

Published in final edited form as:

Res Autism Spectr Disord. 2012 ; 6(1): 96–108. doi:10.1016/j.rasd.2011.03.009.

An Initial Psychometric Evaluation of the CBCL 6–18 in a Sample of Youth with Autism Spectrum Disorders

Vincent Pandolfi^a, Caroline I. Magyar^b, and Charles A. Dill^c

^aRochester Institute of Technology, School Psychology Department, 18 Lomb Memorial Drive, Rochester, N.Y. 14623–5604, United States

^bUniversity of Rochester School of Medicine and Dentistry, Department of Pediatrics, Strong Center for Developmental Disabilities, Golisano Children’s Hospital at Strong, Rochester, N.Y., United States

^cHofstra University, Psychology Department, Hempstead, N.Y., United States

Abstract

Individuals with an autism spectrum disorder (ASD) often present with co-occurring emotional and behavioral disorders (EBD). The Child Behavior Checklist 6–18 (CBCL; Achenbach & Rescorla, 2001) is an EBD measure that contains several norm-referenced scales derived through factor analysis of data from the general pediatric population. The psychometric properties of this widely used and well-researched measure have not been evaluated in samples of youth with ASD. This study evaluated the CBCL’s internal structure, scale reliability, criterion-related validity, and diagnostic accuracy using archival data from a well-characterized sample of youth with ASD ($N = 122$). Confirmatory factor analyses supported the unidimensionality of the CBCL’s syndrome scales and its Internalizing-Externalizing factor structure. Significance tests indicated that many scales discriminated between two subgroups: a group of individuals with ASD+EBD and a group with ASD alone. Diagnostic accuracy analyses indicated that the CBCL had good sensitivity but low specificity for detecting co-occurring disorders. Results supported the use of the CBCL in conjunction with other clinical data when assessing for EBD in youth with ASD.

Keywords

CBCL; Child Behavior Checklist; autism spectrum disorder; reliability; validity; psychometrics

1. Introduction

Individuals with an autism spectrum disorder (ASD) reportedly present with relatively high rates of co-occurring emotional and behavioral disorders¹ (EBDs; see DeBruin et al., 2006; Lainhart, 1999; Leyfer et al., 2006). Because of their age and developmental disability, many youth with ASD cannot provide sufficient information for clinicians to reliably

© 2011 Elsevier Ltd. All rights reserved.

Corresponding Author: Vincent Pandolfi, Tel: +585-475-2875; fax: +585-475-6715. vxppla@rit.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

¹We use this general term to refer to conditions that reflect specific DSM-IV disorders and nonspecific behaviors and syndromes (e.g., aggression, self-injury, withdrawal) and result in functional impairment and/or personal distress.

identify the presence or specific nature of a co-occurring EBD. Thus, third party rating scales are often an important component of clinical assessment. However, few EBD measures have been validated for youth with ASD.

One widely used parent rating scale that assesses for a broad range of emotional and behavioral syndromes is the Child Behavior Checklist 6–18 (CBCL; Achenbach & Rescorla, 2001). The CBCL's norm-referenced scales were developed through factor analysis of data from the general pediatric population. These scales appear to assess for the kinds of EBDs also observed in youth with ASD such as anxiety, depression, withdrawal, social problems, attention problems, and aggression. Although the CBCL's psychometric properties have been evaluated extensively (see Berubé & Achenbach, 2010 bibliography), they have not been comprehensively evaluated in ASD samples. In this study, we evaluated the CBCL's internal structure, scale reliability, criterion-related validity, and diagnostic accuracy using archival data from a well-characterized sample of six to 18 year-olds with ASD. Such evidence is needed to justify the clinical use and interpretation of CBCL scores in the assessment of EBDs in youth with ASD.

1.1 Emotional and Behavioral Disorders in Youth with ASD

High rates of EBD have been reported in youth with ASD. Some of the most commonly occurring disorders include depression (Ghaziuddin, Ghaziuddin, & Greden, 2002), anxiety (Kim et al., 2000), Attention-Deficit/Hyperactivity Disorder (ADHD; Gillberg & Billstedt, 2000), and Oppositional Defiant Disorder (ODD; Gadow, DeVincent, & Drabick, 2008). EBDs have also been reported to persist over time (see Gadow et al., 2004; Mash & Dozois, 2003) suggesting the need for routine screenings in this at risk group.

However, identifying co-occurring EBDs that require specific intervention in youth with ASD is often difficult and challenges in assessment have been described in the literature (e.g., see Matson & Nebel-Schwalm, 2007; Ozonoff, Goodlin-Jones, & Solomon, 2005). First, many of these youth exhibit impairments in social-communication, cognition, self-awareness, and insight that may limit their ability to identify and report alterations in thoughts, feelings, behaviors, levels of personal distress, and/or functional impairment. Second, developmental characteristics may moderate the symptom topography associated with those EBDs that have been observed in the general population. Initial symptom presentation of some disorders may be nonspecific, such as in anxiety disorders (e.g., see White et al., 2009). Atypical presentations and nonspecific symptoms elevate the risk for diagnostic overshadowing whereby behaviors may be attributed to the child's ASD and/or development instead of a co-occurring EBD. For example, for a child who demonstrates some variability in vocalizations but most often presents with low rates, a sustained period of an increased rate of vocalizations may be attributed to development or to the ASD when, in fact, the child might be demonstrating nonspecific anxiety symptoms. Failure to detect co-occurring EBDs can forestall specific intervention and increase the risk for further functional impairment. In fact, EBDs have been associated with poorer outcomes (Howlin et al., 2004) and may moderate response to ASD-specific treatment. To address these issues, best practices in assessment call for a multi-method multi-informant approach, which often includes third-party report such as a parent rating scale. The CBCL represents one such measure that could be included in the assessment process.

1.2 Measures Used to Assess Youth with ASD

Several rating scales are available to assess for EBDs in school-aged youth; however, relatively few have been evaluated in samples of youth with ASD. Some scales were developed for children and adolescents with developmental disabilities. For example, the Aberrant Behavior Checklist (ABC; Aman et al., 1985a; Aman et al., 1985b) and the

Nisonger Child Behavior Rating Form (NCBRF; Aman et al., 1996) assess individuals with an intellectual disability (ID). In addition, the Autism Spectrum Disorder- Comorbid for Children (ASD-CC; Matson et al., 2009) was developed specifically for youth with ASD. The ABC, NCBRF, and ASD-CC scales do not closely correspond to DSM-IV-TR (APA, 2000) diagnostic categories; however, they do assess for specific problems that are often a focus of clinical attention in individuals with ID and ASD. A few studies of these measures in ASD samples have produced some favorable psychometric data with regard to internal structure, reliability, and correlations with other measures (cf. Brinkley et al., 2007; Lecavalier et al., 2004; Matson et al., 2009).

