



Published in final edited form as:

J Abnorm Child Psychol. 2014 October ; 42(7): 1201–1211. doi:10.1007/s10802-014-9865-x.

Testing Models of Psychopathology in Preschool-aged Children Using a Structured Interview-based Assessment

Thomas M. Olino,

Department of Psychology, Temple University, 1701 N. 13th St, Philadelphia, PA 19122, USA
thomas.olino@gmail.com

Lea R. Dougherty,

University of Maryland College Park, College Park, MA, USA

Sara J. Bufferd,

California State University San Marcos, San Marcos, CA, USA

Gabrielle A. Carlson, and

Stony Brook University, Stony Brook, NY, USA

Daniel N. Klein

Stony Brook University, Stony Brook, NY, USA

Abstract

A number of studies have found that broadband internalizing and externalizing factors provide a parsimonious framework for understanding the structure of psychopathology across childhood, adolescence, and adulthood. However, few of these studies have examined psychopathology in young children, and several recent studies have found support for alternative models, including a bi-factor model with common and specific factors. The present study used parents' (typically mothers') reports on a diagnostic interview in a community sample of 3-year old children ($n=541$; 53.9 % male) to compare the internalizing-externalizing latent factor model with a bi-factor model. The bi-factor model provided a better fit to the data. To test the concurrent validity of this solution, we examined associations between this model and paternal reports and laboratory observations of child temperament. The internalizing factor was associated with low levels of surgency and high levels of fear; the externalizing factor was associated with high levels of surgency and disinhibition and low levels of effortful control; and the common factor was associated with high levels of surgency and negative affect and low levels of effortful control. These results suggest that psychopathology in preschool-aged children may be explained by a single, common factor influencing nearly all disorders and unique internalizing and externalizing factors. These findings indicate that shared variance across internalizing and externalizing domains is substantial and are consistent with recent suggestions that emotion regulation difficulties may be a common vulnerability for a wide array of psychopathology.

Keywords

Structure of psychopathology; Preschool psychopathology; Early childhood; Confirmatory factor analysis

Factor analytic studies of the structure of psychopathology have consistently revealed higher-order internalizing and externalizing factors that account for much of the common variance in psychiatric disorders (e.g., Krueger et al. 1998; Slade and Watson 2006; Vollebergh et al. 2001). Most of this work has focused on adults, but a growing number of studies examine school-aged children or adolescents (e.g., Lahey et al. 2008; Martel et al. 2010; Sterba et al. 2007; Trosper et al. in press). However, few studies have examined the structure of psychopathology in preschool-aged children, as research on psychiatric disorders in early childhood is still emerging (Egger and Emde 2011).

Historically, investigations of the structure of psychopathology in very young children utilized parent and teacher questionnaires (Achenbach et al. 1987; Achenbach and Rescorla 2000; Carter et al. 2003; Richman and Graham 1971). The Child Behavior Checklist (CBCL; Achenbach et al. 1987) is, perhaps, the most commonly used measure of youth behavior problems. Factor analyses of the CBCL revealed that the variance in child behavior problems was best summarized by six (or seven) lower-order dimensions (Ivanova et al. 2010). Two of these dimensions loaded on a higher order internalizing dimension, and two loaded on a higher order externalizing dimension. An alternative measure with increasing popularity is the Infant–Toddler Social and Emotional Assessment (ITSEA; Carter et al. 2003). The ITSEA includes items assessing behavior problems more typically found in younger children, including eating problems and sensory sensitivity, and other symptoms from the Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood: 0–3 (DC:0–3, Egger et al. 2005). In a confirmatory factor analysis of the ITSEA problem behavior scales, the best fitting model included higher-order externalizing, internalizing, and dysregulation factors, as well as a total problem factor with loadings to all specific problem behavior domains (Carter et al. 2003). The total problem factor provided a structural explanation for the high degree of covariation between internalizing and externalizing problems.

Although both the CBCL and ITSEA assess many problem behaviors, the domains identified by these instruments and Diagnostic and Statistical Manual-IV diagnoses (DSM-IV; American Psychiatric Association 1994) differ. The Child Symptom Inventory (CSI; Gadow and Sprafkin 1997, 2000) was developed as a parent-report instrument to assess the diagnostic constructs in the DSM. Using the CSI, Strickland et al. (2011) found that a model with distinct dimensions of major depressive disorder (MDD), generalized anxiety disorder (GAD), social phobia, separation anxiety, oppositional defiant disorder (ODD), inattention, and hyperactivity/ impulsivity fit the data better than models with higher-order factors. Thus, their best fitting model was consistent with the DSM-IV classification. However, the authors reported strong associations among the latent dimensions, with 17 of the 21 inter-factor correlations being greater than 0.50. This degree of inter-correlations raises the

possibility that alternative, more parsimonious, models may better describe the structure of preschool psychopathology.

Parent-report questionnaires have a number of virtues, including assessing a large number of youth quickly with limited expense; however, they also have limitations. Parent-reports are susceptible to mood-state and response-style biases; questions and/or responses may be misinterpreted by respondents; and respondents may differ in their thresholds for endorsing items. Interviewer-based structured and semi-structured interviews are an alternative assessment methodology that can reduce the impact of these problems. Interviewers can clarify questions, elicit explicit behavioral examples, and incorporate information from all aspects of the interview to increase the validity of the assessment. Moreover, they can employ standard definitions and thresholds for rating items, rather than relying on respondents' interpretations of the items. However, some limitations remain, including participants' limited recall and interviewer variability in eliciting and rating responses. In addition, the information is still based on parental perception of (problem) behaviors, as opposed to true behavior per se.

Recent work has used interview-based data to examine the structure of psychopathology in young children. Sterba et al. (2007) examined the structure of psychopathology in a sample selected to overrepresent emotional and behavioral problems. The authors found that six separate dimensions - MDD/GAD, social phobia, separation anxiety, ODD/CD, inattention, and hyperactivity/impulsivity - provided the best fit to the data. However, 11 of the 15 associations between pairs of dimensions were greater than 0.50. This finding suggests that there may be a more parsimonious solution; however, model comparisons found that the six-factor model provided a better fit to the data than alternative higher-order models.

In sum, results of studies investigating the structure of preschool psychopathology have varied, with some suggesting a broad hierarchical organization around internalizing and externalizing domains (Achenbach et al. 1987; Achenbach and Rescorla 2000), others suggesting a combination of general and specific domains (Carter et al. 2003), and others support a larger number of specific syndromes (Sterba et al. 2007; Strickland et al. 2011). However, these studies examined a relatively narrow set of models. Additionally, some used exploratory methods (Achenbach et al. 1987; Achenbach and Rescorla 2000); others used confirmatory model fitting approaches with higher-order structures (Sterba et al. 2007), and one study examined a hierarchical model, which was more conceptually complex, but empirically parsimonious (Carter et al. 2003). Importantly, no study compared higher-order and hierarchical models in the same dataset.

