

The Hierarchical Structure and Longitudinal Measurement Invariance of Externalizing Symptoms in the Adolescent Brain and Cognitive Development (ABCD©) Study

Hierarchical Structure and Measurement Invariance of Externalizing Symptoms

Colin Vize^{a,*}, Whitney R. Ringwald^b, Emily R. Perkins^c, Rebecca E. Waller^d, Samuel W. Hawes^e, Amy L. Byrd^f

^aUniversity of Pittsburgh, Psychology, 4113 Sennott Square, Pittsburgh, USA, 15260

^bUniversity of Pittsburgh, Psychology, Pittsburgh, USA,

^cUniversity of Pennsylvania, Psychology, Philadelphia, USA,

^dUniversity of Pennsylvania, Psychology, PA, USA,

^eFlorida International University, Center for Children and Families, Miami, USA,

^fUniversity of Pittsburgh Medical Center, Psychiatry, Pittsburgh, USA,

Abstract

Recent years have seen a greater shift towards alternative nosological frameworks of psychopathology, which frequently include a dimension of externalizing psychopathology. The Hierarchical Taxonomy of Psychopathology (HiTOP) is one such framework. However, the HiTOP has been most often studied in adults and more research is needed in order to better understand the similarities and differences in the structure of externalizing psychopathology earlier in development. This preregistered study sought to examine the validity and utility of extending the HiTOP externalizing dimension and its subdimensions to youth using longitudinal data from the Adolescent Brain and Cognitive Development (ABCD) Study. There were two primary aims: Aim 1: Identify the hierarchical structure of externalizing psychopathology and examine evidence of discriminant validity of the identified dimensions; and Aim 2: Assess the longitudinal measurement invariance of a broad externalizing dimension in the ABCD study, as well as specific underlying symptom dimensions. Results for Aim 1 analyses identified a coherent factor structure comprising three dimensions (narrow externalizing, irritability, and neurodevelopmental problems), and these factors showed important similarities and differences in relation to external correlates. Aim 2 analyses showed that strong invariance was supported for the narrow externalizing and irritability dimensions, while partial strong invariance was supported for broad externalizing and neurodevelopmental problems. Quantification of measurement (non)invariance revealed small effect sizes. Collectively, these results highlight important directions for future research on the HiTOP model in the ABCD study and other youth samples.

Keywords: HiTOP, externalizing, longitudinal invariance, youth psychopathology, ABCD Study

In recent years, researchers have increasingly pursued alternative nosological frameworks of psychopathology with the goal of enhancing etiological models and, ultimately, improving prevention and intervention efforts [?]. One popular framework, the Hierarchical Taxonomy of Psychopathology [(]HiTOP; [?]), has generated significant research interest. The

HiTOP aims to establish an empirically derived structure of psychopathology using techniques, such as factor analysis to identify transdiagnostic dimensions that explain covariation among observed symptoms. Importantly, the hierarchical organization of HiTOP incorporates general and specific symptoms of psychopathology, thereby avoiding the limitations of categorical approaches like that of the Diagnostic and Statistical Manual of Mental Disorders [DSM-5; [?]], including high rates of within-disorder heterogeneity and comorbidity [?]. Thus, the HiTOP aims to generate a nosological framework that is empirically valid and capable of benefiting both researchers and

*Corresponding author

Email addresses: cev18@pitt.edu (Colin Vize),
wringwald@pitt.edu (Whitney R. Ringwald),
perkinse@sas.upenn.edu (Emily R. Perkins),
rwaller@sas.upenn.edu (Rebecca E. Waller), shawes@fiu.edu
(Samuel W. Hawes), byrdal@upmc.edu (Amy L. Byrd)

clinicians [?]. Indeed, across a range of mental health problems, dimensional assessments produce greater validity and reliability compared to categorical assessments (e.g., Markon et al., 2011) and emerging evidence suggests that clinicians find more utility in transdiagnostic, dimensional approaches like HiTOP compared to traditional categorical approaches [e.g., ?].

However, the majority of research examining HiTOP has been conducted in adults, leaving notable gaps in our understanding of how it can be extended to youth during sensitive developmental windows. The present study examined whether components of the HiTOP model, specifically its externalizing dimensions, could be identified in the Adolescent Brain and Cognitive Development (ABCD) Study, a longitudinal, nationally representative study of youth in the United States. The longitudinal nature of the ABCD study allows for tests of whether externalizing dimensions are psychometrically invariant over time—an essential precursor to more advanced modeling techniques (e.g., latent growth models) that can be used to study trajectories of externalizing dimensions over development. Addressing these gaps has the potential to expand the relevance of HiTOP to earlier periods of development, a critical step to further advance HiTOP-focused research [?].

