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

Colin E. Vize1

Amy L. Byrd1

Whitney R. Ringwald1

Emily R. Perkins2

Rebecca Waller2

Samuel W. Hawes3

*University of Pittsburgh1*

*University of Pennsylvania2*

*Florida International University3*

*Funding*: Support for this research was provided by the National Institute of Mental Health (K01 MH130746; PI:Vize);

**Current Study**

Research on the HiTOP model has been primarily conducted in adult populations. The current study sought to build on recent efforts to extend the HiTOP model to developmental samples, while using a large, nationally representative sample of youth followed over multiple years. Collectively, the current study aimed to investigate the validity and utility of disaggregating the externalizing dimension in youth during a developmentally sensitive window to further research on the HiTOP model in non-adult populations. The current study had two aims: **Aim 1**: Identify the hierarchical structure of externalizing psychopathology and examine evidence of discriminant validity of identified dimensions; and **Aim 2**: Assess the longitudinal measurement invariance of a broad-based externalizing dimension in the ABCD study as well as more specific dimensions underlying broad-based externalizing symptoms.

**Method**

**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 three-year follow-up assessment.

**Measures**

**Externalizing Dimensions (Parent-Report; T1-T4).** Externalizing psychopathology was assessed using a subset of items from the parent-report Child Behavior Checklist (CBCL; ), which was administered at each of the four assessments. More specifically, we used the 45 CBCL items and item composites first 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). Table 1 in the preregistration (<https://osf.io/ec36x/?view_only=d63a1452f133421b9e1bef34a41675f1>) provides details for the specific CBCL items that were used for our analyses.

**Diagnostic Constructs (Parent- and Youth-Report; T1).** Clinical correlates at baseline were assessed using clinician ratings of parent- and youth-reported modules of the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS-5). Clinical outcomes were assessed using dimensional symptom counts for the following externalizing, neurodevelopmental, and internalizing diagnostic constructs: conduct disorder, oppositional defiant 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 completed the mood disorder, social anxiety, generalized anxiety disorder, suicidality, and sleep modules at baseline.

**Prosocial Behavior (Youth-Report; T1).** Prosocial behavior was assessed using the 3-item subscale of the youth-reported Strengths and Difficulties Questionnaire (SDQ) (“I try to be nice to other people”; “I care about their feelings”; “I offer to help others”) administered at baseline. 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) at baseline. The scale contained 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 at baseline using an age-corrected composite of 5 tasks from the NIH Toolbox Cognition measures: 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 (Luciana et al., 2018).   

**Preregistered Analyses and Hypotheses[[1]](#footnote-1)**

***Aim 1 Analyses***

To examine the hierarchical structure of externalizing symptoms at baseline, we used the 45 items/composites identified in Michelini et al. (2019) that loaded on the broad-based externalizing dimension identified at the second level of their bass-ackwards analysis. The ‘psych’ package (Revelle, 2023) was used to implement both 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. Forbes’ (2023) extended bass-ackwards approach is similar to the traditional bass-ackwards analysis, where a single factor or principal component is extracted at the first level of the hierarchy, and an additional factor is extracted at each subsequent level of the hierarchy. However, the 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-based externalizing, we expected more fine-grained differences to emerge in the hierarchical structure of the CBCL items/composites. The goal of using such an approach is to have an empirically identified, unidimensional structure for specific externalizing dimensions before moving to more complex modeling approaches (i.e., longitudinal measurement invariance).

After identifying the hierarchical structure of externalizing dimensions, we assessed evidence of the dimensions’ discriminant validity by examining correlations between the identified externalizing dimensions (assessed using factor scores) and external correlates (e.g., conduct disorder symptoms, prosocial behavior, impulsivity, cognitive ability). These analyses were conducted to further characterize the externalizing dimensions derived from the bass-ackwards results and provide evidence of their discriminant validity.

***Aim 2 Analyses***

Following the extended bass-ackwards and correlational analyses, 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, first using a CFA to test for configural measurement invariance and including constraints in a step-by-step fashion. After configural invariance, we examined weak invariance (by adding constraints to factor loadings), then strong invariance (by adding constraints for item intercepts). In some cases, we also examined strict invariance (by adding constraints on item residuals).

For tests of longitudinal measurement invariance, 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; Schermelleh-Engel et al., 2003).