In higher-order models, the broad internalizing and externalizing dimensions are allowed to be correlated, and usually have moderate associations ($r \sim 0.50$; Krueger et al. 1998; Vollebergh et al. 2001). These models posit that any observed associations between specific internalizing and externalizing disorders are due to the associations between the higher-order factors. In addition, they assume that each disorder is part of either the internalizing or the externalizing spectrum, but not both. However, recent work challenges these assumptions. In a study of school-aged children, Lahey et al. (2008) found that a model with higher-order internalizing and externalizing disorders fit the data well; however, MDD and GAD had

strong associations with both the internalizing and externalizing disorder factors. Similarly, in adults, Eaton et al. (2011) found that borderline personality disorder had significant factor loadings on both the distress sub-factor of the internalizing dimension, and the externalizing dimension. These data suggest that additional influences may account for the associations of disorders across the internalizing and externalizing domains.

Hierarchical models, also termed bi-factor models, posit that a single underlying vulnerability may be associated with all relevant dimensions, but individual factors may explain additional, unique variance in a smaller set of dimensions. Bi-factor models have been examined in the psychopathology literature only recently focusing on internalizing (Simms et al. 2012) and externalizing domains (Martel et al. 2010) individually.

In addition to Carter et al. (2003) discussed above, three studies have applied a hierarchical model across the internalizing and externalizing domains in older youth and adults. In a school-aged twin sample, Lahey et al. (2011) examined the heritability of latent factors accounting for variance in disorders. The authors found support for a broad, general factor that influenced all disorders and two additional specific factors, one with loadings to internalizing disorders and the other with loadings to externalizing disorders. The broad, general factor accounted for a moderate amount of genetic variance in disorders. For the individual internalizing and externalizing factors, a higher proportion of variance was explained by unique genetic variance than shared environmental effects. Thomas (2011) examined the Brief Symptom Inventory in a large sample of adult outpatients. Using all internalizing, externalizing, and psychosis items, the author reported that the best fitting model included a general distress factor and specific domain factors. Finally, in a large community sample of adults, Lahey et al. (2012) found that the bi-factor model provided a better fit to the data than a model including fear and distress internalizing domains and a single externalizing domain.

The present study examined the higher-order structure of psychopathology in a community sample of preschool-aged children who were assessed with a structured clinical interview with the primary parent (almost all mothers). As the rates of specific disorders were low (Bufferd et al. 2011), symptomatology was indexed by symptom counts. This maintained model parsimony and a focus on the structure of higher-order factors. We compared a two-factor model specifying internalizing and externalizing dimensions and a model that specified both general and specific factors (i.e., a bi-factor model). As previous studies have found high correlations among individual syndromes (Sterba et al. 2007; Strickland et al. 2011) and among internalizing and externalizing factors (e.g., Krueger and Markon 2006), we hypothesized that a model with a general factor (i.e., a bi-factor model) would provide a superior fit to the data than a higher-order model.

Previous studies of psychopathology structure have largely focused on explaining relationships between disorders using latent dimensions; however, few studies have examined the external validity of the latent factors (e.g., Lahey et al. 2012). This additional step enhances our understanding of the latent structure of psychopathology by describing the nature of the underlying pathological processes. Temperament has garnered much support as a basis for understanding the structure of psychopathology, as different temperament

dimensions have been consistently linked to the internalizing and externalizing spectra (e.g., Clark 2005; Hink et al. 2013). However, these associations have rarely been examined in studies of the latent structure of youth psychopathology and, when they have been examined, have often relied on the same informant for both psychopathology and temperament. Thus, we tested the validity of our latent dimensions of psychopathology by examining associations with temperament measures derived from independent sources: laboratory observations and paternal reports.

Method

Participants

Participants were from a larger sample of 559 3-year old children and their families in a longitudinal study of temperament and risk for psychopathology (Olino et al. 2010). Only one child per family was included. A parent of 541 (96.8 %) of these children was interviewed regarding psychopathology in their child (Bufferd et al. 2011). The child's biological mother was the informant in 98.5 % (533/541) of the assessments. The sample was recruited by contacting families from commercial mailing lists. Families with a 3-year-old child living with at least one biological parent within 20 contiguous miles of Stony Brook, New York were eligible to participate. The primary caregiver was required to speak English, and children with significant medical disorders or developmental disabilities were excluded. The screening identified 815 families as eligible, 66.4 % of whom entered the study. No significant differences were found between families who did and did not participate on child sex and race/ethnicity, and parental marital status and education. Informed consent was obtained from the parent prior to participation. The study was approved by the institutional review board at Stony Brook University, and families were compensated for their participation.

The mean age of the children was 3.56 years ($SD = 0.27$); 247 (53.9 %) were male and 398 (86.9 %) were White/non-Hispanic. Children's receptive language ability, as assessed by the Peabody Picture Vocabulary Test (PPVT; Dunn and Dunn 1997), was in the average range ($M = 102.9$, $SD = 13.9$). The mean ages of mothers and fathers were 36.2 years ($SD = 4.5$) and 38.5 years ($SD = 5.4$), respectively. Most parents were married or cohabiting (96.0 %), approximately half of the parents (56.7 % of the mothers and 46.7 % of the fathers) had graduated from college, and median household income bracket was \$70,000 to \$90,000. Census data suggest the sample is reasonably representative of the surrounding county, where 87.3 % of families with children between 3 and 5 years of age lived with both parents, 79.0 % of individuals were White/non-Hispanic, 48.1 % of adults aged 25–54 graduated from college, and the median household income bracket was \$75,000 to \$99,999.

Child Psychopathology

The Preschool Age Psychiatric Assessment (PAPA; Egger et al. 1999) is a structured, interviewer-based diagnostic interview designed to assess parent-reported psychiatric disorders in preschoolers between the ages of 2 and 6 years. The interviewer must adhere to the protocol and ask all required questions and must also confirm the parent's understanding of the questions, elicit examples of relevant behaviors, and apply a priori guidelines for

rating symptoms using a glossary. To maximize accuracy of recall, symptoms occurring within the 3 months prior to the interview are rated. Satisfactory test-retest reliability has been reported using independent interviews (Egger et al. 2006). The PAPA covers a comprehensive set of symptoms from the DSM-IV. Only 2.2 % of the respondents ($n = 12$) reported that their child had received mental health treatment. Bufferd et al. (2011) reported rates of disorders in this sample: depressive disorders 1.8 %; anxiety disorders 19.6 %; oppositional defiant disorder (ODD) 9.4 %; and attention deficit hyperactivity disorder (ADHD) 2.0 %.

In this study, we used symptom scales for depression, GAD, phobias (the sum of social phobia, specific phobia, and agoraphobia), separation anxiety, panic, inattention, hyperactivity, impulsivity, and ODD to parallel the domains usually examined in previous studies. For each dimension, symptoms were counted as present if they were exceeded a clinically significant threshold. Symptoms (e.g., concentration problems) that are common to multiple disorder dimensions were included for all relevant dimensions.

Interviews were conducted by advanced graduate students in clinical psychology who were trained on administration by one of the original PAPA interviewers. Interviews were conducted by telephone and usually lasted about one hour. To examine inter-rater reliability, a second rater from the pool of interviewers independently rated audiotapes of 21 PAPA interviews. The interviews were randomly selected, but we oversampled participants who reported mental health problems. The intraclass correlations (ICCs) for the symptom scales were 0.85 for depression, 0.93 for GAD, 1.00 for separation anxiety, 1.00 for panic, 0.98 for phobias, 0.99 for inattention, 0.99 for hyperactivity, 0.95 for impulsivity, and 0.99 for ODD.