1. HiTOP as a Model of Externalizing Psychopathology

The HiTOP is organized along continuous dimensions, consistent with evidence showing that differences in psychopathology are a matter of degree as opposed to kind [?]. The HiTOP is also hierarchical, reflecting decades of research establishing that symptoms can be organized into broader dimensions, which in turn can advance research on shared and distinct etiological mechanisms of psychopathology symptoms [?]. At the finest level, the HiTOP provides descriptions of narrow behavioral (e.g., aggression, inattention) and interpersonal (e.g., callousness) symptoms or traits. One can move up the hierarchy to broader dimensions that are called superspectra in the HiTOP (e.g., externalizing and internalizing psychopathology). This dimensional, hierarchical approach allows researchers to study features of psychopathology at varying levels of specificity.

In adults, the externalizing superspectrum encompasses two lower order dimensions (termed spectra in the HiTOP): disinhibited externalizing and antagonistic

externalizing [?]. The disinhibited externalizing spectrum includes the tendency to act impulsively, with little regard for future consequences. The antagonistic externalizing spectrum reflects tendencies to navigate interpersonal situations with little concern for others, including a willingness to exploit and deceive others. The broader externalizing dimension has been linked to various maladaptive outcomes, including aggression, antisocial behavior, and problematic substance use [e.g., ?], while the lower order spectra, including disinhibition and antagonism, offer greater specificity and show differential associations with external correlates [?]. For example, antagonism is more strongly related to various forms of aggression and antisocial behavior relative to disinhibitory symptoms [?], while disinhibitory symptoms are more closely related to problematic substance use [e.g., ?]. While there is substantial evidence supporting the presence of a broad externalizing dimension in youth [e.g., ?], less work has examined whether the antagonistic and disinhibited HiTOP dimensions can be similarly captured in youth samples and if these spectra show unique associations with relevant clinical and psychosocial correlates.

2. Hierarchical Models of Externalizing Psychopathology in Youth

While the HiTOP has seen less direct application in youth samples, there is a rich history of dimensional and hierarchical models of psychopathology in youth. For example, models of externalizing psychopathology have been examined across a wide range of developmental periods, including preschool [e.g., ?] and early adolescence [e.g., ?]. Across these studies, results have consistently shown that a broadband externalizing factor can be reliably assessed from more fine-grained indicators and the externalizing factor is reliably linked to specific problem behaviors (e.g., aggression and rule-breaking). Notably, this work used a wide variety of indicators to model higher-order externalizing dimensions, including DSM-based diagnostic symptoms [e.g., ?], personality features [e.g., ?], and biologically-based indicators like salivary cortisol levels [e.g., ?]. Other relevant examples of externalizing-related research in youth involve research on hierarchical models of maladaptive personality that include antagonistic and disinhibitory personality features, which have also found strong support in youth samples [e.g., ?]. Taken together, this work suggests that dimensional, hierarchical approaches like HiTOP can be successfully applied to youth populations—and, moreover, that Hi-

TOP can serve as a unifying framework for many existing lines of developmental psychopathology research.

Along these lines, more recent work has sought to explicitly examine the HiTOP model in youth samples. For example, using data from the baseline assessment of the Adolescent Brain Cognitive Development (ABCD)TM study, [?] conducted a “bass-ackwards” factor analysis [?] using 102 items of the parent-report Child Behavior Checklist [?]. At the most specific level of the hierarchy, the authors identified five factors that they labeled externalizing, neurodevelopmental, internalizing, somatoform, and detachment. The authors noted that the factors mostly align with the structure of psychopathology as detailed in HiTOP, with the exception that the neurodevelopmental factor that emerged from the broader externalizing factor is not included in the HiTOP model. In a more recent effort, [?] pooled youth self-report data across various early- and mid-adolescent samples (age range=6-18 years old; N=18,290) to identify similarities and differences between the hierarchical structure of symptoms in youth and the HiTOP structure in adults. The broad factors identified were termed externalizing, eating pathology, uncontrollable worry/obsessions and compulsions, and internalizing. Additionally, more specific dimensions of externalizing were noted, including impulsive anger, antagonism, retaliatory anger, and positive psychosis. While the antagonism dimension mirrored the antagonistic externalizing dimension in adults (i.e., consisting of items indexing aggression and cruelty towards others), no clear disinhibited externalizing factor emerged. Instead, items reflecting disinhibitory content (e.g., “Generally, I am an impulsive person”) showed primary loadings on the impulsive anger dimension. Thus, existing research has highlighted similarities and differences between the HiTOP model among adults and youth. Further research is needed to further clarify how the HiTOP could be extended to youth.

