Published in final edited form as:

Clin Psychol Sci. 2021 March ; 9(2): 169–182. doi:10.1177/2167702620959317.

The General Factor of Psychopathology in the Adolescent Brain Cognitive Development (ABCD) Study: A Comparison of Alternative Modeling Approaches

D. Angus Clark, Brian M. Hicks, Mike Angstadt, Saige Rutherford, Aman Taxali, Luke Hyde, Alexander Weigard, Mary M. Heitzeg, Chandra Sripada

University of Michigan

Abstract

Many models of psychopathology include a single general factor of psychopathology (GFP) or “*p* factor” to account for covariation across symptoms. The Adolescent Brain Cognitive Development (ABCD) Study provides a rich opportunity to study the development of the GFP. However, a variety of approaches for modeling the GFP have emerged, raising questions about how modeling choices impact estimated GFP scores. We used the ABCD baseline assessment (ages 9–10 years-old; $N=11,875$) of the parent-rated Child Behavior Checklist (CBCL) to examine the implications of modeling the GFP using items versus scales; using a priori CBCL scales versus data-driven dimensions; and using bifactor, higher-order, or single-factor models. Children’s rank-ordering on the GFP was stable across models, with GFP scores similarly related to criterion variables. Results suggest that while theoretical debates about modeling the GFP continue, the practical implications of these choices for rank-ordering children and assessing external associations will often be modest.

Keywords

General Factor of Psychopathology; *p* factor; Adolescent Brain Cognitive Development Study (ABCD); Child Behavior Checklist (CBCL); Bifactor Model

Introduction

There are well-known concerns about the *Diagnostic and Statistical Manual of Mental Disorders Fifth Ed. (DSM-5)* categorical model of psychiatric nosology, including high rates of comorbidity (Kessler et al., 2005; Krueger & Markon, 2006; Angold, Costello, & Erkanli, 1999), serious symptoms that do not meet the diagnostic thresholds for any individual disorder (Lewinsohn, Shankman, Gau, & Klein, 2004; Rucci et al., 2003), and treatments with efficacy across disorder categories (Barlow et al., 2017; Newby, McKinnon, Kuyken,

Correspondence: Address correspondence to D. Angus Clark, Department of Psychiatry, Addiction Center, University of Michigan, Rachel Upjohn Building, 4250 Plymouth Rd., Ann Arbor, MI 48109. Electronic mail may be sent to cladaavid@med.umich.edu.

Authorship: DAC, BMH, and CS developed the concept of this manuscript. Data collection and preparation was performed by MA, SR, AT, MMH, and CS. DAC conducted the analyses and drafted the paper along with CS. BMH, LH, AW, and MMH provided critical revisions. All authors agree to the author listing and approved the final version of this paper for submission.

Conflict of Interest: The authors declare no conflicts of interest.

Gilbody, & Dalgleish, 2015; Wong, Yocca, Smith, & Lee, 2010). Alongside these concerns, evidence is growing for an alternative, hierarchical conceptualization of psychopathology. A common feature of these models is a single general factor that accounts for variance across psychiatric symptoms. We refer to this single, broad factor as the general factor of psychopathology (GFP; though it is also referred to as the “*p* factor”) (Caspi et al., 2014; Caspi & Moffitt, 2018). The GFP has been identified in multiple datasets, and predicts several clinically-relevant outcomes (Caspi et al., 2014, 2017; Caspi & Moffitt, 2018; Castellanos-Ryan et al., 2016; Lahey et al., 2012; Patalay et al., 2015; Pettersson, Lahey, Larsson, Lundstroem, & Lichtenstein, 2015; Stochl et al., 2015; Waldman, Poore, van Hulle, Rathouz, & Lahey, 2016). However, the factors that shape the development of the GFP—including its genetic, environmental, and neurobiological bases—remain poorly understood.

The Adolescent Brain Cognitive Development (ABCD) Study is a longitudinal study of over 11,000 youth across 21 sites (Volkow et al., 2018) that provides a seminal opportunity to investigate the development of the GFP (Micheline et al., 2019). The ABCD study includes biennial assessments over 10 years beginning at 9–10 years-old that measure psychopathology, neurocognitive abilities, personality, family environment, neighborhood characteristics, substance use, genetics, and structural/functional brain development. ABCD thus represents a unique chance to delineate how genetic, socio-demographic, and neural factors interact to influence the emergence and manifestation of the GFP in youth.

Taking full advantage of the opportunities afforded by ABCD to study the GFP, however, requires methodological rigor, with numerous questions about how to best model the GFP needing to be addressed (van Bork, Epskamp, Rhemtulla, & Borsboom, 2017; Watts, Poore, & Waldman, 2019). One question focuses on the widespread use of bifactor models to estimate GFP scores. Bifactor models include a general factor that all indicator variables load on, and orthogonal specific factors that only subsets of the indicators load on (see <https://osf.io/b8f6d/> for graphic depiction of the bifactor model). The general factor captures non-specific risk for all forms of psychopathology, with specific factors capturing residual clusters of risk for a narrower range of psychiatric symptoms (e.g., internalizing or externalizing problems). Although useful, bifactor models exhibit high fit propensity (Bonifay & Cai, 2017) and as consequence are more likely produce sample-specific results -- especially in the specific factors -- that do not reliably replicate (Markon, 2019; Rodriguez, Reise, & Haviland, 2016; Watts et al., 2019).

Higher-order factor models are an alternative approach to modeling the GFP (Watts et al., 2019). Higher-order models posit an explicitly hierarchical structure wherein observed indicators load on a series of lower order factors, which in turn load on a single higher-order factor (see <https://osf.io/b8f6d/> for graphic depiction of the high-order factor model). The higher-order factor captures covariation between the narrower facets of psychiatric symptoms. The key conceptual difference between the bifactor and higher-order models is the narrower sub-factors. In the bifactor model, specific factors capture residual covariation after accounting for the GFP, while in the higher-order model lower-order factors capture the unresidualized covariation between indicators. These approaches each have theoretical justifications, with disagreements ongoing about which is to be preferred for specific applications. However, the practical implications of choosing a higher-order versus a bifactor

model for rank-ordering youth on the GFP (as opposed to answering questions about specific facets of psychopathology) are unclear.

Another question concerns what observed variables to use in GFP models: item-level data (or symptoms) or item-level data that has been aggregated into scales (or diagnoses). Item-level inputs avoid the need to leverage (possibly false) prior theory in the modeling process, and allow more fine-tuning of model content. However, item-level models are computationally more demanding, more prone to estimation problems, and can be more difficult to interpret (D. A. Clark & Bowles, 2018; Wirth & Edwards, 2007). Both item-level (Lahey et al., 2012; Caspi et al., 2014) and scale-level inputs (Lahey et al., 2015; Castellanos-Ryan et al., 2016; Patalay et al., 2015) have been used to study the GFP, but comparisons between approaches are missing, particularly in large-scale, well-sampled studies.

A third question concerns the number and nature of sub-factors to include. That is, whether to use the existing instrument structure as a basis for sub-factor specification, or take a data-driven approach to identify the optimal covariance structure in the sample data. A priori structures usually have a stronger conceptual basis and facilitate comparability across studies, but may not be supported by dimensionality analyses (Kotelnikova, Olino, Klein, Kryski, & Hayden, 2016), which could undermine substantive conclusions. A more bottom-up, data-driven approach based on exploratory factor analysis will identify a dimensional structure that fits the data better, but the structure may be conceptually ambiguous and fail to generalize across samples (D. A. Clark & Bowles, 2018; Henson & Roberts, 2006; Preacher & MacCallum, 2003; Wirth & Edwards, 2007).

