Conrod P (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology*, 125(8), 1039–1052. 10.1037/abn0000193 [PubMed: 27819466]
- Chang L, Connelly BS, & Geeza AA (2012). Separating method factors and higher order traits of the Big Five: A meta-analytic multitrait–multimethod approach. *Journal of Personality and Social Psychology*, 102(2), 408–426. 10.1037/a0025559 [PubMed: 21967007]
- Clark DA, Hicks BM, Angstadt M, Rutherford S, Taxali A, Hyde L, Weigard AS, ... & 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. 10.1177/2167702620959317 [PubMed: 34621600]
- Connelly BS, & Ones DS (2010). An other perspective on personality: Meta-analytic integration of observers’ accuracy and predictive validity. *Psychological Bulletin*, 136(6), 1092–1122. 10.1037/a0021212 [PubMed: 21038940]
- De Los Reyes A (2011). Introduction to the Special Section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 40(1), 1–9. 10.1080/15374416.2011.533405 [PubMed: 21229439]

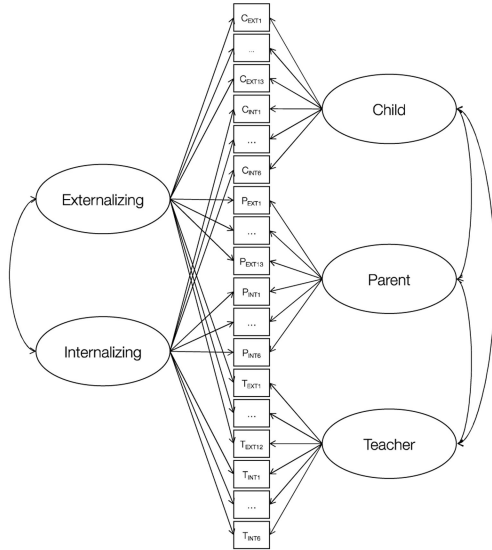
- De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DAG, Burgers DE, & Rabinowitz J (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858–900. 10.1037/a0038498 [PubMed: 25915035]
- De Los Reyes A, Henry DB, Tolan PH, & Wakschlag LS (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, 37(5), 637–652. 10.1007/s10802-009-9307-3 [PubMed: 19247829]
- De Los Reyes A, Lerner MD, Keeley LM, Weber RJ, Drabick DAG, Rabinowitz J, & Goodman KL (2019). Improving interpretability of subjective assessments about psychological phenomena: A review and cross-cultural meta-analysis. *Review of General Psychology*, 23(3), 293–319. 10.1177/1089268019837645
- De Los Reyes A, Thomas SA, Goodman KL, & Kundery SMA (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, 9(1), 123–149. 10.1146/annurev-clinpsy-050212-185617
- DeYoung CG (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, 91(6), 1138–1151. 10.1037/0022-3514.91.6.1138 [PubMed: 17144770]
- Digman JM (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, 73(6), 1246–1256. 10.1037/0022-3514.73.6.1246 [PubMed: 9418278]
- Eid M, Lischetzke T, Nussbeck FW, & Trierweiler LI (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model. *Psychological Methods*, 8(1), 38–60. 10.1037/1082-989X.8.1.38 [PubMed: 12741672]
- Eid M, Nussbeck FW, Geiser C, Cole DA, Gollwitzer M, & Lischetzke T (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, 13(3), 230–253. 10.1037/a0013219 [PubMed: 18778153]
- Feeley TH (2002). Comment on halo effects in rating and evaluation research. *Human Communication Research*, 28(4), 578–586. 10.1111/j.1468-2958.2002.tb00825.x
- Forbes MK, Greene AL, Levin-Aspenson HF, Watts AL, Hallquist M, Lahey BB, Markon KE, ... & Krueger RF (2021). 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. 10.1037/abn0000533 [PubMed: 33539117]
- Forgas JP, & Laham SM (2017). Halo effects. In *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory*, 2nd ed (pp. 276–290). Routledge/Taylor & Francis Group.
- Garner WR, Hake HW, & Eriksen CW (1956). Operationism and the concept of perception. *Psychological Review*, 63(3), 149–159. 10.1037/h0042992 [PubMed: 13323170]
- Geiser C, & Lockhart G (2012). A comparison of four approaches to account for method effects in latent state-trait analyses. *Psychological Methods*, 17(2), 255–283. 10.1037/a0026977 [PubMed: 22309958]
- Glenn LE, Keeley LM, Szollos S, Okuno H, Wang X, Rausch E, Deros DE, ... & De Los Reyes A (2019). Trained observers' ratings of adolescents' social anxiety and social skills within controlled, cross-contextual social interactions with unfamiliar peer confederates. *Journal of Psychopathology and Behavioral Assessment*, 41(1), 1–15. 10.1007/s10862-018-9676-4
- Greene AL, & Eaton NR (2017). The temporal stability of the bifactor model of comorbidity: An examination of moderated continuity pathways. *Comprehensive Psychiatry*, 72, 74–82. 10.1016/j.comppsy.2016.09.010 [PubMed: 27764677]
- Hamlat EJ, Snyder HR, Young JF, & Hankin BL (2019). Pubertal timing as a transdiagnostic risk for psychopathology in youth. *Clinical Psychological Science*, 7(3), 411–429. 10.1177/2167702618810518 [PubMed: 31179165]
- Hopwood CJ, Wright AGC, & Donnellan BM (2011). Evaluating the evidence for the general factor of personality across multiple inventories. *Journal of Research in Personality*, 45(5), 468–478. 10.1016/j.jrp.2011.06.002 [PubMed: 22879686]
- Jackson JJ, Connolly JJ, Garrison SM, Leveille MM, & Connolly SL (2015). Your friends know how long you will live: A 75-year study of peer-rated personality traits. *Psychological Science*, 26(3), 335–340. 10.1177/0956797614561800 [PubMed: 25583945]

