Published in final edited form as:

J Autism Dev Disord. 2009 July ; 39(7): 986–995. doi:10.1007/s10803-009-0716-5.

Confirmatory Factor Analysis of the Child Behavior Checklist 1.5-5 in a Sample of Children with Autism Spectrum Disorders

Vincent Pandolfi,

School Psychology Department, Rochester Institute of Technology, Rochester, N.Y.

Caroline I. Magyar, and

Department of Pediatrics, Division of Neurodevelopmental and Behavioral Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, N.Y.

Charles A. Dill

Psychology Department, Hofstra University, Hempstead, N.Y.

Abstract

Validity studies of measures for emotional and behavioral disorders (EBD) for use with preschool children with autism spectrum disorders (ASD) are lacking. The Child Behavior Checklist 1.5-5 (CBCL; Achenbach & Rescorla, 2000), a widely used measure for EBD, contains several norm-referenced scales derived through factor analysis of data from the general pediatric population. In this study, confirmatory factor analysis of archival data evaluated the adequacy of the CBCL factor model in a well characterized sample of preschoolers with ASD ($N=128$). Psychometric results supported the model and suggested that practitioners can use the CBCL to assess for EBD in young children with ASD in conjunction with other clinical data. This will increase the likelihood of accurate identification and EBD-specific intervention.

Keywords

CBCL; Child Behavior Checklist 1.5-5; autism spectrum disorder; ASD; factor analysis; emotional and behavioral disorder

Individuals with autism spectrum disorders (ASD) present with relatively high rates of co-occurring emotional (internalizing) and behavioral (externalizing) disorders (EBDs; e.g., see Brereton, Tonge, & Einfeld, 2006; Gadow, DeVincent, Pomeroy, & Azizian, 2004; Klinger, Dawson, & Renner, 2003; Lainhart, 1999). It is often difficult to distinguish co-occurring EBDs requiring specific treatment from behaviors reflecting the core diagnostic and associated features of ASD. Failure to accurately identify EBDs forestalls intervention, which may result in greater functional impairment and personal distress for the affected child and family, and may moderate response to interventions that target ASD symptoms (e.g., social communication). No EBD measures have been developed specifically for preschoolers with ASD to assist in diagnostic decision making. In addition, the psychometric properties of existing measures designed for use with the general population have not been studied in ASD samples (Leyfer, Folstein, Bacalman, et al., 2006; Ozonoff, Goodlin-Jones, and Solomon, 2005).

The Child Behavior Checklist 1.5-5 (CBCL; Achenbach & Rescorla, 2000) is a widely used norm-referenced measure that assesses for a wide range of EBDs in children aged 1.5 to 5 years. The CBCL's empirically-derived scales were developed through factor analysis of data from the general pediatric population. However, it is not clear if the CBCL factor model, which represents its scoring structure, is valid for the ASD population. This study evaluated the CBCL factor model using **archival data** from a well characterized sample of preschool children with ASD. Results inform practitioners about the validity of the instrument's scoring structure when assessing for EBDs in this population.

Despite its widespread use, we identified only one study that investigated the CBCL in a sample of children with developmental disabilities. Sikora, Hall, Hartley, Gerrard-Morris, and Cagle (2008) found that the sensitivity of the CBCL's *Withdrawn* and *Pervasive Developmental Problems* scales was superior to that of the Gilliam Autism Rating Scale (Gilliam, 1995) when attempting to identify children with autism. Overall sensitivities were 64.6% for the *Withdrawn* scale, 79.8% for the *Pervasive Developmental Problems* scale, and 53.2% for the *GARS Autism Quotient*. The CBCL scales demonstrated higher sensitivity across gender and levels of cognitive functioning; however, all scales evidenced low specificity. The study represents an initial assessment of the CBCL's effectiveness in screening for the presence of ASD.

The developmental characteristics of children with ASD justify the need to investigate the validity of the CBCL factor model for this group. Their core impairments in social interaction, communication, and behavior, as well as (for most) cognitive delays affect both the qualitative and quantitative dimensions of their emotional and behavioral presentation. For example, many demonstrate inappropriate and/or restricted range of affect, atypical expression of emotions, and difficulty communicating needs or reporting subjective states. Thus, many children with ASD may evidence different phenotypic expressions of disorders identified in the general pediatric population (Matson & Nebel-Schwalm, 2007). When these children are assessed with the CBCL, patterns of covariation among the items may differ from the patterns observed in the test development sample. It is therefore possible that the CBCL factor model might not adequately account for the patterns of covariation evidenced by children with ASD.