Two measures developed for use with the general pediatric population have also recently been evaluated in ASD samples, the Child Symptom Inventory-4 (CSI-4; Gadow & Sprafkin, 2002) and the Behavior Assessment System for Children- Second Edition (BASC-2; Reynolds & Kamphaus, 2004). They assess for a wide range of EBDs, and the CSI-4 is based directly on the DSM-IV. Studies of the BASC-2 and CSI-4 in ASD samples have focused on different psychometric characteristics that included internal structure, score profiles, and the ability of the measures to discriminate between youth with and without an ASD (see Gadow et al., 2008; Lecavalier et al., 2009; Mahan & Matson, 2011; Volker et al., 2010). Although these initial studies on measures developed for youth with developmental disabilities and for the general population have provided important initial reliability and validity data, more research is needed to fully evaluate the psychometric properties of each measure in ASD samples.

1.3 Studies of CBCL Instruments in ASD Samples

Similarly, the current and previous versions of the CBCL preschool and school age forms have not often been investigated in ASD samples (see Pandolfi & Magyar, in press). Previous studies focused on the CBCL's ability to discriminate youth with ASD from other clinical subgroups (see Biederman et al., 2010; Bolte, Dickhut, & Poustka, 1999; Duarte et al. 2003; Ooi et al., 2009; Rescorla, 1988; Sikora et al., 2008) and to help determine prevalence rates and multi-informant agreement for psychiatric syndromes in youth with ASD (see Hurtig et al., 2009; Kanne, Abbacchi, & Constantino, 2009). Collectively, these studies provided initial evidence that the CBCL may have clinical utility for the assessment of youth with ASD. However, these studies were conducted without a complete understanding of the psychometric properties of the CBCL in ASD samples. Furthermore, none of these studies investigated whether the CBCL can discriminate between youth with ASD and a co-occurring EBD (ASD+EBD) from those with ASD alone.

We know of only one study that evaluated the factor structure of a current CBCL instrument in a sample with ASD. A series of confirmatory factor analyses (CFAs) conducted by Pandolfi, Magyar, & Dill (2009) supported the factorial validity of the CBCL 1.5–5 (Achenbach & Rescorla, 2000) in a sample of preschoolers with confirmed diagnoses of ASD. Results indicated that the CBCL 1.5–5 reliably measured two broad dimensions of EBD that reflected the scale's *Internalizing* and *Externalizing Domains*, and seven more narrowly defined dimensions that reflected the CBCL syndrome scales. These findings were consistent with the factor analytic work of the test authors, and supported the use of the instrument when assessing young children with ASD. A similar study is needed for the CBCL 6–18, an instrument that covers most of the child and adolescent age range.

1.4 Present Study

We utilized several methods to evaluate the extent to which psychometric evidence supports the interpretation of CBCL 6–18 scores as indicators of EBDs in youth with an ASD. First, we sought evidence based on *internal structure* through a series of CFAs and scale

reliability analyses. Specifically, each syndrome scale was tested for unidimensionality which is important because each scale is often interpreted separately in clinical practice. We followed these tests with a CFA of the CBCL's factor structure at the scale level. We then performed significance tests across all CBCL scales to determine the extent to which scores differed between a group with ASD alone, and one with ASD+EBD. Finally, we performed diagnostic accuracy tests using ROC analyses to determine how well the CBCL discriminated individuals with ASD and a specific EBD (e.g., a depressive disorder) from the rest of the sample (those without a depressive disorder). Collectively, these analyses expanded upon the work of previous studies of the CBCL in ASD samples by providing an initial comprehensive psychometric evaluation. In addition, because the CBCL measures different constructs than the other measures reviewed here, results of this study can complement ongoing research on the other EBD measures within the ASD population which can further assist practitioners in selecting the most appropriate assessment protocol for individual clients.

2. Method

2.1 Sample

Archival data were analyzed from youth with an ASD ($N = 122$) who participated in one of two large studies. Participant age spanned six to 18 years ($M = 11$ years 3 months, $SD = 3$ years 3 months), with 60.7% falling between 6–11 years, and 39.3% between 12–18 years which reflect the CBCL's normative age ranges. Most ($n = 93$) participated in a large federally funded autism research center in western New York (NIH/NIMH U54 MH066397: Rodier, PI). The center's principal research focus was on the relationship between phenotype and genotype in ASD. The remaining 29 children participated in a statewide prevalence study of ASD in children with Down syndrome (AUCD/RT01 2005-1/2–08: Hyman, PI).

All 122 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 et al., 2002) and expert clinical consensus. Specific cut-off criteria were applied for the ADI-R and ADOS. Participants were considered to have met ADI-R criteria for autism if they obtained a Social score at or above cutoff and a Communication score at least within two points of the cut-off; and to have met ADOS criteria if they met the ASD cut-off for the module administered. For the participants in the prevalence study, the research evaluation team also reviewed assessment data from the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) and the Pervasive Development Disorder-Mental Retardation Scale (Kraijer & de Bildt, 2005) and applied a DSM-IV symptom checklist to guide diagnostic decision-making. The assessment teams for both projects consisted of a licensed doctoral level psychologist, licensed developmental pediatrician, and master's and bachelor's level evaluators. All had training and experience evaluating children with developmental disabilities and ASD and were trained to administer the ADOS and ADI-R reliably within a research protocol.

Developmental testing was completed for all participants including cognitive, language, and adaptive behavior assessments. The 93 participants in the NIH study were also evaluated for a co-occurring psychiatric disorder through a medical history form completed by the parent, direct observation and interview of the participant, and the administration of the Schedule for Affective Disorders and Schizophrenia- Childhood Version (K-SADS; Kaufman et al., 1996) with the parent as reporter. The DSM-IV-TR criteria were used to establish diagnoses. The participants with Down syndrome (DS) were not screened for a co-occurring psychiatric disorder because this was not relevant to the aims of the prevalence study. The data for the children with DS were therefore not used in the significance tests and diagnostic accuracy

analyses; however, their data were used for the CFAs. Table 1 presents participant characteristics.