- Jenness JL, Peverill M, Miller AB, Heleniak C, Robertson MM, Sambrook KA, Sheridan MA, & McLaughlin KA (2020). Alterations in neural circuits underlying emotion regulation following child maltreatment: A mechanism underlying trauma-related psychopathology. *Psychological Medicine*, 1–10. 10.1017/S0033291720000641
- Jernigan TL, Brown SA, & Dowling GJ (2018). The Adolescent Brain Cognitive Development Study. *Journal of Research on Adolescence: The Official Journal of the Society for Research on Adolescence*, 28(1), 154–156. 10.1111/jora.12374 [PubMed: 29460352]
- Klonsky ED, Oltmanns TF, & Turkheimer E (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice*, 9(3), 300–311. 10.1093/clipsy.9.3.300
- Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, & Kupfer DJ (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. 10.1176/appi.ajp.160.9.1566 [PubMed: 12944328]
- Krueger RF, & Tackett JL (2003). Personality and psychopathology: Working toward the bigger picture. *Journal of Personality Disorders*, 17(2), 109–128. 10.1521/pedi.17.2.109.23986 [PubMed: 12755325]
- Laceulle OM, Vollebergh WAM, & 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. 10.1177/2167702614560750
- Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, & Rathouz PJ (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology*, 121(4), 971–977. 10.1037/a0028355 [PubMed: 22845652]
- Lahey BB, Moore TM, Kaczkurkin AN, & Zald DH (2021). Hierarchical models of psychopathology: Empirical support, implications, and remaining issues. *World Psychiatry*, 20(1), 57–63. 10.1002/wps.20824 [PubMed: 33432749]
- Lahey BB, Zald DH, Perkins SF, Villalta-Gil V, Werts KB, Van Hulle CA, Rathouz PJ, ... & Waldman ID (2018). Measuring the hierarchical general factor model of psychopathology in young adults. *International Journal of Methods in Psychiatric Research*, 27(1), e1593. 10.1002/mpr.1593
- Lance CE, Noble CL, & Scullen SE (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, 228–244. [PubMed: 12090412]
- Leising D, Burger J, Zimmermann J, Bäckström M, Oltmanns JR, & Connelly BS (2020). Why do items correlate with one another? A conceptual analysis with relevance for general factors and network models [Preprint]. *PsyArXiv*. 10.31234/osf.io/7c895
- Lenth RV (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means (1.5.5–1) [Computer software].
- Lippold MA, Greenberg MT, & Collins LM (2013). Parental knowledge and youth risky behavior: A person oriented approach. *Journal of Youth and Adolescence*, 42(11), 1732–1744. 10.1007/s10964-012-9893-1 [PubMed: 23269564]
- Littlefield AK, Lane SP, Gette JA, Watts AL, & Sher KJ (2021). The “Big Everything”: Integrating and investigating dimensional models of psychopathology, personality, personality pathology, and cognitive functioning. *Personality Disorders: Theory, Research, and Treatment*, 12(2), 103–114. 10.1037/per0000457
- Makol BA, De Los Reyes A, Garrido E, Harlaar N, & Taussig H (2021). Assessing the mental health of maltreated youth with child welfare involvement using multi-informant reports. *Child Psychiatry & Human Development*, 52(1), 49–62. 10.1007/s10578-020-00985-8 [PubMed: 32279125]
- Makol BA, Youngstrom EA, Racz SJ, Qasmieh N, Glenn LE, & 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. 10.1177/2167702620924439
- Mann FD, Atherton OE, DeYoung CG, Krueger RF, & Robins RW (2020). Big five personality traits and common mental disorders within a hierarchical taxonomy of psychopathology: A