Last, to quantify the *degree* of invariance, rather than solely relying on empirical cutoffs derived from model fit indices (see Nye & Drasgow, 2011; Clark & Donellan, 2023), we computed Cohen’s *d* for mean and covariance structures (*dMACS*; Nye & Drasgow, 2011) using the ‘dmacs’ package (Dueber, 2023). These metrics index the magnitude of noninvariance in the metric of Cohen’s *d* for each indicator and reflect the collective contribution of multiple varieties of noninvariance.

Importantly, using a large number of individual CBCL items would substantially increase the complexity of the longitudinal measurement invariance models and make identifying an invariant set of indicators very difficult, since measurement invariance is difficult to achieve even in simpler scenarios (Putnick & Bornstein, 2016). To mitigate this issue, we created item parcels from the CBCL items based on a set of preregistered criteria. CBCL items with adequate loadings on the respective externalizing dimension (i.e., primary loading ≥ .35) were randomly assigned to parcels, and we aimed to have a minimum of six items be assigned to each parcel while also ensuring that parcels were composed of a similar number of items. The CBCL item parcels served as the observed indicators of externalizing dimensions for all tests of longitudinal measurement invariance.

***Preregistered Hypotheses***

Based on our analytic approach, we had four primary hypotheses. *Hypothesis 1a:* We hypothesized that antagonistic and disinhibited dimensions would be identified at more fine-grained levels of the hierarchy with the same items that have been shown to comprise a broad-based externalizing dimension. *Hypothesis 1b*: 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 antagonism and disinhibition factors would emerge relatively early in the factor extraction process (e.g., at level 2 or level 3). *Hypothesis 1c*: Evidence of discriminant findings would be found for the antagonism and disinhibition externalizing factor scores derived from the bass-ackwards analyses. Specifically, we hypothesized that antagonistic externalizing factor scores would have larger associations with conduct disorder, oppositional defiant disorder, and prosocial behavior (-), compared to disinhibited externalizing. Meanwhile, disinhibited externalizing would have larger effect sizes for ADHD, impulsivity, and fluid intelligence. Furthermore, we expected these correlations to be significantly stronger than the correlations of antagonistic and disinhibited externalizing with internalizing outcomes (e.g., social anxiety, major depression). *Hypothesis 2:* We would be able to establish longitudinal measurement invariance for the measurement models of externalizing symptoms, facilitating future investigations of mean-level change over time in these externalizing dimensions in the ABCD study.

**Results[[2]](#footnote-2)**

**Deviations from Preregistration**

We first note 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.[[3]](#footnote-3) Second, 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-based externalizing, and 4 parcels being used to model the more specific externalizing dimensions identified in our analyses.

**Modified Bass-ackwards Analyses**

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 opted to focus on the 4-factor solution for our subsequent analyses. Standardized factor loadings for the CBCL items at each level of the hierarchy are available on the OSF (<https://osf.io/ec36x/?view_only=d63a1452f133421b9e1bef34a41675f1>).

The first level of the bass-ackwards hierarchy, unsurprisingly, was a broad-based externalizing factor with all 45 CBCL items showing moderately strong to strong standardized loadings on the factor (range=.47-.81). At the subsequent level, a neurodevelopmental problems factor emerged, while the externalizing factor remained relatively unchanged and was strongly correlated with externalizing from level 1 of the hierarchy (*r*=.96). At level 3, the neurodevelopmental factor remain unchanged (*r*=1.00 with the neurodevelopmental factor at level 2). However, externalizing split into an antagonistic externalizing factor characterized by antisocial behavior (e.g., stealing, fighting) and cruelty and meanness to others (e.g., cruelty to animals, bullying, lack of guilt), and a hostility factor indexed by items related to irritability and anger. These three levels of the bass-ackwards hierarchy are displayed in Figure 1.

At the fourth level, an additional factor emerged that was composed only of four items (range of standardized loadings=.32-.61), two of which had notable secondary loadings on other factors. The two items with primary loadings and no secondary loadings on the fourth factor were related to arrogance (“bragging” and “showing off”). Because the fourth factor was not substantively meaningful and overly narrow, we extracted factor scores for the three factors of antagonistic externalizing, hostility, and neurodevelopmental problems and proceeded with our subsequent analyses.