3. Longitudinal Invariance of Externalizing Psychopathology in Youth

In addition to characterizing the hierarchical structure of externalizing psychopathology in youth, it is important to determine whether this structure is consistent over time [?]. Specifically, examining the longitudinal measurement invariance of externalizing psychopathology, or the degree to which the measurement of this latent construct is psychometrically equivalent over time, can help establish whether observed

changes reflect true changes. Because the developmentally relevant behavioral and interpersonal manifestations of externalizing psychopathology are known to change across development [?], establishing longitudinal measurement invariance is necessary to ensure that observed changes are reflective of changes in externalizing rather than developmentally limited indicators. For example, decreases in temper tantrums are likely to reflect an independent developmental process separate from true decreases in a latent externalizing propensity. Longitudinal invariance is particularly important in the ABCD study, given its unique potential to serve as a scientific resource as a nationally representative, longitudinal study of youth and their families. The ABCD study is ideally situated to provide data that can better characterize risk and resilience factors for within-individual change in externalizing psychopathology. Thus, the transparent development of longitudinally invariant measurement models of externalizing dimensions is essential.

4. Current Study

The current preregistered study sought to build on recent efforts to extend the HiTOP model to developmental samples [e.g., ?], while using a large, nationally representative sample of youth followed over multiple years. The current study had two aims: **Aim 1:** Identify the hierarchical structure of externalizing psychopathology and examine evidence of discriminant validity of the identified dimensions; and **Aim 2:** Assess the longitudinal measurement invariance of a broad externalizing dimension in the ABCD study as well as more specific dimensions underlying broad externalizing symptoms. Collectively, the current study aimed to investigate the validity and utility of disaggregating the externalizing dimension specifically in youth during a developmentally sensitive window.

5. Method

Transparency and Openness The analysis plan was preregistered and can be found at [this link](#). Deviations from our preregistered plan are reported in the supplement. R and MPlus code to reproduce our analyses are available at [this link](#). Data are not posted to OSF because ABCD data are restricted to researchers with approved access. However, the code posted to OSF denotes the data files used and how they were cleaned for our analyses, making the results reproducible for researchers with access to ABCD data.

6. Sample

Data were drawn from four waves (baseline, 1-, 2-, and 3-year follow-up) of the Adolescent Brain Cognitive Development (ABCD)TM study (N=11,875 at baseline; Mage=9.51; 48% girls; 57% White; 15% Black; 20% Hispanic/Latino/a). For our primary measures, sample sizes ranged from N=11,862 at baseline to N=10,099 at the 3-year follow-up assessment. A more complete description of the ABCD study, including previous publications using the data, is available at the [ABCD Study site] (<https://abcdstudy.org/>).

7. Measures

Externalizing Dimensions (Parent-Report: T1-T4). Externalizing psychopathology was assessed using a subset of items from the parent-report Child Behavior Checklist [CBCL; [?]], which was administered at all four assessments. Specifically, we used the 45 CBCL items and item composites identified in factor analyses of the ABCD baseline data by [?] that had primary loadings on the externalizing factor and/or the neurodevelopmental factor. These two factors had a total of 45 items/item composites that showed either primary loadings on one of the factors or loaded on both factors without a clear primary loading (e.g., the item “impulsive or acts without thinking” had a loading of .49 on both the externalizing and neurodevelopmental factors). The preregistration provides details for the specific CBCL items that were used for our analyses.

Diagnostic Symptom Counts (Parent- and Youth-Report; T1). Symptom counts were assessed using clinician ratings of parent- and youth-reported modules of the Kiddie Schedule for Affective Disorders and Schizophrenia [KSADS-5; [?]]. Dimensional symptom counts for the following diagnostic constructs were used: conduct disorder (CD), oppositional defiant disorder (ODD), attention deficit hyperactivity disorder (ADHD), major depressive disorder, suicidality/self-harm, generalized anxiety disorder, and social anxiety disorder. Each symptom was indicated as being present (1) or not present (0), and symptoms were summed such that higher scores indicate more symptoms endorsed for the diagnostic construct. Parents completed all KSADS-5 modules at the baseline assessment, while youth only completed mood disorder, social anxiety, generalized anxiety disorder, suicidality, and sleep modules at baseline.

Prosocial Behavior (Parent- and Youth-Report; T1). Prosocial behavior was assessed using the proso-

cial scale of the Strengths and Difficulties Questionnaire (SDQ), which included 3 items (“try to be nice to other people”; “care about their feelings”; “offer to help others”). Items were rated on a 3-point scale ranging from 0 (not true) to 2 (certainly true) and summed such that higher scores represent greater levels of prosocial behavior.

