



Published in final edited form as:

Psychol Assess. 2020 November ; 32(11): 1037–1046. doi:10.1037/pas0000945.

Evaluating maternal psychopathology biases in reports of child temperament: An investigation of measurement invariance

Thomas M. Olino,

Temple University

Karina Guerra-Guzman,

Temple University

Elizabeth P. Hayden,

Western University

Daniel N. Klein

Stony Brook University

Abstract

Parent, especially mothers', reports of child temperament are frequently used in research and clinical practice, but there are concerns that maternal characteristics, including a history of psychopathology, may bias reports on these measures. However, whether maternal reports of youth temperament show structural differences based on mothers' psychiatric history is unclear. We therefore conducted tests of measurement invariance to examine whether maternal psychopathology was associated with structural aspects of child temperament as a means of evaluating potential biases related to mothers' mental disorder history. From two community-based studies of child temperament, 935 mothers completed the Child Behavior Questionnaire (CBQ), as well as semi-structured diagnostic interviews that assessed their own lifetime history of depressive, anxiety, and substance use disorders. Mothers also completed a measure of depressive symptoms concurrent to their completion of the CBQ. We found little evidence that mothers' current depressive symptoms or history of depressive, anxiety, or substance use disorders were associated with the structure of their reports of child temperament. Thus, there is little empirical support for systematic biases in reports of youth temperament as indexed by psychometric modeling.

Public Significance Statement: Biases in parent reports of youth behavior are suspected in the literature, but few studies have examined this issue with rigorous analytic methods. This study did not find evidence that maternal history of psychiatric illness or depressive symptoms led to biases in reports of youth temperament.

Keywords

maternal depression; bias; temperament ratings

Child temperament is an important predictor of developmental outcomes, foreshadowing children's later personality (Caspi & Shiner, 2006; Shiner et al., 2012), psychosocial functioning (Shiner & Masten, 2012), and psychopathology (Bufferd et al., 2014; Caspi et al., 1996; Dougherty et al., 2010). Parent reports are the most commonly used approach to assessing child temperament (Gartstein et al., 2012); for example, in a meta-analysis of sex differences in temperament (Else-Quest et al., 2006) virtually all studies relied on parent reports. Parent reports efficiently provide information about behavior across many different settings, a worthwhile goal for which other assessment strategies (e.g., behavioral observations; Buckley, Klein, Durbin, Hayden, & Moerk, 2002; Durbin, Hayden, Klein, & Olino, 2007) are impractical. However, there are concerns about informant biases in responding (Chilcoat & Breslau, 1997; Kagan, 1997; Seifer et al., 2004), especially those related to parent characteristics such as personality (Hayden et al., 2010), current mood state (McGrath et al., 2008; Youngstrom et al., 1999), history of psychopathology (Whiffen, 1990), and current depressive symptoms (Briggs-Gowan et al., 1996; Clark et al., 2017; Fergusson et al., 1993). The possibility of such biases can be examined using psychometric approaches such as testing measurement invariance across levels of parent characteristics.

There is a large literature studying associations between parental psychopathology and child temperament (Hayden et al., 2005; Pesonen et al., 2006), oftentimes with the goal of linking temperament to an established marker of children's risk for later psychopathology (i.e., parent history of disorder). Associations between parent psychopathology and child temperament could arise from several different processes, including true associations between parent psychopathology and offspring temperament (e.g., heritable psychopathology vulnerability related to individual differences in early temperament). Alternatively, associations could reflect biased reporting of offspring temperament as a function of parents' psychopathology (Kroes et al., 2003), leading to invalid conclusions about the relationship between parental psychopathology and offspring temperament. The relationship between parental psychopathology and biases in parent-reported child temperament can be conceptualized as a lack of measurement invariance between parents with and without a history of disorder.

A small but growing number of studies has examined whether mothers' depression influences their reports of child temperament and related constructs, with inconsistent results (e.g., Clark et al., 2017; Youngstrom et al., 1999). In a seminal study, Youngstrom et al. (1999) asked mothers to rate observed behavior of their own and comparison children, and found that maternal dysphoria was associated with endorsement of greater child negative affect for both their own and comparison children, but not for child positive affect. The authors interpreted their results as indicating the presence of biased reports for child negative affect. In contrast, in an older review of maternal biases in reports of youth behavior, Richters (1992) concluded that evidence of bias related to maternal depression was minimal. However, this conclusion was based on comparisons of offspring characteristics reported on by parents with and without depression. Thus, these conclusions were premature as the review focused on mean-level group comparison. This design is unable to discriminate between group differences and biases in reporting between groups. Here, we focus on biases from a psychometric perspective that can identify systematic differences in reporting of youth behaviors based on parent characteristics. Moreover, although much of this work

focused on the experience of current symptoms and distress, there has been speculation that lifetime experience of depression or other forms of psychopathology may lead to biases in reported youth characteristics.

The literature examining bias has focused on testing whether there are biases that produce test score differences between men and women (e.g., Orri et al., 2018), or between different racial or ethnic groups (e.g., Daneri et al., 2018). In these contexts, bias is identified when the properties of items systematically differ across individual characteristics, a phenomenon that can be examined using a confirmatory factor analytic (CFA) framework (Chiorri et al., 2016; Raykov, 2004; Widaman et al., 2010). Biases may be revealed by differences in how items are related to their underlying factors, and/or levels of traits for which items will be endorsed. When items have biased measurement properties, comparisons between studied groups are not valid tests of group differences. For example, in regard to parent psychopathology-child temperament research, it is conceivable that mothers with a history of depression may rate their children's negative affectivity higher than mothers without a history of depression, even if the target youth do not differ on their actual level of negative affectivity. Studies of bias in reports of youth temperament, however, have not relied on these quantitative tools to examine the presence of bias across parent reporters with and without psychopathology. Evaluation of this psychometric question requires testing of measurement invariance (MI; Millsap, 2011), which involves examining differential item properties across reporter characteristics, using confirmatory factor analysis (CFA).

In one of the few studies using CFA to examine potential biases in reporting, Clark et al. (2017) partitioned variance in scales between maternal, paternal, and consensus informant factors and substantive temperament factors, and found that parent internalizing problems and personality were associated with parent-specific reports of youth temperament. For example, maternal general distress was associated with variance in a factor reflecting maternal-specific reports of child negative affectivity. This suggests that maternal distress may impact the accuracy of reports of child temperament traits. However, this study used scale scores as the manifest indicators. This analytic strategy precludes the identification of the specific items for which biases are manifest and the kinds of biases that are present. Examining MI provides a clearer examination of measurement bias by testing whether items function differently according to rater characteristics. Moreover, even though Clark et al.'s (2017) interpretation of the associations emphasized individual informants, these conclusions were based on models including multiple raters in the same model, an approach that may obscure relationships solely within particular types of informants (e.g., mothers).