The variety of approaches for modelling the GFP increases the “researcher degrees of freedom” for investigating the GFP in ABCD, which raises concerns about how these modeling choices impact scores on the GFP, as well the sub-factors (Gelman & Loken, 2013; Simmons, Nelson, & Simonsohn, 2011). To address these concerns, we fit and compared a series of models in ABCD that varied along the dimensions described above. We used the parent-reported Child Behavior Checklist (Achenbach & Ruffle, 2000) (CBCL)—a widely used measure of child psychopathology—as the basis for GFP model construction (Micheline et al., 2019). We also examined associations between the different GFP factors as well as internalizing and externalizing factor scores, and associations between these factors and a set of external criterion variables.

Methods

Participants

The sample consisted of youth from the ABCD study ($N = 11,875$), a large-scale study of youth aged 9 to 10 years, recruited from 21 research sites across the United States (Barch et al., 2018; Garavan et al., 2018; Volkow et al., 2018). These data were collected from baseline visits between September 1, 2016 and November 15, 2018. The ABCD data repository grows and changes over time; data used in this report came from ABCD Release 2.01, DOI: 10.15154/1504041. We removed 5 participants from the analyses owing to complete nonresponse on the CBCL, leaving a final sample size of 11,869 individuals.

Institutional review board approval was obtained for each site before data collection. All parents provided written informed consent and all children provided assent.

The sample was roughly gender balanced (48.4% female) with a mean age of 9.92 years ($SD = 0.62$ years). Around half (51.6%) of the sample was White, with the remaining participants identifying themselves as Hispanic (20.0%), Black (14.8%), Other/Multi-racial (10.0%), and Asian (2.10%) (data on ethnicity was missing for 1.50% of the sample). Two-thirds of youth (66.6%) came from households in which the parents were married. Most parents (84.3%) reported at least some college, and most households reported an annual income of at least \$50,000 (63.6%).

Child Behavior Checklist

The Child Behavior Checklist (CBCL; age 6 to 18 form) (Achenbach & Ruffle, 2000) contains 119 items in which parents are presented with a particular behavior (e.g., “Destroys others’ things”) and rate the extent to which that behavior was characteristic of their child over the past 6 months. Parents can select one of three response options: 0 (“not true”), 1 (“somewhat or sometimes true”), or 2 (“very true or often true”). Items are combined to create eight composite scale scores: Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior. The Withdrawn, Somatic Complaints, and Anxious/Depressed scales are all conceptualized as pieces of a broader Internalizing Problems composite, while the Delinquent Behavior and Aggressive Behavior scales are conceptualized as pieces of a broader Externalizing Problems composite. Thirty-eight items are not associated with any of the a-priori scales and are classified as “Other Problems.” Three items about substance use were excluded from all major analyses due to extremely low ($n < 5$) rates of endorsement.

Criterion Variables

Several outcome variables tapping different domains of functioning were selected from the ABCD protocol and correlated with GFP, internalizing, and externalizing factor scores. These variables were chosen to capture both alternative means of assessing psychopathology in youth (i.e., different raters and/or inventories), as well as nomologic correlates of psychopathology (e.g., school performance, general cognitive ability). That is, this selection of outcome variables allows the consideration of both convergence in the assessment of youth psychopathology across informants and methods, and the extent to which GFP scores are associated with distinct constructs that are relevant in the study of youth psychopathology. Furthermore, several of the criterion were chosen to be consistent with another recent investigation into the functioning of certain GFP models in ABCD in order to build a broader pool of data about how model choice affects the associations between the GFP and other variables in ABCD (Moore et al., 2020).

Alternative Measures of Psychopathology—First, observed CBCL Total Problems scores were included; two variants of the Total Problems score were computed, the first by averaging together all of the CBCL’s items, and the second by averaging together all of the CBCL’s standard eight scales. These composites effectively reflect a measurement model in which there is a single factor and all indicators are weighted equally. The

Kiddie-Structured Assessment for Affective Disorders and Schizophrenia (KSADS) is a structured, diagnostic interview administered to parents via computer in reference to their child (Kaufman, Birmaher, Brent, & Rao, 1997). The *Brief Problem Monitoring Form* (BPM) is an abbreviated, 18 item version of the CBCL administered to teachers (Piper, Gray, Raber, & Birkett, 2014). Both the KSADS and BPM total problems scores were used here (average factor loadings of .73 and .76, respectively, in unidimensional item response models). The KSADS and BPM total problems scores overlap conceptually with the CBCL total problems scores and GFP scores, but unlike the rating-form based CBCL the KSADS is a structured, clinical interview (also completed by parents), while the BPM is similar to CBCL in structure and content but is completed by teachers instead of parents. Finally, we included a single item from the ABCD protocol about if youth *have ever needed to access mental health services* (yes/no). This is a concise reflection of if youth have levels of psychopathology high enough to warrant potential intervention, and by asking about a single, clearly observable outcome it is less reliant on the idiosyncrasies of any one informant (Clark, Durbin, Hicks, Iacono, & McGue, 2017).

Trait Impulsivity—The 20-item youth short version of the UPPS-P scale (Cyders et al., 2007), which was specifically developed for use in the ABCD study (Barch et al., 2018), was administered via self-report at baseline to index trait impulsivity. The UPPS-P has five subscales: *Negative Urgency*, *Lack of Planning*, *Lack of Perseverance*, and *Sensation Seeking*. The UPPS-P provides a series of more normative trait dimensions that are distinct from, but conceptually related to, psychopathology. Furthermore, all UPPS-P ratings were completed by the youth themselves, which provides another informant perspective.

Cognitive Ability—*General Cognitive Ability* (GCA) scores were computed by fitting a bifactor model to behavioral tasks from the NIH toolbox, the Rey Auditory Verbal Learning Task, the WISC-V, and the “Little Man” task (see Sripada, Angstadt, Rutherford, & Taxali, 2019 for more details on the GCA variables). Exploratory factor analyses suggested that 3 broad factors characterized these tasks, corresponding to memory, speed/flexibility, and reasoning; in the bifactor model these three factors served as the specific factors. The GCA bifactor model fit well by conventional standards: $\chi^2 = 443.16$, $df = 34$, $p < .001$; RMSEA = .032; CFI = .990; TLI = .983; SRMR = .017 (West, Taylor, & Wu, 2012). GCA factor scores (specifically, factor scores for the general factor in the bifactor model) were generated from this model using maximum a posteriori (MAP) scoring (MacCallum, 2009).

Academic Functioning—Selected items from the ABCD Longitudinal Parent Diagnostic Interview for DSM-5 Background were also included that asked parents to rate *how well in general their child does in school*, *what kinds of grades they get on average*, *if they receive any special services at school* (yes/no), and *if they have ever had any detentions or suspensions in the past school year assigned* (yes/no). The first two items tap into parents’ holistic sense of how they see their child functioning in an academic setting, while the second two items provide clearer, observable indicators of potential academic difficulty.

Data Analytic Strategy

Since our aim was to compare alternative approaches to modeling the GFP with the parent-rated CBCL, a preliminary step was to understand the CBCL's general psychometric properties in ABCD. We applied item factor analysis (IFA; Wirth & Edwards, 2007) to the eight canonical CBCL subscales to assess intra-scale dimensionality, and graded item response models (Samejima, 1969) to assess the psychometric properties of the individual CBCL scales. The empirical factor structure of the full CBCL was then evaluated using IFA for the item level, and exploratory factor analysis (EFA) for the scale level. The graded response models helped assess the psychometric quality of the individual scales, which were used as indicators in several models, while the IFAs and EFAs provided a set of bottom-up, sub-factor structures at both the item and scale levels for the model comparisons in the primary analyses.