Impulsivity (Youth-Report; T1). Impulsivity was assessed using the youth-reported Urgency, Premeditation (lack of), Perseverance (lack of), Sensation Seeking, Positive Urgency, Impulsive Behavior Scale (UPPS-P for Children Short Form – ABCD version). The scale had 20 items assessing impulsivity (e.g., “I like to stop and think about things before I do them (reversed)”) and includes 5 subscales: negative urgency, positive urgency, lack of perseverance, lack of planning, and sensation seeking. Items were rated on a 4-point scale from 1 (not at all like me) to 4 (very much like me) and summed such that higher scores represent higher impulsivity.

Fluid Intelligence Composite (T1). Fluid intelligence was assessed using an age-corrected composite of 5 tasks from the NIH Toolbox Cognition measures [see ? , for a detailed description]: List Sorting Working Memory Test, Pattern Comparison Processing Speed Test, Picture Sequence Memory Test, Flanker Task, and Dimensional Change Card Sort Test. These tasks collectively assess abilities related to processing speed, episodic memory, working memory, cognitive control, and cognitive flexibility.

8. Preregistered Analyses and Hypotheses

Aim 1a Analyses: We used parallel analysis and the minimum average partial correlation (MAP) to estimate the number of factors to extract from the 45 items/composites. Next, we used [?] recently developed extension to Goldberg’s “bass-ackwards” analysis [?] to identify the hierarchical structure of externalizing items/composites using the minimum residual factor extraction method and promax rotation. [?] extended bass-ackwards approach differs from the traditional approach in that it 1) identifies redundant components that perpetuate through multiple levels of the hierarchy; 2) aids in identification of artifactual components; and 3) plots the strongest factor correlations among the remaining factors to identify their hierarchical structure. Although past work has used similar factor analytic approaches on CBCL data in the ABCD sample [e.g., ?], by constraining our analyses

to only focus on the 45 CBCL items/composites associated with broad externalizing, we expected more fine-grained differences to emerge in the hierarchical structure of the CBCL items/composites. **Aim 1b Analyses:** After identifying the hierarchical structure of externalizing dimensions, we assessed evidence discriminant validity by examining bivariate and semipartial correlations between the identified externalizing dimensions (assessed using factor scores) and external criterion measures (e.g., diagnostic symptom counts, prosocial behavior, impulsivity, fluid intelligence). Correlations for specific dimensions were also compared to one another using tests of dependent correlations to detect significant differences between correlations.

Aim 2a Analyses: Next, we examined the longitudinal measurement invariance of each externalizing dimension through a series of confirmatory factor analyses (CFA). Longitudinal measurement invariance was examined in typical fashion moving from configural to strong invariance. In some cases, we also examined strict invariance. To test this, we created item parcels from the CBCL items based on a set of preregistered criteria. The item parcels served as the observed indicators of externalizing dimensions for all tests of longitudinal measurement invariance. **Aim 2b Analyses:** To quantify the degree of invariance, we computed Cohen's d for mean and covariance structures [dMACS; [?]]. These metrics index the collective impact of loading and intercept noninvariance in the metric of Cohen's d , with higher values indicating a greater degree of indicator noninvariance.

9. Preregistered Hypotheses

We had three primary hypotheses. Hypothesis 1a: We hypothesized that disinhibition and antagonism dimensions would be identified at more fine-grained levels of the hierarchy with the same items that have been shown to comprise a broad externalizing dimension. If parallel analysis and the minimum average partial correlation test (MAP) suggested a relatively large number of factors could be extracted from the CBCL items (e.g., 5 or more), we expected that the disinhibition and antagonism factors would emerge relatively early in the factor extraction process (e.g., at level 2 or level 3). Hypothesis 1b: We expected evidence of discriminant findings for the derived disinhibition and antagonism factor scores derived. Specifically, we hypothesized that the disinhibition factor would have larger (+) associations with ADHD, impulsivity, and fluid intelligence and the antagonism factor would have larger (-) asso-

ciations with CD, ODD, and prosocial behavior. Furthermore, we expected these correlations to be significantly stronger than the correlations disinhibition and antagonism showed with internalizing outcomes. Hypothesis 2: We would be able to establish longitudinal strong measurement invariance for the measurement models of externalizing factors, facilitating future investigations of mean-level change over time in these externalizing factors in the ABCD study.

10. Results

Aim 1a: Hierarchical structure of externalizing psychopathology Results of the parallel analysis based on polychoric correlations among the CBCL items suggested up to 14 factors could be extracted from the 45 CBCL items, while MAP suggested four factors for extraction. Consistent with our preregistered criteria for factor extraction, we interpreted the 1-, 2-, 3- and 4-factor solutions. However, the fourth factor of the solution was not substantively meaningful, and we thus focus on the 3-factor hierarchy.