In the current study, we examined whether maternal psychopathology leads to systematic differences in the reports of youth behaviors (De Los Reyes & Kazdin, 2004). Specifically, we examined whether maternal psychopathology is associated with the magnitude of associations between underlying latent temperamental traits (i.e., factor loadings) in tests of metric invariance and/or the level of the item that is endorsed (i.e., item intercepts) in tests of scalar invariance, respectively. In the literature on bias in parent reports of youth temperament, the implicit expectation is that relevant temperament items would be common across mothers with and without psychopathology (i.e., support configural invariance), but parental characteristics will influence reported item levels (i.e., intercepts), rather than how

the underlying traits are associated with the items (i.e., factor loadings). The impact of this line of work is critical: If there are systematic differences in either factor loadings or the intercepts for items between parents with and without psychopathology, or between parents with different levels of symptomatology, claims of mean differences in child temperament as a function of these parental characteristics are not valid. The reported latent mean differences in temperament constructs are based on the untested assumption that scalar invariance is supported. Concerns about psychopathology introducing biases to informant reports focus on the notion that disorder leads to a greater probability of endorsing items reflecting problematic behavior in general, or the tendency to perceive problematic behavior as worse than it objectively is. In both cases, item intercepts differ from their predicted value, indicating a direct effect of informant status on item value holding the latent factor constant.

Previous studies have relied extensively on maternal ratings of temperament. Authors express concerns that maternal psychopathology, particularly depression, may lead to biased reports of youth behavior. In our examination of lifetime history of psychopathology, we considered several common forms of psychopathology: maternal depression, anxiety, and substance use disorders. Thus, beyond examining depression, we examine whether biases may be also associated with maternal anxiety and substance use disorders. The results will provide initial evidence about whether the ability to make valid inferences regarding parent-reported child temperament is compromised by the effects of mothers' disorders and symptoms. As our samples included only a modest number of mothers who were currently in a depressive episode, we complemented our analyses using diagnostic data by examining biases as a function of current depressive symptoms using moderated non-linear factor analysis (Bauer, 2017). Finally, we focus on the measurement properties of the newly revised (Clark et al., 2020), briefer Child Behavior Questionnaire (CBQ; Rothbart et al., 2001) major scales as biases at this level of analysis may propagate across higher levels of temperament structure. Accordingly, we do not address the question of whether the posited higher-order three-factor structure of child temperament, broadly, differs depending on maternal psychopathology.

Methods

Participants

Data from this study were collected at two different sites: London, ON, Canada (referred to as the ON sample) and Long Island, New York, USA (referred to as the NY sample). The data sets described in this project were part of larger longitudinal studies conducted at each of these sites. For this report, we include data from the initial assessments when children were 3-years old and have available maternal diagnostic interview and/or self-report of depressive symptomatology. This resulted in a sample of 935 youth (age mean = 38.61 months, $SD = 5.7$ months; 52.4% male). These data were reported on by Kotelnikova et al. (2016) in a paper focused solely on the internal structure of parent-reported child temperament. Human subjects approval was provided by local institutional review boards.

Assessment of Temperament

The biological mothers of the children in both samples completed the Child Behavior Questionnaire (Rothbart et al., 2001) as a measure of their child's temperament at age 3. The standard form of the CBQ consists of 195 items rated on a 7-point Likert scale ranging from 1 (extremely untrue) to 7 (extremely true). In the present study, we relied on the suggested revisions to the CBQ scales from Clark et al. (2020) that were designed to improve unidimensional fit of each scale. There has been surprisingly little psychometric work performed on the CBQ, including tests of MI, possibly due to the large number of items on the CBQ (necessitating large samples for studies of this kind). One study examined MI across parents and teachers (Teglasi et al., 2015) and another study examined MI across child sex (Clark et al., 2016). However, each of these studies examined MI at the level of scale scores as the observed indicators. Thus, this study is unique in examining the fundamental level of items as indicators.

Parental Psychopathology

The Structured Clinical Interview for DSM-IV (SCID; First et al., 1996) was conducted with the biological mothers of the children in the two samples. Interviewers were conducted by master's-level interviewers and advanced clinical psychology graduate students and were trained and supervised by Ph.D.-level psychologists. Interviewers were not involved in collecting any other study data and did not have access to data on the children. At each site there was good reliability for lifetime history of depression ($\kappa_{ON} = 1.00$; $\kappa_{NY} = .93$), anxiety ($\kappa_{ON} = 1.00$; $\kappa_{NY} = .91$), and substance use disorders (SUD; $\kappa_{ON} = .51$ [100% agreement]; $\kappa_{NY} = 1.00$). Lifetime rates of depression (27.5% in ON, 31.7% in NY), anxiety (21.1% in ON, 34.8% in NY), and SUD (21.4% in ON, 22.4% in NY) were generally consistent across sites. Across both samples, there were 271 mothers with a history of depression (39 current episodes); 268 with a history of anxiety (54 current episodes); and 203 with a history of SUD (4 current episodes).

Dimensional Assessment of Depression

Slightly different measures of depression severity were administered in each study, the Inventory to Diagnose Depression (IDD; Zimmerman & Coryell, 1987) in the ON sample and the Diagnostic Inventory for Depression (DID; Zimmerman et al., 2004) in the NY sample. The IDD is a 25-item measure assessing the presence of depressive symptoms and impairment due to symptoms in the past week ($\alpha = .84$ in the ON sample). The DID is a 22-item measure assessing the presence of depressive symptoms and impairment due to symptoms in the past week ($\alpha = .88$ in the NY sample). Each measure includes responses on a 0 to 4 scale (though one item on the IDD has a binary response). For both measures, some items assess frequency and others assess severity of symptoms. Most items are similar, but not identical, across measures; thus, total scores on the IDD and DID were standardized within each sample. Standardized scores were used as a continuous index of depressive symptoms across both samples.

Data Analysis.