**Correlations with External Criterion Measures**

Factor scores for broad-based externalizing (i.e., externalizing at level 1 of the bass-ackwards hierarchy), antagonistic externalizing, hostility, and neurodevelopmental problems were correlated with our external criterion measures assessed at baseline. Results are presented in Table 1. The four factors generally were positively related to clinical criterion measures, though these relations were much stronger for parent-reported symptom counts (*r* range=.07-.66) compared to youth-reported symptom counts (*r* range=.04-.16). Furthermore, the four externalizing factors tended to be more strongly related to symptom counts for conduct disorder, oppositional defiant disorder, and attention deficit-hyperactivity disorder compared to symptom counts related to depression and anxiety diagnostic constructs.

For non-clinical criterion measures, the four factors showed highly similar relations to each of the subscales of the UPPS-P, though these were also small in magnitude (*r*=.04-.16). More notable differences emerged for the fluid intelligence composite, with the neurodevelopmental problems factor showing a stronger negative relation than the other factors (*r*=-.17), though the difference in magnitude was still small. All factors were negatively related to parent-reported prosocial behavior. However, youth-reported prosocial behavior was largely unrelated to the externalizing factors.

Divergent findings among the factors were largely confined to parent-reported symptom counts from the KSADS-5. For example, antagonistic externalizing was more strongly related to conduct disorder symptom counts than both the hostility and neurodevelopmental factors, while the neurodevelopmental factor was more strongly related to ADHD symptom counts. Antagonistic externalizing also showed smaller positive relations with anxiety disorder-related symptom counts compared to other factors. Last, the hostility factor was most strongly related to oppositional defiant disorder symptom counts compared to the other externalizing factors. The other noteworthy divergent finding was for parent-reported prosocial behavior—the neurodevelopmental problems factor showed a slightly smaller relation (*r*=-.22) compared to the other factors (*r* range=-.32 to -.35).

**Longitudinal Measurement Invariance**

After identifying unidimensional externalizing factors from the 45 CBCL items and investigating their divergent relations, we then moved to examining the longitudinal measurement invariance of both broad and specific externalizing factors. For the more specific factors (antagonistic externalizing, hostility, and neurodevelopmental problems), we chose to focus only on antagonistic externalizing and neurodevelopmental problems for our tests of longitudinal measurement invariance. This choice was motivated by a desire to examine a relatively pure dimension of antagonistic externalizing, as content related to hostility and anger tends to be interstitial in nature (i.e., it reflects both antagonistic content but also content relevant to negative affectivity; Watters et al., 2019).

Based on our preregistered criteria for assigning CBCL items to parcels, all 45 items were used to model a latent broad-based externalizing dimension with five parcels of nine randomly assigned items each used as indicators. For antagonistic externalizing, while 19 items had primary loadings on the factor, two items were excluded from parcel assignment based on our preregistered criteria — one due to a below-threshold primary loading (<.35) and the other an above-threshold secondary loading (≤.10 difference from primary loading). Thus, 17 items were assigned to parcels, resulting in four parcels (three four-item parcels, one five-item parcel). Fifteen items were used to model the neurodevelopmental dimension, using three four-item parcels and one three-item parcel. All items were randomly assigned to parcels, and the parcels were used as indicators of the respective latent externalizing dimension at baseline and at one-, two-, and three-year follow-up assessments. Supplementary Tables S1-S3 provide descriptive information for the parcels used as indicators for each externalizing dimension.

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. All models were estimated in MPlus (Version 1.8.10; Muthén & Muthén, 2017).[[4]](#footnote-4)

***Longitudinal Measurement Invariance of Broad-based Externalizing***

Table 2 provides results for the tests of longitudinal measurement invariance. For our model of broad-based externalizing, the model demonstrated excellent absolute fit across separate models at each time point (i.e., configural invariance). Concerning relative fit indices, restricting factor loadings to be equal across all time points (i.e., the test of weak invariance) did not decrement model fit based on our preregistered criteria (ΔCFI=.001; ΔNCI=.005).[[5]](#footnote-5) 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); thus, strong invariance was not achieved. We did not pursue any tests of strict measurement invariance based on the results of the strong invariance test.

To probe the lack of strong invariance, we investigated modification indices for the model since our initial preregistered approach of dropping indicators was not feasible since a smaller number of parcels were used to model the latent dimensions. The modification indices indicated that the intercepts of parcels four and five should be allowed to vary over time, and thus a partial strong measurement invariance model was again tested after freeing the constraints on parcels four and five. This model demonstrated excellent overall fit ( ) and also showed no decrement in model fit compared to the weak invariance model of broad-based externalizing (ΔCFI=.004; ΔNCI=.026) indicating that partial strong longitudinal measurement invariance was achieved.

