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

Colin E. Vize1

Whitney R. Ringwald1

Emily R. Perkins2

Rebecca Waller2

Samuel W. Hawes3

Amy L. Byrd4

*University of Pittsburgh, Department of Psychology1*

*University of Pennsylvania, Department of Psychology2*

*Florida International University, Department of Psychology3*

*University of Pittsburgh, Department of Psychiatry4*

*Funding*: Support for this research was provided by the National Institute of Mental Health (K01 MH130746, Vize; K01119216, Byrd) and a postdoctoral fellowship from the University of Pennsylvania Mind Center for Outreach, Research, and Education (MindCORE) (Perkins).

Corresponding Author: Colin E. Vize, University of Pittsburgh, Pittsburgh, PA 15213. Email: [cev18@pitt.edu](mailto:cev18@pitt.edu)

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentTM (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 U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. 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/>. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from Release 5.0; <https://nda.nih.gov/study.html?id=2147>.

**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*: Hierarchical Taxonomy of Psychopathology; externalizing; youth psychopathology; longitudinal measurement invariance

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 (Eaton et al., 2023). One popular framework, the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017), 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; American Psychiatric Association, 2011), including high rates of within-disorder heterogeneity and comorbidity (Conway et al., 2022). Thus, the HiTOP aims to generate a nosological framework that is empirically valid and capable of benefiting both researchers and clinicians (Kotov et al., 2021). 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., Balling et al., 2023).

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 (Tackett & Hallquist, 2022).

**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 (Haslam et al., 2012). 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 (Lahey et al., 2017). 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 (Krueger et al., 2021). 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., Krueger et al., 2007), while the lower order spectra, including disinhibition and antagonism, offer greater specificity and show differential associations with external correlates (Mullins-Sweatt et al., 2022). For example, antagonism is more strongly related to various forms of aggression and antisocial behavior relative to disinhibitory symptoms (Vize et al., 2018), while disinhibitory symptoms are more closely related to problematic substance use (e.g., Venables & Patrick, 2012). While there is substantial evidence supporting the presence of a broad externalizing dimension in youth (e.g., Bongers et al., 2004), 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.

**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., Hill et al., 2006) and early adolescence (e.g., Goulter et al., 2022). 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., King et al., 2018), personality features (e.g., Tackett et al., 2014), and biologically-based indicators like salivary cortisol levels (e.g., Shirtcliff et al., 2005). 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., De Fruyt & De Clercq, 2014). Taken together, this work suggests that dimensional, hierarchical approaches like HiTOP can be successfully applied to youth populations—and, moreover, that HiTOP 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, Michelini and colleagues (2019) conducted a “bass-ackwards” factor analysis (Goldberg, 2006) using 102 items of the parent-report Child Behavior Checklist (Achenbach, 1999). 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, Forbes and colleagues (2023) 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.[[1]](#footnote-1) 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.

**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 (Vandenberg & Lance, 2000). 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 (Petersen et al., 2020), 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.

**Current Study**

The current preregistered study sought to build on recent efforts to extend the HiTOP model to developmental samples (e.g., Forbes et al., 2023), 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.

**Method**

**Transparency and Openness**

The analysis plan was preregistered and can be found at <https://osf.io/pqdfa/?view_only=a51ffc005e4f41daac4a3b804d23ddb2>. Deviations from our preregistered plan are reported in the supplement. R and MPlus code to reproduce our analyses are available at <https://osf.io/ec36x/?view_only=d63a1452f133421b9e1bef34a41675f1>. 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.

**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/>).

**Measures**

**Externalizing Dimensions (Parent-Report: T1-T4).** Externalizing psychopathology was assessed using a subset of items from the parent-report Child Behavior Checklist (CBCL; Achenbach, 1999), 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 Michelini et al. (2019) 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; Kobak et al., 2013). 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 prosocial 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 Luciana et al., 2018 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.

**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 Forbes’ (2023) recently developed extension to Goldberg’s “bass-ackwards” analysis (Goldberg, 2006) to identify the hierarchical structure of externalizing items/composites using the minimum residual factor extraction method and promax rotation. Forbes’ (2023) 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., Michelini et al., 2019), 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.[[2]](#footnote-2) 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*; Nye & Drasgow, 2011). 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.   

***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 (-) associations 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.

**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.[[3]](#footnote-3) 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 (*rs*=.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 (s=.58-.89) in the configural models, as well as strong stability over time (*r*s=.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.[[4]](#footnote-4) This model demonstrated excellent overall fit (=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 (ΔCFI=.003; ΔNCI=.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 (ΔCFI=.002; ΔNCI=.005), indicating weak measurement invariance. However, after imposing constraints on the intercepts of the parcel indicators (strong invariance), model fit was significantly negatively impacted (ΔCFI=.006; ΔNCI=.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 (=524.25; *df=*88; RMSEA=.020; CFI=.993; TLI=.991) and no decrement in model fit compared to the weak invariance model (ΔCFI=.002; ΔNCI=.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).[[5]](#footnote-5) 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 (Krueger et al., 2021), 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 associated with aggressive behavior and meanness toward others (Forbes et al., 2023). 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 (Burt, 2012). 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 (Leibenluft & Stoddard, 2013). 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 (Mullins-Sweat et al., 2022). 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., Burke, 2012). 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 (Finlay-Jones et al., 2023). 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., Clark et al., 2023). Nonetheless, the presence of such a factor echoes prior work on the structure of psychopathology in youth (Michelini et al., 2019; Holmes et al., 2021) and underscores the importance of this dimension in developmental 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 (Astle et al., 2022). Past research has also emphasized the importance of neurodevelopmental problems in the onset and persistence of antisocial behavior (van Goozen et al., 2022), 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 (Cardinale & Marsh, 2020) and demonstrate negative relations with with anxiety specifically (e.g., Frick et al., 1999). 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 (Finlay-Jones et al., 2023). 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 (Zelazo, 2020).

**Longitudinal Measurement Invariance for Externalizing Dimensions**

Tests of longitudinal measurement invariance showed that strict invariance held for the narrow externalizing dimension[[6]](#footnote-6) 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 (Grimm et al., 2017). 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 (Widaman et al., 2010). 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 (Petersen et al., 2020). 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., King et al. 2018) and simulation-based investigations have found that using partially invariant models for subsequent analyses does not result in biased parameters in common applied settings (Shi et al., 2019). 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.[[7]](#footnote-7) First, all *dMACS* for the broad externalizing and neurodevelopmental problem models (i.e., the two models where noninvariant indicators were detected) fell below our preregistered 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 (Vandenberg & Lance, 2000)—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 (Clark & Donnellan, 2021). 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 (Fabrigar et al., 1999). 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., Forbes et al., 2023) 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 (Reyes & Kazdin, 2004), 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 levels of analysis to more precisely conceptualize psychopathology constructs dimensions will be important (e.g., Joyner & Perkins, 2023).

Finally, while the present study made use of recommended best-practices when using parcels (see Sterba & Rights, 2023), including a transparent preregistration, empirically identified unidimensional factors, and inclusion of sensitivity analyses examining the impact of item-parcel allocation variability, the use of parcels has well-documented limitations. For example, their use necessarily overlooks fine-grained psychometric information about specific items. Though our focus was on latent externalizing dimensions and not item psychometrics per se, item-level characteristics are valuable to understand and future work could aim to better assess this and further refine the assessment of externalizing dimensions.