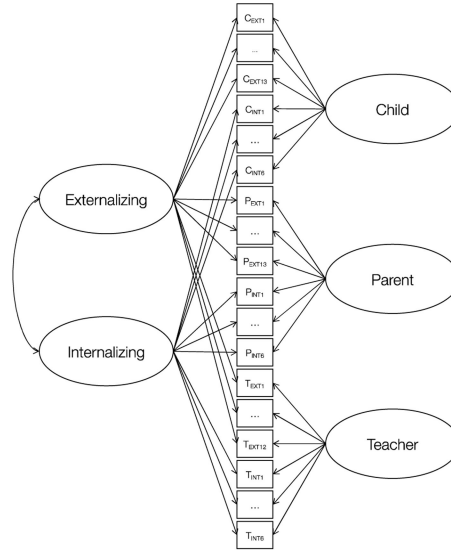
- longitudinal study of Mexican-origin youth. *Journal of Abnormal Psychology*, 129(8), 769–787. 10.1037/abn0000633 [PubMed: 32969675]
- Meehl PE (1945). The dynamics of “structured” personality tests. *Journal of Clinical Psychology*, 1, 296–303.
- Millsap RE (2012). *Statistical approaches to measurement invariance*. Routledge.
- Murray AL, 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. 10.1007/s10802-016-0132-1 [PubMed: 26846993]
- Musek J (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality*, 41(6), 1213–1233. 10.1016/j.jrp.2007.02.003
- Muthén LK and Muthén BO (1998). *Mplus User’s Guide (Eighth Edition)* [Computer software]. Muthén & Muthén.
- Neumann A, Pappa I, Lahey BB, Verhulst FC, Medina-Gomez C, Jaddoe VW, Bakermans-Kranenburg MJ, ... & Tiemeier H (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. 10.1016/j.jaac.2016.09.498
- NICHD Early Child Care Research Network (Ed.). (2005). *Child care and child development: Results from the NICHD study of early child care and youth development*. Guilford Press.
- Noordhof A, Krueger RF, Ormel J, Oldehinkel AJ, & Hartman CA (2015). Integrating Autism-Related Symptoms into the Dimensional Internalizing and Externalizing Model of Psychopathology. The TRAILS Study. *Journal of Abnormal Child Psychology*, 43(3), 577–587. 10.1007/s10802-014-9923-4 [PubMed: 25099360]
- Olino TM, Bufferd SJ, Dougherty LR, Dyson MW, Carlson GA, & Klein DN (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. 10.1007/s10802-018-0398-6 [PubMed: 29359267]
- Oltmanns JR, Smith GT, Oltmanns TF, & Widiger TA (2018). General factors of psychopathology, personality, and personality disorder: Across domain comparisons. *Clinical Psychological Science*, 6(4), 581–589. 10.1177/2167702617750150 [PubMed: 30221082]
- Revelle W, & Wilt J (2013). The general factor of personality: A general critique. *Journal of Research in Personality*, 47(5), 493–504. 10.1016/j.jrp.2013.04.012 [PubMed: 23956474]
- Riem MME, van Hoof MJ, Garrett AS, Rombouts SARB, van der Wee NJA, van IJzendoorn MH, & Vermeiren RRJM (2019). General psychopathology factor and unresolved-disorganized attachment uniquely correlated to white matter integrity using diffusion tensor imaging. *Behavioural Brain Research*, 359, 1–8. 10.1016/j.bbr.2018.10.014 [PubMed: 30315842]
- Rushton P, & Irwing P (2011). The general factor of personality: Normal and abnormal. In *The Wiley-Blackwell handbook of individual differences* (pp. 132–161). Wiley Blackwell.
- Sallis H, Szekely E, Neumann A, Jolicoeur-Martineau A, IJzendoorn M van Hillegers M, Greenwood CMT, ... & 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. 10.1111/jcpp.13067 [PubMed: 31049953]
- Smith GT, Atkinson EA, Davis HA, Riley EN, & Oltmanns JR (2020). The general factor of psychopathology. *Annual Review of Clinical Psychology*, 16(1), 75–98. 10.1146/annurev-clinpsy-071119-115848
- Snyder HR, Young JF, & Hankin BL (2017). Strong homotypic continuity in common psychopathology-, internalizing-, and externalizing-specific factors over time in adolescents. *Clinical Psychological Science*, 5(1), 98–110. 10.1177/2167702616651076 [PubMed: 28239532]
- Suzuki T, Samuel DB, Pahlen S, & Krueger RF (2015). DSM-5 alternative personality disorder model traits as maladaptive extreme variants of the five-factor model: An item-response theory analysis. *Journal of Abnormal Psychology*, 124(2), 343–354. 10.1037/abn0000035 [PubMed: 25665165]
- Tackett JL, Herzhoff K, Reardon KW, De Clercq B, & Sharp C (2014). The externalizing spectrum in youth: Incorporating personality pathology. *Journal of Adolescence*, 37(5), 659–668. 10.1016/j.adolescence.2013.10.009 [PubMed: 24263222]