We then fit 14 distinct GFP models (Table 1). These models varied as a function of whether the GFP was modeled using item or scale level data, whether sub-factors were based on the standard CBCL structure or our IFAs/EFAs, and whether the model form followed a bifactor, a higher-order, or a simple single-factor (all indicators loading on the GFP without sub-factors) structure (see <https://osf.io/b8f6d/> for a graphic depiction of all three model forms). In item level bifactor models all items were included in the model, while in item level higher-order models only the items associated with the relevant sub-factors were included. Thus, in the item level models a slightly larger, more diverse set of items is included in the bifactor models versus the higher-order models (see <https://osf.io/b8f6d/>).

Several pieces of information were computed or collected to characterize the different GFP factors and models. First, the mean standardized factor loading on the GFP (and mean standardized factor loading across all model sub-factors) were computed; all individual factor loadings across models can be found in the supplemental material (<https://osf.io/b8f6d/>). Second, the mean indicator-level explained common variance (I-ECV) was computed for bifactor model based GFPs (Hansen, Cai, Stucky, & Tucker, 2014; Rodriguez et al., 2016). The I-ECV captures the percentage of common variance in an indicator explained by the general versus specific factor. Third, the RMSEA from every model was reported, as the RMSEA was the only consistently provided fit index across estimators and programs (however, given the variation in model structures, programs, and estimators, direct comparisons of fit across models should be made cautiously; these statistics are provided primarily as a reference, not as a way to select one model over another).

Finally, two indicators of latent factor reliability were computed for each GFP factor, the H index of construct replicability and Coefficient Omega (ω) (Rodriguez et al., 2016; Watts et al., 2019). The H index tracks how well a latent factor should replicate across studies, while ω captures the proportion of total score variance in a set of indicators explained by a latent factor, or set of latent factors. Although the standard ω was computed for single-factor and higher-order model derived GFPs (with computations treating the lower-order factors as the full set of indicators, and the GFP as the sole explanatory latent factor), for bifactor model derived GFPs Coefficient Omega Hierarchical (ω_h ; Rodriguez et al., 2016) was computed. This variation specifically captures the proportion of total score variance in the indicators captured by the general factor. Notably, each metric of factor reliability is ultimately a

function of the factor loadings in the model such that factors with several moderate to strong loadings, and/or a large number of indicators, will tend to look stronger (Rodriguez et al., 2016). As the models considered here vary widely in size and structure, this is an important caveat when interpreting factor reliability across models. Furthermore, as noted, summary information regarding factor loadings is also presented, with all individual factor loadings presented online (<https://osf.io/b8f6d/>).

As most GFP models (8 of 14) contained internalizing and externalizing sub-factors, we also examined the robustness of internalizing and externalizing scores across these models. For lower-order factors, the standard ω was computed. For specific factors of bifactor models, Coefficient Omega Hierarchical Subscale was used (ω_h ; Rodriguez et al., 2016), which captures the proportion of total score variance captured by a specific factor among the set of indicators associated with that specific factor (e.g., among internalizing items, what proportion of variance is explained by the internalizing specific factor as opposed to the general factor?). Importantly, the lower-order factor scores from higher-order factor models contain both variance attributable to the GFP as well as the unique variance of the lower-order factor. Therefore, to increase comparability across bifactor and higher-order factor models, we included lower-order factor scores from the higher-order factor models after residualizing the lower-order factors of GFP-related variance (the original lower-order factor scores were also examined). The lower-order factor scores residualized of GFP variance also closely approximate the role of lower-order factors when used as predictors in multiple regression models that also include GFP scores. For researchers interested in how factor scores predict external variables, the lower-order factors residualized of GFP variance are thus of particular interest.

Preliminary item response models, along with the item-level single-factor and bifactor models were estimated using flexMIRT and maximum likelihood estimation (Cai, Maydeu-Olivares, Coffman, & Thissen, 2006). All other measurement models were estimated in Mplus version 8.1 using weighted least squares with mean and variance adjustment (intra-scale IFAs/higher-order factor models), unweighted least squares (full form IFA), or full information maximum likelihood estimation (scale level EFA/single-factor/bifactor/higher-order models) (Muthén & Muthén, 2020). Factor scores were generated using either expected a posteriori (EAP) (all of the item level models estimated in flexMIRT) or maximum a posteriori (MAP) (all of the item and scale level models estimated in Mplus) scoring (MacCallum, 2009).

Correlations between the factor scores, and between the factor scores and outcome variables, were estimated in Mplus version 8.1 using either weighted least squares with mean and variance adjustment (for correlations between the factor scores and categorical outcome variables), or full information maximum likelihood estimation (for correlations between the factor scores, and between the factor scores and continuous outcome variables). Confidence intervals for the correlations between the factor scores and outcome variables were derived via clustered (by family and study site), non-parametric, percentile bootstrapping (with 1,000 draws), which performs well under a variety of data conditions (Falk, 2018). These analyses were also facilitated by the Mplus Automation Package (Hallquist & Wiley, 2018) in R (R Core Team, 2019).

Results

Preliminary Analyses

Results for the preliminary psychometric analysis of the CBCL in ABCD can be found in the online supplemental material (<https://osf.io/b8f6d/>). Overall, the individual, canonical CBCL scales were unidimensional and reliable (especially for children with more problems). The scree plots from the CBCL-wide IFA, and EFA of the scales, each suggested the presence of a strong first factor (<https://osf.io/b8f6d/>). All of the factor loadings from the most coherent IFA and EFA solutions can be found in the online supplemental material (<https://osf.io/b8f6d/>). At the item level, factor structures were relatively coherent until the 5-factor model. The factors that reliably emerged corresponded to internalizing problems, externalizing problems, attention problems, and somatic problems. At the scale level, a two-factor solution was most coherent, with the delinquent and aggressive behavior scales loading onto a single externalizing factor.

Fourteen Distinct GFP Models

The distinguishing characteristics of the 14 different GFP models are summarized in Table 1, with a fuller narrative description provided in the supplemental material (<https://osf.io/b8f6d/>). The mean standardized factor loadings on the GFP and model sub-factors are also presented in Table 1, along with the mean indicator-level explained common variance (I-ECV) and RMSEAs. Latent factor reliability indices and correlations between GFP factor scores are presented in Table 2. Associations between the GFP scores and the outcome variables are presented in Table 3.

Factor loadings on the GFP were consistently strong (mean $\lambda = .74$), indicating a high degree of shared variance across the CBCL items and scales. On average, factor loadings on the bifactor model specific factors were small to moderate (mean $\lambda = .34$), while factor loadings on the lower-order factors of the higher-order factor model were moderate to strong (mean $\lambda = .70$). I-ECV values tended to be high across the bifactor models (mean I-ECV of 75%), suggesting that the GFP accounted for the majority of the shared variance across indicators. With the exception of GFP-10 and GFP-14, all models demonstrated adequate fit on the basis of conventional fit thresholds for the RMSEA (i.e., RMSEA $\leq .08$).

Consistent with the generally strong factor loadings reported in Table 1, all GFPs demonstrated a high degree of reliability (mean $H = .93$; mean $\omega = .92$; mean $\omega_h = .90$). Higher order models, and scale level models, tended to have relatively lower reliability, which in part reflects the smaller number of indicators loading on the GFP itself (see Table 2). The GFP factor scores were all strongly intercorrelated (mean $r = .92$). The smallest correlations were observed between the item-level and scale-level models, but none of these associations fell below $r = .85$.

All GFP scores were strongly associated with the observed CBCL total problems scores (mean $r = .92$). Beyond this, the GFP scores were moderately associated with the convergent measures of psychopathology from other instruments and raters (mean $r = .49$), weakly associated with the facets of trait impulsivity (mean $r = .12$) and general cognitive ability (mean $r = -.16$), and moderately predictive of school performance (mean $r = .32$).

Associations with the criterion variables were highly consistent across models, with the standard deviation for correlation coefficients across GFP scores ranging from below .01 (Negative Urgency and Sensation Seeking) to .05 (CBCL scale based total score) (mean SD across outcome variables = .03).