In this study, confirmatory factor analysis tested the adequacy of the CBCL factor model in a sample of preschool children with ASD. The CBCL factor model represents the instrument's scoring structure which is widely used in clinical practice. A confirmatory procedure, as opposed to an exploratory factor analysis, was used to directly evaluate whether the empirically-derived factor model is valid for this group of children. Results have direct implications for the relevance of the CBCL in diagnostic decision-making.

Method

Sample

The **archival data** analyzed were collected from voluntary participants ($N=128$) presenting to a large federally funded autism research center in Western New York between November, 2003 and October, 2007 (NIH/NIMH U54 MH066397, Rodier, PI). All participants met research criteria for an ASD, as determined by algorithms from the Autism Diagnostic Interview-Revised (ADI-R; Rutter, LeCouteur, & Lord, 2003) and the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2002) as well as expert consensus. The center's assessment team consisted of a licensed doctoral level psychologist, and master's and bachelor's level clinicians, who had training and experience evaluating children with developmental disabilities and ASD. All evaluators were trained to administer the ADOS and ADI-R within a research protocol, and met rigorous reliability standards. Table 1 presents sample demographic data.

The sample was predominantly White, male, and from middle- and upper-middle class backgrounds. Mean cognitive and adaptive behavior standard scores were more than two standard deviations below the population mean ($M=100$, $SD=15$). Many children earned composite standard scores <70 on the cognitive (66.4%) and adaptive behavior (73.0%) measures. Of those with valid cognitive and adaptive behavior composite scores ($n=105$), 61.9% had standard scores <70 on both measures. Thus, a majority evidenced developmental delays.

Materials

The CBCL is a paper and pencil measure completed by primary caregivers. Ratings of problem behaviors describe a child's functioning during the last two months. The CBCL's empirically-derived scales were developed through factor analyses of data from the general pediatric population. Achenbach & Rescorla (2000) described characteristics of their sample as well as the exploratory and confirmatory factor analyses used during scale development. Factor analysis identified seven first order factors which represented separate EBD syndromes and accounted for the observed covariation among CBCL items. Each syndrome consisted of emotional and behavioral symptoms (items) that were more highly related to each other than to items that comprised other syndromes. Similarly, two higher order factors emerged: one consisted of internalizing syndrome scales and the other consisted of externalizing syndromes. The organization of the CBCL is given in Figure 1.

Six syndrome scales contribute to either the broad *Internalizing* or *Externalizing* problem domains. This factor model represents the CBCL scoring structure for these domains. Note that one syndrome scale, *Sleep Problems*, **does not** contribute to either broad domain, but does contribute to the *Total Problems* score.

Raw scores for the syndrome scales, broad domains, and *Total Problems* are converted to normalized *T*-scores (see Achenbach & Rescorla, 2000 pp.62-65 for a description). Significant scale elevations strongly suggest the presence of an EBD. Achenbach & Rescorla (2000) reported that "borderline" and "clinically significant" elevations (see manual for *T*-score thresholds) discriminated those referred for mental health/special education services from those who were not. The manual presents additional evidence attesting to the adequacy of the CBCL's psychometric properties.

In this study, one primary caregiver completed the 100 item CBCL for each participant. Most respondents were the child's mother. Each person indicated how often their child displayed emotional or behavioral problems by endorsing one of three item response options: 0 "*Not true*," 1 "*Somewhat or Sometimes True*," or 2 "*Very True or Often True*." All protocols were scored by members of the assessment team.

Tested Models

Sample size did not allow for one CFA of the entire hierarchical CBCL model, so the factor model was evaluated in two phases. First, *seven* item-level CFAs helped determine whether the *item scores* within each of the syndrome scales were accounted for by a single underlying latent factor. There were enough participants in this study to perform tests of single factor models given the relatively small number of items defining a syndrome scale (5 to 19). Each of the seven syndrome scales were analyzed separately. In each model the latent factor represented the CBCL syndrome and thus, its items served as indicators. In the second phase, a correlated *Internalizing-Externalizing* model was tested (given in Figure 1) because Achenbach & Rescorla (2000) reported a moderate correlation between these domains. For all CFAs each indicator loaded on only one factor and measurement errors were uncorrelated. In any CFA, a unit of measurement must be assigned to the latent factors since unobserved

variables do not have a scale of measurement, so a value of 1.0 was assigned to the variance of all latent factors.

Method of Analysis

Confirmatory factor analysis (CFA) comprised the majority of the analyses. As noted above, CFAs were conducted in two phases which represented two levels of analysis: *item* (ordered-categorical data) and *scale score* (continuous data). Issues and problems associated with ordered-categorical and non-normal data, and the relatively small sample size employed here, were addressed before the CFAs were performed using PRELIS 2 and LISREL 8.80 (Jöreskog & Sörbom, 2006) software. Issues pertaining to missing data, multivariate normality, empty cells in 2×2 tetrachoric correlation tables, and the estimation methods that were used are addressed below.