**Future Directions**

An important goal is to further expand investigations into the HiTOP in youth to better understand the lower-order structure of psychopathology symptoms. Our findings are in line with emerging evidence suggesting that the broad externalizing dimension of HiTOP is compatible across adults and youth (e.g., Michelini et al., 2019; Forbes et al., 2023); however, organization of finer-grained dimensions is less consistent. A valuable question for further research would be to examine the replicability of the finer-grained dimensions of externalizing symptoms identified here, using alternative assessments, informants, and developmental timepoints.

Future research can also examine other factors relevant to measurement invariance (e.g., sex, ethnicity) using the measurement models developed in the present study. However, subgroup analyses using the parcel developed in the present study would necessitate further sensitivity analyses, as within-sample item-parcel allocation variability is proportional to sampling error (Sterba & MacCallum, 2010), meaning that our use of the total sample in the ABCD study helped minimize item-parcel allocation variability.

Last, future work can also examine alternative hierarchical modeling frameworks. Our bass-ackwards approach can be contrasted with higher-order factor models estimated from ABCD study data (e.g., Clark et al., 2023) in terms of how variance is partitioned—in higher-order factor models, each level of the hierarchy is based on different variances. For example, the broad externalizing dimension in a higher-order model would represent the shared variance among narrow externalizing, irritability, and neurodevelopmental problems dimensions. In the bass-ackwards approach, the broad externalizing factor reflects the shared variance of all CBCL items. These approaches provide different information and future work can expand on the present results by investigating the use of higher-order factor models.

**References**

Achenbach, T. M. (1999). The Child Behavior Checklist and related instruments. In *The use of psychological testing for treatment planning and outcomes assessment, 2nd ed* (pp. 429–466). Lawrence Erlbaum Associates Publishers.

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry*, *63*(4), 397–417.

Balling, C. E., South, S. C., Lynam, D. R., & Samuel, D. B. (2023). Clinician Perception of the Clinical Utility of the Hierarchical Taxonomy of Psychopathology (HiTOP) System. *Clinical Psychological Science*.

Bongers, I. L., Koot, H. M., Van Der Ende, J., & Verhulst, F. C. (2004). Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development*, *75*(5), 1523–1537.

Burke, J. D. (2012). An affective dimension within oppositional defiant disorder symptoms among boys: Personality and psychopathology outcomes into early adulthood. *Journal of Child Psychology and Psychiatry*, *53*(11), 1176–1183.

Burt, S. A. (2012). How do we optimally conceptualize the heterogeneity within antisocial behavior? An argument for aggressive versus non-aggressive behavioral dimensions. *Clinical Psychology Review*, *32*(4), 263–279.

Cardinale, E. M., & Marsh, A. A. (2020). The reliability and validity of the Inventory of Callous Unemotional Traits: A meta-analytic review. *Assessment, 27*(1), 57-71.

Clark, D. A., Hicks, B. M., Angstadt, M., Rutherford, S., Taxali, A., Hyde, L., Weigard, A., Heitzeg, M. M., & Sripada, C. (2023). The General Factor of Psychopathology in the Adolescent Brain Cognitive Development (ABCD) Study: A Comparison of Alternative Modeling Approaches. *Clinical Psychological Science*, *9*(2), 169-182.

Clark, D. A., & Donellan, M. B. (2021). What if apples become oranges? A primer on measurement invariance in repeated measures research. In J. F. Rauthmann (Ed.) *The Handbook of Personality Dynamics and Processes*, pp. 837-854. Elsivier.

Conway, C. C., Kotov, R., Krueger, R., & Caspi, A. (2022). Translating the Hierarchical Taxonomy of Psychopathology (HiTOP) from potential to practice: Ten research questions. *American Psychologist*.

De Fruyt, F., & De Clercq, B. (2014). Antecedents of personality disorder in childhood and adolescence: Toward an integrative developmental model. *Annual Review of Clinical Psychology*, *10*, 449–476.

Eaton, N. R., Bringmann, L. F., Elmer, T., Fried, E. I., Forbes, M. K., Greene, A. L., Krueger, R. F., Kotov, R., McGorry, P. D., Mei, C., & Waszczuk, M. A. (2023). A review of approaches and models in psychopathology conceptualization research. *Nature Reviews Psychology*.

Fabrigar, L. R., Wegener, D. T., Link to external site, this link will open in a new tab, MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299.

Finlay-Jones, A. L., Ang, J. E., Brook, J., Lucas, J. D., MacNeill, L. A., Mancini, V. O., Kottampally, K., Elliott, C., Smith, J. D., & Wakschlag, L. S. (2023). Systematic review and meta-analysis: Early irritability as a transdiagnostic neurodevelopmental vulnerability to later mental health problems. *Journal of the American Academy of Child & Adolescent Psychiatry*.

Forbes, M. K. (2023). Improving hierarchical models of individual differences: An extension of Goldberg’s bass-ackward method. *Psychological Methods*, 1–20.

Forbes, M. K., Watts, A. L., Twose, M., Barrett, A., Hudson, J., Lyneham, H., McLellan, L., Newton, N., Sicouri, G., Chapman, C., McKinnon, A., Rapee, R., Slade, T., Teesson, M., Markon, K., & Sunderland, M. (2023). A Hierarchical Model of the Symptom-Level Structure of Psychopathology in Youth. *PsyArXiv*. <https://doi.org/10.31234/osf.io/7kcfz>

Frick, P. J., Lilienfeld, S. O., Ellis, M., Loney, B., & Silverthorn, P. (1999). The association between anxiety and psychopathy dimensions in children. *Journal of Abnormal Child Psychology, 27*, 383-392.

Goldberg, L. R. (2006). Doing it all Bass-Ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality*, *40*(4), 347–358.

Goulter, N., McMahon, R. J., Lansford, J. E., Bates, J. E., Dodge, K. A., Max Crowley, D., & Pettit, G. S. (2022). Externalizing psychopathology from childhood to early adolescence: Psychometric evaluation using latent variable and network modeling. *Psychological Assessment*, *34*(11), 1008–1021.

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth Modeling: Structural Equation Modeling and Multilevel Modeling Approaches*. Guilford Press.

Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine*, *42*, 903–920.

Hill, A. L., Degnan, K. A., Calkins, S. D., & Keane, S. P. (2006). Profiles of externalizing behavior problems for boys and girls across preschool: The roles of emotion regulation and inattention. *Developmental Psychology*, *42*(5), 913–928.

Holmes, J., Link to external site, this link will open in a new tab, Mareva, S., Link to external site, this link will open in a new tab, Bennett, M. P., Link to external site, this link will open in a new tab, Black, M. J., & Guy, J. (2021). Higher-order dimensions of psychopathology in a neurodevelopmental transdiagnostic sample. *Journal of Abnormal Psychology*, *130*(8), 909–922.

Joyner, K. J., & Perkins, E. R. (2023). Challenges and ways forward in bridging units of analysis in clinical psychological science. *Journal of Psychopathology and Clinical Science*, *132*(7), 888–896.

King, K. M., Luk, J. W., Witkiewitz, K., Racz, S., McMahon, R. J., & Wu, J. (2018). Externalizing Behavior Across Childhood as Reported by Parents and Teachers: A Partial Measurement Invariance Model. *Assessment*, *25*(6), 744–758.

Kobak, K. A., Kratochvil, C. J., Stanger, C., & Kaufman, J. (2013). Computerized screening of comorbidity in adolescents with substance or psychiatric disorders. Paper presented: Anxiety and Depression: Technology and New Media in Practice and Research; La Jolla, CA.