Internalizing and Externalizing Sub Factors

GFP models 3, 4, 5, 7, 8, 9, 11, and 13 all contained clear internalizing and externalizing sub-factors. The other models either did not include any sub-factors (GFP-1 and 10), included sub-factors that were too narrow (GFPs 2 and 6), or included sub-factors that did not clearly represent internalizing and externalizing (GFPs 12 and 14, which included an externalizing and an “all other problems” factor). Factor reliability indices, as well as correlations between the internalizing factor scores, and externalizing factor scores, are presented in Table 4. Associations between the internalizing factor scores, externalizing factor scores, and all of the outcome variables can be found in the online supplement (<https://osf.io/b8f6d/>).

All of the lower order factors from higher-order factor models demonstrated a high degree of reliability, while evidence was more mixed for the bifactor model specific factors. Item level specific factors appeared reliable according to H , but no bifactor specific factors appeared reliable on the basis of ω_{hs} . This discrepancy is likely due to the higher number of indicators in the item level models, and the strength of the general factor, which enters into the computation of ω_{hs} , but not H . The bifactor and higher-order factor model internalizing factor scores (mean $r = .78$) and externalizing factor scores (mean $r = .72$) were strongly correlated across models after the lower order factor scores were residualized of GFP variance; associations were only moderately sized based on the original lower order factors (mean r s of .46 and .45 for internalizing and externalizing, respectively).

Across the bifactor specific factors and GFP residualized lower-order factors, associations with the outcome variables tended to be small in magnitude (around $r \sim .10 - .20$) but consistent across models, with an average standard deviation in correlation coefficients across models and outcomes of .04 for the internalizing factor scores, and .05 for the externalizing factor scores. Associations between the original lower-order factor scores and the outcome variables were more consistent with what was observed with the GFP factor scores in terms of effect size, direction, and estimate precision, which is expected since these lower order factor scores are saturated with GFP variance.

Discussion

The ABCD longitudinal study of over 11,000 youth presents a unique opportunity to investigate the development and etiology of the GFP. Taking full advantage of this opportunity, however, requires thoughtfulness in modeling the GFP. Here, we conducted a systematic, side-by-side comparison of 14 different approaches to constructing GFPs in the ABCD dataset. Results support two main conclusions. First, regardless of the features that distinguished our 14 GFP models, GFP factor scores were highly inter-correlated, and were similarly predictive of external variables in ABCD. Second, across GFP models, internalizing and externalizing factors were also fairly stable and similarly predictive of

external variables. Overall, results suggest that participants' standings on the GFP, as well as internalizing and externalizing dimensions, are robust across many alternative model specifications.

CBCL GFP Factors

Fourteen models that included a latent factor representing the GFP were estimated that varied as a function of whether the GFP was modeled at the item or scale level, whether the sub-factors were based on the standard CBCL structure or IFAs/EFAs, and whether the model was consistent with a bifactor, a higher-order, or single-factor structure. Indicator variables generally loaded strongly on the GFP factor, and consequently GFP factors tended to demonstrate a high degree of reliability based on the indices considered. Although bifactor and item level GFPs tended to demonstrate the highest reliability (Table 2), as noted above, GFP factors with more indicators will often tend to look more reliable based on the formulas used here (Rodriguez et al., 2016). Although the higher-order and scale-level GFPs had fewer indicators, factor loadings were still generally strong (Table 1), indicating coherent, reliable factors.

Relatedly, models tended to fit well based on the conventional thresholds for the RMSEA. As noted, comparisons across models should be made with caution given the differences in indicators and estimators used. However, a somewhat cleaner comparison is possible between GFPs 10, 12, and 14 as these are all scale level models (single factor, bifactor, and higher order, respectively) drawing on the same set of indicators. Fit was similar across these models, but as has often been observed, the bifactor model had the best fit. The RMSEA includes a penalty for model complexity, and so in a sense should favor the more parsimonious higher-order models (even if a higher-order model is effectively a more constrained bifactor model). Bifactor models are exceptionally effective at fitting data, however, thus potentially overriding this preference (Bonifay & Cai, 2017). Furthermore, there is evidence that the RMSEA is also sensitive to a variety of data and model features (e.g., sample size, non-normality, strength of factor loadings), and will struggle to differentiate models when data are categorical (Clark & Bowles, 2018). These model properties, and the known limitations of model fit thresholds, help highlight how fit statistics alone do not necessarily provide a strong basis to decisively prefer one approach to modeling the GFP over another.

The rank-ordering of children's estimated GFP scores was remarkably stable across models. That is, youth scoring high on the GFP factor in one model (e.g., a scale level single-factor model) were very likely to score high on GFP in other models. Correspondingly, the GFP factor scores were similarly correlated with the criterion variables. The strong associations between the GFP factor scores and the observed CBCL total problems scores were particularly useful, highlighting the fact that ultimately GFP factor scores function as differentially weighted total scores, and that the rank ordering of youth demonstrating a substantial number of problems is generally robust to the specifics of how those scores are derived.

Our results with the outcome variables also dovetail with those of other recent investigations into how specific modeling choices (e.g., bifactor versus high-order factor models) affect

the associations between the GFP and other variables (Forbes et al., 2020; Moore et al., 2020). That is, the consistency of GFP scores across the models observed here suggests that even specific operationalizations of the GFP not considered in the present investigation should lead to similar conclusions, which is what has been found. For example, several of the external outcome variables included here were also considered in the recent investigation of the GFP in ABCD by Moore and colleagues (2020). Despite differences in methods, all of the GFP scores across both papers were similarly related to the external variables. This speaks to the robustness of the general trends demonstrated here in terms of the consistency of GFP factor score estimates across modeling choices, and in the associations between the GFP and outcome variables in ABCD.

Internalizing and Externalizing Dimensions

As expected, the unadjusted associations between internalizing and externalizing factors across bifactor and higher-order factor models were moderate. However, when we residualized the internalizing and externalizing lower-order factor scores of variance from the GFP, making them more directly comparable to specific factors from bifactor models, the rank-ordering of the children was much more stable across all eight models. The factor scores residualized of GFP variance were also similarly associated to the criterion variables. This highlights the close statistical relatedness between bifactor and higher-order models, reinforcing how the practical implications of the conceptual differences between these models and their sub-factors should not be overstated (Rijmen, 2010). Lower-order and specific factors are unique statistical entities with distinct interpretations, making meaningful direct comparisons between bifactor and higher-order model sub-factors difficult. However, when these factors are made more conceptually analogous the differences between models are much less notable.

This is especially relevant as one enduring concern about bifactor models specifically is that their specific factors are unstable and less trustworthy than the lower-order factors in higher-order models. Factor reliability for the bifactor sub-factors was indeed mixed (Table 3), with much of the heterogeneity due to the specific characteristics of the individual indices. The core issue with specific factors is that after the general factor is extracted there is less common variance to parse and so the loadings on specific factors are smaller, and thus—as is the case for relatively smaller effect sizes broadly, holding all else equal—are estimated with less precision. Furthermore, smaller factor loadings imply that indicators are less informative for discriminating between individuals at different levels of the latent trait. Thus, parameters associated with specific factors in a bifactor model will tend to be estimated less precisely within a given model, and the specific factor indicators will typically be less effective at making fine grained distinctions between people. Indeed, in a sample as large as ABCD most parameters here were estimated precisely, but the specific factor loadings were estimated with larger standard errors (compared to the general factor), and the confidence intervals for the criterion associations for the bifactor specific factors and GFP-residualized lower-order factors were wider than those observed for the GFP factor scores and the original lower order factor scores.