The first level of the bass-ackwards hierarchy was a broad externalizing factor with all 45 CBCL items showing moderately strong to strong standardized loadings on the factor (range=.47-.81). At the second level, a neurodevelopmental problems factor emerged, characterized by distractibility, poor motor coordination, and hyperactivity. The broad externalizing factor remained virtually unchanged and was strongly correlated with the broad externalizing factor from level 1 of the hierarchy ($r=.96$). At level 3, the neurodevelopmental problems factor remained unchanged ($r=1.00$ with the level 2 factor). However, the broad externalizing factor split into a narrower externalizing and an irritability factor, both of which were strongly related to the broad externalizing factor ($r_s=.91$). The narrow externalizing factor was characterized by aggression (e.g., threatening others, fighting), meanness to others (e.g., cruelty, bullying, lack of guilt), and rule-breaking behaviors (e.g., stealing, destroys composite), and the irritability factor was indexed by items related to affective reactivity and lability, hostility, and distrust of others. These three levels of the hierarchy are displayed in Figure 1. The extracted factor scores from level 3 were strongly related to the broad externalizing factor score (r range=.80-.87) from level 1 and positively related to one another (r range =.52-.62). Aim 1b: Discriminant validity of the identified dimensions

We extracted scores for the factors identified at level 3 of the bass-ackwards hierarchy: narrow externalizing,

irritability, and neurodevelopmental problems. Factor scores were also extracted for broad externalizing (i.e., externalizing at level 1 of the bass-ackwards hierarchy). Correlations between factor scores and external criterion measures are presented in **Table 1**.

Diagnostic Symptom Counts (Parent- and Youth-Report). At the bivariate level, all factors were positively associated with symptom counts across all disorders and informants, and associations were stronger for parent-reported symptoms (r range=.07-.66) compared to youth-reported symptoms (r range=.04-.16). Furthermore, these factors tended to be more strongly related to symptoms of CD, ODD, and ADHD (r range=.31-.60) compared to symptoms of depression and anxiety (r range=.04-.32).

Semipartial correlations accentuated differences observed at the bivariate level, most prominently for parent-report, providing strong evidence for discriminant validity. Compared to other factors, narrow externalizing showed the strongest associations with CD symptoms (sr =.59) and the irritability factor with ODD symptoms (sr =.43). The neurodevelopmental problems factor was unrelated to CD (sr =-.02) and ODD (sr =-.04) symptoms, and strongly associated with ADHD symptoms (sr =.50). Associations with depression and anxiety symptoms were notably smaller, and after accounting for overlap, the narrow externalizing factor demonstrated a negative association with generalized (sr =-.12) and social (sr =-.05) anxiety disorder symptoms. The irritability factor showed the strongest association with depression (sr =.20) and generalized anxiety symptoms (sr =.21) and the neurodevelopmental problems factor with social anxiety disorder (sr =.13). Notably, there was little difference across the three factors in the strength of associations with suicidality/self-harm (sr range=.08-.11).

Prosocial Behavior (Parent- and Youth-Report). At the bivariate level, all factors were negatively associated with prosocial behavior, and these associations were stronger for parent-report (r range=-.22 to -.35) compared to youth-report (r range=-.06 to -.09). Semipartial correlations were smaller for narrow externalizing (sr =-.21) and irritability (sr =-.18), and reduced to non-significance for neurodevelopmental problems (sr =-.01).

Impulsivity. All factors showed similarly sized positive associations to all indices of impulsivity, though these were notably small in magnitude (r =.04-.16). The narrow externalizing and irritability factors showed the strongest associations with negative urgency (.16

and .15, respectively), and the neurodevelopmental problems factor was most strongly associated with lack of perseverance (r =.18). Semipartial correlations were smaller in magnitude, but similar to one another. The exception was the relation between neurodevelopmental problems and lack of perseverance (sr =.18), compared to the relation to the other factors (sr range=-.02-.03).

Fluid Intelligence. All factors were negatively associated with fluid intelligence, though these associations were small in magnitude. The neurodevelopmental problems factor demonstrated the strongest negative association (r =-.17), followed by the narrow externalizing factor (r =-.14), and the irritability factor (r =-.08). Semipartial correlations were similar, though the relation between irritability and fluid intelligence became positive (sr =.08).