The Early Childhood Inventory-4 (ECI-4) is a parent rating scale used to screen DSM-IV emotional and behavioral disorders in 3- to 6-year-olds (Gadow and Sprafkin 2000). In this study, parents completed the ADHD and ODD sections of the inventory. Sprafkin and colleagues (2002) reported that the correct classification rates for ADHD and ODD with respect to chart diagnoses were 60 % and 74 %, respectively. In the present sample, internal consistency estimates (α) were 0.79 for ADHD-Inattention, 0.82 for ADHD-Hyperactivity/Impulsivity, and 0.85 for ODD.

Due to concerns about administration time, in the first 60 % of the sample ($n=324$), the interviewer used the ECI-4 ADHD and ODD scales as a screen to help determine whether to complete the ADHD and ODD sections of the PAPA. All ECI ODD and ADHD items were reviewed by the interviewers. When parent-reports indicated a low likelihood of ODD or ADHD symptoms (i.e., most items were endorsed as “never” or “sometimes”), interviewers probed the respective broad domains to confirm the absence of symptoms before skipping out. When parent-reports indicated a potential likelihood of ODD or ADHD symptoms (i.e., items endorsed as “often” or “very often”), the corresponding PAPA sections were administered in their entirety. In the remaining 40 % of the sample ($n=217$), the PAPA ADHD and ODD sections were administered to all parents. ECI-4 ADHD and ODD item scores were used to impute missing PAPA ADHD and ODD item scores using maximum likelihood imputation estimation procedures for missing values (Acock 2005). This is less

biased than pairwise and listwise deletion procedures, even with large amounts of missing data (i.e., upwards of 70 % missing data; Graham et al. 2007; Navarro Pastor 2003).

Observations of Child Temperament

Laboratory Temperament Assessment Battery (Lab-TAB)—These procedures have been reported previously (Dougherty et al. 2011; Dyson et al. 2012; Olino et al. 2010), hence we summarize them briefly. Each child and a parent visited the laboratory for a 2-hour observational assessment of temperament that included a standardized set of 12 episodes selected to elicit a range of temperament-relevant behaviors. Tasks were selected from the Lab-TAB (Goldsmith et al. 1995). To prevent carry-over effects, no episodes presumed to evoke similar affective responses occurred consecutively, and each episode was followed by a brief play break. Each episode was videotaped and later coded. Experimenters/coders were unaware of all data on child psychopathology. The episodes were selected to elicit a number of affective and behavioral dimensions, including positive affect (*Make that Car Go, Pop-up Snakes, Popping Bubbles*), fear (*Risk Room, Stranger Approach, Exploring New Objects*), anger/frustration and sadness (*Transparent Box, Impossibly Perfect Green Circles, Box Empty*), and impulsivity/inhibitory control (*Tower of Patience, Arc of Toys, Snack Delay*).

Coding Procedures—Coding procedures followed those reported in a previous study (Durbin et al. 2005). The scoring developed for these assessments has been related to home observations and demonstrated moderate stability over time in an independent sample (Durbin et al. 2007). The following temperament traits, emotional displays, and behaviors were rated: positive affect (PA), fear, sadness, anger, BI, inhibitory control, interest, activity, sociability, impulsivity, initiative, anticipatory PA, dominance, warmth, social interest, affiliation, assertiveness, clinginess, pushiness, hostility, noncompliance, avoidance, and social fear. Each variable was rated for all episodes except BI and inhibitory control.

Each display of facial, bodily and vocal affect (PA, fear, sadness, anger) in each episode was rated on a three-point scale (low, moderate, high). Ratings were summed separately within each channel (facial, bodily, vocal) across the 12 episodes, standardized, and averaged across the three channels to derive total scores for each affect domain. With the exception of BI and inhibitory control, the other variables were rated once per episode on a 4-, 5-, or 10-point scale, and averaged across episodes. Coefficient alphas ranged from 0.50–0.87 (median= 0.70), and intraclass coefficients (ICCs) for interrater reliability ranged from 0.40–0.92 (median=0.75; $n=35$).

BI was coded in the three episodes designed to assess this behavior: *Risk Room, Stranger Approach*, and *Exploring New Objects*. Specific behaviors, including latency to touch objects, tentative play, gaze aversion, latency to vocalize, and approach to and avoidance of stranger were coded for each epoch, which spanned 20–30 seconds depending on the particular episode. Within each epoch, a maximum intensity rating of facial, bodily, and vocal fear was also coded. Except for latencies, behaviors were coded on three- or four-point scales. BI was computed as the average standardized ratings across epochs and episodes. Coefficient alpha for the BI scale was 0.80 and the interrater ICC was 0.88 ($n = 35$).

Inhibitory control was coded in two episodes: *Tower of Patience* and *Snack Delay*. The coding system was adapted from Carlson (2005) and involved tallying the number of times a child failed to wait his or her turn during the episode. *Tower of Patience* consisted of 14, and *Snack Delay* consisted of 7, trials. The global inhibitory control variable ($\alpha=0.70$, interrater ICC=0.98, $n=8$) was constructed by adding the standardized scores for the two episodes.

Principal Components Analysis (PCA) of Lab-TAB Variables—We (Dougherty et al. 2011) estimated a PCA to obtain a smaller number of temperament scores for additional analyses. Based on the eigenvalue > 1.0 rule and inspection of the scree plot, five components, consistent with results from Dyson et al. (2012), were extracted. Each component was easily interpretable and included several theoretically relevant variables with loadings greater than 0.40 and few cross-loadings on multiple factors. Unit-weighted scores were derived for each child to yield five temperament scales: sociability/assertiveness ($\alpha=0.93$); dysphoria ($\alpha=0.80$); fear ($\alpha=0.71$); exuberance ($\alpha=0.88$); and disinhibition ($\alpha=0.70$). Items were not double-counted on scales: in the few cases where there were cross-loadings, the item was added to the scale with the higher loading. Inter-rater ICCs were acceptable for all scales.

Paternal Report of Child Temperament

Child Behavior Questionnaire (CBQ; Rothbart et al. 2001)—Fathers completed the CBQ, a widely used 195-item caregiver report measure of temperament for 3- to 7-year-old children. Paternal reports were available for 402 children. Maternal reports were not used in order to reduce informant effects on PAPA and CBQ associations. However, in the eight cases when fathers were the informants for the PAPA, maternal CBQ reports were used. We relied on the three broadband factors identified by Rothbart et al. (2001): surgency (activity level, high intensity pleasure, impulsivity, and shyness [reversed]), negative affect (anger-frustration, discomfort, fear, sadness, and soothability [reversed]); and effortful control (attention focusing, inhibitory control, low intensity pleasure, and perceptual sensitivity). Coefficient alphas for surgency, negative affect, and effortful control were 0.71, 0.77, and 0.67, respectively.

Data Analysis

As would be expected in a community sample, symptom scores were positively skewed. Hence, all analyses were evaluated using robust maximum likelihood estimation methods (MLR). We examined a single-factor model; a two-factor model with correlated internalizing and externalizing broad band factors; and a bi-factor model, including a single underlying factor with loadings to all symptom score indicators and unique internalizing and externalizing factors with loadings to internalizing and externalizing score indicators, respectively.