The gender distribution approximates that observed in the general ASD population (see APA, 2000). The sample was predominantly middle- to upper-middle class, white, and non-Hispanic. Different cognitive tests were used in the research protocols and included the WISC-IV (Wechsler, 2003), WAIS-III (Wechsler, 1997), Stanford-Binet-5 (Roid, 2003), and the Leiter-R (Roid & Miller, 1997). The participants evidenced a range of cognitive functioning on these measures, with full scale IQs (FSIQ) ranging from moderate/severe intellectual disability to superior functioning, and 31.3% were intellectually disabled. Most had significant adaptive behavior deficits. The most common co-occurring psychiatric disorders were anxiety disorders (38.5%; Specific Phobia, Obsessive-Compulsive Disorder, Generalized Anxiety Disorder), ADHD (35.2%), depressive disorders (16.5%; Major Depressive Disorder, Depressive Disorder NOS), tic disorders (14.5%), and ODD (6.6%). Most participants (60.3%) had more than one co-occurring disorder. None of them evidenced a psychotic disorder.

2.2 Instrument

The norm-referenced CBCL is completed by parents and caregivers, and it describes a child's functioning during the previous six months. The items measure specific emotional and behavioral problems on a three point Likert scale (0= "Not True," 1= "Somewhat or Sometimes True," or 2= "Very True or Often True"). The technical manual reports good to excellent psychometric properties (see Achenbach & Rescorla, 2001). In this study, the CBCL was most often completed by the child's mother.

The CBCL contains two empirically-derived broadband scales and eight syndrome scales.² The test manual addresses the item development procedures used to help ensure the content validity of each scale. The broadband *Internalizing Domain* is a measure of emotional problems and contains three syndrome scales: *Anxious/Depressed*, *Withdrawn/Depressed*, and *Somatic Complaints*. The broadband *Externalizing Domain* measures behavioral problems and contains the *Rule Breaking Behavior* and *Aggressive Behavior* syndrome scales. Three other syndrome scales do not belong to either broadband scale: *Social Problems*, *Thought Problems*, and *Attention Problems*. These are considered mixed syndrome scales because they had sizeable factor loadings on both broad domains in the Achenbach & Rescorla (2001) factor analyses. A *Total Problems* scale quantifies overall impairment and is derived from the raw score sum of all eight syndrome scales, and a group of 17 "Other Problems" items that do not belong to any specific syndrome scale.

Raw scores for each scale are converted to norm-referenced *T*-scores ($M = 50$, $SD = 10$). Separate norms are provided for each gender within the 6–11 and 12–18 year age ranges. "Clinically significant" elevations are indicated by *T*-scores ≥ 64 on the broadband scales, and ≥ 70 on the syndrome scales. "Borderline" elevations range from 60–63 and 65–69 on the broadband and syndrome scales, respectively. These qualitative categories reflect symptom severity and scores falling within either category suggest the need for a more comprehensive diagnostic assessment.

2.3 Tested Models in CFAs

The CBCL's internal structure was evaluated in two phases using methods similar to those of Pandolfi et al. (2009). First, eight *item-level* CFAs helped determine whether the

²The CBCL also has conceptually-derived *DSM Oriented Scales*, which were not the focus of this study. See Achenbach and Rescorla (2001) for a description.

relationships among *item scores* within each of the syndrome scales were accounted for by a single latent factor. CFA has become a widely used method to assess for scale unidimensionality (Netemeyer, Bearden, & Sharma, 2003). Each syndrome scale was analyzed separately. In each model, the latent factor represented the CBCL syndrome and its items served as indicators. Support for the idea that one factor adequately accounts for the interrelationships among the items within each syndrome scale was necessary prior to evaluating the CBCL's factor structure at the *scale level*.³

In the second phase, the *Internalizing* and *Externalizing* domains were modeled as correlated latent factors because Achenbach & Rescorla (2001) reported a moderate correlation between these domains. Each domain's syndromes served as indicators. In addition to the empirically-derived two-factor model, we also evaluated a more parsimonious single factor model which contained the same five syndrome scales. For all CFAs each indicator was specified to load on only one factor, measurement errors were uncorrelated, and a value of 1.0 was assigned to the variance of all latent factors.

2.4 Method of Analysis

As noted above, CFAs were conducted at two levels: *item* (ordered-categorical data) and *scale score* (continuous data). At the scale score level raw scores were analyzed so we could collapse data across gender and the 6–11 and 12–18 age ranges. All analyses were completed using PRELIS 2 and LISREL 8.80 software (Jöreskog & Sörbom, 2006).

2.4.1 Missing Data—There were 28 missing data points for 13 subjects across 20 items that were required for the CFAs.⁴ These missing data exhibited no discernible pattern across items or participants and represented *only .2 percent* of all data points used in the CFAs. An attempt to recover missing data was made through PRELIS' matching imputation procedure. Participants were matched on age and level of adaptive behavior functioning.⁵ All but one missing data point were successfully imputed. The case with the missing data point was omitted from the analyses to avoid unequal sample sizes across items.⁶

2.4.2 Multivariate Normality—Neither item nor scale score distributions exhibited multivariate normality. Item score distributions were not normal because all CBCL items were dichotomized for the item level analyses (“Not True” = 0; and “Somewhat or Sometimes True” and “Very True or Often True” = 1), which was consistent with Achenbach & Rescorla (2001). The continuous scale score distributions were positively skewed for the five scales that contributed to the *Internalizing* and *Externalizing* domains. Because chi-square tests (χ^2) are typically reported in CFA research, we presented the Satorra-Bentler chi-square statistic ($SB\chi^2$; Satorra & Bentler, 1994) which is preferred for ordinal and non-normal data (see Curran, West, & Finch, 1996).

2.4.3 Empty Cells—Tetrachoric correlation matrices for CBCL items were analyzed in the item-level CFAs. Despite dichotomizing the items, we observed item pairs with at least one empty cell in their 2×2 bivariate frequency table, which is used to compute the tetrachoric correlations. Empty cells can bias the correlations (see Greer, Dunlap, & Beatty, 2003), and items contributing to several such cells were not included in the analyses. Item

³We performed exploratory analyses to assess multi-factor solutions for all syndrome scales. Analyses failed to converge or resulted in out of range parameter estimates which indicated that multi-factor models were inadequate.

⁴Missing data were also observed for Items 24, 44, 56h, and 109. These “Other Problems” items were not in any of the scales assessed by the CFAs. These items do contribute to the *Total Problems* score.

⁵Qualitative adaptive behavior categories were used for matching because 93 participants were evaluated with the Vineland Adaptive Behavior Scales and 29 with the Vineland Adaptive Behavior Scales-II. The categories correspond to those presented in Table 1.