Aim 2a: Longitudinal Measurement Invariance

All items were randomly assigned to parcels, and the parcels were used as indicators of a given latent factor at baseline, one-year follow-up, two-year follow-up, and three-year follow-up assessments. **Supplementary Table S1** provides details on the number of items assigned to each parcel and item content for each parcel. **Supplementary Tables S2-S5** provide descriptive statistics for the parcel indicators for each factor. **Table 2** provides results for all tests of longitudinal measurement invariance. All models of the broad externalizing factor and the three narrower factors showed excellent absolute fit across all invariance tests, and strong standardized factor loadings for the parcels (λ s=.58-.89) in the configural models, as well as strong stability over time (rs =.68-.81). Model comparison results are summarized below (see Table 2).

Broad Externalizing Factor. Relative to the configural model, restricting factor loadings to be equal (i.e., test of weak invariance) did not decrement model fit (ΔCFI =.001; ΔNCI =.005). However, constraining the parcel indicator intercepts to be equal across time (i.e., the test of strong invariance) did result in a significant decrement in model fit (ΔCFI =.004; ΔNCI =.026). In an attempt to establish partial strong invariance, we freed the intercept constraints on parcels 4 and 5 based on modification indices. This model demonstrated excellent overall fit (χ^2 =947.21; df =152; RMSEA=.021; CFI =.994; TLI =.992) and showed no decrement in model fit compared to the weak invariance model of broad externalizing (ΔCFI =.000; ΔNCI =.001), indicating that partial strong longitudinal measurement invariance was supported. Given that full strong

invariance was not supported, we did not test for strict measurement invariance.

Narrow Externalizing Factor. Model fit was not significantly impacted by adding constraints for weak, strong, or strict measurement invariance. Thus, we were able to demonstrate strict longitudinal measurement invariance for the narrow externalizing factor.

Irritability. Model fit was not significantly impacted by adding constraints for weak or strong measurement invariance. However, the constraints for strict invariance resulted in a significant decrement in model fit ($\Delta\text{CFI}=.003$; $\Delta\text{NCI}=.009$). Thus, we were only able to demonstrate strong longitudinal measurement invariance.

Neurodevelopmental Problems. Relative to the configural model, model fit was not significantly impacted after constraining factor loadings to be equal ($\Delta\text{CFI}=.002$; $\Delta\text{NCI}=.005$), indicating weak measurement invariance. However, after imposing constraints on the intercepts of the parcel indicators (strong invariance), model fit was significantly negatively impacted ($\Delta\text{CFI}=.006$; $\Delta\text{NCI}=.018$). Based on modification indices, the intercepts of parcels 3 and 4 were freed at baseline and three-year follow-up. The resulting partial strong measurement invariance model showed excellent overall fit ($\chi^2=524.25$; $\text{df}=88$; $\text{RMSEA}=.020$; $\text{CFI}=.993$; $\text{TLI}=.991$) and no decrement in model fit compared to the weak invariance model ($\Delta\text{CFI}=.002$; $\Delta\text{NCI}=.005$), indicating that partial strong longitudinal measurement invariance was supported.

Aim 2b: Effect Sizes for Longitudinal Measurement Noninvariance Table 3 provides results for dMACS and additional effect sizes (i.e., proportion of mean change in the latent factor attributable to noninvariance) for the broad externalizing factor and neurodevelopmental problems factor (i.e., the two factors for which noninvariance was detected). Results showed that the effect sizes of indicator noninvariance were small, with dMACS ranging from .00-.18. While dMACS summarized the magnitude of noninvariance at the indicator level, there was a more noticeable impact of noninvariance at the latent level of the broad externalizing dimension, with noninvariance accounting for between 25-30% of observed mean change in the dimension across the pairwise comparisons.

Sensitivity Analyses We conducted a series of sensitivity analyses to examine the robustness of our results relative to other analytic choices. This included 1) examining alternative factoring and rotation methods

for our bass-ackwards analysis, 2) exploring the impact of variability in item-parcel allocation on measurement model parameters (i.e., loadings) and model fit statistics, and 3) examining the impact of cluster effects of family and study site in the ABCD study. Overall, sensitivity analyses showed that all results were robust to these factors. An overview of these analyses and results are provided in Supplementary Materials (see **Section II** and **Table S6**).

Discussion The present study sought to delineate the hierarchical structure of externalizing symptoms in the ABCD study, and evaluate the longitudinal measurement invariance of the identified dimensions. Overall, there was mixed support for our preregistered hypotheses. Specifically, in our bass-ackwards analysis, we did not find evidence for easily interpretable disinhibition or antagonism dimensions (*Hypothesis 1a*) and instead identified three factors: narrow externalizing, irritability, and neurodevelopmental problems. There was support for the discriminant validity of the specific externalizing dimensions (*Hypothesis 1b*) and, consistent with our hypothesis (*Hypothesis 2*), strong measurement invariance was established for two of these factors (narrow externalizing and irritability). While strong invariance could not be established for the broad externalizing and neurodevelopmental problems factors, a less restrictive model (partial strong invariance) was supported. Importantly, effect size metrics for longitudinal invariance highlighted the necessity of examining the impact of noninvariance at the level of indicators and latent dimensions, as noninvariance showed the greatest impact on the latent level here. Together, results have implications for extending the HiTOP model to youth and for future research on the development of externalizing problems in the ABCD study as well as other prospective, longitudinal studies.