All models were estimated in Mplus 8.3 (Muthén & Muthén, 1998) and model execution and summaries of results were facilitated using the MplusAutomation package (Hallquist & Wiley, 2018) in R (R Development Core Team, 2011). As the CBQ includes seven response options without disproportional endorsements of extreme responses (i.e., evidence of skew), models were estimated specifying continuous indicator variables. All models were estimated using maximum likelihood estimation. We evaluated models on two goodness of fit indices. Specifically, we used the comparative fit index (CFI; Bentler, 1990) and Root Mean Square Error of Approximation (RMSEA; Steiger, 1990) presented with the ninety percent confidence intervals. Although cut-offs are somewhat arbitrary (Marsh et al., 2004), current conventions suggest that excellent model fit is indicated by CFI values $\geq .95$ (Hu & Bentler, 1999) and RMSEA values $\leq .05$ (MacCallum et al., 2006); good fit is indicated by CFI greater than $.90$ and a RMSEA between $.05$ and $.10$. We report model chi-square values for completeness

We focused on analyses of the individual scale scores for which biases would threaten validity of the measures, rather than on the broad structure of child temperament. Tests of MI (Millsap, 2011) between mothers with and without depressive, anxiety, and SUDs were examined using the analytic features of testing the configural, metric, and scalar invariant models within the model command in Mplus. Configural invariance examines whether the pattern of loadings is consistent across groups. Metric invariance examines whether the relationship between the latent factor and the observed indicator variable differs across groups (i.e., equivalence of factor loadings). Scalar invariance examines whether intercepts of indicator variables differs across groups. Differences in intercept(s) provide information about differences in the propensity to endorse items (e.g., groups having different thresholds for evaluating behavior on a given item), holding the position on the latent variable continuum constant. Model fit comparisons were evaluated by investigating change in both CFI and RMSEA using Chen's (2007) guidelines. Chen (2007) recommended interpreting reductions in CFI and RMSEA of $.015$ as indicating non-invariance (i.e., failure to demonstrate measurement invariance). When the RMSEA and CFI changes led to different conclusions, we relied on the more conservative index to inform interpretations. We describe measures of effect size (dMACS; Nye & Drasgow, 2011) to provide complementary information about the magnitude in differences averaged across items for each factor. This metric integrates differences in factor loadings and intercepts into a single metric that is keyed in the positive direction.

Tests of MI across dimensionally assessed depressive symptoms were conducted following methods described by Bauer (2017). These models permit testing of measurement differences using continuous variables, expanding *multiple indicator-multiple cause* models to include associations between continuous exogenous variables (i.e., maternal depressive symptoms) and indicators of latent variables. Briefly, separate models regress the individual indicator variable, the factor loading to that same indicator variable, and the latent variable on maternal depressive symptoms in a single group model. The moderation of the factor loading from maternal depressive symptoms is a test of the equality of loadings (akin to metric invariance). The prediction of the indicator from maternal depressive symptoms is

a test of the equality of intercepts (akin to scalar invariance). Thus, there are not distinct models testing individual levels of invariance. Due to the number of tests, we used a more stringent alpha adjustment of $p < .01$ for these analyses.

Results

Tests of single factor models.

Single factor models were fit to each of the CBQ dimensions individually. Nearly all of the models (Table 1) demonstrated at least adequate fit on at least one primary indicator of fit. The model for impulsivity had poor fit across both primary fit indicators. Thus, we present further analyses including impulsivity for completeness, but qualify any conclusions based on the poor fit of the initial models. Moreover, throughout tests of invariance when full support for fit in configural invariance models is not satisfied, we provide further tests of invariance for completeness.

Tests of measurement invariance between mothers with and without history of depressive, anxious, and substance use disorders.

Again, we relied on single factor models for each CBQ dimension and tested for configural, scalar, and metric invariance between mothers with and without a lifetime history of depressive, anxiety, or SUDs. For models focusing on maternal depressive disorders (Table 2), configural invariance was supported for most models. Consistent with the initial CFAs, the impulsivity model was a poor fit to the data. In the configural models, high-intensity pleasure also showed poor fit to the data for both the CFI and RMSEA. Fear, low-intensity pleasure, and sadness demonstrated poor fit to the data according to the CFI, but adequate fit according to the RMSEA. In tests of metric invariance, there were no scales for which there were reductions in CFI or RMSEA greater than .015, showing support for metric invariance. Moreover, in tests of scalar invariance, no scales had a reduction in CFI or RMSEA greater than .015. Thus, across mothers with and without lifetime history of depressive disorders, scalar invariance was supported. Relying on the dMACS metric, 20.0% of the Activity scale items, 27.3% of items on the anger, 33.3% on the Attention Shifting, 25.0% on the Discomfort, 30.0% on the Fear, 30.0% on the High-intensity pleasure, 36.4% on the Inhibitory Control, 30.0% on the Low-intensity pleasure, 25.0% on the Perceptual Sensitivity, 27.3% on the Sadness, 22.2% on the Smiling/Laughter, and 66.7% of items on the Soothability factors showed between small and medium effect size differences (i.e., $.20 < |d| < .50$). No items showed more than moderate effects (i.e., $d > .50$).

For models focusing on maternal anxiety disorders (Table 3), configural invariance was supported for most models. Consistent with the initial CFAs, the impulsivity model was a poor fit to the data. In the configural models, high-intensity pleasure also demonstrated poor fit to the data according to the CFI and the RMSEA. The fear and Low-intensity pleasure scales demonstrated poor fit to the data according to the CFI, but adequate fit according to the RMSEA. In tests of metric invariance, no scales demonstrated deterioration in fit, indexed by worsening of fit based on CFI or RMSEA, showing support for metric invariance. Similarly, in tests of scalar invariance, no scales had worsening of CFI or RMSEA exceeding .015. Thus, across mothers with and without lifetime history of anxiety

disorders, scalar invariance was generally supported. Relying on the dMACS metric, 27.3% of items on the anger, 25.0% on the Attention Focusing, 33.3% on the Attention Shifting, 25.0% on the Discomfort, 40.0% on the Fear, 20.0% on the High-intensity pleasure, 22.2% on the Impulsivity, 27.3% on the Inhibitory Control, 30.0% on the Low-intensity pleasure, 45.5% on the Sadness, 22.2% on the Smiling/Laughter, and 44.4% of items on the Soothability factors showed between small and medium effect size differences (i.e., $.20 < |d| < .50$). No items showed more than moderate effects (i.e., $d > .50$).

For models focusing on maternal SUDs (Table 5), configural invariance was supported for most models. Consistent with the initial CFAs, the impulsivity model was a poor fit to the data. In the configural models, high-intensity pleasure also demonstrated poor fit to the data according to the CFI and the RMSEA. The fear and Low-intensity pleasure scales demonstrated poor fit to the data according to the CFI, but adequate fit according to the RMSEA. In tests of metric invariance, no scales demonstrated deterioration in fit, indexed by worsening of fit based on CFI or RMSEA, showing support for metric invariance. Similarly, in tests of scalar invariance, no scales had worsening of CFI or RMSEA exceeding .015. Thus, across mothers with and without lifetime history of SUDs, scalar invariance was generally supported. Relying on the dMACS metric, 20.0% of items on the Activity, 63.6% of items on the anger, 30.0% on the Fear, High-intensity pleasure 20.0% on the Low-intensity pleasure, 27.3% on the Sadness, 22.2% on the Smiling/Laughter, and 33.3% of items on the Soothability factors showed at least small effect size differences. We provide comparisons of latent mean differences across maternal report of youth temperament dimensions by maternal history of depression, anxiety, and SUDs in Supplementary Table 4.