All models were estimated using Mplus, version 7.0 (Muthén and Muthén 1998–2012). Models were evaluated on several indices of goodness of fit, as well as whether or not theoretical predictions, as indicated by specific paths within the model, were supported. To evaluate the overall fit of the models, three indices were selected: the comparative fit index (CFI; Bentler 1990), the root mean square error of approximation (RMSEA; Steiger 1990),

and the Bayesian Information Criteria (BIC). Although existing guidelines are arbitrary (Marsh et al. 2004), current conventions suggest that good fit is indicated by a CFI greater than 0.95 (Hu and Bentler 1999) and a RMSEA below 0.05 (MacCallum et al. 1996); adequate fit is indicated by a CFI greater than 0.90 and a RMSEA between 0.05 and 0.10. The BIC is a relative fit index, with lower values indicating better fit. Sattora-Bentler chi-square difference tests were conducted to examine differences in model fit. For completeness, we also report chi-square tests, but do not interpret them as minor deviations indicate poor fit in larger samples. Models were fit using full data (including data imputed from the ECI) yielding results for the full sample and for participants who had complete PAPA data.

Results

Confirmatory factor analysis (CFA) was used to determine whether the ADHD and ODD scales are better represented as single or multiple dimensions. As all indicators for these models were dichotomous, a robust weighted least squares estimator was employed (WLSMV) and the DIFFTEST procedure was used to evaluate differences in model fit. For ADHD, CFA models were estimated for a one-factor ($\chi^2[135]=378.88$, $P<0.001$; CFI=0.94; RMSEA=0.06 (0.05–0.06)); two-factor (inattention vs. hyperactivity/ impulsivity; $\chi^2[134]=327.79$, $P < 0.001$; CFI=0.96; RMSEA =0.05 (0.04–0.06)); and three-factor solutions (inattention vs. hyperactivity vs. impulsivity; $\chi^2[132]=314.76$, $P < 0.001$; CFI=0.96; RMSEA=0.05 [0.04–0.06]). Specific model comparisons found that the two-factor model fit better than the one-factor model ($\chi^2[1]=37.47$, $P < 0.001$) and the three-factor model fit better than the two-factor model ($\chi^2[2] = 16.49$, $P < 0.001$). Thus, we retained three dimensions for ADHD.

For ODD, CFA models were estimated for one- and a two-factor models differentiating across emotional (e.g., temper outbursts, touchy/easily annoyed, angry/resentful, and spiteful) and non-emotional symptoms (e.g., argue with grownups, refusal to follow rules, annoy others on purpose, and blame others for poor behavior). Three-factor models have also been suggested in the literature (e.g., Burke et al. 2010; Stringaris and Goodman 2009), but we did not examine a three factor model as some factors would have included only one or two items, yielding unstable solutions. The one-factor ($\chi^2[20]=35.63$, $p = 0.02$, CFI = 0.99, and RMSEA= 0.04 [0.02–0.06]) and the two-factor ($\chi^2[19]=31.37$, $p=0.04$, CFI=0.99, and RMSEA=0.04 [0.01–0.06]) models both provided excellent fits to the data, although the two-factor model provided a significantly better fit ($\chi^2[1] =5.79$, $p = 0.02$). However, the association between the two latent factors was extremely strong ($r = 0.93$), suggesting that the factors were highly overlapping. Given the magnitude of this correlation, we estimated ODD as a single dimension.

Table 1 displays the full correlation matrix and sample statistics for the symptom scales for the full sample (i.e., including all imputed data). All scale scores were computed by summing the endorsed items for each disorder (or symptom dimensions for ADHD; i.e., inattention, hyperactivity, and impulsivity). The strongest bivariate association was between depression and GAD. Depression and GAD symptoms were also associated with symptoms of all other disorders. Symptoms of externalizing disorders were all associated with each

other. Anxiety symptoms, other than GAD, tended to be associated with each other, but were not significantly associated with externalizing symptoms. Table 2 provides overall fit information for the models; the top panel of the table provides information for the full sample and the bottom panel provides fit information for models based on participants with complete data in the PAPA ADHD and ODD sections. Conclusions were substantively similar across the two sets of analyses.

One-factor Model

The one-factor model specified that all disorders were due to a single underlying vulnerability factor. Inattention, hyperactivity, ODD, GAD, and impulsivity had significant factor loadings, but depression, separation anxiety, panic, and phobias did not. Overall, this model provided a very poor fit to the data (Table 2). Standardized factor loadings ranged from 0.13 (phobias) to 0.68 (GAD).

Two-factor Model

The two-factor model specified that depression, GAD, separation anxiety disorder, panic disorder, phobic disorder scores loaded on an internalizing factor and inattention, hyperactivity, impulsivity, and oppositional-defiant disorder scores loaded on an externalizing factor. Factor loadings for all disorders on their respective factors were highly significant ($p < 0.001$). The correlation between the internalizing and externalizing factors was of a moderate magnitude ($r = 0.54$, $p < 0.001$). Overall, model fit was marginal (Table 2), but significantly better than for the one-factor model. Standardized factor loadings ranged from 0.10 (panic) to 0.68 (GAD) for the internalizing factor and from 0.35 (impulsivity) to 0.93 (ODD) for the externalizing factor.

Bi-factor model

The bi-factor model specified that all symptom scores loaded on a general factor; GAD, depression, separation anxiety disorder, panic disorder, and phobic disorder scores loaded on an internalizing factor; and inattention, hyperactivity, impulsivity, and oppositional-defiant disorder scores loaded on an externalizing factor. Correlations between the common latent factor and the internalizing and externalizing latent factors were constrained at zero to permit model estimation. The association between the internalizing and externalizing factor was freely estimated.

The association between the internalizing and externalizing factors was non-significant ($r = -0.05$, $p = 0.77$). As model fit was only adequate, we inspected modification indices for residual correlations that could be added to improve model fit. The largest modification index indicated that including a residual correlation between the error terms of the depression and inattention indicators would improve model fit. As some of the PAPA items used to assess concentration problems were similar for both disorders, we included this correlation in the model. Thus, we re-estimated the model after fixing the association between the internalizing and externalizing factors to zero and freely estimating the residual correlation between the error terms of the depression and inattention indicators.

This model provided an excellent fit to the data (Table 2). Factor loadings from the common latent factor to depression, GAD, inattention, hyperactivity, impulsivity, and ODD were highly significant ($p < 0.001$); loadings for separation anxiety and panic were also significant ($p < 0.05$); however, the factor loading for phobias was non-significant ($p = 0.70$). Factor loadings for internalizing and externalizing indicator scores on their respective factors were all highly significant ($p < 0.001$). As the factor loading from the common latent factor to the phobias score was non-significant, a final model was estimated after removing this path. Model fit was nearly identical to the previous model with the same pattern of factor loading strengths. In addition, the BIC for the bi-factor models (both the full and trimmed versions) were lower than any of the other models tested. Factor loading parameters are displayed in Fig. 1.