Kotov, R., Krueger, R. F., Watson, D., Bagby, M., Carpenter, W. T., & Caspi, A. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477.

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., Deyoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweatt, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. C. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A Quantitative Nosology Based on Consensus of Evidence. *Annual Review of Clinical Psychology*, *17*, 83–108.

Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Keyes, K. M., Latzman, R. D., Michelini, G., Patrick, C. J., Sellbom, M., Slade, T., South, S. C., Sunderland, M., Tackett, J., Waldman, I., Waszczuk, M. A., … Kotov, R. (2021). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry*, *20*, 171–193.

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking Antisocial Behavior, Substance Use, and Personality: An Integrative Quantitative Model of the Adult Externalizing Spectrum. *Journal of Abnormal Psychology*, *116*, 645–666.

Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). A hierarchical causal taxonomy of psychopathology across the life span. *Psychological Bulletin*, *143*(2), 142–186.

Leibenluft, E., & Stoddard, J. (2013). The developmental psychopathology of irritability. *Development and Psychopathology, 25*(402), 1473–1487.

Luciana, M., Bjork, J. M., Nagel, B. J., Barch, D. M., Gonzalez, R., Nixon, S. J., & Banich, M. T. (2018). Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Developmental Cognitive Neuroscience*, *32*, 67–79.

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

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance. *Journal of Applied Psychology*, *93*(3), 568–592.

Michelini, G., Barch, D. M., Tian, Y., Watson, D., Klein, D. N., & Kotov, R. (2019). Delineating and validating higher-order dimensions of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study. *Translational Psychiatry*, *9*(1), 21–25.

Mullins-Sweatt, S. N., Bornovalova, M. A., Carragher, N., Clark, L. A., Corona Espinosa, A., Jonas, K., Keyes, K. M., Lynam, D. R., Michelini, G., Miller, J. D., Min, J., Rodriguez-Seijas, C., Samuel, D. B., Tackett, J. L., & Watts, A. L. (2022). HiTOP Assessment of Externalizing Antagonism and Disinhibition. *Assessment*, *29*(1), 34–45.

Muthén, B. O., & Muthén, L. K. (2017). *Mplus User’s Guide*. Version 8.7.

Nye, C. D., & Drasgow, F. (2011). Effect Size Indices for Analyses of Measurement Equivalence: Understanding the Practical Importance of Differences Between Groups. *Journal of Applied Psychology*, *96*(5), 966–980.

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

Reyes, A., & Kazdin, A. E. (2004). Measuring Informant Discrepancies in Clinical Child Research. Psychological Assessment, 16(3), 330–334.

Shi, D., Song, H., & Lewis, M. D. (2019). The Impact of Partial Factorial Invariance on Cross-Group Comparisons. *Assessment*, *26*(7), 1217–1233.

Shirtcliff, E. A., Granger, D. A., Booth, A., & Johnson, D. (2005). Low salivary cortisol levels and externalizing behavior problems in youth. *Development and Psychopathology*, *17*(01).

Sterba, S. K., & MacCallum, R. C. (2010). Variability in parameter estimates and model fit across repeated allocations of items to parcels. *Multivariate Behavioral Research*, *45*(2), 322–358.

Sterba, S. K., & J. D. Rights. (2023). Item parceling in SEM: A researcher degree‑of‑freedom ripe for opportunistic use. In Hoyle, R. H. (Ed.). *Handbook of Structural Equation Modeling*, pp. 296-315. Guilford Publications.

Tackett, J. L., Hallquist, M., & Link to external site, this link will open in a new tab. (2022). The need to grow: Developmental considerations and challenges for modern psychiatric taxonomies. *Journal of Psychopathology and Clinical Science*, *131*(6), 660–663.

Tackett, J. L., Herzhoff, K., Reardon, K. W., De Clercq, B., & Sharp, C. (2014). The externalizing spectrum in youth: Incorporating personality pathology. *Journal of Adolescence*, *37*(5), 659–668.

van Goozen, S. H. M., Langley, K., & Hobson, C. W. (2022). Childhood Antisocial Behavior: A Neurodevelopmental Problem. *Annual Review of Psychology*, *73*(1), 353–377.

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70.

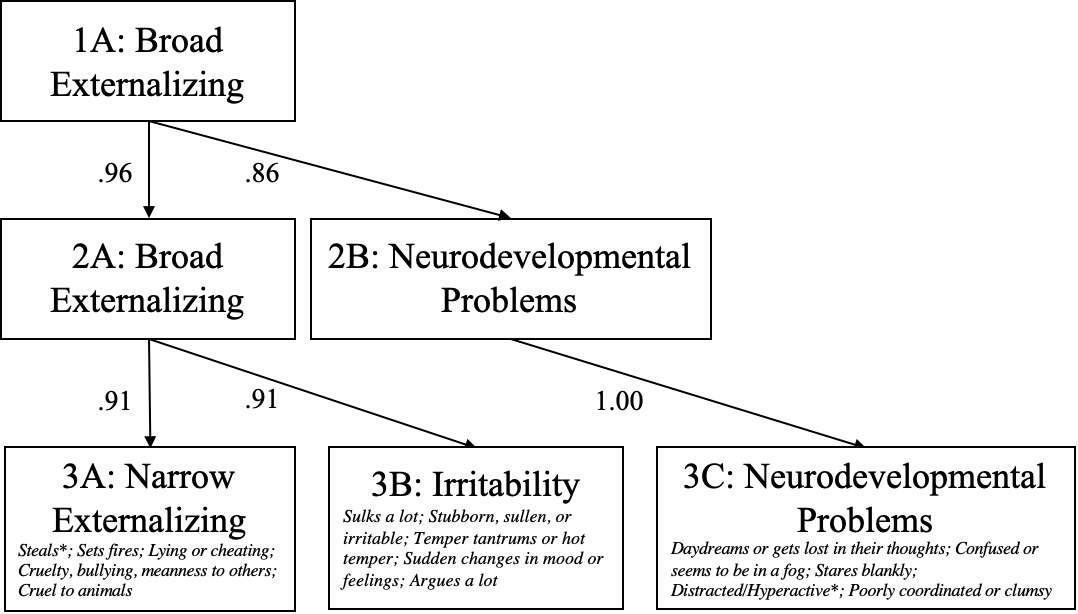
Venables, N. C., & Patrick, C. J. (2012). Validity of the Externalizing Spectrum Inventory in a Criminal Offender Sample: Relations with Disinhibitory Psychopathology, Personality, and Psychopathic Features. *Psychological Assessment*, *24*(1), 88–100.

Vize, C. E., Miller, J. D., & Lynam, D. R. (2018). FFM facets and their relations with different forms of antisocial behavior: An expanded meta-analysis. *Journal of Criminal Justice*, *57*, 67–75.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial Invariance Within Longitudinal Structural Equation Models: Measuring the Same Construct Across Time. *Child Development Perspectives*, *4*(1), 10–18.

Zelazo, P. D. (2020). Executive Function and Psychopathology: A Neurodevelopmental Perspective. *Annual Review of Clinical Psychology*, *16*(1), 431–454.

Figure 1. Results of Bass-ackwards Analyses Using CBCL Externalizing Items



*Note*: Correlations between factor scores at subsequent levels are included in the figure; the five CBCL items with highest loadings on respective factors at level 3 are in italics; \*=composite of multiple CBCL items.