Missing Data—There were 32 missing data points for CBCL items. Fifteen of these were required for the CFAs.¹ These missing data represented *only .3 percent* of all data points used in the CFAs. The missing data had no systematic pattern across participants or items: missing data were observed across 24 different items, and 18 subjects had at least one missing data point. This random loss of data was recovered through PRELIS' matching imputation procedure. The authors matched subjects on gender and two developmental variables: age and ADOS social-communication total score. These variables were selected because not all subjects had a valid cognitive score and little is known about the relationship between CBCL scores and autism symptomatology. In all, 31 of 32 missing data points were successfully imputed. Item 57, *problems with eyes (without medical cause)*, could not be imputed. This was not problematic since it did not contribute to any of the scales tested by the CFAs.

Multivariate Normality—The multivariate normality assumption for Maximum Likelihood CFA was not met for item and scale scores. Consistent with Achenbach & Rescorla (2000), all CBCL items were dichotomized for the item level analyses (*Not True* = 0; and *Somewhat/Sometimes True and Very True/Often True* = 1) and therefore item distributions were not normal. For continuous scale score data, histograms and significance tests indicated a violation of multivariate normality ($z = 3.38, p = .001$). Positively skewed distributions were evident for the *Emotionally Reactive, Anxious/Depressed, and Somatic Complaints* scales.

The measurement properties of the variables and departures from normality affected the statistical test of model fit, the overall χ^2 -test. This test is biased when data distributions are not normal: true models are rejected too often. In contrast, the Satorra-Bentler chi-square statistic ($SB\chi^2$; Satorra & Bentler, 1994) is satisfactory when multivariate normality is not tenable (Curran, West, & Finch, 1996). Therefore, all CFAs used the Satorra-Bentler correction.

Empty Cells—Tetrachoric correlation matrices for CBCL items were analyzed in the item-level CFAs. Despite dichotomizing the items, several item pairs evidenced at least one "zero frequency" cell in the 2×2 tetrachoric table, which can bias estimates of the correlations (see Greer, Dunlap, & Beatty, 2003). Items contributing to several zero frequency cells were omitted from the analyses, and included: '4' *avoids looking others in the eye*, '23' *doesn't answer when people talk to him/her*, '35' *gets in many fights*, '68' *self-conscious or easily embarrassed*, '86' *too concerned with neatness*, and '93' *vomiting, throwing up (without medical cause)*.² This resulted in an absence of zero frequency cells for all item level CFAs.

¹Thirty-three items do not contribute to the seven syndrome scales and are labeled "Other Problems" on the CBCL protocol. These items contribute to the *Total Problems* score.

²The specific zero frequency cells are available upon request from the first author.

There was no reason to delete items for the two-factor CFA. The analyses were based on continuous data and all scale scores were computed by summing their items which retained their original 0-2 metric.

Estimation Methods—The Robust Diagonally Weighted Least Squares (DWLS) estimator was used for the item-level CFAs. DWLS is appropriate for ordered-categorical data and small-to-moderate sample sizes (Flora & Curran, 2004; Wirth & Edwards, 2007). Robust Maximum Likelihood (RML) was used for the non-normally distributed continuous data. LISREL 8.80 provided robust test statistics for both estimation methods.

Assessing Model Fit—Multiple methods evaluated model fit. First, models were inspected for out of range parameter estimates which suggest an “incorrect” model (e.g., negative error variances). Next, several model fit indices were examined. The $SB\chi^2$ is a goodness-of-fit statistic and evaluates how well a model reproduces the pattern of correlations observed among variables. The hypothesized model produces a correlation matrix that would be expected if the model is “correct.” This *model-implied* matrix is then subtracted from the sample’s *observed* correlation matrix and a *residual* matrix results. To the extent the model is correct (i.e., fits the data well) the residual matrix should be essentially a null matrix. If the $SB\chi^2$ statistic is *not significant*, it suggests that the residual matrix is a null one, and it can be concluded that the model fits the data. If it is significant, one would conclude the residual matrix is *not* null and that the model does *not* fit the data.