External Validity of Factors

After identifying the best fitting model, we estimated associations between the common, internalizing, and externalizing latent factors and dimensions of observed and paternal-reported temperament (Table 3). In these models, individual temperament dimensions were regressed on all three latent factors and effect sizes were computed from t -statistics [$r = t^2 / (t^2 + df)$]. This estimation method permitted examining associations for the full sample, capitalizing on robust maximum likelihood estimation. For observed temperament dimensions, we found that sociability and exuberance were negatively associated, and fear/inhibition was positively associated, with the internalizing factor score. We also found that disinhibition was positively associated with the externalizing factor score. While there was a weak trend for disinhibition, no observed temperament dimension was significantly associated with the common latent factor score.

Using paternal reports of temperament, we found that CBQ surgency was negatively associated with the internalizing factor score. CBQ surgency was positively associated, and CBQ effortful control was negatively associated, with the externalizing factor score. Finally, CBQ surgency and CBQ negative affect were positively associated, and CBQ effortful control was negatively associated, with the common factor score.¹

Discussion

Recent work indicates that psychopathology in young children is more common than had been previously assumed (Bufferd et al. 2011; Egger and Emde 2011). However, few studies have examined the structure of psychopathology in young children, particularly using interviewer-based diagnostic interviews. The present study examined the structure of psychopathology in an unselected community sample of preschool aged children using a semi-structured diagnostic interview. Based on the strong associations between factors in

¹In addition to relying on the CBQ factor scores, we conducted secondary analyses using a rationally constructed positive emotionality scale. The CBQ surgency scale largely reflects impulsivity and sensation seeking and limits our understanding of how a more traditional positive emotionality construct is associated with the latent dimensions. We examined a composite of two scales – smiling/laughter and shyness-reverse scored ($r = 0.35, p < 0.001$) – to index positive emotionality. Using this scale, positive emotionality was negatively associated with the internalizing factor ($r = -0.19, p < 0.001$) and positively associated with the externalizing factor ($r = 0.11, p < 0.01$), however it was not related to the common factor ($r = 0.04, p = 0.45$)

previous studies, we hypothesized that a bi-factor model would provide the best fit to the data.

Using symptom scores across a total of seven internalizing and externalizing disorders (with ADHD represented by three dimensions, resulting in a total of nine symptom scales), we found that a bi-factor model best fit the data. Our best fitting model found that all disorders, except for the phobias, loaded on the latent common vulnerability factor, all internalizing symptom scores loaded on a latent internalizing factor, and all externalizing symptom scores loaded on an externalizing latent factor. After taking the common factor into account, the internalizing and externalizing factors were no longer associated. This suggests that the well-replicated correlation between the internalizing and externalizing dimensions may be due to a common vulnerability. Our findings are consistent with the structure identified by the ITSEA in young children (Carter et al. 2003) and a diagnostic interview in the NESARC study of adults (Lahey et al. 2012). Similar to our results, these studies found that different forms of psychopathology, across both internalizing and externalizing dimensions, were explained by a single underlying factor with additional factors for the internalizing and externalizing domains.

Our results stand in contrast to some other studies of the structure of preschool psychopathology (Sterba et al. 2007; Strickland et al. 2011). In a sample of youth recruited to have higher levels of psychopathology (Sterba et al. 2007) and in a sample of unselected youth (Strickland et al. 2011) models consisting of individual DSM-IV disorders best fit the data. However, most factor correlations in both studies exceeded 0.50, suggesting that the relationships among factors may be explained more parsimoniously. In both studies, the authors examined a number of higher-order factor models that sought to explain the correlations among the factors. However, neither tested a bi-factor model.

In our data, the two-factor model, comprised of distinct internalizing and externalizing factors, provided a marginal fit to the data. The magnitude of the association between these broad factors was similar to a number of previous studies (Krueger et al. 1998; Krueger and Markon 2006; Slade and Watson 2006; Vollebergh et al. 2001). However, a bi-factor model provided a significantly better fit, raising the theoretically provocative possibility that a general vulnerability to psychopathology may explain the full complement of comorbidity.

Interestingly, we found that the phobias (social and specific) did not load on the general factor, but loaded solely on the internalizing factor. This suggests that the internalizing factor largely reflects a fear component, which is consistent with the association with the observed fear/inhibition dimension. Likewise, Lahey et al. (2011) found that specific phobia and agoraphobia had modest loadings ($B_s=0.19$ and 0.15 , respectively) on the general shared genetic factor, but stronger loadings on the shared genetic internalizing factor ($B_s=0.27$ and 0.52 , respectively). Alternatively, these findings may be accounted for by normative fears in children, which may not be influenced by the general vulnerability factor. In contrast, GAD had a very high loading on the common factor. This is consistent with the view that GAD is the disorder that is most highly saturated with negative emotionality/neuroticism, which is thought to contribute to most forms of psychopathology.

The bi-factor model explains the covariation of the symptom data using latent factors. However, the underlying processes that give rise to these vulnerabilities are not specified. To test the validity of the model and better understand the nature of these latent factors and processes, we estimated associations of the latent factors with observational laboratory-based and paternal-report measures of temperament. Some theorists have suggested that the internalizing and externalizing factors both are influenced by neuroticism, some of the internalizing disorders are influenced by low extraversion, and the externalizing factor is influenced by low constraint (Clark 2005; Krueger et al. 1998). Consistent with this work, we found that the internalizing factor was associated with lower levels of observed exuberance and sociability and father-reported surgency and higher levels of observed fear/inhibition. The externalizing factor was associated with higher levels of observed disinhibition and paternal-reported surgency and lower levels of paternal-reported effortful control. Importantly, the CBQ surgency scale largely includes scales related to impulsivity and sensation seeking, which may account for the association between surgency scale and the externalizing factor score.

We also found that fathers' reports of surgency, negative affect, and low effortful control were associated with the common factor. This suggests that the common factor includes variance pertaining to all three major dimensions of temperament, as observed by children's fathers. However, we did not find any significant associations between laboratory-assessed temperament and the common factor. Paternal reports of temperament may be associated with the common factor because parents see children across multiple contexts over long periods that are quite different from structured laboratory observations. Hence, it is possible that parents are more sensitive than lab observations to detecting temperamental links to a nonspecific dimension of child psychopathology. As described above, however, associations were found between laboratory-based indices of temperament and specific internalizing and externalizing factors.

Previous studies could not rule out the possibility that the common factor reflects method factors, such as a bias to report high levels of psychopathology (e.g., Carter et al. 2003). We are also unable to rule out a common method variance explanation. However, we believe that the links between paternal reports of temperament and the common factor argue against this interpretation. It is unlikely that method variance can entirely account for associations between two variables tapping different constructs (psychopathology versus temperament), assessed using different measurement approaches (structured, interviewer-based, diagnostic interview versus self-administered questionnaire), and different informants (mothers versus fathers). Rather, the common factor is likely to reflect some systematic variability in behavior. Recently, conceptual work has argued (Clark 2005), and empirical work (Hink et al. 2013; Kotov et al. 2010) has found, that neuroticism has strong and consistent associations with all common mental disorders, suggesting that the common factor could reflect neuroticism. In preschool-aged children, this common factor may manifest as negative emotionality (NE), which may be less influenced by cognitive factors implicated in neuroticism. Indeed, Carter et al. (2003) found that the strongest factor loading for their general factor was for NE. However, the present data are not fully consistent with this conjecture. The common factor was associated with paternal reported surgency (although not observed positive emotionality), negative affect, and low effortful control. Thus, it may

be accurate to characterize the common factor in preschool-aged children as reflecting dysregulated emotionality. This finding suggests a general problem with emotion regulation that cuts across multiple forms of psychopathology and is consonant with recent views of emotion regulation deficits as a transdiagnostic construct (Aldao and Nolen-Hoeksema 2010; Nolen-Hoeksema and Watkins 2011).