- van der Linden D, Pekaar KA, Bakker AB, Schermer JA, Vernon PA, Dunkel CS, & Petrides KV (2017). Overlap between the general factor of personality and emotional intelligence: A meta-analysis. *Psychological Bulletin*, 143(1), 36–52. 10.1037/bul0000078 [PubMed: 27841449]
- Watts AL, Lane SP, Bonifay W, Steinley D, & Meyer FAC (2020). Building theories on top of, and not independent of, statistical models: The case of the p-factor. *Psychological Inquiry*, 31(4), 310–320. 10.1080/1047840X.2020.1853476 [PubMed: 33510565]
- Watts AL, Poore HE, & Waldman ID (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, 7(6), 1285–1303. 10.1177/2167702619855035
- Weissman DG, Bitran D, Miller AB, Schaefer JD, Sheridan MA, & McLaughlin KA (2019). Difficulties with emotion regulation as a transdiagnostic mechanism linking child maltreatment with the emergence of psychopathology. *Development and Psychopathology*, 31(3), 899–915. 10.1017/S0954579419000348 [PubMed: 30957738]
- Widiger TA, Sellbom M, Chmielewski M, Clark LA, DeYoung CG, Kotov R, Krueger RF, ... & Wright AGC (2019). Personality in a Hierarchical Model of Psychopathology. *Clinical Psychological Science*, 7(1), 77–92. 10.1177/2167702618797105
- Wood S, & Scheipl F (2020). Generalized Additive Mixed Models using “mgcv” and “lme4” (0.2–6) [Computer software].