Tests of invariance across mothers' current depressive symptoms

In these models, we followed recommendations from Bauer (2017), which specify associations between maternal depression symptoms and the latent CBQ scale factor, the factor loading to a target CBQ item, and the observed CBQ target item. The focus of the analyses for invariance focus on the latter two parameters, as these analyses provide information about measurement bias. Across the sixteen scales, there were 118 parameters estimated for the loadings and 118 estimated for intercepts (see Supplementary Table 1). Maternal depressive symptoms moderated the magnitude of four factor loadings. These were for two items on the Low-intensity pleasure scale and two items on the smiling/laughter scale, with depression associated with a weaker loading for the item. Maternal depression was associated with the intercepts for twelve CBQ items: activity level (1 of 8 items), anger (1 of 9 items), high-intensity pleasure (1 of 8 items), impulsivity (4 of 7 items), Low-intensity pleasure (1 of 8 items), sadness (1 of 9 items), shyness (1 of 8 items), and soothability (2 of 7 items). For the scales in which there were at least two significant associations between maternal depression severity and the CBQ item intercepts, the direction of the associations was mixed. Thus, mothers' depressive symptoms were associated with both higher and lower rating of items, with no clear pattern accounting for why certain items tended to be rated more highly or lower as mothers' symptoms increased. Moreover, the average effect sizes for associations between maternal depressive symptoms and intercepts and loadings across all items were small (mean $r_s = .04$ [SD = .03] and $.03$ [SD = .02], respectively). Thus, maternal depressive symptoms did not

consistently influence loadings or intercepts for maternal temperament ratings. Moreover, after applying a stricter correction ($\alpha_{\text{corrected}} = .05/236$) for multiple comparisons, only four intercepts were associated with maternal self-reported depression and no factor loadings were significantly moderated.

Discussion

Studies of youth temperament frequently rely on parent, especially maternal, reports (Else-Quest et al., 2006; Garstein & Rothbart, 2003). However, in studies that rely on parent reports of temperament, authors frequently note that the conclusions of their studies are tentative due to potential bias in reports based on parental characteristics such as history of psychopathology and/or current symptoms. If mothers' psychopathology significantly influences structural factors, mean comparisons of temperament in children of mothers with and without disorder are invalid, rendering linkages between early temperament and risk indexed by maternal disorder uninterpretable. Importantly, however, our results do not support systematic bias in terms of MI in maternal reports of child temperament as related to lifetime history of MDD, anxiety, or SUDs, or mothers' current depressive symptoms.

We explored associations between maternal history of depression and factor loadings (i.e., metric invariance) and intercepts (i.e., scalar invariance) after estimating models evaluating configural invariance as a baseline model. We found support for configural, metric, and scalar invariance across dimensions of temperament between mothers with and without a lifetime history of depression. This indicates that comparisons of youth temperament, as indexed by maternal report, is not biased by maternal history of depression. Though most of the concerns about biases in parental reports of temperament have focused on the role of maternal depression, we extended this line of inquiry to address maternal anxiety and SUD. Results of analyses for maternal history of anxiety and SUD followed the same overall pattern as those for maternal history of MDD. Thus, several common forms of psychopathology in addition to depression showed MI when comparing those with and without a history of that disorder.

Failure to identify measurement differences in child temperament related to mothers' lifetime history of depression does not rule out the possibility that concurrent depressive symptoms may influence structure. We therefore examined associations between current maternal depressive symptoms and relevant parameters reflecting bias in child temperament reports. Once again, in these models, there was little evidence of bias related to current maternal depressive symptoms using MI, and when associations were found, they were typically small in size and were in both positive and negative directions. Thus, in terms of measurement variability, our findings do not indicate the presence of systematic bias in maternal reports of youth behavior related to mothers' symptoms.

Our findings have implications for past and future research regarding informant reports on youth behavior. Research on multi-informant agreement and discrepancy (e.g., De Los Reyes et al., 2015) has proceeded under the assumption that tests of mean differences in child behavior in mothers with and without disorder are valid (i.e., that measurement properties are similar regardless of mothers' psychopathology history). The present work

provides evidence supporting this assumption, at least in the case of a widely used index of child temperament, the CBQ. Thus, previous work in this area is not undermined by measurement biases in subsets of maternal raters. Although this work does not find evidence of measurement bias in reports of temperament across maternal psychopathology, there are other indices of validity that require further attention (e.g., multimethod associations; Durbin et al., 2007).

This study benefitted from a large sample of youth in early childhood with mothers who were well-characterized on psychopathology. We examined both the influence of maternal lifetime history of multiple forms of psychopathology and, using a novel analytic procedure, mothers' current depressive symptoms, with results providing converging evidence for very modest effects. However, the study also had several limitations. First, both studies included modest racial and ethnic diversity. Thus, further work in this area with greater diversity would enhance generalizability of the findings. Second, although inter-rater reliability was excellent for depressive and anxiety disorders, agreement on SUDs was poor in the ON sample, largely due to low base-rates in the reliability sample. However, percent agreement for this diagnosis was very high. Third, although we examined both lifetime diagnosis of depression and current symptomatology, there are many other dimensions of heterogeneity for depression (e.g., chronicity, age of onset) that may lead to more sensitive tests of biases. Moreover, we lacked complementary dimensional symptom measures for anxiety and substance use. Thus, we are unable to speak to whether current experience of anxiety or substance use may bias ratings of youth temperament. Fourth, the analyses focused on a single measure of youth temperament. Further analyses relying on different measures of temperament and related behavioral outcomes will be important to test. Fifth, the samples both came from the community. Thus, there were few cases of current maternal depression and modest maternal symptom levels. Further work with samples enriched for higher levels of parental psychopathology are needed for this critical area of research. Sixth, some of the CBQ scales provided only modest fit to the data in initial and configural invariance models. Thus, the results of these models are interpreted with caution.

This study tested the assumption that maternal psychopathology biases reports of child temperament. In rigorous tests of MI, we found that neither maternal history of common forms of psychopathology or current depression symptoms are associated with biases in mothers' rating of items in a widely used temperament questionnaire. Thus, previously reported findings of associations between maternal depression and mothers' ratings of child temperament appear to be only minimally influenced by bias. Further work is needed to examine alternative methods of evaluating bias (e.g., predictive validity), in addition to tests of MI. Future work should examine the role of related constructs, including parent temperament and personality, in biases in reports of child temperament.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute of Mental Health Grants R01 MH069942 (Dr. Klein) and R01 MH107495 (Dr. Olino) and Canadian Institutes of Health Research (Operating Grant; Dr. Hayden). Supplementary materials includes all relevant covariance matrices to permit replication of analyses.