⁶See Brown (2006) for a discussion on problems associated with factor analyses of correlation matrices with unequal *n*-sizes across item pairs.

omission is not unprecedented in factor analyses of the CBCL (e.g., see Ivanova, Achenbach, Dumenci et al., 2007; Pandolfi et al., 2009), and it mostly affected the *Thought Problems* and *Rule Breaking Behavior* scales. For all but two of these items, 92–100% of respondents scored the omitted item “0” which indicated that such behaviors were observed very infrequently in this sample.⁷

Continuous data were used in the CFAs of the one- and two-factor models. The raw scores for each syndrome reflected the sum of *all* item scores: no items were omitted from these analyses. Items retained their original 0, 1, 2 metric.

2.4.4 Estimation Method—The Robust Diagonally Weighted Least Squares (DWLS) estimator was used for all CFAs and LISREL 8.80 provided robust test statistics. DWLS is appropriate for ordered-categorical data (Flora & Curran, 2004; Wirth & Edwards, 2007), as well as for non-normal continuous data. Yang-Wallentin, Jöreskog, and Luo (2010) found that DWLS worked well for models with six to 16 ordinal variables in sample sizes ranging from 100–1600. Our sample size and the number of ordinal variables used in the item-level CFAs all fell well within these parameters.

2.4.5 Assessing Model Fit—Multiple methods evaluated model fit and included inspection of model parameters, statistical tests, and psychometric indices. In all CFAs, adequacy of a model was determined by examining the pattern of results across all fit measures. First, models were inspected for out of range parameter estimates (e.g., negative error variances), which suggest an “incorrect” model. Several model fit indices were then examined. The $SB\chi^2$ is a statistical test of model fit and a nonsignificant test ($p > .05$) indicates good fit. Because the outcomes of statistical tests are often related to sample size (Bentler & Bonett, 1980) psychometric fit indices were also used and were considered the primary fit measures because they are less dependent on sample size.

We used three commonly reported psychometric fit indices: the Root Mean Square Error of Approximation (*RMSEA*; Steiger & Lind, 1980), the Comparative Fit Index (*CFI*; Bentler, 1990), and the Standardized Root Mean Square Residual (*SRMR*; see Bentler, 1995). The *RMSEA* is sensitive to misspecified factor loadings (Hu & Bentler, 1999). A range of *RMSEA* values have been considered to evaluate model fit: values $\leq .05$ indicate a *good* fit, values greater than .05 and $\leq .10$ indicate an *acceptable* fit, and values $> .10$ indicate models that should be rejected (MacCallum, Browne, & Sugawara, 1996). The *CFI* is also sensitive to misspecified factor loadings and values close to .95 suggest good fit (Hu & Bentler, 1999). The *SRMR* is sensitive to misspecified factor covariances and values $\leq .08$ are considered evidence of good fit (Hu & Bentler, 1999).

Although it would have been preferable to use the same set of fit indices for tests of scale unidimensionality and the scale level factor structure, this was not possible. First, the *SRMR* is not recommended for CFAs of dichotomous variables (see Yu, 2002). Second, Curran et al. (2003) reported that the *RMSEA* tends to over-reject good models when $N < 200$, and it adjusts for parsimony by incorporating a penalty function for complex models with few degrees of freedom. This was the case here for the one- and two-factor models using continuous data. Thus, our analysis plan proceeded as follows: (a) because the *SRMR* could not be used in the *item-level* CFAs, the *CFI* and *RMSEA* were the primary fit measures, and (b) in CFAs of the continuous data we used the *CFI* and *SRMR* as primary fit measures. The latter approach incorporated one index that was sensitive to misspecified factor covariances which was appropriate for a test of the two-factor model.

⁷The specific zero frequency cells are available upon request from the first author.

2.5 Additional Analyses

The CFAs were followed by additional analyses using IBM SPSS Statistics 18.0.2 (2010). We examined criterion-related validity through significance tests that examined mean differences in CBCL scale scores between a group with ASD alone and one with ASD +EBD. Next, ROC analyses were used to determine diagnostic accuracy of each CBCL subscale. For these analyses, we considered the following disorders: depressive disorders (which included individuals with Major Depressive Disorder and Depressive Disorder-NOS), anxiety disorders (those with Specific Phobia, Obsessive-Compulsive Disorder, and Generalized Anxiety Disorder), ADHD, and ODD.

3. Results

3.1 CFA Tests for Unidimensionality

Table 2 presents results for the item level CFAs.

No out of range parameters were observed. The *RMSEA* and *CFI* supported the unidimensionality of all initially tested models with the exception of *Thought Problems* and *Aggressive Behavior*. Modification indices suggested that these models could be improved by allowing for correlated disturbance (error) terms for some item pairs. First, disturbance terms for two item pairs on *Thought Problems* should be correlated: *plays with own sex parts in public* with *plays with own sex parts too much*; and *sleeps less than most kids with trouble sleeping*. Correlated disturbances were also suggested for two item pairs on *Aggressive Behavior*: *destroys his/her own things* with *destroys things belonging to his/her family or others*; and *disobedient at home* with *disobedient at school*. Modeling correlated disturbances is appropriate when it is substantively meaningful or if systematic error in item responses can be reasonably attributed to certain item characteristics (see Byrne, 1998). This was the case here because the item pairs reflected highly similar wording and content. The *Thought Problems* model with only one correlated disturbance was retained because the correlation between the other disturbance terms was not statistically significant. For these models CFAs indicated marginal fit for the *Thought Problems* as the CFI fell just below .95, and acceptable fit was found for *Aggressive Behavior*.

Table 2 shows that the median factor loadings for all models ranged from .34 to .75. This means that 11.6 to 56.3% of a typical item's variance was accounted for by its latent factor. Scale reliabilities were generally high, with a median of .85 (range .69 to .94). CFA parameters were used to compute this index and it reflects the proportion of true score variance measured by the scale (see Brown, 2006; Raykov, 1997; 2001). This measure of scale reliability was preferred to coefficient α since CFA parameters were available and it has less restrictive assumptions than the internal consistency models of reliability (see Green & Yang, 2009; Sijtsma, 2009).

Nine items exhibited non-significant factor loadings. These were observed across four scales, with one on *Somatic Complaints*, two on *Social Problems*, four on *Thought Problems*, and two on *Rule Breaking Behavior*. These included: (a) *problems with eyes (not if corrected by glasses)* on *Somatic Complaints*, (b) *clings to adults or too dependent* and *prefers being with younger kids* on *Social Problems*, (c) *can't get his/her mind off certain thoughts, obsessions; plays with own sex parts in public; stores up too many things he/she doesn't need; and talks or walks in sleep* on *Thought Problems*, and (d) *prefers being with older kids* and *runs away from home* on *Rule Breaking Behavior*.