Unfortunately, statistical tests such as the $SB\chi^2$ are dependent upon sample size (Bentler & Bonett, 1980): very large samples increase statistical power, which can increase the likelihood of rejecting a true model. Therefore, two psychometric indices of fit, which are less dependent on sample size, were employed here: the Root Mean Square Error of Approximation (*RMSEA*; Steiger & Lind, 1980) and the Comparative Fit Index (*CFI*; Bentler, 1990). The *RMSEA* addresses how well the model would fit the population correlation matrix if it were available (Byrne, 1998), and it represents a *lack* of fit. *RMSEAs* $\leq .05$ indicate a *good* fit and values greater than .05 but less than .10 indicate an *acceptable* fit (MacCallum, Browne, & Sugawara, 1996). The *CFI* quantifies the amount of variation and covariation accounted for by the proposed model by comparing its fit to the fit of a baseline model of uncorrelated variables (Bentler, 1992). Hu and Bentler (1999) recommended that *CFI* values $\geq .95$ be considered evidence of good model fit. Adequacy of all tested models was determined by examining the pattern of results across all measures of fit.

Results

Syndrome Scales: Item Level Analyses

No out of range parameter estimates were observed in any of the single factor models. *CFA* results for the syndrome scales are given in Table 2.

With one exception, the *CFI* and *RMSEA* indicated that all items within each scale are accounted for by one factor. The *RMSEA* was not acceptable for *Sleep Problems*.³ Fortunately, this scale *does not* contribute to either of the higher order *Internalizing* and *Externalizing* domains. The $SB\chi^2$ indicated an acceptable fit for three scales: *Somatic Complaints*, *Withdrawn*, and *Attention Problems*. However, as noted above, the $SB\chi^2$ is sensitive to sample size and has a tendency to reject good models. Thus, the preponderance of evidence supported

³An exploratory Full Information Maximum Likelihood analysis (Jöreskog & Moustaki, 2006) with promax rotation identified two factors that were labeled *Dyssomnia* and *Parasomnia*. All factor loadings were of a relatively large magnitude. Except for *wakes up often at night*, each item exhibited a strong relationship with only one factor. Complete information about this analysis is available upon request from the first author.

all six scales that contribute to the *Internalizing* and *Externalizing* domains. Because evaluation of the CBCL factor model occurred in two phases, support for a single factor underlying each syndrome scale was an important prerequisite in establishing the validity of the higher-order two-factor model.⁴

Median factor loadings for each scale are also presented in Table 2. These loadings ranged from .52 (*Emotionally Reactive*) to .72 (*Withdrawn* and *Sleep Problems*). This indicated that 27-52% of a typical item's variance was attributable to the single underlying factor. In all, 56 of the 61 factor loadings (92%) were statistically significant ($\alpha = .05$). Three items from *Somatic Complaints* were not significant: *can't stand things out of place* (.28), *diarrhea* (.16), and *doesn't eat well* (.34). One item from the *Withdrawn* scale, *acts too young* (.16), and one item from the *Attention Problems* scale, *clumsy* (.18), were also not significant.⁵

Internal Consistency—Coefficient α (unstandardized) is presented in Table 3 for each of the seven syndrome scales.

The Fisher-Bonnett significance test for independent α -coefficients (Bonnett, 2003a; see also Kim & Feldt, 2008) compared the α s in this study to those reported by Achenbach and Rescorla (2000) for the normative sample. In general, the values obtained in the present study were similar to those reported by Achenbach and Rescorla, except for the *Somatic Complaints* scale. In this case, internal consistency was lower for the ASD sample.

The Fisher-Bonnett test is partly dependent upon sample size, the number of items within a scale, and size of the α coefficients. Given that the sample size in this study and that of Achenbach and Rescorla are quite large *in toto*, it is not surprising that very small discrepancies will be statistically significant. This was the case for the *Aggressive Behavior* scale.

Two Factor Model: Syndrome Scale Analyses

No out of range parameter estimates were observed. The overall fit indices and standardized factor loadings are given in Figure 2.

The overall results supported the two-factor model. Specifically, both the *RMSEA* and *CFI* indices indicated an acceptable fit but the statistical measure ($SB\chi^2$) did not. All factor pattern loadings were statistically significant ($\alpha = .05$). The amount of each scale's variance due to its factor ranged from moderate to large effects. In addition, the correlation between the *Internalizing* and *Externalizing* factors (.73) was statistically significant. This result supported the possibility of another higher order factor (*Total Problems*) underlying the two domains.⁶

Internal Consistency—Coefficient α (unstandardized) is presented in Table 3 for the two domains and *Total Problems* scales. The tests of significance indicated that, when compared to the norm sample, both the *Internalizing* and *Total Problems* scales had lower coefficient α s for the ASD sample, but the magnitude of the differences were quite small.

CBCL Raw Scores: ASD Sample vs. Normative Sample

Table 4 presents significance tests of mean CBCL raw scores obtained by this ASD sample and those obtained by the normative sample, whose values were treated as population parameters.

⁴Tests of invariance were conducted (tau and parallel models) but resulted in improper solutions, most likely due to misspecification. Thus, congeneric solutions were accepted.