Table 1. Correlations Between Bass-ackwards Factor Scores and External Criterion Measures

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Factor Scores** | | | |
|  | *A1: Broad Externalizing* | *C1: Narrow Externalizing* | *C2: Irritability* | *C3: Neurodevelopmental* |
| *Symptom Counts (Parent-Report)* |  |  |  |  |
| Conduct Disorder | .52 | .60a (.59) | .39b (.04) | .31c (-.02) |
| Oppositional Defiant Disorder | .52 | .45a (.21) | .53b (.43) | .32c (.-.04) |
| ADHD | .53 | .41a (.16) | .38b (-.01) | .57c (.50) |
| Major Depressive Disorder | .34 | .26a (.07) | .32b (.20) | .28a (.12) |
| Suicidality(Self-Harm | .23 | .20a (.11) | .20a (.09) | .19a (.08) |
| Generalized Anxiety Disorder | .21 | .09a (-.12) | .23b (.21) | .22b (.16) |
| Social Anxiety Disorder | .09 | .04a (-.05) | .08b (.03) | .13c (.14) |
| *Symptom Counts (Youth-Report)* |  |  |  |  |
| Major Depressive Disorder | .13 | .13a (.10) | .09b (.00) | .10b (.05) |
| Suicidality (Self-Harm | .12 | .11a (.07) | .09a (.01) | .11a (.06) |
| Generalized Anxiety Disorder | .06 | .04a (.00) | .05a (.03) | .06a (.04) |
| Social Anxiety Disorder | .05 | .04a (.02) | .04a (.01) | .05a (.03) |
| *Prosocial Behavior* (*Parent-Report*) | -.35 | -.33a (-.21) | -.32a (-.18) | -.22b (-.01) |
| *Prosocial Behavior* (*Youth-Report*) | -.09 | -.09a (-.07) | -.07ab (-.01) | -.06b (-.02) |
| *Impulsivity (Youth-Report)* |  |  |  |  |
| Negative Urgency | .17 | .16a (.10) | .15a (.08) | .11b (.02) |
| Positive Urgency | .15 | .15a (.13) | .10b (-.02) | .12b (.07) |
| Lack of Premeditation | .16 | .15a (.09) | .12b (.02) | .14ab (.08) |
| Lack of Perseverance | .15 | .12a (.03) | .10a (-.02) | .18b (.18) |
| Sensation Seeking | .04 | .06a (.08) | .02b (-.03) | .02b (-.01) |
| *Fluid Intelligence* | -.15 | -.14a (-.10 | -.08b (.08 | -.17c (-.17) |
| *Note*: Symptom counts are based on past and current symptom counts endorsed from the KSADS-5 interview (ADHD=attention deficit hyperactivity disorder); Prosocial Behavior is from the SDQ; Impulsivity is from the UPPS-P; Fluid Intelligence=age-corrected composite of 5 tasks from the NIH Toolbox Cognition measures (see Method); *N*=11,470-11,840 across reported correlations; semipartial correlations (i.e., relation between the externalizing dimensions and external criteria controlling for overlap among externalizing dimensions) are in parentheses; All bivariate and semipartial correlations larger than .04 are significant at *p*<.01; Mismatching superscripts indicate that the zero-order correlations are significantly different from each other at *p* <.01 based on tests of dependent correlations. | | | | |

Table 2. Results for Longitudinal Measurement Invariance

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Broad Externalizing** | | | | | | | | | | |
| **Model** |  | *df* | AIC | BIC | RMSEA | TLI | CFI | NCI | Δ | ΔCFI | ΔNCI |
| Configural | 767.72 | 134 | 688492.79 | 689201.41 | .021 | .993 | .995 | .974 | -- | -- | -- |
| Weak | 910.07 | 146 | 688751.37 | 689371.41 | .021 | .992 | .994 | .966 | 126.20\* | .001 | .005 |
| Strong | 1569.46 | 158 | 689790.69 | 690322.15 | .027 | .987 | .989 | .898 | 1065.82\* | **.005** | **.026** |
| *Partial Strong* | 947.21 | 152 | 688775.91 | 689351.66 | .021 | .992 | .994 | .967 | 26.95\* | .000 | .000 |
|  | **Narrow Externalizing** | | | | | | | | | | |
| **Model** |  | *df* | AIC | BIC | RMSEA | TLI | CFI | NCI | Δ | ΔCFI | ΔNCI |
| Configural | 312.54 | 74 | 258611.20 | 259186.95 | .016 | .988 | .993 | .990 | -- | -- | -- |
| Weak | 338.15 | 83 | 258739.44 | 259248.76 | .016 | .989 | .992 | .989 | 30.25\* | .001 | .001 |
| Strong | 371.44 | 92 | 258749.45 | 259192.33 | .016 | .989 | .991 | .988 | 27.33\* | .001 | .001 |
| Strict | 418.21 | 104 | 259021.15 | 259375.46 | .016 | .989 | .990 | .987 | 47.40\* | .001 | .001 |
|  | **Irritability** | | | | | | | | | | |
| **Model** |  | *df* | AIC | BIC | RMSEA | TLI | CFI | NCI | Δ | ΔCFI | ΔNCI |
| Configural | 166.25 | 74 | 347238.66 | 347814.41 | .010 | .998 | .999 | .996 | -- | -- | -- |
| Weak | 222.48 | 83 | 347325.62 | 347834.94 | .012 | .997 | .998 | .994 | 50.40\* | .001 | .002 |
| Strong | 283.71 | 92 | 347393.79 | 347836.67 | .013 | .996 | .997 | .991 | 56.22\* | .001 | .003 |
|  | **Neurodevelopmental Problems** | | | | | | | | | | |
| **Model** |  | *df* | AIC | BIC | RMSEA | TLI | CFI | NCI | Δ | ΔCFI | ΔNCI |
| Configural | 264.19 | 74 | 409170.94 | 409746.69 | .015 | .995 | .997 | .992 | -- | -- | -- |
| Weak | 387.31 | 83 | 409373.45 | 409882.76 | .018 | .993 | .995 | .987 | 108.73\* | .002 | .005 |
| Strong | 840.08 | 92 | 410069.50 | 410512.39 | .026 | .985 | .989 | .969 | 720.23\* | **.006** | **.018** |
| *Partial Strong* | 524.25 | 88 | 409576.47 | 410048.88 | .020 | .991 | .993 | .982 | 110.15\* | .002 | .005 |
| *Note:* *Partial Strong=*denotes a measurement model where some model constraints of the strong invariance model have been freed (see Results); *df*=degrees of freedom; AIC=Akaike information criterion; BIC: Bayesian information criterion; RMSEA: Root mean square error of approximation; TLI: Tucker Lewis index; CFI: comparative fit index; NCI: McDonald’s non-centrality index; \*= Δ test is significant at *p*<.05; differences in exact values for model fit indices are due to rounding; all models were estimated using robust maximum likelihood, and model fit indices are based on the scaled values; bolded ΔCFI and ΔNCI values are above thresholds for decrement in model fit. | | | | | | | | | | | |