3.2 CFA Tests of Scale Level Factor Structure

A competing and more parsimonious one-factor model was evaluated in addition to the two-factor *Internalizing-Externalizing* model. None of the fit measures supported the one-factor model ($SB\chi^2 = 32.49$, $df = 5$, $p < .001$; $CFI = .873$; $SRMR = .093$). Thus, this model was rejected.

Results for the *Internalizing-Externalizing* model are given in Figure 1.

The CFI and $SRMR$ indices supported the model. All factor loadings were statistically significant and of moderate to high magnitude. Thus, indicators were strongly related to their purported underlying factor. Scale reliabilities were computed using Guttman's $\lambda-2$ (Guttman, 1945) for the *Internalizing* and *Externalizing Domains*, and these estimates were .90 for both factors.⁸

The factor correlation was moderately high and statistically significant. The factors shared 34.8% of the variance which indicated that they do not provide redundant information. This result appeared to provide evidence for a higher order *Total Problems* factor, which also demonstrated a high level of reliability ($\lambda-2 = .94$).

3.3 Criterion-related Validity

The meaning of CBCL scores and scale elevations in youth with ASD, for the purposes of screening and diagnostic assessments, can be substantially enhanced if scores can be shown to differ in expected ways between clinical subgroups. Here we examined mean differences in CBCL scores obtained by one group with ASD alone and one with ASD+EBD. These groups displayed very similar means across age, FSIQ, and adaptive behavior, and very similar gender ratios, percentages receiving community psychological services and percentages using medication.⁹

Means and standard deviations for the CBCL raw scores are presented in Table 3 for both groups. Given the sample sizes, the independent t -test is generally robust to departures from normality. Because nonparametric tests produced identical results, we report the t -test results. The ASD+EBD group had significantly higher mean scores on the *Anxious/Depressed*, *Withdrawn/Depressed*, *Somatic Complaints*, *Thought Problems*, *Internalizing Domain*, and *Total Problems* scales. Effect sizes were generally moderate to large, and most persons in the ASD+EBD group had scores exceeding the mean scores obtained by the ASD only group across all CBCL scales.

3.4 Diagnostic Accuracy

ROC analyses were conducted to determine how well CBCL scales identified individuals with ASD+ a specific co-occurring EBD from those without the co-occurring EBD. Table 4 presents the best discriminating scales among both the broadband and syndrome scales for the most commonly observed disorders in the sample.

The best discriminating scales were those with the largest statistically significant areas under the curve (AUC). AUCs of 1.00 indicate perfect classification whereas values of .50 reflect chance level classification. The results were generally consistent with the validity studies of Achenbach & Rescorla (2001): the best discriminating scales were conceptually consistent with the disorder of interest. The *Internalizing Domain*, *Anxious/Depressed*, and *Withdrawn/*

⁸Drinks alcohol without parents' approval; smokes, chews, or sniffs tobacco; and uses drugs for nonmedical purposes were omitted from the *Externalizing* analysis due to zero item variances.

⁹Interested readers may contact the first author for the specific means and percentages.

Depressed scales best discriminated those with a depressive disorder from those without. The *Anxious/Depressed* and *Internalizing Domain* were best in identifying those with an anxiety disorder. *Attention Problems* and *Total Problems* were best for ADHD. *Total Problems*, the *Externalizing Domain*, and *Aggressive Behavior* were the best for ODD. As Table 4 shows, the empirically-derived cut-scores obtained from these analyses demonstrated acceptable sensitivity but specificity was low for many scales. Better specificities were obtained for the *Internalizing Domain* for depressive disorders and *Total Problems* for ODD. The patterns of positive and negative predictive values suggested that the empirically-derived cut scores were more useful for ruling out specific disorders than they were for positively identifying them.

4. Discussion

4.1 General Findings

The CBCL is one of the most widely used and well-researched behavior rating scales for youth available. Our study represents an initial psychometric evaluation of the CBCL in a sample of youth with ASD. Results provided evidence pertaining to the interpretation of CBCL scores as indicators of EBDs in youth with an ASD. This study addressed a gap in the evidence-based assessment literature pertaining to psychometric support for measures used with specific subgroups in the population (see Mash & Hunsely, 2005).

On the whole, the CBCL demonstrated favorable psychometric properties in our sample of youth with ASD. We found initial support for both the unidimensionality of the syndrome scales and support for the CBCL factor structure at the scale level. Several mean CBCL scale scores obtained by a group with ASD+EBD were significantly higher than those obtained by the group with ASD alone. With respect to diagnostic accuracy, sensitivity was generally acceptable but many scales exhibited low specificity. We now elaborate on the evidence pertaining to the validity of CBCL scores as measures of EBDs in ASD, address study limitations, and discuss implications for research and practice.

4.2 Syndrome and Broadband Scales

CFAs directly tested the CBCL's empirically established internal structure. Results indicated that: (a) the interrelationships among test items within each syndrome scale could be accounted for by a single latent factor, although weaker support was found for *Thought Problems*, and (b) the interrelationships among the internalizing and externalizing syndrome scales could be accounted for by two correlated but nonredundant latent factors which were the *Internalizing* and *Externalizing Domains*. Thus, we found that the CBCL measured two broad and eight narrow dimensions of EBD in our sample of youth with ASD, consistent with Achenbach & Rescorla (2001). Scale reliabilities were good to excellent.

The group with ASD+EBD scored significantly higher than the one with ASD alone across *Anxious/Depressed*, *Withdrawn/Depressed*, *Somatic Complaints*, *Thought Problems*, the *Internalizing Domain*, and *Total Problems*. All of these differences were in the expected direction. Thus, most of the differences were found on those scales that assessed internalizing problems, which is not surprising since depressive and anxiety disorders were two of the most commonly diagnosed disorders in the sample. Although many participants were diagnosed with ADHD, group differences on *Attention Problems* were not observed. These results lend initial support for the CBCL's ability to discriminate between groups of individuals with ASD vs. those with ASD+EBD.

The ROC analyses indicated that the CBCL may have utility in discriminating individuals with an ASD and a specific co-occurring EBD from those without the specific EBD. The specific EBDs evaluated included depressive disorders, anxiety disorders, ADHD, and

ODD. These disorders are cited in the literature as some of the most commonly co-occurring EBDs in youth with ASD. The scales that discriminated the best were those that were conceptually consistent with each EBD analyzed. AUCs were statistically significant, and sensitivity values were moderate to high; however, specificity values were generally low.