***Antagonistic Externalizing***

We followed the same procedure to test the longitudinal measurement invariance of the antagonistic externalizing dimension. Similar to broad-based externalizing, antagonistic externalizing showed excellent absolute model fit across the tests of invariance. Furthermore, tests of measurement invariance showed that model fit was not negatively impacted by adding constraints for weak, strong, or strict measurement invariance models. Thus, we were able to demonstrate strict longitudinal measurement invariance for the antagonistic externalizing measurement model.

***Neurodevelopmental Problems***

The measurement model for neurodevelopmental problems similarly showed excellent absolute fit across models. We were also able to establish weak measurement invariance, with model fit not being negatively impacted after constraining factor loadings to be equal (ΔCFI=.002; ΔNCI=.005). However, after imposing constraints on the intercepts of the parcel indicators, model fit was negatively impacted, indicating that strong longitudinal measurement invariance could not be established.

As before with our model of broad-based externalizing, we investigated whether partial strong measurement invariance could be established by examining modification indices for the neurodevelopmental model. Modification indices suggested freeing the intercepts of parcels three and four at baseline and the three-year follow up to improve model fit. After freeing these parameters and re-estimating the model, the partial strong measurement invariance model showed no decrement in model fit compared to the weak invariance model ( ), indicating that partial strong longitudinal measurement invariance was achieved.

**Effect Sizes for Longitudinal Measurement Noninvariance**

As previously noted, Cohen’s *d* for mean and covariance structures (*dMACS*) provides an effect size to quantify the impact of measurement noninvariance, and is interpreted in the same way as in other contexts (i.e., generally, an effect is considered small if *d*=.20 or lower). *dMACS* reflects the degree of noninvariance at the indicator level, and captures differences between indicators due to all sources of measurement noninvariance (weak, strong, strict). Importantly, *dMACS* quantifies the impact of noninvariance at the level of indicators. In certain scenarios, measurement noninvariance at the indicator level may not translate to bias at the level of the latent dimension. For example, two indicators may be biased in opposite directions, and thus when combined do not lead to bias at the level of the latent dimension. Thus, Table 3 provides results for *dMACS* and additional effect sizes (i.e., proportion of mean change attributable to noninvariance) for externalizing dimensions at baseline assessment compared to the one-, two-, and three-year follow-up assessments.[[6]](#footnote-6)

Results showed that across models and time point comparisons, the impact of indicator noninvariance was not very large, with *dMACS* ranging from .01-.23. Nonetheless, when examining the degree of bias due to noninvariance at the level of the latent dimension, measurement noninvariance accounted for a notable degree of observed mean change when comparing baseline scores to later timepoints. For example, comparing the broad-based externalizing dimension at baseline to broad-based externalizing at the three-year follow-up assessment (i.e., Time 1-Time 4 comparison), the estimated mean difference due to noninvariance was -.91, suggesting that measurement noninvariance resulted in a higher mean at baseline compared to the three-year follow-up. This value can be compared with the observed mean difference, in this case 2.11, which reflects the difference between the sum score of the five parcels at baseline versus three-year follow-up. Thus, about 43% of the decrease in broad-based externalizing from baseline to the three-year follow-up is attributable to measurement noninvariance (see Table 3). Across the other models, the impact of noninvariance was largely observed for latent mean differences when comparing baseline to other time points; differences in the variances of the latent externalizing dimensions tended to be less impacted by measurement noninvariance.

**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. When using weighted least squares extraction and an oblique equamax 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 when comparing the original solution to the weighted least squares solution.

Second, researchers have called attention to the fact 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 present results. To do so, we estimated models using the baseline ABCD data for each of the externalizing dimensions (broad-based externalizing, antagonistic externalizing, and neurodevelopmental problems), specifying the same number of parcel-based indicators for each dimension (i.e., five parcels for broad-based externalizing, and four parcels for the other two dimensions). Next, the respective items used for the parcels were randomly assigned to parcels 500 separate times, and after each assignment, the CFA model was fit to the data. We then extracted the average values, standard deviations, and minimum and maximum values for factor loadings as well as for overall model fit (based on RMSEA, CFI, and TLI). These values are reported in Supplementary Table S4, and show that while there indeed was variability in parameter estimates and model fit indices due to the sampling variability of random assignment, factor loadings and model fit indices were consistent and did not indicate that our results were dependent on the particular item-parcel assignment strategy we used.