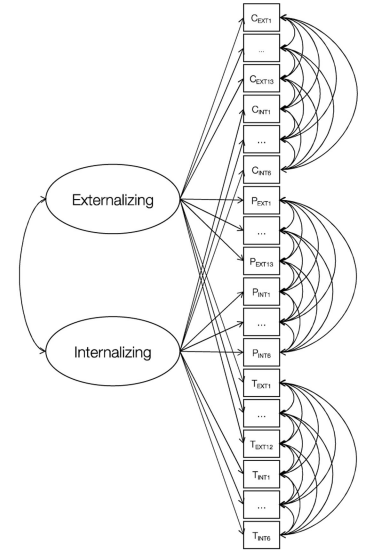
a. Correlated traits, correlated methods model.



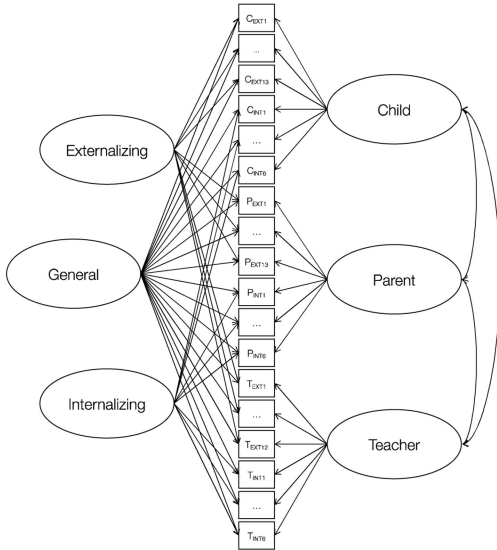
b. Correlated traits, orthogonal methods model.



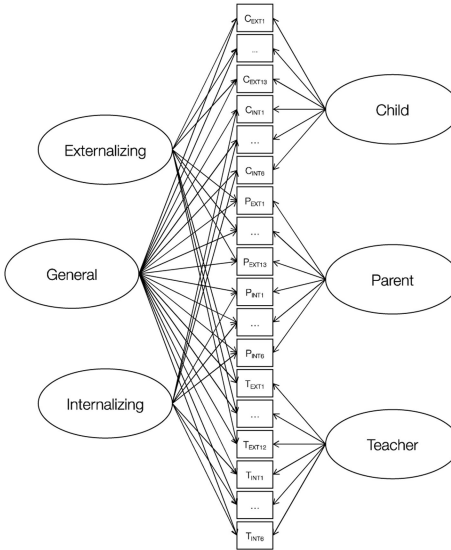
c. Correlated traits, correlated uniquenesses model.



d. Bifactor correlated methods model.



e. Bifactor orthogonal methods model.



f. Bifactor correlated uniquenesses model.

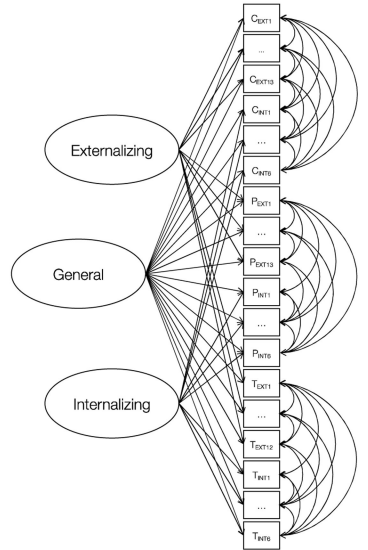


Figure 1. Example schematics of multi-informant models tested.

Covariances between method factors are represented by dashed lines. They are retained in the correlated traits/correlated method model and are dropped in the correlated traits/orthogonal methods model (and the same goes for the bifactor models with and without correlated methods).