Table 3. Effect Sizes to Quantify the Impact of Measurement Invariance for Noninvariant Dimensions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Broad Externalizing** | | | | |
| **Comparison** | Δ*Mean* | Observed Mean Difference | % of Mean Difference | *dMACS* range |
| Time 1-Time 2 | -.24 | -.85 | 28% | .00-.06 |
| Time 1-Time 3 | -.41 | -1.57 | 26% | .00-.10 |
| Time 1-Time 4 | -.66 | -2.11 | 31% | .00-.18 |
| **Neurodevelopmental Problems** | | | | |
| **Comparison** | Δ*Mean* | Observed Mean Difference | % of Mean Difference | *dMACS* range |
| Time 1-Time 2 | .00 | -.25 | 0% | .00-.06 |
| Time 1-Time 3 | .00 | -.54 | 0% | .00-.06 |
| Time 1-Time 4 | .02 | -.71 | 3% | .00-.12 |
| *Note*: Δ*Mean =* amount of bias in latent mean differences between time points due to measurement noninvariance; Observed Mean Difference=observed mean difference between time points in the sum scores of the parcels; % of Mean Difference=percent of observed mean difference attributable to measurement noninvariance; *dMACS =* Cohen’s *d* for Mean and Covariance Structure; *dMACS* are effect sizes for indicator noninvariance and are computed based on the partial strong invariance models. Therefore, they describe the impact of allowing specific parcels to be noninvariant in their intercepts (see Results) | | | | |

**Supplementary Materials**

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

**Table of Contents**

1. **Deviations from Preregistration**
2. **Overview of 4- and 14-factor Bass-ackwards Solutions**
3. **Overview of Sensitivity Analyses**
4. **Interpreting Effect Sizes for Measurement Noninvariance**

**Supplementary Tables**

* 1. **Table S1.** CBCL Item Assignments to Item-parcels for Externalizing Dimensions
  2. **Table S2.** Descriptive Information for Broad Externalizing Item Parcels Across Assessment Timepoints
  3. **Table S3.** Descriptive Information for Narrow Externalizing Item Parcels Across Assessment Timepoints
  4. **Table S4.** Descriptive Information for Irritability Item Parcels Across Assessment Timepoints
  5. **Table S5.** Descriptive Information for Neurodevelopmental Item Parcels Across Assessment Timepoints
  6. **Table S6.** Results of Sensitivity Analyses to Examine Impact of Item-parcel Allocation Variability

**Deviations from Preregistration**

We note a few deviations from our preregistered analytical plan. First, we incorrectly stated in the preregistration that 44 items from the CBCL would be used—the correct number should have been 45. The CBCL item that was mistakenly omitted from Table 1 of the preregistration was the ‘bragging/boasting’ item. Second, nearly all parcels showed strong evidence of non-normality and were positively skewed. As a result, we used robust maximum likelihood to estimate all models and we report the appropriately scaled model fit statistics in our results. Missing data were handled using full-information maximum likelihood. Though all models were estimated in MPlus, we chose to also estimate all models using the ‘lavaan’ package because MPlus is proprietary software and not freely available. Last, our initial proposal for the number of items to be assigned to each parcel did not account for the possibility that a smaller number of items (e.g., 15 items) would meet our criteria to be used to create parcels for externalizing dimensions. Given that our proposal for the number of items per parcel was somewhat arbitrary, we opted to use a parceling strategy that ensured at least 4 indicators were used to estimate latent factors (i.e., to avoid a fully saturated model) and that each model had a similar amount of parcel indicators. This approach resulted in 5 parcels being used to model broad externalizing, and 4 parcels being used to model the more specific externalizing dimensions identified in our analyses.

We also note the use of numerous additional packages for our analyses conducted in R (R Core Team, 2023; version 4.3.1) with RStudio (Posit Team, 2023; version 2023.6.1.2023). They include the ‘MPlusAutomation’ package (Hallquist & Wiley, 2018) ‘psych’ (Revelle, 2023), ‘tidyverse’ (Wickham et al., 2018), ‘lavaan’ (Rosseel, 2012), ‘dmacs’ (Dueber, 2023) and ‘semTools’ (Jorgensen et al., 2022).

**Overview of 4- and 14-factor Bass-ackwards Solutions**

As noted in the manuscript, MAP suggested four factors be extracted. Examination of the fourth factor showed that it was comprised of four items (range of standardized loadings=.32-.61), two of which (*Talks too much; Impulsive or acts without thinking*) had notable secondary loadings on other factors. The other two items with primary loadings and no secondary loadings were related to arrogance (*Bragging, boasting; Showing off or clowning*).

We hypothesized that disinhibition and antagonism factors would emerge in our analysis, but these factors were not present in the three- or four-factor bass-ackwards solutions. We explored the separate 14-factor solution (i.e., the number of factors parallel analysis suggested for extraction) to see if antagonism or disinhibition factors emerged at finer levels of the hierarchy. The 14-factor solution also showed that no such factors could be identified at any level of the bass-ackwards hierarchy. However, at the eighth level, a factor emerged comprised of five items more reflective of disinhibitory content (e.g., *Distracted/Hyperactive* composite, *fails to finish things they start, impulsive or acts without thinking*) but these items also showed equivalent loadings on the neurodevelopmental factor at the eight level, and the neurodevelopmental problems factor remained consistent across all levels of the hierarchy. Similarly, an interpretable antagonism factor did not emerge in the 14-factor solution, contrary to our hypothesis. Instead, the narrow externalizing factor identified at level 3 remained consistent across nearly every level of the hierarchy.

**Overview of Sensitivity Analyses**

We conducted a series of sensitivity analyses to examine the robustness of our results to various analytic choices. First, we conducted our bass-ackwards analysis using an alternative factoring method and alternative rotation given that there is some debate about how to best model ordinal data in factor analytic frameworks. Using weighted least squares extraction and an oblique equamax rotation (compared to our original analysis using minimum residual extraction and a promax rotation), we extracted four factors and compared their similarity to our original four-factor solution using congruence coefficients. The results showed near perfect factor congruence (all Tucker Congruence Coefficients=.99) when comparing the original solution to the weighted least squares solution.

Second, researchers have highlighted that variability in item-parcel allocation (i.e., variability that arises from the random assignment of items to parcels for structural equation modeling) can have a sizeable impact on model parameters and overall model fit (e.g., Sterba & Rights, 2017). Thus, we examined how variability in item-parcel assignments may have influenced our results. To do so, we estimated models using the baseline parent-reported CBCL data for each of the externalizing dimensions (broad externalizing, narrow externalizing, irritability, and neurodevelopmental problems), specifying the same number of parcel-based indicators for each dimension (i.e., five parcels for broad externalizing, and four parcels for the other three dimensions). Next, the respective CBCL items were randomly assigned to parcels, and the CFA model was fit to the data using robust maximum likelihood estimation. This process was repeated 500 times, and factor loadings and model fit indices were saved after each iteration. Average values, standard deviations, and minimum and maximum values for parcel factor loadings as well as for overall model fit (RMSEA, CFI, and TLI) are reported in **Supplementary Table S6**. Results showed modest variability in estimates of factor loadings and model fit indices, and there was strong consistency in factor loadings and model fit indices for each externalizing dimension indicating that our results were not dependent on the particular item-parcel assignment used in our primary analyses.

Third, we also examined whether our tests of longitudinal measurement invariance were impacted by the clustering effects in the ABCD study. Specifically, we reexamined our measurement invariance models including clustering effects of family and study site (see MPlus output on OSF page: <https://osf.io/ec36x/?view_only=d63a1452f133421b9e1bef34a41675f1>). These clustering variables were examined separately using MPlus. There was no effect of family on our results, and while incorporating the clustering effect of site did result in small improvements in overall model fit, it did not have an impact on relative fit comparisons for our tests of longitudinal measurement invariance.