References

- Bauer DJ (2017). A More General Model for Testing Measurement Invariance and Differential Item Functioning. *Psychological Methods*, 22(3), 507–526. [PubMed: 27266798]
- Bentler PM (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. [PubMed: 2320703]
- Briggs-Gowan MJ, Carter AS, & Schwab-Stone M (1996). Discrepancies among mother, child, and teacher reports: Examining the contributions of maternal depression and anxiety. *Journal of Abnormal Child Psychology*, 24(6), 749–765. 10.1007/BF01664738 [PubMed: 8970908]
- Buckley ME, Klein DN, Durbin CE, Hayden EP, & Moerk KC (2002). Development and validation of a q-sort procedure to assess temperament and behavior in preschool-age children. *Journal of Clinical Child & Adolescent Psychology*, 31(4), 525–539. [PubMed: 12402571]
- Bufferd SJ, Dougherty LR, Olino TM, Dyson MW, Lappook RS, Carlson GA, & Klein DN (2014). Predictors of the onset of depression in young children: A multimethod, multi-informant longitudinal study from ages 3 to 6. *Journal of Child Psychology and Psychiatry*, 55(11), 1279–1287. [PubMed: 24828086]
- Caspi A, Moffitt TE, Newman DL, & Silva PA (1996). Behavioral observations at age 3 years predict adult psychiatric disorders: Longitudinal evidence from a birth cohort. *Archives of General Psychiatry*, 53(11), 1033–1039. [PubMed: 8911226]
- Caspi A, & Shiner RL (2006). Personality Development. In Eisenberg N, Damon W, & Lerner RM (Eds.), *Handbook of Child Psychology: Vol. 3, Social, Emotional, and Personality Development* (pp. 265–286). John Wiley & Sons Inc.
- Chen FF (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504.
- Chilcoat HD, & Breslau N (1997). Does psychiatric history bias mothers' reports? An application of a new analytic approach. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(7), 971–979. [PubMed: 9204676]
- Chiorri C, Hall J, Casely-Hayford J, & Malmberg L-E (2016). Evaluating Measurement Invariance Between Parents Using the Strengths and Difficulties Questionnaire (SDQ). *Assessment*, 23(1), 63–74. 10.1177/1073191114568301 [PubMed: 25604631]
- Clark DA, Donnellan MB, Durbin CE, Brooker RJ, Neppl TK, Gunnar M, Carlson SM, Le Mare L, Kochanska G, Fisher PA, Leve LD, Rothbart MK, & Putnam SP (2020). Using item response theory to evaluate the Children's Behavior Questionnaire: Considerations of general functioning and assessment length. *Psychological Assessment*. 10.1037/pas0000883
- Clark DA, Durbin CE, Donnellan MB, & Neppl TK (2017). Internalizing Symptoms and Personality Traits Color Parental Reports of Child Temperament. *Journal of Personality*, 85(6), 852–866. 10.1111/jopy.12293 [PubMed: 27897316]
- Clark DA, Listro CJ, Lo SL, Durbin CE, Donnellan MB, & Neppl TK (2016). Measurement invariance and child temperament: An evaluation of sex and informant differences on the Child Behavior Questionnaire. *Psychological Assessment*, 28(12), 1646–1662. 10.1037/pas0000299 [PubMed: 26914022]
- Daneri MP, Sulik MJ, Raver CC, & Morris PA (2018). Observers' reports of self-regulation: Measurement invariance across sex, low-income status, and race/ethnicity. *Journal of Applied Developmental Psychology*, 55, 14–23. 10.1016/j.appdev.2017.02.001
- De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabick DA, Burgers DE, & Rabinowitz J (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858–900. 10.1037/a0038498 [PubMed: 25915035]

- De Los Reyes A, & Kazdin AE (2004). Measuring Informant Discrepancies in Clinical Child Research. *Psychological Assessment*, 16(3), 330–334. 10.1037/1040-3590.16.3.330 [PubMed: 15456389]
- Dougherty LR, Klein DN, Durbin CE, Hayden EP, & Olino TM (2010). Temperamental positive and negative emotionality and children's depressive symptoms: A longitudinal prospective study from age three to age ten. *Journal of Social and Clinical Psychology*, 29(4), 462–488. 10.1521/jscp.2010.29.4.462
- Durbin CE, Hayden EP, Klein DN, & Olino TM (2007). Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion*, 7(2), 388–399. [PubMed: 17516816]
- Else-Quest NM, Hyde JS, Goldsmith HH, & Van Hulle CA (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, 132(1), 33–72. [PubMed: 16435957]
- Fergusson DM, Lynskey MT, & Horwood LJ (1993). The effect of maternal depression on maternal ratings of child behavior. *Journal of Abnormal Child Psychology*, 21(3), 245–269. [PubMed: 8335763]
- First MB, Spitzer RL, Gibbon M, & Williams JBW (1996). *The Structured Clinical Interview for DSM-IV Axis I Disorders – Non-patient edition*. Biometrics Research Department, New York State Psychiatric Institute.
- Garstein MA, & Rothbart MK (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior and Development*, 26, 64–86.
- Gartstein MA, Bridgett DJ, & Low CM (2012). Asking Questions about Temperament. In Zenter M & Shiner RL (Eds.), *Handbook of Temperament* (pp. 183–209). Guilford Press.
- Hallquist MN, & Wiley JF (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in M plus. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 621–638. [PubMed: 30083048]
- Hayden EP, Durbin CE, Klein DN, & Olino TM (2010). Maternal personality influences the relationship between maternal reports and laboratory measures of child temperament. *Journal of Personality Assessment*, 92(6), 586–593. [PubMed: 20954060]
- Hayden EP, Klein DN, & Durbin CE (2005). Parent reports and laboratory assessments of child temperament: A comparison of their associations with risk for depression and externalizing disorders. *Journal of Psychopathology & Behavioral Assessment*, 27(2), 89–100.
- Hu L, & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Kagan J (1997). Temperament and the reactions to unfamiliarity. *Child Development*, 68.
- Kotelnikova Y, Olino TM, Klein DN, Krysti KR, & Hayden EP (2016). Higher- and lower-order factor analyses of the Children's Behavior Questionnaire in early and middle childhood. *Psychological Assessment*, 28(1), 92–108. 10.1037/pas0000153 [PubMed: 26029946]
- Kroes G, Veerman JW, & De Bruyn EEJ (2003). Bias in parental reports? Maternal psychopathology and the reporting of problem behavior in clinic-referred children. *European Journal of Psychological Assessment*, 19(3), 195–203. 10.1027//1015-5759.19.3.195
- MacCallum RC, Browne MW, & Sugawara HM (2006). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149.
- Marsh HW, Hau KT, & Wen Z (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320–341. 10.1207/s15328007sem1103_2
- McGrath JM, Records K, & Rice M (2008). Maternal depression and infant temperament characteristics. *Infant Behavior and Development*, 31(1), 71–80. 10.1016/j.infbeh.2007.07.001 [PubMed: 17714790]
- Millsap RE (2011). *Statistical approaches to measurement invariance*. Taylor and Francis Group.
- Muthén LK, & Muthén BO (1998). *Mplus User's Guide*. Eighth Edition. Muthén & Muthén.
- Nye CD, & 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.