This does not imply that bifactor specific factor scores cannot or should not be used, only that they should be used judiciously with these caveats in mind. Furthermore, when lower-order factors are made conceptually equivalent to bifactor specific factors they function similarly to bifactor specific factors. In other words, if scientific interest is specifically in the unique variance associated with internalizing and externalizing problems (i.e., variance that does not overlap with the GFP), bifactor and higher-order models lead to similar conclusions with similarly precise (or imprecise) estimates.

Which GFP Model is Best?

Results inform recent questions about how best to construct GFP models. Existing work has employed different model types, different numbers of sub-factors, and different inputs for model estimation. Overall, our results suggest that when interest is in the core dimensions of psychopathology, specifically the GFP and internalizing and externalizing dimensions, the effect of modeling choice should be relatively modest. Across a variety of models, factor scores for the GFP, internalizing, and externalizing dimensions are highly inter-correlated and are associated similarly with external criterion variables of interest. The small differences among these models in factor scores for GFP, internalizing, and externalizing are unlikely to have much practical significance. Thus, our analyses provide researchers interested in GFP modeling—particularly within the context of the ABCD study—with a “menu” of plausible GFP models. This compliments the recent results of Michelini and colleagues (2019). They provided researchers with a number of useful factor structures for psychopathology at different levels of abstraction, while the current results highlight that once researchers have selected a conceptually useful factor structure, there is considerable flexibility in terms of how to derive scores and rank-order children on the selected dimensions of interest.

Still, each of the different approaches we considered, even if they overall yield broadly similar GFP factors and internalizing and externalizing dimensions, may nonetheless still have certain advantages for different situations or for addressing certain questions. As noted earlier, fit statistics alone do not provide a strong basis to decisively prefer one approach over another. Thus, researchers may prefer different GFP modeling specifications based on their own prior theoretical commitments or research questions. For example, while internalizing and externalizing dimensions were present in eight of the 13 models we studied, many of these models included additional specific factors unique to those models. Additionally, four models lacked internalizing and/or externalizing dimensions, with single factor models lacking facets altogether (though indicator specific residuals may be viewed as playing analogous roles). And as noted above, the type of facet researchers are interested in (e.g., GFP residualized or saturated) is also related to model choice.

Similarly, the choice of modelling GFPs with item versus scale-level inputs will in part depend on researchers' interests and questions. Item-level inputs avoid the need to rely on an established dimensional structure, allow considerable flexibility in fine-tuning models, and tend to appear more reliable due to the larger number of indicators. However, item-level models are more likely to encounter estimation problems and often present interpretational difficulties (D. A. Clark & Bowles, 2018; Wirth & Edwards, 2007). In the present analyses,

the IFAs and item-based GFP models did not provide meaningfully different representations of the GFP compared to the models based on CBCL subscales—despite some differences in factor reliability—and the latter were computationally simpler and easier to interpret. Notably, this held despite the fact that at the item level, the standard CBCL scale structure did not clearly emerge. That is, although the empirical factor structure of the CBCL in the ABCD sample diverged from the standard organization of the form, the a priori scale structure was nevertheless practically useful and psychometrically defensible, and should be adequate for modeling the GFP in other applications (though the present results regarding the utility of scale level modeling likely also extend to empirically-derived composites).

Limitations and Future Directions

This study has several limitations. First, we focused on the parent-rated CBCL for modeling GFPs. This choice was guided by the widespread use of the CBCL in the developmental psychopathology literature. Furthermore, other plausible instruments that are part of the ABCD dataset lacked coverage across the full sample at the current time. As these additional instruments become available in the coming years, future studies can jointly model GFPs in a multiple rater framework. Another limitation is that we focused on effects of modeling choices on GFPs and internalizing and externalizing dimensions specifically. Future studies in ABCD should examine the psychometric properties of other more specific facets that emerge from alternative modeling choices, such as the attention and somatization, including their stability in held out samples and predictive validity with respect to major criterion variables.

Other important future directions concern the development of the GFP and related sub-factors across adolescence. The current analyses were all cross-sectional and at age 9 – 10 and thus may not generalize to other ages or developmental periods. There is evidence in adults though that GFP scores derived via different methods are stable over time (Forbes et al., 2020). Given the amount of rapid change in childhood and adolescence it is critical to understand how the assessment of the GFP changes across time, and interacts with general maturational trends. The nature of the latent variable models considered, and high-degree of correspondence across GFP factors, suggest that similar trends across time would be identified across models, but this leaves open the possibility of important changes and trends within models (e.g., measurement invariance over time, increasing or decreasing absolute levels of psychopathology). Relatedly, one strength of ABCD is the inclusion of several hundred twin pairs, which allows biometric twin modeling to explore the genetic and environmental origins of the GFP, and the origins of phenotypic associations like those observed here. Again, the extremely high correlations across GFP scores imply that conclusions would be similar across the types of models considered here, though this is an important empirical question.

Conclusion

Overall, results point to the substantial robustness of a “core structure” of psychopathology as measured by the CBCL—encompassing the GFP along with internalizing/externalizing dimensions—that emerges across a number of different models that are specified in different ways. By showcasing a variety of methods for calculating GFP scores in ABCD, these

findings lay a foundation for future investigations into the development of psychopathology in early life and beyond.

Acknowledgments

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This 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.

This work was supported by the following grants from the United States National Institutes of Health, the National Institute on Drug Abuse, and the National Institute on Alcohol Abuse and Alcoholism: R01MH107741 (CS), U01DA041106 (CS, LH, MH, BH), T32 AA007477 (FB). In addition, CS was supported by a grant from the Dana Foundation David Mahoney Neuroimaging Program. This research was supported in part through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor.

References

- Achenbach TM, & Ruffle TM (2000). The child behavior checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review*, 21, 265–271. [PubMed: 10922023]
- Angold A, Costello EJ, & Erkanli A (1999). Comorbidity. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40, 57–87.
- Barch DM, Albaugh MD, Avenevoli S, Chang L, Clark DB, Glantz MD, ... Yurgelun-Todd D (2018). Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Developmental Cognitive Neuroscience*, 32, 55–66. [PubMed: 29113758]
- Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, ... Boswell JF (2017). The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: A randomized clinical trial. *JAMA Psychiatry*, 74, 875–884. [PubMed: 28768327]
- Bonifay W, & Cai L (2017). On the Complexity of Item Response Theory Models. *MULTIVARIATE BEHAVIORAL RESEARCH*, 52, 465–484. [PubMed: 28426237]
- Cai L, Maydeu-Olivares A, Coffman DL, & Thissen D (2006). Limited-information goodness-of-fit testing of item response theory models for sparse 2P tables. *British Journal of Mathematical and Statistical Psychology*, 59, 173–194.
- Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, ... Poulton R (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2, 119–137. [PubMed: 25360393]
- Caspi A, Houts RM, Belsky DW, Harrington H, Hogan S, Ramrakha S, ... Moffitt TE (2017). Childhood forecasting of a small segment of the population with large economic burden. *Nature Human Behaviour*, 1, 0005.
- Caspi A, & Moffitt TE (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, 175, 831–844.
- Castellanos-Ryan N, Brière FN, O'Leary-Barrett M, Banaschewski T, Bokde A, Bromberg U, ... Gallinat J (2016). The structure of psychopathology in adolescence and its common personality and cognitive correlates. *Journal of Abnormal Psychology*, 125, 1039. [PubMed: 27819466]