Thus, the cut scores derived from the ROC analyses identified most individuals with the EBD evaluated, but many false positives were observed. This was especially true for anxiety disorders and ADHD. Higher specificities were obtained for the *Internalizing Domain* when identifying depressive disorders, and the *Total Problems* scale for identifying ODD. Perhaps the low specificities could be expected given that: (a) the CBCL scales reflect dimensional EBD *syndromes* and not specific categorical DSM *disorders* and (b) high rates of co-morbidity were observed in this sample which is consistent with other reports in the literature (e.g., Leyfer et al., 2006). The findings suggest the need for the use of multiple assessment methods in clinical practice to better understand why youth with ASD score at or above the empirically-derived cutoffs on the syndrome and broad domain scales.

4.3 Limitations

We needed to omit items from the item-level CFAs due to their extremely low rate of endorsement. Although not desirable, item omission is not unprecedented in factor analytic research of the CBCL with new populations (e.g., Ivanova et al., 2007). This mostly affected the *Thought Problems* and *Rule Breaking Behavior* scales. Most omitted items exhibited little or no variance, which can result in biased tetrachoric correlations and affect CFA results. The extent to which item omission changes the fundamental nature of each construct should be considered both conceptually and empirically in future studies.

Several factors might have contributed to this lack of item variance. Such may include ASD-specific factors, low base rates of certain problems in the general population, and/or characteristics unique to this sample. *Rule Breaking Behavior*, for example, assesses for behaviors associated with ODD and Conduct Disorder (CD). In our sample the respondent rarely, if ever, endorsed items reflecting problems related to alcohol and drug use, sexual problems, fire setting, stealing, truancy, and vandalism. Perhaps the social-cognitive deficits evidenced by youth with ASD act as a protective factor for socially mediated behavior problems (see Guttman-Steinmetz, Gadow, & DeVincent, 2009). Items omitted from the *Thought Problems* scale may be related to low base rates of some problems in the general population. Items assessing perceptual distortions, as observed in schizophrenia (e.g., auditory/visual hallucinations), and those assessing self-harm/suicide were seldom observed in this sample that ranged in age from six to 18 years. This is not surprising since schizophrenia has a typical onset in *adolescence* and *early adulthood* and has a lifetime prevalence rate of approximately 1% (Gur, et al., 2005). In addition, the 2006 suicide rate for the general population was less than 1% for the child and adolescent age ranges (Heron et al., 2009). Therefore, one could reasonably expect these items to be endorsed infrequently in children and adolescents with ASD.

Another limitation pertains to generalization. Our data were gathered from predominately white, middle-class voluntary research participants. The extent to which the results apply to more diverse samples including those who do not volunteer for research is unclear. Further, our significance tests and diagnostic accuracy analyses did not include the data from the 29 participants with both ASD +DS so it is unclear how these scales would perform for these individuals. The issue here relates to the *magnitude* of scores for persons with DS and not for the *internal structure* of the CBCL. Our CFA findings provide preliminary support for the validity of the CBCL's internal structure for a wide range of youth with disabilities, all of whom share a diagnosis of ASD.

4.4 Implications for Research and Practice

Replication of our analyses on more diverse samples is needed. The CBCL's psychometric properties should be evaluated across several variables that could be related to the presentation of EBD in youth with ASD. Such might include gender, age (e.g., the 6–11 and 12–18 CBCL normative age ranges), and level of cognitive functioning. This information would greatly advance our understanding about the appropriate uses and interpretations of CBCL scores for more specific subgroups within the ASD population.

Additional research is also needed to determine the extent to which scale elevations reflect ASD-specific problems vs. co-occurring EBDs (see also Georgiades et al., 2010). For example, to what extent do elevations on the *Withdrawn/Depressed* subscale reflect depression vs. social impairment in ASD? To what extent do elevations on the *Attention Problems* subscale reflect ADHD-specific problems vs. ASD-specific attention problems? Multiple regression analysis could be used to determine the relative proportions of variance in CBCL scale scores that are accounted for by EBD- vs. ASD-specific symptoms. In addition, profile analysis may assist by demonstrating the extent to which various clinical and nonclinical subgroups obtain different profiles on the CBCL. Profiles can be compared across typically developing controls, individuals with ASD alone, those with ASD+EBD, and other clinical groups.

More psychometric and assessment research is needed to further develop evidence-based clinical assessment guidelines. Such guidelines are needed to recommend evidence-based measures to practitioners that can assist them in diagnostic decision-making, intervention planning, and monitoring response to intervention. Standardized evidence-based assessment protocols can also assist in promoting more uniformity in diagnostic assessment and eligibility determination procedures for school- and community-based programs and services. We believe that this study reflects a step in that direction.

We recommend that practitioners consider using the CBCL 6–18 as part of a multi-method assessment protocol for screening and ongoing monitoring of co-occurring EBD in youth with ASD. Although more research on the CBCL is needed, our initial psychometric results support the position that CBCL scores can be interpreted as indicators of two broad and eight narrow dimensions of EBD for this population. The *Thought Problems* scale received only modest support here, so scores should be interpreted with caution. In general, scale elevations can be taken as evidence of a significant emotional and/or behavioral problem requiring further diagnostic assessment and specific intervention. Because youth with ASD may not display the full range of socially-mediated ODD/CD behaviors, they may not often have elevated scores on *Rule Breaking Behavior* and, in turn, the *Externalizing Domain*. Therefore, item scores and other data sources (e.g., the NCBRF, ASD-CC) should be inspected to help identify externalizing problems that may be related to functional impairment. Finally, the generally low specificity values obtained in the diagnostic accuracy analyses indicates that CBCL data should be combined with other ASD-specific data such as the ADOS, ADI-R, and Childhood Autism Rating Scale- Second Edition (Schopler et al., 2010) and EBD-specific data such as the K-SADS and CSI-4 to help determine whether CBCL scale elevations reflect an exacerbation of ASD symptoms and/or the presence of a co-occurring EBD requiring specific treatment (e.g., treatment for anxiety, depression, etc.). The selection of additional measures will depend on factors such as the nature of the presenting problem, the child's age and functioning level, and the amount of psychometric support for various measures. This practice will help guide diagnostic decision-making and the selection of appropriate interventions and support services.

Acknowledgments

This study was supported in part by NIH grant U54MH066397 (Rodier, PI Studies to Advance Autism Research and Treatment (STAART) Center; Magyar, PI STAART Assessment Core), General Clinical Research Center grant 5 MO1RR0044, NIH, National Center for Research Resources, and AUCD/RTO1 2005-1/2-08 (Hyman, PI).