- Orri M, Rouquette A, Pingault J-B, Barry C, Herba C, Côté SM, & Berthoz S (2018). Longitudinal and Sex Measurement Invariance of the Affective Neuroscience Personality Scales. *Assessment*, 25(5), 653–666. 10.1177/1073191116656795 [PubMed: 27343247]
- Pesonen A-K, Räikkönen K, Heinonen K, Järvenpää A-L, & Strandberg TE (2006). Depressive vulnerability in parents and their 5-year-old child's temperament: A family system perspective. *Journal of Family Psychology*, 20(4), 648–655. 10.1037/0893-3200.20.4.648 [PubMed: 17176200]
- R Development Core Team, R. (2011). R: A language and environment for statistical computing. R foundation for statistical computing Vienna, Austria.
- Raykov T (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, 35(2), 299–331. 10.1016/S0005-7894(04)80041-8
- Richters JE (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin*, 112(3), 485–499. 10.1037/0033-2909.112.3.485 [PubMed: 1438639]
- Rothbart MK, Ahadi SA, Hershey KL, & Fisher P (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, 72(5), 1394–1408. 10.1111/1467-8624.00355 [PubMed: 11699677]
- Seifer R, Sameroff A, Dickstein S, Schiller M, & Hayden LC (2004). Your own children are special: Clues to the sources of reporting bias in temperament assessments. *Infant Behavior and Development*, 27(3), 323–341. 10.1016/j.infbeh.2003.12.005
- Shiner RL, Buss KA, McClowry SG, Putnam SP, Saudino KJ, & Zentner M (2012). What Is Temperament Now? Assessing Progress in Temperament Research on the Twenty-Fifth Anniversary of Goldsmith et al.(1987). *Child Development Perspectives*, 6(4), 436–444.
- Shiner RL, & Masten AS (2012). Childhood personality as a harbinger of competence and resilience in adulthood. *Development and Psychopathology*, 24(02), 507–528. [PubMed: 22559127]
- Steiger JH (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25(2), 173–180. [PubMed: 26794479]
- Teglasi H, Schussler L, Gifford K, Annotti LA, Sanders C, & Liu H (2015). Child Behavior Questionnaire–Short Form for Teachers: Informant Correspondences and Divergences. *Assessment*, 22(6), 730–748. 10.1177/1073191114562828 [PubMed: 25573857]
- Whiffen VE (1990). Maternal Depressed Mood and Perceptions of Child Temperament. *The Journal of Genetic Psychology*, 151(3), 329–339. 10.1080/00221325.1990.9914621 [PubMed: 2266357]
- Widaman KF, Ferrer E, & Conger RD (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1), 10–18. 10.1111/j.1750-8606.2009.00110.x [PubMed: 20369028]
- Youngstrom E, Izard C, & Ackerman B (1999). Dysphoria-related bias in maternal ratings of children. *Journal of Consulting and Clinical Psychology*, 67(6), 905–916. [PubMed: 10596512]
- Zimmerman M, & Coryell W (1987). The Inventory to Diagnose Depression (IDD): A self-report scale to diagnose major depressive disorder. *Journal of Consulting and Clinical Psychology*, 55(1), 55–59. [PubMed: 3571659]
- Zimmerman M, Sheeran T, & Young D (2004). The Diagnostic Inventory for Depression: A self-report scale to diagnose DSM-IV major depressive disorder. *Journal of Clinical Psychology*, 60(1), 87–110. 10.1002/jclp.10207 [PubMed: 14692011]

Table 1.

Model fit information for one-factor models of individual CBQ scales

	χ^2	df	CFI	RMSEA (90% CI)
Attention Shifting	19.69***	2	0.96	0.097 (0.061-0.139)
Attention Focusing	58.02***	9	0.95	0.076 (0.058-0.096)
Discomfort	45.21***	9	0.97	0.066 (0.047-0.085)
Perceptual Sensitivity	36.67***	9	0.98	0.057 (0.039-0.077)
Impulsivity	281.18***	14	0.82	0.143 (0.129-0.158)
Smiling	36.28***	14	0.97	0.042 (0.025-0.058)
Soothability	135.16***	14	0.91	0.096 (0.082-0.111)
Activity	96.08***	20	0.93	0.064 (0.051-0.077)
Anticipation	120.87***	20	0.92	0.074 (0.061-0.086)
Fear	107.80***	20	0.86	0.069 (0.056-0.082)
High-intensity pleasure	203.38***	20	0.89	0.099 (0.087-0.112)
Low-intensity pleasure	124.18***	20	0.84	0.075 (0.062-0.088)
Shyness	358.01***	20	0.93	0.135 (0.123-0.147)
Anger	159.35***	27	0.91	0.072 (0.062-0.084)
Inhibitory Control	79.93***	27	0.97	0.046 (0.034-0.058)
Sadness	89.74***	27	0.89	0.05 (0.039-0.062)

Note:

*** $p < .001$. CFI = Comparative Fit Index; RMSEA = Root mean square error of approximation; CI = Confidence interval.

Table 2.

Tests of invariance across mothers with and without depressive disorders.

Title	Model	χ^2	df	CFI	RMSEA	CFI	RMSEA
Activity (8)	Configural	123.26	40	0.92	0.068 (0.054-0.082)		
	Metric	129.39	47	0.92	0.062 (0.050-0.075)	0.001	-0.006
	Scalar	134.61	54	0.92	0.057 (0.045-0.070)	0.002	-0.005
Anger (9)	Configural	188.35	54	0.90	0.074 (0.063-0.086)		
	Metric	191.20	62	0.91	0.068 (0.057-0.079)	0.004	-0.006
	Scalar	202.52	70	0.90	0.065 (0.054-0.075)	-0.003	-0.003
Anticipation (8)	Configural	136.27	40	0.92	0.073 (0.060-0.087)		
	Metric	145.34	47	0.92	0.068 (0.056-0.081)	-0.002	-0.005
	Scalar	153.93	54	0.92	0.064 (0.052-0.076)	-0.001	-0.004
Attention Focusing (6)	Configural	64.00	18	0.95	0.075 (0.056-0.096)		
	Metric	66.36	23	0.95	0.065 (0.047-0.083)	0.003	-0.01
	Scalar	73.82	28	0.95	0.060 (0.044-0.077)	-0.002	-0.005
Attention Shifting (4)	Configural	18.85	4	0.97	0.091 (0.052-0.134)		
	Metric	20.67	7	0.97	0.066 (0.034-0.099)	0.003	-0.025
	Scalar	28.80	10	0.96	0.064 (0.038-0.093)	-0.012	-0.002
Discomfort (6)	Configural	56.02	18	0.97	0.068 (0.049-0.089)		
	Metric	70.13	23	0.96	0.067 (0.050-0.086)	-0.008	-0.001
	Scalar	72.61	28	0.96	0.059 (0.043-0.076)	0.002	-0.008
Fear (8)	Configural	118.95	40	0.87	0.066 (0.053-0.080)		
	Metric	128.53	47	0.86	0.062 (0.049-0.075)	-0.004	-0.004
	Scalar	144.08	54	0.85	0.061 (0.049-0.073)	-0.014	-0.001
High Intensity Pleasure (8)	Configural	237.09	40	0.88	0.104 (0.092-0.117)		
	Metric	250.36	47	0.88	0.098 (0.086-0.110)	-0.004	-0.006
	Scalar	265.57	54	0.87	0.093 (0.082-0.104)	-0.005	-0.005
Impulsivity (7)	Configural	289.81	28	0.82	0.144 (0.129-0.159)		
	Metric	300.53	34	0.82	0.132 (0.118-0.146)	-0.003	-0.012
	Scalar	301.34	40	0.82	0.120 (0.108-0.133)	0.003	-0.012