- Clark DA, & Bowles RP (2018). Model Fit and Item Factor Analysis: Overfactoring, Underfactoring, and a Program to Guide Interpretation. *MULTIVARIATE BEHAVIORAL RESEARCH*, 53, 544–558. [PubMed: 29683723]
- Clark D, Angus, Durbin CE, Hicks BM, Iacono WG, & McGue M (2017). Personality in the age of industry: Structure, heritability, and correlates of personality in middle childhood from the perspective of parents, teachers, and children. *Journal of Research in Personality*, 67, 132–143. [PubMed: 28408770]
- Cyders MA, Smith GT, Spillane NS, Fischer S, Annus AM, & Peterson C (2007). Integration of impulsivity and positive mood to predict risky behavior: Development and validation of a measure of positive urgency. *Psychological Assessment*, 19, 107. [PubMed: 17371126]
- Falk CF (2018). Are robust standard errors the best approach for interval estimation with nonnormal data in structural equation modeling? *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 244–266.
- Forbes MK, Greene AL, Levin-Aspenson HF, Watts AL, Hallquist M, Lahey BB, ... Waldman ID. (2020). A Comparison of the Reliability and Validity of the Predominant Models Used in Research on the Empirical Structure of Psychopathology.
- Garavan H, Bartsch H, Conway K, Decastro A, Goldstein RZ, Heeringa S, ... Zahs D (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience*, 32, 16–22. [PubMed: 29703560]
- Gelman A, & Loken E (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Department of Statistics, Columbia University.
- Hallquist MN, & Wiley JF (2018). MplusAutomation: An R package for facilitating large-scale latent variable analyses in M plus. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 621–638.
- Hansen M, Cai L, Stucky BD, & Tucker JS (2014). Methodology for developing and evaluating the PROMIS® smoking item banks. *Nicotine and Tobacco Research*, 16, S175–S189. [PubMed: 23943843]
- Henson RK, & Roberts JK (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, 66, 393–416.
- Kaufman J, Birmaher B, Brent D, & Rao U (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 980–988. [PubMed: 9204677]
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, & Walters EE (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 593–602. [PubMed: 15939837]
- Kotelnikova Y, Olino TM, Klein DN, Kryski KR, & Hayden EP (2016). Higher-and lower-order factor analyses of the Children’s Behavior Questionnaire in early and middle childhood. *Psychological Assessment*, 28, 92. [PubMed: 26029946]
- Krueger RF, & Markon KE (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annu. Rev. Clin. Psychol*, 2, 111–133. [PubMed: 17716066]
- 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, 971. [PubMed: 22845652]
- Lahey BB, Rathouz PJ, Keenan K, Stepp SD, Loeber R, & Hipwell AE (2015). Criterion validity of the general factor of psychopathology in a prospective study of girls. *Journal of Child Psychology and Psychiatry*, 56, 415–422. [PubMed: 25052460]
- Lewinsohn PM, Shankman SA, Gau JM, & Klein DN (2004). The prevalence and comorbidity of subthreshold psychiatric conditions. *Psychological Medicine*, 34, 613–622. [PubMed: 15099416]
- MacCallum RC (2009). Factor Analysis. In *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 124–147). London: SAGE Publications Ltd.

- Markon KE (2019). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. In *Annual Review of Clinical Psychology: Vol. 15. ANNUAL REVIEW OF CLINICAL PSYCHOLOGY, VOL 15* (pp. 51–69). ANNUAL REVIEWS.
- Micheline G, Barch DM, Tian Y, Watson D, Klein DN, & Kotov R (2019). Delineating and validating higher-order dimensions of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study. *Translational Psychiatry*, 9, 1–15. [PubMed: 30664621]
- Moore TM, Kaczurkin AN, Durham EL, Jeong HJ, McDowell MG, Dupont RM, ... Kardan O (2020). Criterion Validity and Relationships between Alternative Hierarchical Dimensional Models of General and Specific Psychopathology. *BioRxiv*.
- Muthén LK, & Muthén BO (2020). *Mplus User's Guide* (Version 8th Edition). Los Angeles, CA: Muthén & Muthén.
- Newby JM, McKinnon A, Kuyken W, Gilbody S, & Dalgleish T (2015). Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. *Clinical Psychology Review*, 40, 91–110. [PubMed: 26094079]
- Patalay P, Fonagy P, Deighton J, Belsky J, Vostanis P, & Wolpert M (2015). A general psychopathology factor in early adolescence. *The British Journal of Psychiatry*, 207, 15–22. [PubMed: 25906794]
- Pettersson E, Lahey B, Larsson H, Lundstroem S, & Lichtenstein P (2015). The general factor of psychopathology in childhood and maladaptive outcomes in young adulthood: A Swedish population twin study. *Behavior Genetics*, 45, 679–679. SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA.
- Piper BJ, Gray HM, Raber J, & Birkett MA (2014). Reliability and validity of brief problem monitor, an abbreviated form of the child behavior checklist. *Psychiatry and Clinical Neurosciences*, 68, 759–767. [PubMed: 24735087]
- Preacher KJ, & MacCallum RC (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, 2, 13–43.
- R Core Team. (2019). *R: A language and environment for statistical computing* (3.6.0). Vienna, Austria: R Foundation for Statistical Computing.
- Rijmen F (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47, 361–372.
- Rodriguez A, Reise SP, & Haviland MG (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21, 137–150. [PubMed: 26523435]
- Rucci P, Gherardi S, Tansella M, Piccinelli M, Berardi D, Bisoffi G, ... Pini S (2003). Subthreshold psychiatric disorders in primary care: Prevalence and associated characteristics. *Journal of Affective Disorders*, 76, 171–181. [PubMed: 12943947]
- Samejima F (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34, 1–97.
- Simmons JP, Nelson LD, & Simonsohn U (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366. [PubMed: 22006061]
- Stochl J, Khandaker GM, Lewis G, Perez J, Goodyer IM, Zammit S, ... Jones PB (2015). Mood, anxiety and psychotic phenomena measure a common psychopathological factor. *Psychological Medicine*, 45, 1483–1493. [PubMed: 25394403]
- van Bork R, Epskamp S, Rhemtulla M, & Borsboom D (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology*, 27, 759–773.
- Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, ... Conway K (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental Cognitive Neuroscience*, 32, 4–7. [PubMed: 29051027]
- Waldman ID, Poore HE, van Hulle C, Rathouz PJ, & Lahey BB (2016). External validity of a hierarchical dimensional model of child and adolescent psychopathology: Tests using confirmatory factor analyses and multivariate behavior genetic analyses. *Journal of Abnormal Psychology*, 125, 1053. [PubMed: 27819467]

- Watts AL, Poore HE, & Waldman ID (2019). Riskier Tests of the Validity of the Bifactor Model of Psychopathology. *Clinical Psychological Science*, 7, 1285–1303.
- Wirth RJ, & Edwards MC (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, 12, 58–79. [PubMed: 17402812]
- Wong EH, Yocca F, Smith MA, & Lee C-M (2010). Challenges and opportunities for drug discovery in psychiatric disorders: The drug hunters' perspective. *International Journal of Neuropsychopharmacology*, 13, 1269–1284.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Fourteen Different General Factor of Psychopathology (GFP) Models