The authors thank Courtney McGuire who assisted in managing portions of the database for this project. We also thank Dr. Thomas Achenbach and Dr. Leslie Rescorla for their feedback on an earlier version of the manuscript. All CBCL items in this manuscript were reprinted with permission: Copyright 2001 by T.M. Achenbach and L.A. Rescorla.

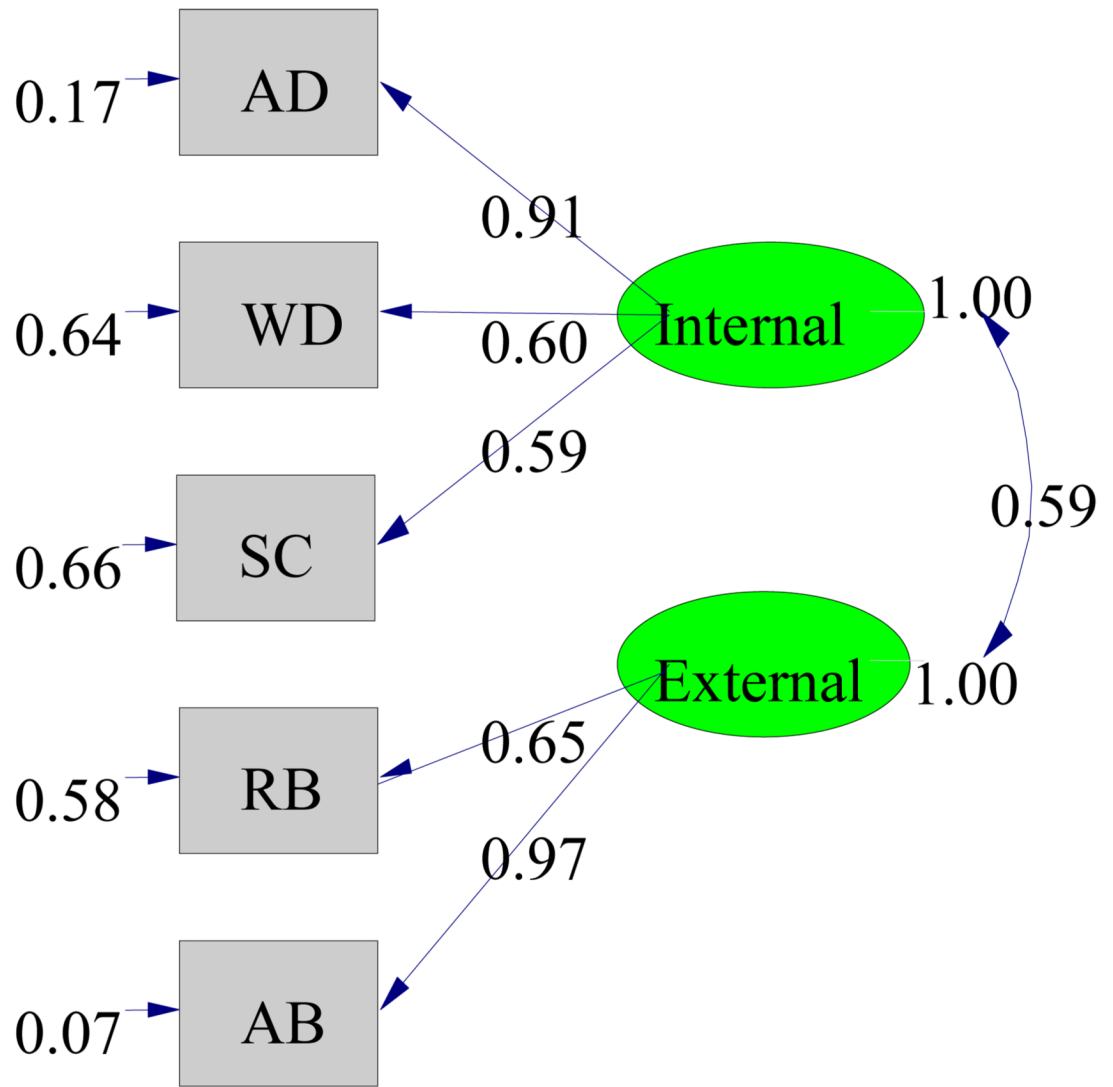
References

- Achenbach, T.M.; Rescorla, L.A. Manual for the ASEBA Preschool Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families; 2000.
- Achenbach, T.M.; Rescorla, L.A. Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families; 2001.
- Aman MG, Singh NN, Stewart AW, Field CJ. The aberrant behavior checklist: A behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency*. 1985a; 89:485–491. [PubMed: 3993694]
- Aman MG, Singh NN, Stewart AW, Field CJ. Psychometric characteristics of the aberrant behavior checklist. *American Journal of Mental Deficiency*. 1985b; 89:492–502. [PubMed: 3158201]
- Aman MG, Tasse MJ, Rojahn J, Hammer D. The Nisonger CBRF: A child behavior rating form for children with developmental disabilities. *Research in Developmental Disabilities*. 1996; 17:41–57. [PubMed: 8750075]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders: Fourth Edition Text Revision. Arlington, VA: Author; 2000.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. [PubMed: 2320703]
- Bentler, PM. EQS structural equations program manual. Encino, CA: Multivariate Software; 1995.
- Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*. 1980; 88(3):588–606.
- Berubé, R.L.; Achenbach, T.M. Bibliography of published studies using the Achenbach System of Empirically Based Assessment: 2006 Edition. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2010. Available online at www.ASEBA.org
- Biederman J, Petty CR, Fried R, Wozniak J, Micco JA, Henin A, Faraone SV. Child behavior checklist clinical scales discriminate referred youth with autism spectrum disorder : A preliminary study. *Journal of Developmental and Behavioral Pediatrics*. 2010; 31(6):485–490. [PubMed: 20585266]
- Bolte S, Dickhut H, Poustka F. Patterns of parent-reported problems indicative in autism. *Psychopathology*. 1999; 32:93–97. [PubMed: 10026453]
- Brinkley J, Nations L, Abramson RK, Hall A, Wright HH, Gabriels R, Cuccaro ML. Factor analysis of the Aberrant Behavior Checklist in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2007; 37:1949–1959. [PubMed: 17186368]
- Brown, T.A. Confirmatory factor analysis for applied research. New York: Guilford Press; 2006.
- Byrne, B.M. Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming. Mahwah, N.J: Lawrence Erlbaum Associates; 1998.
- Curran PJ, Bollen KA, Chen F, Paxton P, Kirby JB. Finite sampling properties of the point estimates and confidence intervals of the RMSEA. *Sociological Methods & Research*. 2003; 32(2):208–252.
- 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.
- DeBruin EI, Ferdinand RF, Meester S, deNijs PFA, Verheij F. High rates of psychiatric co-morbidity in PDD-NOS. *Journal of Autism and Developmental Disorders*. 2006; 37:877–886.
- Duarte CS, Bordin IAS, de Oliveira A, Bird H. The CBCL and the identification of children with autism and related conditions in Brazil: Pilot findings. *Journal of Autism and Developmental Disorders*. 2003; 33(6):703–707. [PubMed: 14714937]