Title	Model	χ^2	df	CFI	RMSEA	CFI	RMSEA
Inhibitory Control (9)	Configural	103.11	54	0.97	0.045 (0.031-0.058)		
	Metric	115.57	62	0.97	0.044 (0.031-0.056)	-0.002	-0.001
	Scalar	126.00	70	0.97	0.042 (0.030-0.054)	-0.002	-0.002
Low Intensity Pleasure (8)	Configural	124.95	40	0.87	0.069 (0.055-0.082)		
	Metric	139.18	47	0.86	0.066 (0.053-0.079)	-0.011	-0.003
	Scalar	147.90	54	0.85	0.062 (0.050-0.074)	-0.003	-0.004
Perceptual Sensitivity (6)	Configural	41.75	18	0.98	0.054 (0.033-0.076)		
	Metric	45.56	23	0.98	0.047 (0.026-0.066)	0.001	-0.007
	Scalar	58.56	28	0.98	0.049 (0.031-0.067)	-0.006	0.002
Sadness (9)	Configural	116.20	54	0.89	0.05 (0.038-0.063)		
	Metric	123.68	62	0.89	0.047 (0.035-0.059)	0.001	-0.003
	Scalar	133.26	70	0.89	0.045 (0.033-0.056)	-0.003	-0.002
Shyness (8)	Configural	373.86	40	0.93	0.136 (0.123-0.149)		
	Metric	385.35	47	0.92	0.126 (0.115-0.138)	-0.001	-0.01
	Scalar	394.86	54	0.92	0.118 (0.107-0.129)	-0.001	-0.008
Smiling (7)	Configural	64.20	28	0.95	0.054 (0.036-0.071)		
	Metric	74.85	34	0.94	0.052 (0.036-0.068)	-0.007	-0.002
	Scalar	81.83	40	0.94	0.048 (0.033-0.063)	-0.001	-0.004
Soothability (7)	Configural	145.29	28	0.91	0.096 (0.081-0.112)		
	Metric	154.59	34	0.91	0.089 (0.075-0.103)	-0.002	-0.007
	Scalar	166.49	40	0.90	0.084 (0.071-0.097)	-0.005	-0.005

Note: Numbers in parentheses indicates number of items on the scale. CFI = Comparative Fit Index; RMSEA = Root mean square error of approximation; CI = Confidence interval. Change in CFI and RMSEA are relative to the previous model (i.e., Metric vs. Configural; Scalar vs. Metric).

Table 3.

Tests of invariance across mothers with and without anxiety disorders.

Title	Model	χ^2	df	CFI	RMSEA	CFI	RMSEA
Activity (8)	Configural	125.43	40	0.92	0.069 (0.055-0.082)		
	Metric	136.55	47	0.91	0.065 (0.052-0.078)	-0.004	-0.004
	Scalar	141.99	54	0.92	0.06 (0.048-0.072)	0.002	-0.005
Anger (9)	Configural	192.77	54	0.90	0.075 (0.064-0.087)		
	Metric	199.35	62	0.90	0.070 (0.059-0.081)	0.001	-0.005
	Scalar	205.14	70	0.90	0.065 (0.055-0.076)	0.001	-0.005
Anticipation (8)	Configural	133.69	40	0.92	0.072 (0.059-0.086)		
	Metric	145.68	47	0.92	0.068 (0.056-0.081)	-0.004	-0.004
	Scalar	159.43	54	0.91	0.066 (0.054-0.078)	-0.006	-0.002
Attention Focusing (6)	Configural	65.29	18	0.95	0.076 (0.057-0.097)		
	Metric	70.11	23	0.95	0.067 (0.050-0.086)	0.000	-0.009
	Scalar	76.38	28	0.95	0.062 (0.045-0.079)	-0.001	-0.005
Attention Shifting (4)	Configural	19.59	4	0.96	0.093 (0.054-0.136)		
	Metric	22.06	7	0.97	0.069 (0.038-0.102)	0.002	-0.024
	Scalar	23.57	10	0.97	0.055 (0.026-0.084)	0.003	-0.014
Discomfort (6)	Configural	48.04	18	0.97	0.061 (0.040-0.082)		
	Metric	50.39	23	0.98	0.051 (0.032-0.071)	0.003	-0.010
	Scalar	57.53	28	0.97	0.048 (0.030-0.066)	-0.002	-0.003
Fear (8)	Configural	120.93	40	0.86	0.067 (0.053-0.081)		
	Metric	128.09	47	0.86	0.062 (0.049-0.075)	-0.001	-0.005
	Scalar	136.42	54	0.86	0.058 (0.046-0.070)	-0.002	-0.004
High Intensity Pleasure (8)	Configural	229.05	40	0.89	0.102 (0.09-0.115)		
	Metric	233.75	47	0.89	0.094 (0.082-0.106)	0.001	-0.008
	Scalar	241.68	54	0.89	0.088 (0.077-0.099)	0.000	-0.006
Impulsivity (7)	Configural	300.35	28	0.81	0.147 (0.132-0.162)		
	Metric	308.70	34	0.81	0.134 (0.120-0.147)	-0.001	-0.013
	Scalar	318.54	40	0.81	0.124 (0.112-0.137)	-0.003	-0.010