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 while including clustering effects for whether youth were part of the same family, and also examined the impact of study site as a clustering variable. 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 results in some small improvements in overall model fit, it did not have an impact on relative fit comparisons for our tests of longitudinal measurement invariance.

**Discussion**

The present preregistered study sought to delineate the hierarchical structure of externalizing symptoms in the ABCD study, and subsequently evaluate the longitudinal measurement invariance of the identified dimensions.

Figure 1

*Results of Bass-ackwards Analyses Using CBCL Externalizing Items*



*Note*: EXT=Externalizing; HOS=Hostility; NEURO=Neurodevelopmental; correlations between factor scores at subsequent levels are included in the figure.

Table 1

*Correlations Between Bass-ackwards Factor Scores and External Criterion Measures*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Factor Scores** | | | |
|  | *A1: Externalizing* | *C1: Ant. Externalizing* | *C2: Hostility* | *C3: Neurodevelopmental* |
| *Symptom Counts-Parent Report* |  |  |  |  |
| Conduct Disorder | .58 | .66 | .43 | .35 |
| Oppositional Defiant Disorder | .63 | .55 | .65 | .40 |
| ADHD | .61 | .49 | .43 | .66 |
| Major Depressive Disorder | .42 | .31 | .39 | .38 |
| Suicidality/Self-harm | .34 | .28 | .31 | .26 |
| Generalized Anxiety Disorder | .29 | .15 | .30 | .30 |
| Social Anxiety Disorder | .15 | .07 | .15 | .20 |
| *Symptom Counts-Youth Report* |  |  |  |  |
| Major Depressive Disorder | .13 | .12 | .10 | .12 |
| Suicidality/Self-harm | .16 | .14 | .13 | .14 |
| Generalized Anxiety Disorder | .06 | .04 | .06 | .07 |
| Social Anxiety Disorder | .06 | .04 | .05 | .06 |
| *Impulsivity-Youth Report* |  |  |  |  |
| Negative Urgency | .17 | .16 | .15 | .11 |
| Positive Urgency | .15 | .15 | .10 | .12 |
| Lack of Premeditation | .16 | .15 | .12 | .14 |
| Lack of Perseverance | .15 | .12 | .10 | .18 |
| Sensation Seeking | .04 | .06 | .02 | .02 |
| Prosocial Behavior-Parent Report | -.35 | -.33 | -.32 | -.22 |
| Prosocial Behavior-Youth Report | -.09 | -.09 | -.07 | -.06 |
| Fluid Intelligence | -.15 | -.14 | -.08 | -.17 |
| *Note*: Ant. Externalizing*=*Antagonistic Externalizing factor; symptom count measures are based on past and current symptom counts endorsed from the KSADS-5 interview; impulsivity is from the UPPS-P; prosocial behavior is from the SDQ; Fluid Intelligence=age-corrected composite of 5 tasks from the NIH Toolbox Cognition measures (see Method). | | | | |

Table 2

*Results for Longitudinal Measurement Invariance*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Broad-based Externalizing** | | | | | | | | | | |
| **Model** |  | *df* | AIC | BIC | RMSEA | TLI | CFI | NCI | Δ | ΔCFI | ΔNCI |
| Configural | 844.18 | 134 | 692176.75 | 692885.37 | .021 | .992 | .995 | .971 | -- | -- | -- |
| Weak | 974.54 | 146 | 692412.91 | 693032.95 | .028 | .992 | .994 | .966 | 133.34\* | .001 | .005 |
| Strong | 1624.08 | 158 | 693421.95 | 693953.41 | .028 | .987 | .989 | .898 | 1040.38\* | .004 | .026 |
|  | **Antagonistic Externalizing** | | | | | | | | | | |
| **Model** |  | *df* | AIC | BIC | RMSEA | TLI | CFI | NCI | Δ | ΔCFI | ΔNCI |
| Configural | 131.79 | 74 | 302646.22 | 303221.97 | .008 | .998 | .999 | .998 | -- | -- | -- |
| Weak | 176.92 | 83 | 302745.51 | 303254.83 | .010 | .997 | .998 | .996 | 37.42\* | .001 | .002 |
| Strong | 245.13 | 92 | 302853.36 | 303296.24 | .012 | .996 | .997 | .994 | 125.36\* | .001 | .002 |
| Strict | 347.15 | 104 | 303079.24 | 303433.55 | .014 | .994 | .995 | .990 | 87.05\* | .002 | .004 |
|  | **Neurodevelopmental** | | | | | | | | | | |
| **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 |
| *Note:* *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. | | | | | | | | | | | |