**Interpreting Effect Sizes for Measurement Noninvariance**

Measurement invariance effect sizes (see Meade, 2010 and Gunn et al., 2020, for overviews) quantify the degree of noninvariance at either the level of the indicator or the level of the latent dimension. More specifically, while typical confirmatory factor analysis methods for testing noninvariance can detect nonequivalence in measurement across groups or across time, they cannot determine the magnitude of the nonequivalence nor its impact on specific research questions (e.g., how much will noninvariance bias a mean comparison across time?). One specific effect size measure for continuous indicators is *dMACS* (Nye & Drasgow, 2011), which quantifies the shared impact of weak and strong noninvariance at the level of an indicator. In other words, *dMACS* describes how large of a difference there is in an indicator’s factor loading *plus* the difference in the indicator’s interceptor. This combined difference is expressed in the metric of Cohen’s *d*, and researchers have provided informed benchmarks for interpreting small to large effects in the context of measurement noninvariance testing (Nye et al., 2019). The *dMACS* effect size complements typical measurement invariance testing (i.e., nested model comparisons) and provides additional information about the magnitude of noninvariance. This is akin to using effect sizes to complement null hypothesis significance testing, since statistical significance is not the same as practical significance.

Importantly, the *dMACS* effect sizes reported in the manuscript are derived from partial strong invariance models for the broad externalizing factor and neurodevelopmental problems factor. Since weak invariance was established for these models (i.e., all factor loadings were constrained to be equal in the partial strong invariance models), the *dMACS* effect sizes only reflect differences in intercepts. Moreover, only intercepts for indicators that were allowed to freely vary could have non-zero *dMACS* because all other intercepts had been constrained to be equal, and thus no differences could be present across time points. The parcels with freely estimated intercepts were parcels 4 and 5 for broad externalizing, and parcels 3 and 4 for neurodevelopmental problems. Thus, the results in **Table 3** in the manuscript describe the impact of allowing these parcels to be noninvariant in their intercepts.

However, one shortcoming of *dMACS* is that it is always a positive value (Nye & Drasgow, 2011; but see Nye et al., 2019 and Gunn et al., 2020 for effect sizes that can be either positive or negative). Thus, *dMACS* does not provide information about the direction of bias due to noninvariance. This may obscure how noninvariance at the indicator level impacts the latent dimension and subsequent mean comparisons across time. To examine the effect of noninvariance at the latent level in the partial strong measurement models, we computed the amount of observed mean difference between time points that could be attributed to noninvariance (Δ*Mean* in **Table 3**).[[8]](#footnote-8) Δ*Mean* reflects the sum of all indicator-level differences, so bias in opposite directions can cancel out at the latent level, or bias in similar directions can accumulate at the latent level. Both patterns occurred in our analyses.

For example, when comparing broad externalizing at Time 1 to Time 4, the observed *dMACS* were .16 and .18 for parcels 4 and 5, respectively. Though these effect sizes are small and similar in magnitude to one another, bias accumulated because the intercepts of both parcels 4 and 5 were noninvariant in the same direction.[[9]](#footnote-9) That is, when the impact of noninvariance for these two parcels is combined at the latent level of broad externalizing, noninvariance is more impactful. To estimate the true mean change over time, the degree of change due to invariance needs to be taken into account. For example, the observed change in latent broad externalizing between Time 1 and Time 4 was a decrease of 2.11 points, but .66 points of that decrease was due to noninvariant intercepts of parcels 4 and 5. Thus, the adjusted degree of change, taking into measurement noninvariance, is a decrease of 1.45 points.

This can be contrasted with the results for the neurodevelopmental problems model. In this model, parcels 3 and 4 also showed small *dMACS* effect sizes and they were also of equal magnitude (both parcels had *dMACS* =.12). However, the intercepts of these parcels were biased in opposite directions. When summed, the biasing effects of noninvariant intercepts canceled out resulting in almost no bias at the latent level. For example, mean differences between Time 1 and Time 4 showed that only 2% of the mean difference was attributable to measurement noninvariance. This is a trivial degree of impact, and it is unlikely to influence any substantive inferences involving mean change over time. Thus, while the *dMACS* effect sizes for the broad externalizing and neurodevelopmental problems dimensions were small and seemingly comparable to one another, the direction of the noninvariance resulted in notable differences at the latent level that have important implications for comparing means over time. These results highlight the benefit of considering effect sizes of measurement noninvariance and their impact on mean comparisons across groups or time.

References

Clark, D. A., Listro, C. J., Lo, S. L., Durbin, C. E., Donnellan, M. B., & Neppl, T. K. (2016). Measurement invariance and child temperament: An evaluation of sex and informant differences on the Child Behavior Questionnaire. *Psychological Assessment, 28*(12), 1646–1662.

Dueber, D. (2023). dmacs: Measurement Nonequivalence Effect Size Calculator (R package version 0.1.0.9002). Retrieved from <https://github.com/ddueber/dmacs>.

Gunn, H. J., Grimm, K. J., & Edwards, M. C. (2020). Evaluation of Six Effect Size Measures of Measurement Non-Invariance for Continuous Outcomes. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(4), 503–514.

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural Equation Modeling, 25*, 621-638.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling (R package version 0.5-6). Retrieved from [https://CRAN.R-project.org/package=semTools](https://cran.r-project.org/package=semTools).

Meade, A. W. (2010). A taxonomy of effect size measures for the differential item functioning of items and scales. *Journal of Applied Psychology, 95*, 728–743.

Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., & Drasgow, F. (2019). How Big Are My Effects? Examining the Magnitude of Effect Sizes in Studies of Measurement Equivalence. *Organizational Research Methods, 22*(3), 678–709.

Nye, C., & Drasgow, F. (2011). Effect Size Indices for Analyses of Measurement Equivalence: Understanding the Practical Importance of Differences Between Groups. *The Journal of Applied Psychology*, *96*, 966–980.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36.

Sterba, S. K., & Rights, J. D. (2017). Effects of parceling on model selection: Parcel-allocation variability in model ranking. *Psychological Methods, 22*(1), 47–68.

Revelle, W. (2023). psych: Procedures for Psychological, Psychometric, and Personality Research (R package version 2.3.6). Retrieved from https://CRAN.R-project.org/package=psych.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel, DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software, 4(*43), 1686.