Our study had a reasonably large sample assessed with an interviewer-based, semi-structured diagnostic interview. We also examined theoretically motivated models to explain the observed data, and tested the validity and explored the nature of the latent variables using independent validators. However, the study should be considered in light of a number of limitations. First, as we used a community-based, rather than clinical, sample, symptom levels were modest. Associations among disorders may differ when symptom levels are higher. Second, we examined symptom scores, as opposed to diagnoses or individual criteria-level data. However, the modest level of symptomatology meant that there would have been too few cases of some disorders to examine covariation between full syndromes. Alternatively, using criteria-level data would require estimating many more parameters, which would reduce the stability of the model solutions. Third, ADHD and ODD symptoms were assessed using different approaches for portions of the sample. For approximately half the sample, a screener and skip-out strategy was employed and for the other half skip-outs were not used and full information was obtained through the interview. Thus, data for many participants in the first half of the sample were imputed based on available responses. However, overall model fit for the two halves of the sample was quite comparable. Fourth, a number of less common disorders were not assessed, including post-traumatic stress, obsessive compulsive, bipolar, conduct, pervasive developmental, and elimination disorders. Fifth, although our sample was fairly representative of the local area, our recruitment base consisted of individuals from a commercial mailing list. That may have introduced some degree of non-representativeness to the sample. Sixth, despite relying on a diagnostic interview to assess preschool psychopathology, the data are based on parents' perceptions of children's behavior. Incorporating data from other sources (e.g., a second caretaker, observation of the child) would provide a more comprehensive assessment and minimize biases associated with a single informant. Finally, although adequate fit was achieved without reliance on post-hoc modifications to our model, we implemented one post-hoc modification to achieve good fit (i.e., the residual correlation between depression and inattention). Thus, this modification may have capitalized on chance to improve model fit. Future work is necessary to determine whether this modification is sample specific, or if more robust models can be identified to explain the covariation among symptoms in very young children.

In conclusion, we found that the dimensions underlying preschool psychopathology are best described by a bi-factor model, including a general factor and two specific dimensions reflecting internalizing and externalizing disorders. In future studies, it will be important to investigate continuities and discontinuities in the structure of psychopathology across development as the incidence of, and relationships between disorders change. Understanding the nature of these latent factors and whether and how they change over the course of development may provide critical information about specific targets for prevention efforts

and how and when to intervene to prevent the onset of a wider array of specific clinical conditions.

Acknowledgments

This work was supported by R01 MH069942 (PI: Klein), K01 MH092603 (PI: Olino), F31 MH084444 (PI: Bufferd), and M01 RR10710 (Stony Brook University General Clinical Research Center). Dr. Carlson receives funding from Glaxo Smith Kline and Bristol Myers Squibb.

References

- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA Preschool Forms & Profiles. Burlington: University of Vermont, Research Center for Children, Youth, and Families; 2000.
- Achenbach TM, Edelbrock C, Howell CT. Empirically based assessment of the behavioral/emotional problems of 2-and 3-year-old children. *Journal of Abnormal Child Psychology*. 1987; 15(4):629–650. [PubMed: 3437096]
- Acock AC. Working with missing values. *Journal of Marriage and Family*. 2005; 67(4):1012–1028.
- Aldao A, Nolen-Hoeksema S. Specificity of cognitive emotion regulation strategies: A transdiagnostic examination. *Behaviour Research and Therapy*. 2010; 48(10):974–983. [PubMed: 20591413]
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th edition, rev.. Washington, DC: Author; 1994.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107(2):238–246. [PubMed: 2320703]
- Bufferd SJ, Dougherty LR, Carlson GA, Klein DN. Parent-reported mental health in preschoolers: Findings using a diagnostic interview. *Comprehensive Psychiatry*. 2011; 52(4):359–369. [PubMed: 21683173]
- Burke JD, Hipwell AE, Loeber R. Dimensions of oppositional defiant disorder as predictors of depression and conduct disorder in preadolescent girls. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2010; 49(5):484–492. [PubMed: 20431468]
- Carlson SM. Developmentally sensitive measures of executive function in preschool children. *Developmental neuropsychology*. 2005; 28(2):595–616. [PubMed: 16144429]
- Carter AS, Briggs-Gowan MJ, Jones SM, Little TD. The infant-toddler social and emotional assessment (ITSEA): Factor structure, reliability, and validity. *Journal of Abnormal Child Psychology*. 2003; 31(5):495–514. [PubMed: 14561058]
- Clark LA. Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology*. 2005; 114(4):505–521. [PubMed: 16351374]
- Dougherty LR, Bufferd SJ, Carlson GA, Dyson M, Olino TM, Durbin CE, Klein DN. Preschoolers' observed temperament and psychiatric disorders assessed with a parent diagnostic interview. *Journal of Clinical Child and Adolescent Psychology*. 2011; 40(2):295–306. [PubMed: 21391025]
- Dunn, LM.; Dunn, LM. Peabody picture vocabulary Test. 3rd eds.. Circle Pines: American Guidance Service; 1997.
- Durbin CE, Klein DN, Hayden EP, Buckley ME, Moerk KC. Temperamental emotionality in preschoolers and parental mood disorders. *Journal of Abnormal Psychology*. 2005; 114(1):28–37. [PubMed: 15709809]
- Durbin CE, Hayden EP, Klein DN, Olino TM. Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion*. 2007; 7(2):388–399. [PubMed: 17516816]
- Dyson MW, Olino TM, Durbin CE, Goldsmith HH, Klein DN. The structure of temperament in preschoolers: A two-stage factor analytic approach. *Emotion*. 2012; 12(1):44–57. [PubMed: 21859196]
- Eaton N, Krueger R, Keyes K, Skodol A, Markon K, Grant B, Hasin D. Borderline personality disorder co-morbidity: relationship to the internalizing-externalizing structure of common mental disorders. *Psychological Medicine*. 2011; 41(05):1041–1050. [PubMed: 20836905]