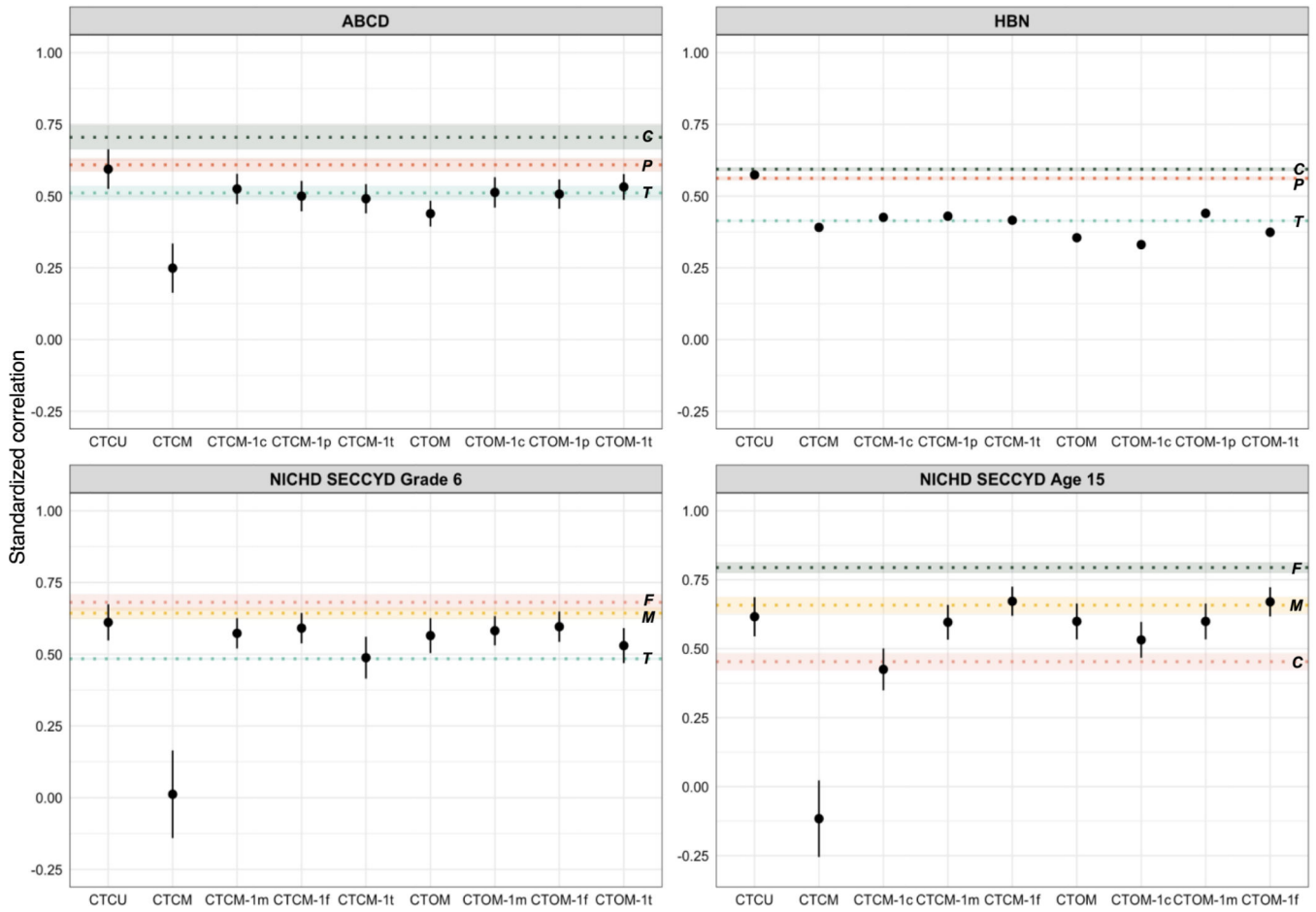


Figure 2. Correlations between latent externalizing and internalizing factors across mono- and multi-informant models.

All correlations are displayed with standard error bars. Multi-informant correlations are displayed as points, whereas mono-informant correlations are displayed as dotted reference lines. CTCU=correlated traits, correlated uniquenesses; CTCM=correlated traits, correlated methods; CTOM=correlated traits, orthogonal methods; c=child; p=parent; m=mother; f=father; t=teacher.