⁵Tables of these factor loadings are available upon request from the first author.

⁶As was the case for the syndrome scales, tests of invariance were problematic.

Raw scores were used for significance testing because the *T*-score distributions of the syndrome scales were truncated at 50 (Achenbach & Rescorla, 2000). With the exception of the *Anxious/Depressed* scale, the ASD sample's CBCL raw scores were significantly higher than those of the normative sample. Hedges' *g* (Hedges & Olkin, 1985) effect sizes ranged from moderate to large. A significant percentage of scores obtained by children with ASD fell above the mean scores of the normative sample (see the last column in Table 4).

Discussion

Overall Findings

This is the first study to evaluate the factor structure of the CBCL 1.5-5 in a well-characterized sample of children with ASD. This study specifically investigated the adequacy of the current CBCL factor model because it represents the scoring structure of the test which is widely used in clinical practice. CFA results indicated that the CBCL 1.5-5 measures the same constructs in children with ASD as it does in the general population. The internal consistencies of most scales were very similar to those reported by Achenbach & Rescorla (2000). Except for the *Anxious/Depressed* scale, the mean raw scores of the present sample were (expectedly) significantly higher than those of the normative sample. The present results support use of the CBCL to assess for EBDs in preschoolers with ASD.

Construct Validity

Item-level CFAs supported the six factors contributing to the *Internalizing* and *Externalizing* domains. Items within the Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Attention Problems, and *Aggressive Behavior* syndrome scales are each accounted for by a single construct. Results did not support a unidimensional *Sleep Problems* factor; however, this scale does not contribute to either broad domain. CFA results indicated that two higher order *Internalizing* and *Externalizing* factors underlie the syndrome scales. The moderately high correlation between these two factors provides indirect evidence for a third order factor, which may represent the *Total Problems* scale. These findings support the scoring structure of the CBCL for use with preschoolers with ASD. It is recommended that practitioners consider significant elevations on either the syndrome scales or the broad domains as suggestive of emotional and/or behavioral difficulties that may be associated with functional impairment, and therefore require specific treatment.

Reliability

With the exception of *Somatic Complaints*, alpha coefficients observed in this study were very similar to those obtained by Achenbach & Rescorla (2000). Interestingly, the *Sleep Problems* scale demonstrated good internal consistency ($\alpha = .83$), which means the items share a substantial amount of common variance; however, high internal consistency is necessary but not sufficient for unidimensionality. Although more research is needed to clarify the nature of the sleep construct(s) measured by the CBCL it is recommended that practitioners consider a scale elevation as a general indicator of possible sleep problems. Internal consistency was best for the *Internalizing*, *Externalizing*, and *Total Problems* scales, which would be expected due to their relatively large number of items. The α coefficients for each of these three scales were of relatively large magnitude ($\geq .80$) which further supports their use in assessment (see Salvia & Ysseldyke, 2001).

Analysis of Raw Scores

Except for the *Anxious/Depressed* scale, the ASD sample's mean syndrome and domain raw scores were significantly higher than those obtained by the normative sample. This result is consistent with findings that children with ASD demonstrate higher rates of EBD when

compared to the general population (see e.g., Gadow et al., 2004). Particularly striking is the large percentage of children with ASD who scored above the mean of the normative sample, ranging from 45 to 99% across all scales (median = 76.5%). Future research should investigate *why* children with ASD exhibit scale score elevations. For example, such research could inform practitioners as to whether elevations on the *Withdrawn* scale reflect a co-occurring emotional problem such as depression, or social interaction deficits that are a defining feature of ASD. The Sikora et al. (2008) findings suggest that the *Withdrawn* scale may be sensitive to the presence of autism; however, their report did not state whether study participants were screened for co-occurring EBD. Until more data become available, practitioners are encouraged to continue using multiple assessment methods to better understand the specific reasons for scale elevations for each individual child. This practice will increase diagnostic accuracy and the likelihood of providing appropriate specific intervention.

Limitations

Six items were omitted from the CFAs of the syndrome scales. These items displayed insufficient variance in this sample and, as stated above, could have biased estimates of several tetrachoric correlations. Interestingly, two items were clearly related to ASD (*avoids looking others in the eye*, and *doesn't answer when people talk to him/her*) and, expectedly, an overwhelming majority of participants had nonzero scores on these items. The remaining four items were endorsed for a substantial minority of participants. Item omission is not unprecedented in CFAs of instruments within the Achenbach System of Empirically Based Assessment (see Ivanova, Achenbach, Dumenci, et al. 2007 and Ivanova, Achenbach, Rescorla, et al., 2007 for examples). We believe that omission of the items did not substantially affect our results because our CFA findings are consistent with the Achenbach & Rescorla (2000) model, our methodology was similar to that used in other studies, and relatively few items were omitted.