- Egger HL, Emde RN. Developmentally sensitive diagnostic criteria for mental health disorders in early childhood: DSM-IV, RDC-PA, and the revised DC: 0-3. *American Psychologist*. 2011; 66(2):95–106. [PubMed: 21142337]
- Egger, HL.; Ascher, BH.; Angold, A. *The Preschool Age Psychiatric Assessment: Version 1.1*. Durham, NC: Center for Developmental Epidemiology, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center; 1999.
- Egger, HL.; Fenichel, E.; Guedeney, A.; Wise, BK.; Wright, HH. *Diagnostic classification of mental health and developmental disorders of infancy and early childhood: Revised edition (DC: 0-3R)*. Washington: Zero To Three Press; 2005.
- Egger HL, Erkanli A, Keeler G, Potts E, Walter BK, Angold A. Test-retest reliability of the preschool age psychiatric assessment (PAPA). *Journal of the American Academy of Child and Adolescent Psychiatry*. 2006; 45(5):538–549. [PubMed: 16601400]
- Gadow, K.; Sprafkin, J. *Child Symptom Inventory 4, Norms manual*. Stony Brook: Checkmate Plus; 1997.
- Gadow, K.; Sprafkin, J. *Early childhood inventory-4 screening manual*. Stony Brook: Checkmate Plus; 2000.
- Goldsmith, HH.; Reilly, J.; Lemery, KS.; Longley, S.; Prescott, A. *Laboratory Temperament Assessment Battery: Preschool Version*. 1995. Unpublished Manuscript
- Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*. 2007; 8(3):206–213. [PubMed: 17549635]
- Hink LK, Rhee SH, Corley RP, Cosgrove VE, Hewitt JK, Schulz-Heik RJ, Waldman ID. Personality dimensions as common and broadband-specific features for internalizing and externalizing disorders. *Journal of Abnormal Child Psychology*. 2013; 41:939–957. [PubMed: 23474797]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6(1):1–55.
- Ivanova MY, Achenbach TM, Rescorla LA, Harder VS, Ang RP, Bilenberg N, Dias P. Preschool psychopathology reported by parents in 23 societies: testing the seven-syndrome model of the child behavior checklist for ages 1.5-5. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2010; 49(12):1215–1224. [PubMed: 21093771]
- Kotov R, Gamez W, Schmidt F, Watson D. Linking “big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*. 2010; 136(5):768–821. [PubMed: 20804236]
- Krueger RF, Markon KE. Reinterpreting comorbidity: A model-based Approach to Understanding and Classifying Psychopathology. *Annual Review of Clinical Psychology*. 2006; 2:111–133.
- Krueger RF, Caspi A, Moffitt TE, Silva PA. The structure and stability of common mental disorders (DSM-III-R): A longitudinal-epidemiological study. *Journal of Abnormal Psychology*. 1998; 107(2):216–227. [PubMed: 9604551]
- Lahey BB, Rathouz PJ, Van Hulle C, Urbano RC, Krueger RF, Applegate B, Waldman ID. Testing structural models of DSM-IV symptoms of common forms of child and adolescent psychopathology. *Journal of Abnormal Child Psychology*. 2008; 36(2):187–206. [PubMed: 17912624]
- Lahey BB, Van Hulle CA, Singh AL, Waldman ID, Rathouz PJ. Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Archives of General Psychiatry*. 2011; 68(2):181–189. [PubMed: 21300945]
- Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is There a General Factor of Prevalent Psychopathology During Adulthood? *Journal of Abnormal Psychology*. 2012; 121(4): 971–977. [PubMed: 22845652]
- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*. 1996; 1:130–149.
- Marsh HW, Hau KT, Wen Z. 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*. 2004; 11(3):320–341.

- Martel MM, Gremillion M, Roberts B, von Eye A, Nigg JT. The structure of childhood disruptive behaviors. *Psychological Assessment*. 2010; 22(4):816–826. [PubMed: 21133546]
- Muthén, LK.; Muthén, BO. *Mplus User's Guide*. 7th ed.. Los Angeles: Muthén & Muthén; 1998–2012.
- Navarro Pastor JB. Methods for the analysis of explanatory linear regression models with missing data not at random. *Quality & quantity*. 2003; 37(4):363–376.
- Nolen-Hoeksema S, Watkins ER. A Heuristic for Developing Transdiagnostic Models of Psychopathology Explaining Multifinality and Divergent Trajectories. *Perspectives on Psychological Science*. 2011; 6(6):589–609.
- Olino TM, Klein DN, Dyson MW, Rose SA, Durbin CE. Temperamental emotionality in preschool-aged children and depressive disorders in parents: Associations in a large community sample. *Journal of Abnormal Psychology*. 2010; 119(3):468–478. [PubMed: 20677836]
- Richman N, Graham P. A behavioural screening questionnaire for use with 3-year-old children: Preliminary findings. *Journal of Child Psychology and Psychiatry*. 1971; 12:5–33. [PubMed: 23289166]
- Rothbart MK, Ahadi SA, Hershey KL, Fisher P. Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*. 2001; 72(5):1394–1408. [PubMed: 11699677]
- Simms L, Prisciandaro J, Krueger R, Goldberg D. The structure of depression, anxiety and somatic symptoms in primary care. *Psychological Medicine*. 2012; 42(1):15–28. [PubMed: 21682948]
- Slade T, Watson D. The structure of common DSM-IV and ICD-10 mental disorders in the Australian general population. *Psychological Medicine*. 2006; 36(11):1593–1600. [PubMed: 16882356]
- Sprafkin J, Volpe RJ, Gadow KD, Nolan EE, Kelly K. A DSM-IV-referenced screening instrument for preschool children: the Early Childhood Inventory-4. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2002; 41(5):604–612. [PubMed: 12014793]
- Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990; 25(2):173–180.
- Sterba SK, Egger HL, Angold A. Diagnostic specificity and nonspecificity in the dimensions of preschool psychopathology. *Journal of Child Psychology and Psychiatry*. 2007; 48(10):1005–1013. [PubMed: 17915001]
- Strickland J, Keller J, Lavigne JV, Gouze K, Hopkins J, LeBailly S. The structure of psychopathology in a community sample of preschoolers. *Journal of Abnormal Child Psychology*. 2011; 39(4):601–610. [PubMed: 21271285]
- Stringaris A, Goodman R. Three dimensions of oppositionality in youth. *Journal of Child Psychology and Psychiatry*. 2009; 50(3):216–223. [PubMed: 19166573]
- Thomas ML. Rewards of bridging the divide between measurement and clinical theory: Demonstration of a bifactor model for the brief symptom inventory. *Psychological Assessment*. 2011; 24(1):101–113. [PubMed: 21767026]
- Trosper SE, Whitton SW, Brown TA, Pincus DB. Understanding the Latent Structure of the Emotional Disorders in Children and Adolescents. *Journal of Abnormal Child Psychology*. in press
- Vollebergh WAM, Iedema J, Bijl RV, de Graaf R, Smit F, Ormel J. The structure and stability of common mental disorders: the NEMESIS study. *Archives of General Psychiatry*. 2001; 58(6):597–603. [PubMed: 11386990]