Hierarchical Structure of Externalizing Psychopathology: Links to the HiTOP Model While the current study identified a broad externalizing dimension that strongly aligns with the externalizing superspectrum of HiTOP [?], the specific externalizing dimensions showed greater divergence from what has been found in adults. Contrary to hypotheses, clear disinhibition or antagonism dimensions were not identified. Instead, three dimensions with notable links to prior work and some similarities to the HiTOP model emerged. The first was a narrow externalizing dimension, consisting of aggression, meanness to others, and rule breaking behaviors. This factor mirrored antagonism dimensions identified in similar HiTOP-focused analyses in youth samples, which included items as-

sociated with aggressive behavior and meanness toward others [?]. In our sample, this dimension also contained content aligned with rule-breaking behavior (e.g., stealing, vandalism), which stands somewhat in contrast to past research highlighting the importance of distinguishing rule-breaking behaviors from more overtly antagonistic behaviors like aggression [?]. Interestingly, even at subsequent levels of the bass-ackwards hierarchy, this narrow externalizing factor did not split into more more specific antagonistic (e.g., aggression, callousness) or rule-breaking factors, suggesting that these features could not be empirically separated. While this may be related to developmental timing and/or the nature of this community sample, it will be important to continue examining the hierarchical structure of externalizing pathology across development, and empirically determine if and when these dimensions differentiate in the ABCD Study.

Results also identified an irritability dimension, characterized by affective reactivity and lability, hostility, and distrust of others. While not predicted, this dimension echoes prior work documenting the central role of irritability in early manifestations of externalizing behavior in youth [?]. Notably, characteristics of the irritability dimension found here – hostility and distrust of others – are similar to features of the antagonistic externalizing dimension that has been identified in adult samples [?]. However, this dimension was also characterized by a general moodiness that is not explicitly interpersonal in nature, and may point to affective lability as a developmental precursor to interpersonal hostility. In addition, extant research shows that irritability is an important subdimension of externalizing behaviors in youth that predicts subsequent internalizing psychopathology [e.g., ?]. Moreover, irritability is a transdiagnostic factor observed in an array of psychiatric disorders, and can help to explain the co-occurrence of externalizing and internalizing problems [?]. Thus, irritability represents an important dimension within developmental models of externalizing behavior and extensions of the HiTOP to youth.

A neurodevelopmental problems dimension defined by inattention and hyperactivity emerged and remained intact across all levels of our bass-ackwards hierarchy. Notably, much of the factor content aligns with ADHD symptoms, and past structural work in the ABCD study has labeled related factors “inattention” or “ADHD” factors [e.g., ?]. Nonetheless, the presence of such a factor echoes prior work on the structure of psychopathology in youth [? ?] and underscores the importance of this dimension in developmen-

tal models of externalizing behavior. Moreover, the emergence of this dimension dovetails with other research emphasizing the importance of understanding neurodevelopmental problems as a transdiagnostic dimension in youth psychopathology [?]. Past research has also emphasized the importance of neurodevelopmental problems in the onset and persistence of antisocial behavior [?], which further contextualizes the presence of neurodevelopmental problems within the domain of youth externalizing problems.

Evidence of Discriminant Validity for Externalizing Dimensions The three identified externalizing dimensions demonstrated evidence of discriminant validity based on their bivariate correlations. This was most notable for parent-reported diagnostic symptom counts and prosocial behavior, which may be unsurprising given shared method variance. The largest differences between dimensions were observed for disorders that are typically conceptualized as externalizing, with the narrow externalizing factor demonstrating the strongest association with CD symptoms, the irritability factor with ODD symptoms, and the neurodevelopmental problems factor with ADHD symptoms.