Table S1. CBCL Item Assignments to Item-parcels for Externalizing Dimensions

|  |  |
| --- | --- |
| **Broad Externalizing** | |
| *Parcel Number (Number of items)* | *CBCL Item Content* |
| 1 (9 items) | Repeats certain acts over and over, compulsions |
| Disobey composite |
| Sets fires |
| Stares blankly |
| Can’t get their mind off certain thoughts, obsessions |
| Strange ideas |
| Poor school work |
| Cruel to animals |
| Runs away from home |
| 2 (9 items) | Daydreams or gets lost in their thoughts |
| Teases a lot |
| Thinks about sex too much |
| Sudden changes in mood or feelings |
| Acts too young for their age |
| Prefers being with younger kids |
| Impulsive or acts without thinking |
| Lying or cheating |
| Sulks a lot |
| 3 (9 items) | Fails to finish things they start |
| Argues a lot |
| Poorly coordinated or clumsy |
| Stubborn, sullen, or irritable |
| Attack/Threatens composite |
| Gets in many fights |
| Bragging, boasting |
| Cruelty, bullying, or meanness to others |
| Hangs around with others who get in trouble |
| 4 (9 items) | Swearing or obscene language |
| Steals composite |
| Nervous movements or twitching |
| Whining |
| Distracted/Hyperactive composite |
| Showing off or clowning |
| Suspicious of others |
| Confused or seems to be in a fog |
| Demands a lot of attention |
| 5 (9 items) | Temper tantrums or hot temper |
| Feels others are out to get them |
| Doesn’t seem to feel guilty after misbehaving |
| Screams a lot |
| Talks too much |
| Gets hurt a lot, accident prone |
| Destroys composite |
| Peer problems composite |
| Easily jealous |
| **Narrow Externalizing** | |
| 1 (4 items) | Hangs around with others who get in trouble |
| Lying or cheating |
| Showing off or clowning |
| Swearing or obscene language |
| 2 (4 items) | Teases a lot |
| Disobeys composite |
| Cruelty, bullying, or meanness to others |
| Cruel to animals |
| 3 (4 items) | Runs away from home |
| Gets in many fights |
| Doesn’t seem to feel guilty after misbehaving |
| Attacks/Threatens composite |
| 4 (4 items) | Steals composite |
| Sets fires |
| Destroys composite |
| Thinks about sex too much |
| **Irritability** | |
| 1 (3 items) | Sudden changes in mood or feelings |
| Whining |
| Stubborn, sullen, or irritable |
| 2 (3 items) | Demands a lot of attention |
| Screams a lot |
| Sulks a lot |
| 3 (3 items) | Easily jealous |
| Feels others are out to get them |
| Argues a lot |
| 4 (2 items) | Temper tantrums or hot temper |
| Suspicious of others |
| **Neurodevelopmental Problems** | |
| 1 (4 items) | Acts too young for their age |
| Distracted/Hyperactive composite |
| Nervous movements or twitching |
| Can’t get their mind off certain thoughts; obsessions |
| 2 (4 items) | Prefers being with younger kids |
| Repeats certain acts over and over; compulsions |
| Stares blankly |
| Strange ideas |
| 3 (4 items) | Confused or seems to be in a fog |
| Poor school work |
| Gets hurt a lot, accident prone |
| Fails to finish things they start |
| 4 (3 items) | Poorly coordinated or clumsy |
| Daydreams or gets lost in their thoughts |
| Talks too much |
| *Note*: Some items included in the broad externalizing parcels are not included in other parcels (e.g., Impulsive or acts without thinking) due to either low primary loadings or high secondary loadings in the bass-ackwards results (see Method). | |

Table S2. Descriptive Information for Broad Externalizing Item Parcels Across Assessment Timepoints

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Baseline (*N*=11,862)** | | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 | 5 |
| 1 | .99 | 1.53 | -- |  |  |  |  |
| 2 | 1.77 | 2.20 | .74 | -- |  |  |  |
| 3 | 2.02 | 2.25 | .68 | .75 | -- |  |  |
| 4 | 1.68 | 2.08 | .74 | .77 | .73 | -- |  |
| 5 | 1.63 | 2.18 | .68 | .75 | .76 | .74 | -- |
| **1 Year Follow-up (*N*=11,203)** | | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 | 5 |
| 1 | .97 | 1.51 | -- |  |  |  |  |
| 2 | 1.74 | 1.74 | .73 | -- |  |  |  |
| 3 | 1.97 | 1.97 | .68 | .75 | -- |  |  |
| 4 | 1.52 | 1.52 | .74 | .76 | .74 | -- |  |
| 5 | 1.48 | 1.48 | .66 | .73 | .76 | .73 | -- |
| **2 Year Follow-up (*N*=10,899)** | | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 | 5 |
| 1 | .91 | 1.47 | -- |  |  |  |  |
| 2 | 1.62 | 2.14 | .73 | -- |  |  |  |
| 3 | 1.90 | 2.18 | .69 | .75 | -- |  |  |
| 4 | 1.35 | 1.92 | .74 | .77 | .74 | -- |  |
| 5 | 1.31 | 1.99 | .67 | .73 | .75 | .73 | -- |
| **3 Year Follow-up (*N*=10,099)** | | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 | 5 |
| 1 | .96 | 1.51 | -- |  |  |  |  |
| 2 | 1.64 | 2.14 | .73 | -- |  |  |  |
| 3 | 1.94 | 2.18 | .69 | .76 | -- |  |  |
| 4 | 1.29 | 1.87 | .73 | .77 | .73 | -- |  |
| 5 | 1.19 | 1.91 | .67 | .73 | .75 | .74 | -- |
| *Note*: Values are based on sum scores of the CBCL items included in the parcels; SD=Standard deviation | | | | | | | |

Table S3. Descriptive Information for Narrow Externalizing Item Parcels Across Assessment Timepoints

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Baseline (*N*=11,862)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | .24 | .64 | -- |  |  |  |
| 2 | .38 | .69 | .54 | -- |  |  |
| 3 | .56 | .97 | .57 | .60 | -- |  |
| 4 | .57 | .98 | .55 | .60 | .63 | -- |
| **1 Year Follow-up (*N*=11,203)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | .22 | .61 | -- |  |  |  |
| 2 | .34 | .65 | .52 | -- |  |  |
| 3 | .55 | .98 | .58 | .59 | -- |  |
| 4 | .51 | .93 | .55 | .59 | .65 | -- |
| **2 Year Follow-up (*N*=10,899)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | .21 | .59 | -- |  |  |  |
| 2 | .34 | .68 | .52 | -- |  |  |
| 3 | .54 | .98 | .55 | .61 | -- |  |
| 4 | .46 | .91 | .55 | .61 | .64 | -- |
| **3 Year Follow-up (*N*=10,099)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | .21 | .59 | -- |  |  |  |
| 2 | .33 | .67 | .53 | -- |  |  |
| 3 | .52 | .97 | .54 | .62 | -- |  |
| 4 | .44 | .90 | .56 | .62 | .64 | -- |
| *Note:* Values are based on sum scores of the CBCL items included in the parcels; SD=Standard deviation | | | | | | |

Table S4. Descriptive Information for Irritability Item Parcels Across Assessment Timepoints

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Baseline (*N*=11,862)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | .88 | 1.21 | -- |  |  |  |
| 2 | .56 | .98 | .65 | -- |  |  |
| 3 | .84 | 1.07 | .62 | .64 | -- |  |
| 4 | .34 | .63 | .61 | .61 | .58 | -- |
| **1 Year Follow-up (*N*=11,203)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | .82 | 1.17 | -- |  |  |  |
| 2 | .51 | .94 | .64 | -- |  |  |
| 3 | .81 | 1.05 | .62 | .63 | -- |  |
| 4 | .31 | .62 | .60 | .60 | .58 | -- |
| **2 Year Follow-up (*N*=10,899)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | .74 | 1.11 | -- |  |  |  |
| 2 | .45 | .88 | .64 | -- |  |  |
| 3 | .76 | 1.03 | .61 | .63 | -- |  |
| 4 | .27 | .58 | .59 | .62 | .59 | -- |
| **3 Year Follow-up (*N*=10,099)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | .77 | 1.13 | -- |  |  |  |
| 2 | .42 | .84 | .63 | -- |  |  |
| 3 | .75 | 1.00 | .61 | .62 | -- |  |
| 4 | .26 | .57 | .59 | .60 | .58 | -- |
| *Note:* Values are based on sum scores of the CBCL items included in the parcels; SD=Standard deviation | | | | | | |

Table S5. Descriptive Information for Neurodevelopmental Problems Item Parcels Across Assessment Timepoints