Model Number	Model Type	Items or Scales	CBCL or EFA Based Sub-Factors	Number of Sub-Factors	Mean (SD) Loading on GFP Factor	Mean (SD) Loading on Facet Factors	Mean (SD) I-ECV	RMSEA
GFP-1	Single Factor	Items	N/A	0	.62 (.14)	N/A	N/A	.08 ¹
GFP-2	Bifactor	Items	CBCL (Scales)	8	.61 (.13)	.36 (.20)	.72 (.22)	.08 ¹
GFP-3	Bifactor	Items	CBCL (Int, Ext)	2	.61 (.14)	.34 (.14)	.74 (.17)	.08 ¹
GFP-4	Bifactor	Items	EFA (Int, Ext, Atten)	3	.61 (.13)	.30 (.16)	.79 (.17)	.07 ¹
GFP-5	Bifactor	Items	EFA (Int, Ext, Atten, Somat)	4	.61 (.12)	.34 (.18)	.75 (.20)	.07 ¹
GFP-6	Higher-Order	Items	CBCL (Scales)	8	.83 (.10)	.72 (.09)	N/A	.03 ²
GFP-7	Higher-Order	Items	CBCL (Int, Ext, Atten)	3	.85 (.05)	.67 (.12)	N/A	.04 ²
GFP-8	Higher-Order	Items	EFA (Int, Ext, Atten)	3	.89 (.04)	.65 (.12)	N/A	.03 ²
GFP-9	Higher-Order	Items	EFA (Int, Ext, Atten)	4	.85 (.06)	.66 (.11)	N/A	.03 ²
GFP-10	Single Factor	Scales	N/A	0	.71 (.12)	N/A	N/A	.11 ³
GFP-11	Bifactor	Scales	CBCL (Int, Ext)	2	.69 (.13)	.43 (.07)	.68 (.06)	.06 ³
GFP-12	Bifactor	Scales	EFA (Other, Ext)	2	.70 (.13)	.29 (.17)	.82 (.14)	.05 ³
GFP-13	Higher-Order	Scales	CBCL (Int, Ext)	2	.83 (.00)	.76 (.16)	N/A	.07 ³
GFP-14	Higher-Order	Scales	EFA (Other, Ext)	2	.90 (.00)	.74 (.13)	N/A	.11 ³

GFP = General Factor of Psychopathology; CBCL = Child Behavior Checklist; EFA = exploratory factor analysis; Scales = sub-factors based on individual CBCL scales; Int = internalizing; Ext = externalizing; Atten = attention problems; Somat = somatic problems; Other = all other problems; I-ECV = indicator level explained common variance; N/A = not applicable. Items and scales that did not load on a specific factor in a bifactor model were excluded from the computations of the mean loading on the sub-factors and the mean I-ECV. Shared superscripts denotes that models and fit statistics were derived via the same program and estimator (1 = full information maximum likelihood estimation in flexMIRT (only the full information fit statistics could be computed), 2 = weighted least squares with mean and variance adjustments in Mplus, 3 = full information maximum likelihood in Mplus); differences between fit statistics in different clusters should be interpreted with caution. Sub-factor refers to either specific factors in a bifactor model or lower-order factors in a higher-order factor model; sub-factor construction was based on the a priori CBCL scales, CBCL internalizing and externalizing factors, or EFAs conducted in the current sample. The variability in loadings on the GFP factor was 0 in GFP-13 and GFP-14 because with only two lower-order factors the loadings on the GFP had to be fixed to equality for proper estimation.

Table 2

Correlations between General Factor of Psychopathology (GFP) Factor Scores

Models	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2. GFP-2 (B/I/C/8)														
2. GFP-2 (B/I/C/8)	.99													
3. GFP-3 (B/I/C/2)	.99	.99												
4. GFP-4 (B/I/E/3)	.98	.98	.96											
5. GFP-5 (B/I/E/4)	.98	.97	.96	.99										
6. GFP-6 (H/I/C/8)	.98	.97	.96	.97	.97									
7. GFP-7 (H/I/C/3)	.97	.96	.95	.95	.94	.98								
8. GFP-8 (H/I/E/3)	.99	.99	.99	.97	.97	.98	.97							
9. GFP-9 (H/I/E/4)	.99	.99	.98	.97	.97	.98	.97	.99						
10. GFP-10 (F/S/C/0)	.87	.86	.86	.86	.86	.90	.87	.89	.89					
11. GFP-11 (B/S/C/2)	.86	.86	.87	.85	.85	.90	.86	.89	.89	.99				
12. GFP-12 (B/S/E/2)	.86	.86	.87	.85	.85	.90	.86	.89	.89	.99	.99			
13. GFP-13 (H/S/C/2)	.86	.85	.82	.86	.86	.89	.90	.86	.87	.92	.90	.90		
14. GFP-14 (H/S/E/2)	.88	.88	.87	.87	.86	.91	.91	.90	.90	.97	.97	.96	.97	
Number of Indicators	116	116	116	116	116	8	3	3	4	8	8	8	2	2
<i>H</i>	.99	.99	.99	.99	.99	.96	.91	.91	.90	.91	.91	.91	.82	.90
ω	.99	-	-	-	-	.95	.89	.90	.89	.89	-	-	-	-
ω_h	-	.97	.95	.93	.93	-	-	-	-	-	.85	.87	.82	.90

GFP = general factor of psychopathology; B = bifactor model structure; H = higher order model structure; F = single factor structure; I = item level model; S = scale level model; C = CBCL based sub-factors; E = EFA based sub-factors; # = number of sub-factors in model; Number of Indicators = how many items, scales, or lower order factors loaded on the GFP; *H* = *H* index of construct replicability; ω = coefficient omega; ω_h = coefficient omega-hierarchical. Coefficient omega used for computing GFP latent factor reliability in single factor confirmatory factor analytic and higher order models; coefficient omega-hierarchical used for computing GFP latent factor reliability in bifactor models.

Table 3

Correlations between Internalizing and Externalizing Sub-Factor Scores

Externalizing Sub-Factors																		
Internalizing Sub-Factors										Externalizing Sub-Factors								
GFP Residualized Lower Order Factors	GFP 3	GFP 4	GFP 5	GFP 7	GFP 8	GFP 9	GFP 11	GFP 13		GFP 3	GFP 4	GFP 5	GFP 7	GFP 8	GFP 9	GFP 11	GFP 13	
GFP-3 (B/I/C/2) GFP-4 (B/I/E/3)	.91									.61								
GFP-5 (B/I/E/4)	.84	.92								.66	.99							
GFP-7 (H/I/C/3)	.90	.82	.70							.86	.58	.64						
GFP-8 (H/I/E/3)	.86	.78	.64	.84						.87	.83	.86	.85					
GFP-9 (H/I/E/4)	.81	.72	.77	.82	.83					.83	.81	.87	.86	.95				
GFP-11 (B/S/C/2)	.87	.75	.68	.84	.85	.79				.57	.59	.61	.63	.64	.63			
GFP-13 (H/S/C/2)	.78	.75	.67	.88	.74	.74	.85			.61	.61	.68	.69	.63	.72	.57		
Original Lower Order Factors	GFP 3	GFP 4	GFP 5	GFP 7	GFP 8	GFP 9	GFP 11	GFP 13		GFP 3	GFP 4	GFP 5	GFP 7	GFP 8	GFP 9	GFP 11	GFP 13	
GFP-3 (B/I/C/2) GFP-4 (B/I/E/3)	.91									.61								
GFP-5 (B/I/E/4)	.84	.92								.66	.99							
GFP-7 (H/I/C/3)	.56	.49	.38							.47	.40	.46						
GFP-8 (H/I/E/3)	.46	.42	.30	.97						.44	.48	.53	.98					
GFP-9 (H/I/E/4)	.49	.44	.38	.97	.98					.44	.46	.51	.98	.99				
GFP-11 (B/S/C/2)	.87	.75	.68	.55	.47	.49				.57	.59	.61	.40	.40	.39			
GFP-13 (H/S/C/2)	.53	.45	.38	.92	.89	.90	.62			.51	.40	.46	.91	.89	.89	.45		
Number of Indicators	31	43	22	31	43	22	3	3	29	32	38	29	32	38	29	32	2	
H	.85	.84	.76	.96	.97	.96	.37	.82	.82	.86	.89	.97	.97	.98	.98	.37	.96	
-	-	-	-	.95	.96	.95	-	.76	-	-	-	-	.97	.97	.97	-	.84	
ω _{HS}	.29	.17	.14	-	-	-	.24	.18	.18	.21	.22	-	-	-	-	.27	-	

GFP = general factor of psychopathology; B = bifactor model structure; H = higher order model structure; I = item level model; S = scale level model; C = CBCL based sub-factors; E = EFA based sub-factors; # = number of sub-factors in model; Number of Indicators = how many items or scales loaded on the factor; *H* = *H* index of construct replicability; ω = coefficient omega; ω_{HS} = coefficient omega-hierarchical, subscale. Coefficient omega used for computing latent factor reliability in single and higher order models; coefficient omega-hierarchical, subscale used for computing latent factor reliability in bifactor models. Correlations with lower-order factors after being residualized of GFP variance presented on top of table; correlations with original lower-order factors presented on bottom of table. Correlations between bifactor and higher order models bolded and italicized.