Controlling for overlap among the dimensions generally enhanced this pattern of discriminant validity, but several results are noteworthy. First, the unique variance in the narrow externalizing dimension showed the smallest relations to internalizing symptoms and was negatively related to anxiety disorder symptoms, while maintaining similar relations with conduct disorder symptoms and prosocial behavior. Thus, the unique aspects of this dimension appear to be aligned with callous-unemotional features—past research has demonstrated how these features are weakly related to broad internalizing problems [?] and demonstrate negative relations with with anxiety specifically [e.g., ?]. Second, the irritability dimension maintained the strongest associations with mood and anxiety symptoms even after accounting for overlap between factors, echoing research pointing to irritability as a transdiagnostic indicator of psychopathology [?]. Finally, the unique variance in neurodevelopmental problems was unrelated to CD, and ODD symptoms as well as prosociality. This finding has implications for research on “cool” (abstract-cognitive) and “hot” (reward-related) executive functioning systems, and suggests that the unique variance in the neurodevelopmental problems dimension appears to largely reflect the “cold” end of the hot-cold continuum in executive function [?].

Longitudinal Measurement Invariance for Ex-

ternalizing Dimensions Tests of longitudinal measurement invariance showed that strict invariance held for the narrow externalizing dimension and strong invariance held for the irritability dimension. Statistically, strong invariance indicates that all longitudinal changes in the means and covariances of the indicators can be attributed to changes in their respective latent dimensions across time [?]. Substantively, strong invariance indicates that researchers can conduct comparisons across time for these externalizing dimensions and investigate their associations with other variables without worrying that the indicators have been systematically influenced by unmeasured factors [?]. Notably, partial strong invariance was met for the broad externalizing and neurodevelopmental problems factors, indicating that longitudinal changes in observed indicator means and covariances are attributable to changes in the latent means, *but only for the longitudinally invariant indicators*. Given that the neurodevelopmental problems factor, in particular, focuses on developmentally specific content that likely manifests differently as youth mature, some degree of noninvariance is unsurprising [?]. Though strong invariance is the typical benchmark that must be met before making mean comparisons across time, researchers commonly use partial strong invariance models of externalizing [e.g., ?] and simulation-based investigations have found that using partially invariant models for subsequent analyses does not result in biased parameters in common applied settings [?]. Taken together, findings provide empirical support for the examination of between- and within-person change in these latent externalizing dimensions over time.

Additionally, our investigation of invariance effect sizes, while not commonly assessed or reported, points to important additional areas of consideration. First, all *dMACS* for the broad externalizing and neurodevelopmental problem models (i.e., the two models where non-invariant indicators were detected) fell below our pre-registered threshold for inferring meaningful impact of measurement invariance ($dMACS < .20$). This suggests that even though our previous tests indicated that constraining intercepts significantly worsened model fit, the difference in indicator intercepts across time was very small — providing further support for the notion that our findings of only partial strong invariance for certain factors do not preclude their use in subsequent analyses. Second, despite the consistently small *dMACS* effect sizes, there was a noteworthy impact of noninvariance but only at the latent level of broad externalizing. Specifically, between 25-30% of observed

mean differences were due to measurement invariance. Substantively, this suggests that comparisons between the broad externalizing dimension at baseline and later time points will overestimate mean differences if this bias is not taken into account. These results highlight a well-known, but infrequently examined, implication of measurement noninvariance [?]—that the cumulative effects of noninvariant indicators can lead to larger biasing effects at the latent level (observed for broad externalizing), or in other cases, indicators can cancel out, leading to almost no bias at the latent level (observed for neurodevelopmental problems). Results underscore the importance of examining the effect of longitudinal noninvariance at the latent *and* indicator level [?]. Future work examining the latent broad externalizing factor over time would need to consider the biasing impact of noninvariance. For other externalizing dimensions, the impact of noninvariance was trivial (neurodevelopmental problems) or there was evidence in support of strong measurement invariance (narrow externalizing, irritability), suggesting the examination of their trajectories will reflect true changes and not measurement artifacts.

Limitations While the current study addresses notable gaps in the literature, results should be considered in light of several noteworthy limitations. First, as is the case with any factor analytic technique, the solution is dependent on the input [?]. That is, our results assume that the CBCL items provide sufficient coverage of the externalizing dimension. While the CBCL is well-validated and extensively used in developmental samples, insufficient coverage of any specific content (e.g., lying) and/or low base rates of less developmentally salient behaviors (e.g., substance use) limit our ability to identify such dimensions. This point is important to keep in mind when drawing comparisons with other studies that have utilized different measures of externalizing behaviors [e.g., ?] as they may differ in content coverage that generate divergent factor solutions, particularly at more fine-grained levels.

Second, externalizing dimensions were assessed using a single informant (i.e., parent-report) and it is perhaps unsurprising that associations with external clinical correlates were strongest for parent-reported indices. Given that cross-informant reliability for externalizing psychopathology in youth is consistently low [?], it will be particularly important to examine the structure of externalizing dimensions in the ABCD study using alternative informants (e.g., youth, teacher). Additionally, expanding beyond monomethod assessments and incorporating multiple