We acknowledge the relatively small sample size and issues pertaining to generalization. Sample size precluded one CFA of the entire CBCL model. However, the sample size was sufficient to evaluate each syndrome scale separately. Support for the *Internalizing-Externalizing* model was strengthened given that all of the single factor models contributing to the two broad domains were supported by the item-level CFAs. In addition, the CFA methods used were those recommended for ordinal data and small to moderate sample size. With respect to generalization, it is not clear if the results apply to ASD subgroups with different demographic and developmental characteristics (e.g., cognitive levels, severity of autism). It is also not known if the present findings generalize to children presenting for evaluations outside of research centers. Replication of this study with other ASD samples would be informative.

Future Directions

As in any CFA, support for a factor model does not mean that it is the *single best fitting model* for a target population. It is possible that a competing model, that has not yet been tested, will prove to be better. Future replication studies may suggest model modifications that provide better fit for data within ASD samples. At present, this issue is probably of little practical concern for the CBCL because it has a long history of clinical use and psychometric support.

We did not investigate the CBCL's DSM Oriented scales which were meant to align with broad DSM-IV-TR diagnostic categories. These scales were developed through expert consensus and complement the findings obtained from the empirically-derived scales. Validity studies of these scales are needed, and the authors are preparing a criterion-related validity study of both the DSM Oriented and empirically-derived scales to see how strongly they are related to EBD status.

Conclusion

To our knowledge, this is the only study to evaluate the factor structure of the CBCL 1.5-5 in a well characterized sample of preschoolers with ASD. The present findings supported the CBCL factor model which suggests that its scoring structure is appropriate for this subgroup. Our findings suggest that practitioners can use CBCL scores, in conjunction with other clinical data, to assist in diagnostic decisions and to test hypotheses regarding the nature of EBDs observed in preschoolers with ASD.

Acknowledgments

This study was supported in part by NIH grant U54MH066397 (Rodier, PI; Magyar, PI Core Assessment Center) and General Clinical Research Center grant 5 MO1RR0044, NIH, National Center for Research Resources. The authors thank Courtney McGuire for her assistance in developing the database for this study.

References

- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA Preschool Forms & Profiles. University of Vermont, Research Center for Children, Youth, and Families; Burlington, VT: 2000.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin* 1990;107:238–246. [PubMed: 2320703]
- Bentler PM. On the fit of models to covariances and methodology to the *Bulletin*. *Psychological Bulletin* 1992;112(3):400–404. [PubMed: 1438635]
- Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 1980;88(3):588–606.
- Bonnett DG. Sample size requirements for comparing two alpha coefficients. *Applied Psychological Measurement* 2003a;27:72–74.
- Brereton AV, Tonge BJ, Einfeld SL. Psychopathology in children and Adolescents with autism compared to young people with intellectual disability. *Journal of Autism and Developmental Disorders* 2006;36(7):863–870. [PubMed: 16897401]
- Byrne, BM. Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming. Lawrence Erlbaum Associates; Mahwah, N.J.: 1998.
- Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods* 1996;1(1):16–29.
- Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods* 2004;9(4):466–491. [PubMed: 15598100]
- Gadow KD, DeVincent CJ, Pomeroy J, Azizian A. Psychiatric symptoms in preschool children with PDD and clinic and comparison samples. *Journal of Autism and Developmental Disorders* 2004;34(4):379–393. [PubMed: 15449514]
- Gilliam, JE. Gilliam Autism Rating Scale. Pro-Ed; Austin, TX: 1995.
- Greer T, Dunlap WP, Beatty GO. A monte carlo evaluation of the tetrachoric correlation coefficient. *Educational and Psychological Measurement* 2003;63(6):931–950.
- Hedges, LV.; Olkin, I. Statistical methods for meta-analysis. Academic Press; San Diego, CA: 1985.
- Hollingshead, AA. Four-factor index of social status. Yale University; New Haven, CT: 1975. Unpublished manuscript
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;6:1–55.
- Ivanova MY, Achenbach TM, Dumenci L, Rescorla LA, Almqvist F, Weintraub S, et al. Testing the 8-syndrome structure of the Child Behavior Checklist in 30 societies. *Journal of Clinical Child and Adolescent Psychology* 2007;36(3):405–417. [PubMed: 17658984]
- Ivanova MY, Achenbach TM, Rescorla LA, Dumenci L, Almqvist F, Bilenberg N, et al. The generalizability of the Youth Self Report syndrome structure in 23 societies. *Journal of Consulting and Clinical Psychology* 2007;75(5):729–738. [PubMed: 17907855]