Table 4
Correlations between General Factor of Psychopathology (GFP) Factor Scores and Criterion Variables

	Total Problems (CBCL) _T	Total Problems (CBCL) _S	Total Problems (KSAD)	Total Problems (BPM)	Mental Health Services	Negative Urgency	Lack of Planning	Sensation Seeking	Positive Urgency	Lack of Perseverance	General Cognitive Ability	Well in School	Grades	Special Services	Detention
GFP-1	.90 [.90,.90]	.87 [.87,.87]	.51 [.50,.52]	.40 [.38,.43]	.54 [.51,.56]	.15 [.13,.17]	.15 [.13,.17]	.02 [.00,.04]	.12 [.11,.14]	.16 [.15,.18]	-.14 [-.16,-.13]	-.38 [-.35,-.40]	-.19 [.17,-.21]	.32 [.30,.35]	.43 [.39,.46]
GFP-2	.90 [.90,.90]	.87 [.87,.87]	.51 [.50,.52]	.41 [.39,.44]	.53 [.50,.55]	.15 [.13,.17]	.16 [.14,.18]	.02 [.01,.04]	.13 [.12,.15]	.17 [.15,.19]	-.16 [-.18,-.14]	-.39 [-.37,-.41]	-.20 [.18,-.22]	.34 [.31,.36]	.43 [.39,.47]
GFP-3	.89 [.89,.89]	.86 [.86,.86]	.51 [.50,.52]	.43 [.41,.46]	.52 [.50,.55]	.15 [.13,.17]	.17 [.15,.19]	.03 [.01,.05]	.14 [.09,.16]	.18 [.17,.20]	-.17 [-.19,-.16]	-.42 [.40,-.43]	-.22 [.20,-.23]	.36 [.34,.38]	.43 [.39,.47]
GFP-4	.89 [.89,.90]	.87 [.86,.86]	.49 [.47,.50]	.37 [.34,.39]	.51 [.48,.53]	.14 [.12,.16]	.14 [.12,.16]	.01 [.01,.03]	.11 [.08,.13]	.15 [.13,.17]	-.13 [.15,-.11]	-.33 [.31,-.35]	-.17 [.15,-.18]	.30 [.27,.32]	.39 [.35,.43]
GFP-5	.89 [.89,.90]	.87 [.87,.87]	.49 [.47,.50]	.35 [.32,.38]	.51 [.49,.54]	.13 [.11,.15]	.13 [.11,.15]	.00 [.02,.02]	.10 [.11,.12]	.15 [.13,.17]	-.12 [.14,-.10]	-.33 [.30,-.35]	-.16 [.14,-.18]	.30 [.27,.32]	.36 [.32,.40]
GFP-6	.92 [.92,.92]	.91 [.91,.91]	.53 [.52,.54]	.41 [.39,.44]	.52 [.50,.54]	.15 [.13,.17]	.15 [.13,.17]	.01 [.01,.03]	.13 [.12,.15]	.16 [.14,.18]	-.16 [.18,-.15]	-.38 [.35,-.39]	-.19 [.17,-.21]	.33 [.31,.35]	.42 [.39,.46]
GFP-7	.91 [.91,.91]	.88 [.88,.88]	.53 [.51,.54]	.42 [.39,.44]	.52 [.50,.54]	.16 [.14,.18]	.16 [.14,.18]	.03 [.01,.05]	.13 [.11,.15]	.16 [.14,.18]	-.16 [.17,-.14]	-.37 [.35,-.39]	-.20 [.18,-.21]	.31 [.29,.33]	.43 [.40,.47]
GFP-8	.92 [.91,.92]	.90 [.89,.89]	.53 [.52,.54]	.43 [.40,.45]	.53 [.50,.55]	.15 [.14,.17]	.16 [.14,.18]	.02 [.00,.04]	.13 [.11,.15]	.18 [.16,.20]	-.16 [.18,-.15]	-.40 [.38,-.42]	-.21 [.19,-.22]	.35 [.33,.37]	.41 [.38,.45]
GFP-9	.92 [.92,.92]	.90 [.90,.90]	.53 [.52,.54]	.41 [.38,.43]	.53 [.51,.55]	.15 [.13,.17]	.15 [.13,.17]	.02 [.00,.04]	.12 [.11,.14]	.17 [.15,.19]	-.15 [.17,-.13]	-.38 [.36,-.40]	-.20 [.18,-.21]	.33 [.31,.35]	.40 [.37,.44]
GFP-10	.98 [.98,.98]	.99 [.99,.99]	.62 [.60,.63]	.44 [.41,.47]	.47 [.44,.48]	.15 [.13,.17]	.14 [.12,.16]	.01 [.01,.03]	.13 [.11,.15]	.16 [.14,.18]	-.18 [.20,-.16]	-.36 [.34,-.37]	-.20 [.18,-.22]	.33 [.31,.35]	.34 [.32,.37]
GFP-11	.96 [.96,.96]	.98 [.98,.98]	.61 [.60,.63]	.45 [.42,.48]	.46 [.44,.48]	.15 [.13,.17]	.14 [.12,.17]	.01 [.01,.03]	.13 [.11,.15]	.17 [.15,.19]	-.20 [.21,-.18]	-.37 [.35,-.39]	-.22 [.20,-.23]	.34 [.32,.37]	.34 [.31,.36]
GFP-12	.96 [.96,.96]	.99 [.99,.98]	.61 [.59,.63]	.44 [.41,.47]	.47 [.44,.48]	.14 [.12,.16]	.14 [.12,.16]	.00 [.02,.02]	.12 [.10,.14]	.17 [.15,.19]	-.18 [.20,-.16]	-.37 [.35,-.39]	-.22 [.19,-.23]	.34 [.32,.36]	.32 [.30,.35]
GFP-13	.96 [.96,.96]	.94 [.94,.94]	.58 [.56,.60]	.38 [.34,.41]	.46 [.43,.48]	.15 [.13,.17]	.13 [.11,.15]	.01 [.01,.03]	.11 [.09,.13]	.12 [.10,.14]	-.12 [.14,-.10]	-.27 [.25,-.29]	-.15 [.13,-.16]	.24 [.22,.26]	.33 [.30,.36]
GFP-14	.98 [.98,.98]	.97 [.97,.97]	.62 [.60,.63]	.44 [.41,.47]	.46 [.44,.48]	.16 [.14,.18]	.15 [.13,.17]	.02 [.00,.04]	.14 [.12,.16]	.15 [.13,.17]	-.18 [.20,-.16]	-.34 [.32,-.36]	-.20 [.18,-.21]	.31 [.28,.33]	.36 [.34,.39]

GFP = general factor of psychopathology; CBCL = Child Behavior Checklist; I = item-based CBCL total problems score; KSADS = Kiddie-Structured Assessment for Affective Disorders and Schizophrenia; BPM = teacher reported Brief Problem Monitoring Form; Mental Health Services = have youth ever accessed mental health services; Well in School = how to parents rate youth performance in school; Grades = how do parents rate the grades youth obtain in school; Special Services = has a youth required special services in school; Detention = has a youth ever been assigned detention. 95% confidence intervals presented in brackets under estimates; confidence intervals derived via clustered, non-parametric percentile bootstrap with 1,000 draws.