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Baseline (*N*=11,862)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | 1.11 | 1.44 | -- |  |  |  |
| 2 | .44 | .88 | .57 | -- |  |  |
| 3 | .85 | 1.14 | .64 | .51 | -- |  |
| 4 | .76 | 1.05 | .58 | .52 | .57 | -- |
| **1 Year Follow-up (*N*=11,203)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | 1.12 | 1.46 | -- |  |  |  |
| 2 | .40 | .83 | .59 | -- |  |  |
| 3 | .87 | 1.15 | .63 | .49 | -- |  |
| 4 | .70 | 1.00 | .59 | .52 | .56 | -- |
| **2 Year Follow-up (*N*=10,899)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | 1.03 | 1.40 | -- |  |  |  |
| 2 | .35 | .78 | .57 | -- |  |  |
| 3 | .86 | 1.16 | .62 | .48 | -- |  |
| 4 | .61 | .95 | .57 | .51 | .55 | -- |
| **3 Year Follow-up (*N*=10,099)** | | | | | | |
| Parcel | *Mean* | *SD* | 1 | 2 | 3 | 4 |
| 1 | 1.07 | 1.40 | -- |  |  |  |
| 2 | .32 | .75 | .55 | -- |  |  |
| 3 | .92 | 1.19 | .63 | .47 | -- |  |
| 4 | .58 | .92 | .56 | .49 | .55 | -- |
| *Note:* Values are based on sum scores of the CBCL items included in the parcels; SD=Standard deviation | | | | | | |

Table S6. Results of Sensitivity Analyses to Examine Impact of Item-parcel Allocation Variability

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Broad-based Externalizing** | | | | | | | | | | | | | | | | |
|  | *Loadings* | | | | *Overall Model Fit* | | | | | | | | | | | |
|  |  | | | | RMSEA | | | | TLI | | | | CFI | | | |
|  | Mean | *SD* | Min. | Max. | Mean | SD | Min. | Max. | Mean | SD | Min. | Max. | Mean | SD | Min. | Max. |
| Parcel 1 | .85 | .03 | .73 | .92 | .05 | .02 | .01 | .12 | .99 | .01 | .93 | 1.00 | .99 | .01 | .97 | 1.00 |
| Parcel 2 | .85 | .03 | .67 | .92 |
| Parcel 3 | .86 | .03 | .76 | .91 |
| Parcel 4 | .85 | .03 | .77 | .91 |
| Parcel 5 | .85 | .03 | .73 | .91 |
| **Antagonistic Externalizing** | | | | | | | | | | | | | | | | |
|  | *Loadings* | | | | *Overall Model Fit* | | | | | | | | | | | |
|  |  | | | | RMSEA | | | | TLI | | | | CFI | | | |
|  | Mean | *SD* | Min. | Max. | Mean | SD | Min. | Max. | Mean | SD | Min. | Max. | Mean | SD | Min. | Max. |
| Parcel 1 | .73 | .07 | .47 | .85 | .04 | .02 | .00 | .14 | .98 | .03 | .83 | 1.00 | .99 | .01 | .94 | 1.00 |
| Parcel 2 | .74 | .07 | .48 | .86 |
| Parcel 3 | .74 | .06 | .52 | .86 |
| Parcel 4 | .74 | .07 | .47 | .87 |
| **Irritability** | | | | | | | | | | | | | | | | |
|  | *Loadings* | | | | *Overall Model Fit* | | | | | | | | | | | |
|  |  | | | | RMSEA | | | | TLI | | | | CFI | | | |
|  | Mean | *SD* | Min. | Max. | Mean | SD | Min. | Max. | Mean | SD | Min. | Max. | Mean | SD | Min. | Max. |
| Parcel 1 | .79 | .04 | .63 | .87 | .04 | .02 | .00 | .11 | .99 | .01 | .93 | 1.00 | 1.00 | .004 | .98 | 1.00 |
| Parcel 2 | .79 | .04 | .63 | .88 |
| Parcel 3 | .79 | .04 | .63 | .87 |
| Parcel 4 | .72 | .06 | .49 | .82 |
| **Neurodevelopmental** | | | | | | | | | | | | | | | | |
|  | *Loadings* | | | | *Overall Model Fit* | | | | | | | | | | | |
|  |  | | | | RMSEA | | | | TLI | | | | CFI | | | |
|  | Mean | *SD* | Min. | Max. | Mean | SD | Min. | Max. | Mean | SD | Min. | Max. | Mean | SD | Min. | Max. |
| Parcel 1 | .77 | .05 | .62 | .89 | .05 | .03 | .00 | .12 | .98 | .02 | .91 | 1.00 | .99 | .01 | .96 | 1.00 |
| Parcel 2 | .76 | .05 | .63 | .88 |
| Parcel 3 | .77 | .05 | .62 | .88 |
| Parcel 4 | .72 | .07 | .56 | .88 |
| *Note:* RMSEA: Root mean square error of approximation; TLI: Tucker Lewis index; CFI: comparative fit index; Mean=average value across 500 random item-parcel allocations; SD=standard deviation of value across 500 random item-parcel allocations; Min.=Minimum value observed across the 500 random-item parcel allocations; Max.=Maximum value observed across the 500 random-item parcel allocations. | | | | | | | | | | | | | | | | |

1. See Forbes et al. (2023) for a discussion of factors that may have contributed to the emergence of a positive psychosis dimension within externalizing, since this result departs from findings from the HiTOP model in adults. [↑](#footnote-ref-1)
2. As noted in the preregistration, criteria for significant decrements in model fit were based on the recommendations of past simulation studies (Meade et al., 2008) focused on changes in the comparative fit index (CFI; ∆CFI critical value = .002) and McDonald’s non-centrality index (NCI; ∆NCI critical values for our measurement models = .0065 and .0067) (Meade et al. 2008). We also examined changes in χ2 (i.e., ∆χ2) for each model comparison, although ∆χ2 is well known to be overly sensitive to sample size and thus was not given as much weight as other indices. We used the following indices to examine relative and absolute model fit at each time point: AIC, BIC, RMSEA (acceptable fit: < .08, good fit: < .05), and CFI and TLI (acceptable fit: .95-97, good fit: >.97). [↑](#footnote-ref-2)
3. Standardized factor loadings for the CBCL items at each level of the 4-factor hierarchy are available on the OSF page for the project. We also provide a more in-depth overview of the 4-factor and 14-factor solutions in the supplementary material. [↑](#footnote-ref-3)
4. Our initial preregistration proposed dropping indicators; however, this was not feasible since fewer parcels were used to model the latent factors. [↑](#footnote-ref-4)
5. At present, dMACS can only be used to compare two groups or time points at a time (Nye & Drasgow, 2011). Thus, effect sizes describe pairwise comparisons between baseline and subsequent time points. [↑](#footnote-ref-5)
6. We note that the narrow externalizing dimension used for our analyses of longitudinal measurement invariance was further shifted towards antagonistic content, since items with low primary loadings (*bragging, boasting*) or high secondary loadings (*peer problems* composite, *impulsive/acts without thinking*) were excluded from the narrow externalizing item-parcels. [↑](#footnote-ref-6)
7. Because effect sizes of measurement noninvariance are not commonly reported in measurement invariance testing, we provide a more thorough conceptual and statistical overview of these analyses in **Section III** of the supplement. [↑](#footnote-ref-7)
8. See equation 6 in Nye & Drasgow, 2011 for how these values are computed, and Clark et al., 2016 for an applied example in developmental research. [↑](#footnote-ref-8)
9. The direction of bias is given by the ‘$ItemDeltaMean’ in the ‘dmacs’ output for our models, which provides the expected bias in the indicator mean due to measurement noninvariance and can be either positive or negative. [↑](#footnote-ref-9)