- Jöreskog, K.; Moustaki, I. Factor analysis of ordinal variables with full information maximum likelihood. 2006. <http://www.ssicentral.com/lisrel/techdocs/orfimpl.pdf>
- Jöreskog, K.; Sörbom, D. LISREL 8.80. Scientific Software International, Inc.; Lincolnwood, IL: 2006.
- Kim S, Feldt LS. A comparison of tests for equality of two or more independent alpha coefficients. *Journal of Educational Measurement* 2008;45(2):179–193.
- Klinger, LG.; Dawson, G.; Renner, P. Autistic disorder. In: Mash, EJ.; Barkley, RA., editors. *Child Psychopathology*. Vol. 2nd Ed.. 2003. p. 409–454.
- Lainhart JE. Psychiatric problems in individuals with autism, their parents, and siblings. *International Review of Psychiatry* 1999;11:278–298.
- Leyfer OT, Folstein SE, Bacalman S, Davis NO, Dinh E, Morgan J, et al. Comorbid psychiatric disorders in children with autism: Interview development and rates of disorders. *Journal of Autism and Developmental Disorders* 2006;36:849–861. [PubMed: 16845581]
- Lord, C.; Rutter, M.; DiLavore, P.; Risi, S. *Autism Diagnostic Observation Schedule: Manual*. Western Psychological Services; Los Angeles: 2002.
- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* 1996;1:130–149.
- Matson JL, Nebel-Schwalm MS. Comorbid psychopathology with autism spectrum disorder in children: An overview. *Research in Developmental Disabilities* 2007;28:341–352. [PubMed: 16765022]
- Mullen, EM. *Mullen Scales of Early Learning*. American Guidance Service; Circle Pines, MN: 1995.
- Ozonoff S, Goodlin-Jones BL, Solomon M. Evidence-based assessment of autism spectrum disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology* 2005;34(3):523–540. [PubMed: 16083393]
- Rutter, M.; LeCouteur, AL.; Lord, C. *Autism Diagnostic Interview- Revised*. Western Psychological Services; Los Angeles: 2003.
- Salvia, JS.; Ysseldyke, JE. *Assessment*. Vol. 8th Ed.. Houghton Mifflin; Boston: 2001.
- Satorra, A.; Bentler, PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye, A.; Clogg, CC., editors. *Latent variable analysis: Applications to developmental research*. Sage; Thousand Oaks, CA: 1994. p. 399–419.
- Sikora DM, Hall TA, Hartley SL, Gerrard-Morris AE, Cagle S. Does parent report of behavior differ across ADOS-G classifications: Analysis of scores from the CBCL and GARS. *Journal of Autism and Developmental Disorders* 2008;38:440–448. [PubMed: 17619131]
- Sparrow, SS.; Balla, DA.; Cicchetti, DV. *Vineland Adaptive Behavior Scales- Interview Edition*. American Guidance Service; Circle Pines, MN: 1984.
- Steiger, JH.; Lind, JM. Statistically based tests for the number of common factors; Paper presented at the meeting of the Psychometric Society; Iowa City, IA. 1980;
- Wirth RJ, Edwards MC. Item factor analysis: Current approaches and future directions. *Psychological Methods* 2007;12(1):58–79. [PubMed: 17402812]

Figure 1.
CBCL factor model. ER=*Emotionally Reactive*, AD=*Anxious/Depressed*, SC=*Somatic Complaints*, WD=*Withdrawn*, AP=*Attention Problems*, AB=*Aggressive Behavior*, Internal=*Internalizing Domain*, External=*Externalizing Domain*.

Figure 2.

CFA Results for the two factor model. ER=Emotionally Reactive, AD=Anxious/Depressed, SC=Somatic Complaints, WD=Withdrawn, AP=Attention Problems, AB=Aggressive Behavior, Internal=Internalizing Domain, External=Externalizing Domain. Factor pattern loadings are embedded in the path connecting a latent factor with its indicators. Disturbance values (error variances) are presented immediately to the left of each indicator. Satorra-Bentler Chi-Square=17.02, $df=8$, $p=.03$, RMSEA=0.09, CFI=.97

Table 1**Sample Demographics**

	Percent
Gender: Male	89
SES ^a	
Major Business/Professional	31
Medium Business/Minor Professional	49
Skilled Craftsman/Clerical/Sales	12
Machine Operator/Semi-skilled	6
Unskilled Laborers/Menial Service	2
Race ^b	
Asian	1
Black/African-American	5
White	89
More than one category	5
Not reported	1
Ethnicity: Hispanic/Latino	3
Not reported	1
Developmental Data	<i>M</i> <i>SD</i>
Age (Months)	42.43 10.19
Mullen Scales of Early Learning ^c	65.25 18.96
Vineland Adaptive Behavior Scales ^d	
Adaptive Behavior Composite	65.11 12.33
Communication	70.06 17.13
Daily Living Skills	65.57 10.03
Socialization	67.50 13.07

Note. *N*=128.