Table 3

*Effect Sizes to Quantify the Impact of Measurement Invariance*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Broad-based Externalizing** | | | | | | | |
| **Comparison** | Δ*Mean* | Observed Mean Difference | % of Mean Difference | Δ*Var* | Observed Variance Difference | % of Var. Difference | *dMACS* range |
| Time 1-Time 2 | -.36 | -.85 | 42% | 1.72 | 5.37 | 15% | .01-.08 |
| Time 1-Time 3 | -.92 | -1.57 | 56% | 3.59 | 10.89 | 33% | .05-.17 |
| Time 1-Time 4 | -.91 | -2.11 | 43% | 3.91 | 14.73 | 27% | .01-.23 |
| **Antagonistic Externalizing** | | | | | | | |
| **Comparison** | Δ*Mean* | Observed Mean Difference | % of Mean Difference | Δ*Var* | Observed Variance Difference | % of Var. Difference | *dMACS* range |
| Time 1-Time 2 | -.11 | -.22 | 50% | .17 | .69 | 25% | .02-.07 |
| Time 1-Time 3 | -.17 | -.33 | 52% | .14 | .90 | 16% | .03-.11 |
| Time 1-Time 4 | -.19 | -.48 | 40% | .11 | 1.35 | 8% | .04-.12 |
| **Neurodevelopmental** | | | | | | | |
| **Comparison** | Δ*Mean* | Observed Mean Difference | % of Mean Difference | Δ*Var* | Observed Variance Difference | % of Var. Difference | *dMACS* range |
| Time 1-Time 2 | -.05 | -.25 | 20% | .08 | .59 | 14% | .02-.07 |
| Time 1-Time 3 | -.27 | -.54 | 50% | .61 | 1.87 | 33% | .02-.16 |
| Time 1-Time 4 | -.23 | -.71 | 32% | .72 | 2.41 | 30% | .03-.21 |
| *Note*: Δ*Mean =* bias in latent mean due to measurement invariance; Observed Mean Difference=observed difference in the sum scores of the parcels; % of Mean Difference=percent of observed mean difference attributable to measurement invariance; Δ*Var =* bias in latent variance due to measurement invariance; Observed Variance Difference=observed difference in the variance of the sum scores of the parcels; % of Var. Difference=percent of observed variance difference attributable to measurement invariance; *dMACS =* Cohen’s *d* for Mean and Covariance Structure. | | | | | | | |

1. The preregistration for the current study can be found at <https://osf.io/ec36x/?view_only=d63a1452f133421b9e1bef34a41675f1>. [↑](#footnote-ref-1)
2. R and MPlus code to reproduce our analyses are available at <https://osf.io/ec36x/?view_only=d63a1452f133421b9e1bef34a41675f1>. Data are not posted on the OSF page because ABCD data is restricted to qualified researchers with approved access. However, the code posted to OSF highlights which data files were used and how they were cleaned for our analyses. Therefore, researchers with access to the most recent ABCD data release (Release 5.0) will be able to reproduce our results. [↑](#footnote-ref-2)
3. The CBCL item that was mistakenly omitted from Table 1 of the preregistration was the ‘bragging/boasting’ item. [↑](#footnote-ref-3)
4. Because MPlus is propriety software and not freely available, all models were also estimated using the ‘lavaan’ package in R. Code for the tests of longitudinal measurement invariance (for both MPlus and lavaan) are available on the OSF page for the project (<https://osf.io/ec36x/?view_only=d63a1452f133421b9e1bef34a41675f1>). [↑](#footnote-ref-4)
5. Based on the number of latent factors and number of indicators used for our models, the appropriate cutoff for ΔNCI for the broad-based externalizing model is .0065, and .0067 for the antagonistic and neurodevelopmental models (see Table 12 in Meade et al., 2008). We also note that all Δ tests were statistically significant, and these tests are typically consulted in measurement invariance tests. However, because of our very large sample, while we present the results of the Δ tests we focus on ΔCFI and ΔNCI. [↑](#footnote-ref-5)
6. At present, *dMACS* can only be used to compare two groups or time points at a time. [↑](#footnote-ref-6)