Title	Model	χ^2	df	CFI	RMSEA	CFI	RMSEA
Inhibitory Control (9)	Configural	109.42	54	0.97	0.048 (0.035-0.060)		
	Metric	121.46	62	0.96	0.046 (0.034-0.058)	-0.003	-0.002
	Scalar	128.68	70	0.96	0.043 (0.031-0.055)	0.001	-0.003
Low Intensity Pleasure (8)	Configural	138.09	40	0.85	0.074 (0.060-0.087)		
	Metric	147.60	47	0.85	0.069 (0.056-0.081)	-0.004	-0.005
	Scalar	154.81	54	0.85	0.064 (0.053-0.076)	-0.001	-0.005
Perceptual Sensitivity (6)	Configural	48.27	18	0.98	0.061 (0.041-0.082)		
	Metric	57.27	23	0.98	0.057 (0.039-0.076)	-0.003	-0.004
	Scalar	69.67	28	0.97	0.057 (0.041-0.074)	-0.006	0.000
Sadness (9)	Configural	111.46	54	0.90	0.048 (0.036-0.061)		
	Metric	118.51	62	0.90	0.045 (0.032-0.057)	0.002	-0.003
	Scalar	134.78	70	0.89	0.045 (0.034-0.057)	-0.014	0.000
Shyness (8)	Configural	356.62	40	0.93	0.132 (0.12-0.145)		
	Metric	358.49	47	0.93	0.121 (0.109-0.133)	0.001	-0.011
	Scalar	362.22	54	0.93	0.112 (0.101-0.123)	0.001	-0.009
Smiling (7)	Configural	50.95	28	0.97	0.043 (0.023-0.061)		
	Metric	63.40	34	0.96	0.044 (0.027-0.060)	-0.010	0.001
	Scalar	66.91	40	0.96	0.039 (0.021-0.055)	0.004	-0.005
Soothability (7)	Configural	152.48	28	0.91	0.099 (0.084-0.115)		
	Metric	160.80	34	0.91	0.091 (0.077-0.105)	-0.002	-0.008
	Scalar	167.52	40	0.90	0.084 (0.071-0.097)	-0.001	-0.007

Note: Numbers in parentheses indicates number of items on the scale. CFI = Comparative Fit Index; RMSEA = Root mean square error of approximation; CI = Confidence interval. Change in CFI and RMSEA are relative to the previous model (i.e., Metric vs. Configural; Scalar vs. Metric).

Table 4.

Tests of invariance across mothers with and without SUDs.

Title	Model	χ^2	df	CFI	RMSEA	CFI	RMSEA
Activity (8)	Configural	119.42	40	0.92	0.066 (0.053-0.080)		
	Metric	125.41	47	0.92	0.061 (0.048-0.074)	0.001	-0.005
	Scalar	131.53	54	0.93	0.056 (0.044-0.069)	0.001	-0.005
Anger (9)	Configural	184.58	54	0.90	0.073 (0.062-0.085)		
	Metric	189.55	62	0.91	0.067 (0.057-0.079)	0.002	-0.006
	Scalar	204.68	70	0.90	0.065 (0.055-0.076)	-0.005	-0.002
Anticipation (8)	Configural	148.86	40	0.91	0.078 (0.065-0.091)		
	Metric	152.51	47	0.91	0.070 (0.058-0.083)	0.003	-0.008
	Scalar	158.06	54	0.91	0.065 (0.054-0.077)	0.001	-0.005
Attention Focusing (6)	Configural	68.94	18	0.94	0.079 (0.060-0.099)		
	Metric	71.82	23	0.95	0.069 (0.051-0.087)	0.002	-0.010
	Scalar	74.18	28	0.95	0.060 (0.044-0.077)	0.003	-0.009
Attention Shifting (4)	Configural	21.06	4	0.96	0.097 (0.059-0.140)		
	Metric	22.34	7	0.97	0.070 (0.038-0.103)	0.004	-0.027
	Scalar	23.00	10	0.97	0.054 (0.025-0.083)	0.005	-0.016
Discomfort (6)	Configural	53.43	18	0.97	0.066 (0.046-0.087)		
	Metric	57.41	23	0.97	0.058 (0.039-0.076)	0.001	-0.008
	Scalar	60.21	28	0.97	0.050 (0.033-0.068)	0.002	-0.008
Fear (8)	Configural	137.43	40	0.84	0.073 (0.060-0.087)		
	Metric	144.08	47	0.84	0.068 (0.055-0.080)	0.000	-0.005
	Scalar	152.50	54	0.84	0.064 (0.052-0.076)	-0.002	-0.004
High Intensity Pleasure (8)	Configural	241.29	40	0.88	0.106 (0.093-0.119)		
	Metric	242.62	47	0.88	0.096 (0.084-0.108)	0.003	-0.010
	Scalar	249.77	54	0.88	0.090 (0.078-0.101)	0.000	-0.006
Impulsivity (7)	Configural	290.78	28	0.82	0.144 (0.129-0.159)		
	Metric	294.30	34	0.82	0.130 (0.117-0.144)	0.002	-0.014
	Scalar	299.10	40	0.82	0.120 (0.107-0.133)	0.001	-0.010

Title	Model	χ^2	df	CFI	RMSEA	CFI	RMSEA
Inhibitory Control (9)	Configural	125.17	54	0.96	0.054 (0.042-0.066)		
	Metric	133.22	62	0.96	0.050 (0.039-0.062)	0.000	-0.004
	Scalar	140.93	70	0.96	0.047 (0.036-0.059)	0.000	-0.003
Low Intensity Pleasure (8)	Configural	161.99	40	0.82	0.082 (0.069-0.096)		
	Metric	172.39	47	0.81	0.077 (0.065-0.089)	-0.005	-0.005
	Scalar	180.11	54	0.81	0.072 (0.060-0.084)	-0.001	-0.005
Perceptual Sensitivity (6)	Configural	49.77	18	0.98	0.062 (0.042-0.083)		
	Metric	56.30	23	0.98	0.057 (0.038-0.076)	-0.001	-0.005
	Scalar	63.73	28	0.97	0.053 (0.036-0.070)	-0.002	-0.004
Sadness (9)	Configural	109.56	54	0.90	0.048 (0.035-0.061)		
	Metric	117.58	62	0.90	0.045 (0.032-0.057)	0.000	-0.003
	Scalar	127.03	70	0.90	0.042 (0.030-0.054)	-0.002	-0.003
Shyness (8)	Configural	374.20	40	0.92	0.136 (0.124-0.149)		
	Metric	382.47	47	0.92	0.126 (0.114-0.137)	0.000	-0.010
	Scalar	383.64	54	0.93	0.116 (0.105-0.127)	0.002	-0.010
Smiling (7)	Configural	52.47	28	0.96	0.044 (0.025-0.062)		
	Metric	65.25	34	0.95	0.045 (0.028-0.062)	-0.010	0.001
	Scalar	69.52	40	0.96	0.041 (0.024-0.056)	0.003	-0.004
Soothability (7)	Configural	135.75	28	0.92	0.092 (0.077-0.108)		
	Metric	157.87	34	0.91	0.090 (0.076-0.104)	-0.012	-0.002
	Scalar	162.39	40	0.91	0.082 (0.069-0.096)	0.001	-0.008

Note: Numbers in parentheses indicates number of items on the scale. CFI = Comparative Fit Index; RMSEA = Root mean square error of approximation; CI = Confidence interval. Change in CFI and RMSEA are relative to the previous model (i.e., Metric vs. Configural; Scalar vs. Metric).