^a Hollingshead (1975) scale.

^b Total exceeds 100 percent due to rounding error.

^c *n*=107, reflects number of valid results for the Mullen Scales (Mullen, 1995)

^d *n*=126, two Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) unable to be administered.

Table 2

CFA Results For Syndrome Scales

Model	$SB\chi^2_a$	df	RMSEA	CFI	Median Factor Loading
Emotionally Reactive	56.09 *	27	.09	.95	.52
Anxious/Depressed ^b	23.38 *	14	.07	.97	.54
Somatic Complaints ^c	37.24	27	.06	.98	.66
Withdrawn ^d	14.82	9	.07	.98	.72
Sleep Problems	43.49 **	14	.13	.95	.72
Attention Problems	4.59	5	.00	1.00	.62
Aggressive Behavior ^e	184.37 **	135	.05	.99	.70

Notes. N=128. CFAs used Diagonally Weighted Least Squares estimation.

^a Satorra-Bentler chi-square.

^b sans Item 68 *self-conscious or easily embarrassed*.

^c sans Item 86 *too concerned with neatness or cleanliness* and Item 93 *vomiting, throwing up (without medical cause)*.

^d sans Item 4, *avoids looking others in the eye* and Item 23 *doesn't answer when people talk to him/her*.

^e sans Item 35 *gets in many fights*.

* $p < .05$

** $p < .01$.

Table 3
Significance Tests on Alpha Coefficients Obtained in the Present Study and by Achenbach & Rescorla (2000)

Scale	No. Items	α Present study ^a	α Achenbach & Rescorla (2000) ^b	z
Emotionally Reactive	9	.67	.73	1.42
Anxious/Depressed	8	.63	.66	0.60
Somatic Complaints	11	.49	.80	6.78***
Withdrawn	8	.73	.75	0.55
Sleep Problems	7	.83	.78	1.78
Attention Problems	5	.68	.68	0.00
Aggressive Behavior	19	.89	.92	2.31*
Internalizing	36	.80	.89	4.46***
Externalizing	24	.90	.92	1.62
Total Problems ^c	100	.93	.95	2.51**

Notes. Data from column 4 are from the *Manual for the ASEBA Preschool Forms & Profiles* (p.155), by T.M. Achenbach and L.A. Rescorla, 2000, Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families. Copyright 2000 by T.M. Achenbach and L.A. Rescorla. Reprinted with permission. All significance tests used the Fisher-Bonnett approach.

^a N=128

^b N=1126

^c Missing data for one subject on Item 57 was not successfully imputed, α is based on $n=127$. Item 100 is an open-ended item and was not imputed. When a parent left this item blank, it was scored 0.

* $p < .05$

** $p < .01$

*** $p < .005$

Table 4
Significance Tests for CBCL Raw Scores: Present sample vs. normative sample

Scale	ASD Sample		Norm Sample		g^a	Percentage of ASD scores above norm sample mean
	M	SD	M	SD		
Emotionally Reactive	3.88*	2.78	2.40	2.20	0.68	75.17
Anxious/Depressed	2.62	2.21	2.90	2.30	-0.13	44.83
Somatic Complaints	2.95*	2.15	1.80	1.90	0.63	73.57
Withdrawn	6.24*	2.98	1.50	1.70	2.75	99.70
Sleep Problems	4.04*	3.53	2.80	2.40	0.50	69.15
Attention Problems	5.15*	2.40	2.50	1.90	1.41	92.07
Aggressive Behavior	13.77*	7.35	10.40	6.40	0.53	70.19
Internalizing Domain	15.69*	7.09	8.60	6.20	1.14	87.29
Externalizing Domain	18.91*	8.89	12.90	7.70	0.77	77.94
Total Problems ^b	55.35*	21.43	33.30	18.70	1.17	87.90

Note. $N=128$. Data from columns 4 and 5 are from the *Manual for the ASEBA Preschool Forms & Profiles* (p.153), by T.M. Achenbach and L.A. Rescorla, 2000, Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families. Copyright 2000 by T.M. Achenbach and L.A. Rescorla. Reprinted with permission.

^a Hedges' g (Hedges & Olkin, 1985) reflects magnitude of the differences.

^b Raw score data based on $n=127$ as Item 57 could not be imputed for one participant.

* $p < .001$.