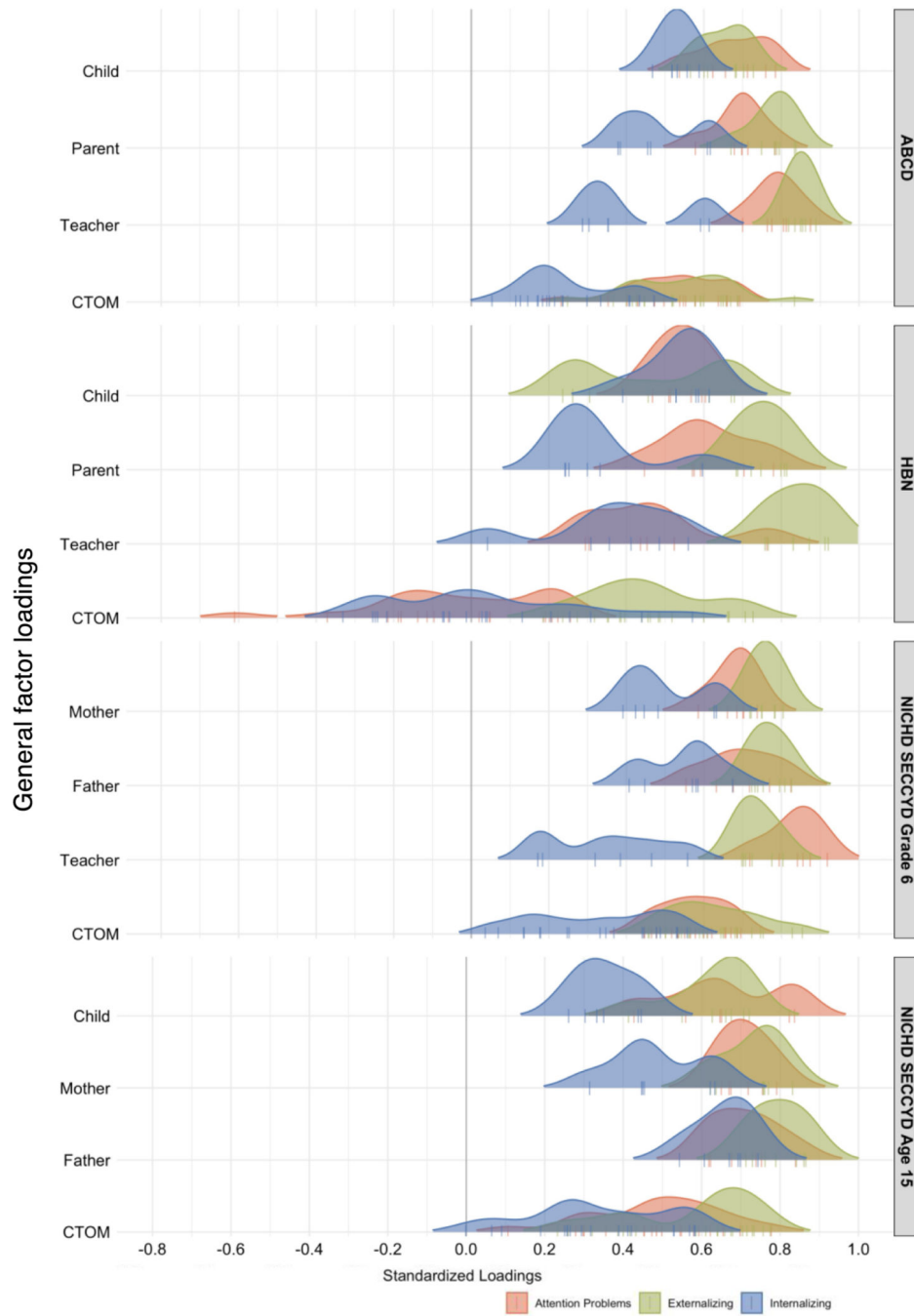


Figure 3. Density plots of loadings on general factors across mono- and multi-informant models. CTOM = general factor loadings from the correlated traits-orthogonal methods model; all other sets of loadings are from their respective mono-informant model.