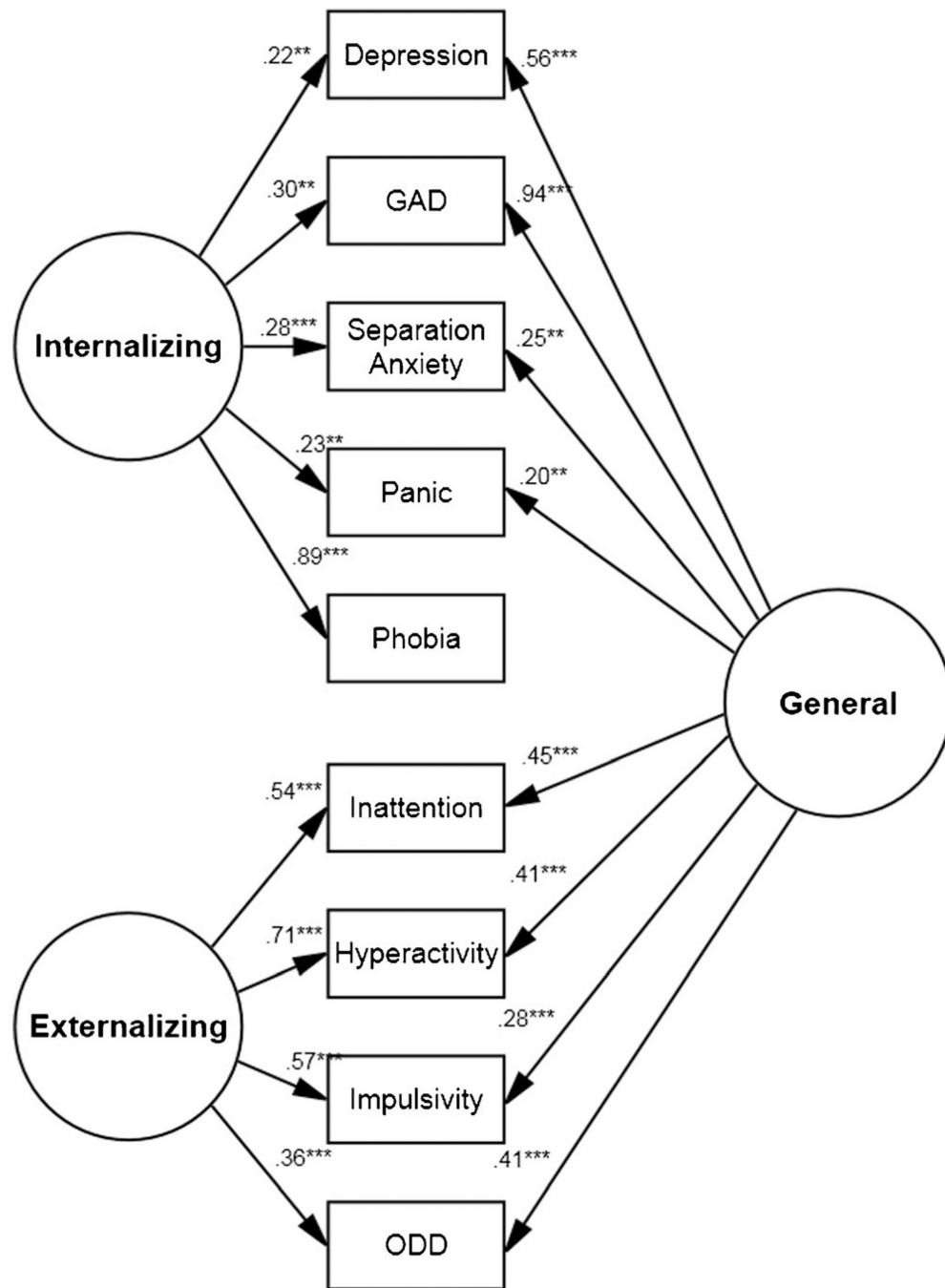


Fig. 1. Bi-factor model of preschool aged psychopathology. Standardized factor loadings are presented. The residual error terms are not displayed. The residual covariance path between the depression and inattention dimension is not displayed ($r = 0.36, p < 0.001$). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 1

Associations among symptom counts for disorders.

	DEP	GAD	SEP	PANIC	PHOB	INAT	HYP	IMP	ODD
DEP	-	0.59***	0.18***	0.11***	0.18***	0.26***	0.23***	0.19***	0.44***
GAD	0.59***	-	0.25***	0.24***	0.22***	0.37***	0.31***	0.21***	0.33***
SEP	0.24***	0.32***	-	0.07	0.15***	0.05	0.01	0.05	0.10*
PANIC	0.16***	0.25***	0.09*	-	0.20***	0.07	0.05	0.07	0.04
PHOB	0.19***	0.26***	0.25***	0.21***	-	0.01	0.00	0.00	0.04
INAT	0.28***	0.42***	0.05	0.06	0.00	-	0.56**	0.49**	0.32**
HYP	0.21***	0.38***	0.10*	0.03	-0.03	0.59***	-	0.53**	0.40**
IMP	0.18***	0.25***	0.04	0.09*	-0.04	0.45***	0.50***	-	0.39**
ODD	0.47***	0.39***	0.15***	0.05	-0.01	0.31***	0.43***	0.41***	-
Mean	0.80	0.73	0.58	0.10	0.55	0.30	0.45	0.22	1.09
SD	1.00	0.97	0.94	0.40	1.16	0.97	1.16	0.56	1.70

* $p < 0.05$ *** $p < 0.001$

Pearson correlations are displayed below the diagonal and Spearman correlations are displayed above the diagonal

DEP Depression Symptoms, GAD Generalized Anxiety Disorder Symptoms, SEP Separation Anxiety Disorder Symptoms, PANIC Panic Disorder Symptoms, PHOB Social and Specific Phobia Total Symptoms, INAT Inattention Symptoms, HYP Hyperactivity Symptoms, IMP Impulsivity Symptoms

Table 2

Fit for evaluated models

	Chi-square	df	CFI	RMSEA (90 % CI)	BIC	S-B χ^2_{diff}
Full Sample						
1. One Factor Model	316.46***	27	0.58	0.14(0.13–0.16)	12311.34	-
2. Two-Factor Model	125.60***	26	0.85	0.08(0.07–0.10)	12089.99	2 vs. 1:62.53***
3. Bi-Factor Model	77.30***	17	0.91	0.08 (0.06–0.10)	12056.65	3 v. 2:46.53***
4. Final Bi-Factor Model	29.74	18	0.98	0.03 (0.01–0.06)	11995.05	4 v. 3: 18.31***
Complete PAPA Data						
1. One Factor Model	170.81***	27	0.76	0.10(0.08–0.11)	9953.39	-
2. Two-Factor Model	104.37***	26	0.87	0.07 (0.06–0.09)	9874.20	2 vs. 1:44.78***
3. Bi-Factor Model	81.72***	17	0.89	0.08(0.07–0.10)	9869.20	3 v. 2: 30.72***
4. Final Bi-Factor Model	32.75*	18	0.98	0.04 (0.02–0.06)	9678.23	4 v. 3:5.65*

*** $p < 0.001$

* $p < 0.05$

S-B χ^2_{diff} : Sattora-Bentler chi-square difference test for the specified model comparison

Table 3

Associations between factor scores and observed and paternal-reported temperament.

	Internalizing Factor	Externalizing Factor	Common Factor
<u>Observed</u>			
Sociability	-0.17***	-0.02	0.06
Dysphoria	0.02	0.02	0.06
Fear/inhibition	0.13**	-0.07	-0.03
Exuberance	-0.15**	-0.06	0.06
Disinhibition	-0.03	0.11**	0.07 ⁺
<u>Paternal-report</u>			
Surgency	-0.20***	0.20***	0.12**
Negative Affect	0.07	-0.02	0.13**
Effortful Control	0.06	-0.18***	-0.13**

⁺
 $p < 0.10$ *
 $p < 0.05$ **
 $p < 0.01$ ***
 $p < 0.001$