- 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, Drabick DAG. Oppositional defiant disorder as a clinical phenotype in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2008; 38:1302–1310.
- 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]
- Gadow KD, Schwartz J, DeVincent C, Strong G, Cuva S. Clinical utility of the autism spectrum disorder scoring algorithms for the Child Symptom Inventory-4. *Journal of Autism and Developmental Disorders*. 2008; 38:419–427. [PubMed: 17616796]
- Gadow, KD.; Sprafkin, J. *Child symptom inventory-4 screening and norms manual*. Stony Brook, N.Y.: Checkmate Plus; 2002.
- Georgiades S, Szatmari P, Duku E, Zwaigenbaum L, Bryson S, Roberts W, Thompson A. Phenotypic overlap between core diagnostic features and emotional/behavioral problems in preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2010
- Ghaziuddin M, Ghaziuddin N, Greden J. Depression in persons with autism: Implications for research and clinical care. *Journal of Autism and Developmental Disorders*. 2002; 32(4):299–306. [PubMed: 12199134]
- Gillberg C, Billstedt E. Autism and Asperger syndrome: Coexistence with other clinical disorders. *Acta Psychiatrica Scandinavica*. 2000; 102:321–330. [PubMed: 11098802]
- Green SB, Yang Y. Commentary on coefficient alpha: A cautionary tale. *Psychometrika*. 2009; 74(1): 121–135.
- Greer T, Dunlap WP, Beatty GO. A monte carlo evaluation of the tetrachoric correlation coefficient. *Educational and Psychological Measurement*. 2003; 63(6):931–950.
- Gur, RE.; Andreasen, N.; Asarnow, R.; Gur, R.; Jones, P.; Kendler, K.; Weinberger, D. Defining schizophrenia. In: Evans, DL.; Foa, EB.; Gur, RE.; Hendin, H.; O'Brien, CP.; Seligman, MEP.; Walsh, BT., editors. *Treating and preventing adolescent mental health disorders: What we know and what we don't know*. New York: Oxford University Press; 2005. p. 77-107.
- Guttman L. A basis for analyzing test-retest reliability. *Psychometrika*. 1945; 10:255–282. [PubMed: 21007983]
- Guttman-Steinmetz S, Gadow KD, DeVincent CJ. Oppositional defiant and conduct disorder behaviors in boys with autism spectrum disorder with and without attention-deficit hyperactivity disorder versus several comparison samples. *Journal of Autism and Developmental Disorders*. 2009; 39:976–985. [PubMed: 19288296]
- Hedges, LV.; Olkin, I. *Statistical methods for meta-analysis*. San Diego, CA: Academic Press; 1985.
- Heron, MP.; Hoyert, DL.; Murphy, SL.; Xu, JQ.; Kochanek, KD.; Tejada-Vera, B. Deaths: Final data for 2006. *National vital statistics reports*. 2009. Retrieved from Centers for Disease Control and Prevention website: http://www.cdc.gov/nchs/data/nvsr/nvsr57/nvsr57_14.pdf
- Hollingshead, AA. *Four-factor index of social status*. New Haven, CT: Yale University; 1975. Unpublished manuscript
- Howlin P, Goode S, Hutton J, Rutter M. Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 2004; 45:212–229.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6:1–55.
- Hurtig T, Kuusikko S, Mattila M, Haapsamo H, Ebeling H, Jussila K, Moilanen I. Multi-informant reports of psychiatric symptoms among high-functioning adolescents with Asperger syndrome or autism. *Autism*. 2009; 13(6):583–598. [PubMed: 19933765]
- IBM SPSS Statistics 18.0.2. Chicago, IL: SPSS, Inc; 2010.
- 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]
- Jöreskog, K.; Sörbom, D. *LISREL 8.80*. Lincolnwood, IL: Scientific Software International, Inc; 2006.

- Kanne SM, Abbacchi AM, Constantino JN. Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: The importance of environmental context. *Journal of Autism and Developmental Disorders*. 2009; 39(6):856–864. [PubMed: 19191016]
- Kaufman, J.; Birmaher, B.; Brent, D.; Rao, U.; Ryan, N. Kiddie-Sads-Present and Lifetime Version; Version 1.0 of October, 1996. 1996. <http://www.wpic.pitt.edu/ksads>
- Kim JA, Szatmari P, Bryson SE, Steiner DL, Wilson FJ. The prevalence of anxiety and mood problems among children with autism and Asperger syndrome. *Autism*. 2000; 4:117–131.
- Kraijer D, de Bildt A. The PDD-MRS: An instrument for identification of autism spectrum disorders in persons with mental retardation. *Journal of Autism and Developmental Disorders*. 2005; 35(4): 499–513. [PubMed: 16134035]
- Lainhart JE. Psychiatric problems in individuals with autism, their parents, and siblings. *International Review of Psychiatry*. 1999; 11:278–298.
- Lecavalier L, Aman MG, Hammer D, Stoica W, Mathews GL. Factor analysis of the Nisonger Child Behavior Rating Form in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2004; 34(6):709–721. [PubMed: 15679190]
- Lecavalier L, Gadow KD, DeVincent CJ, Edwards MC. Validation of the DSM-IV model of psychiatric syndromes in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009; 39:278–289. [PubMed: 18654843]
- 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*. Los Angeles: Western Psychological Services; 2002.
- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*. 1996; 1:130–149.
- Mahan S, Matson JL. Children and adolescents with autism spectrum disorders compared to typically developing controls on the Behavioral Assessment System for Children, Second Edition (BASC-2). *Research in Autism Spectrum Disorders*. 2011; 5:119–125.
- Mash, EJ.; Dozois, DJA. Child psychopathology: A developmental-systems perspective. In: Mash, EJ.; Barkley, RA., editors. *Child psychopathology*. 2nd Ed. New York: Guilford; 2003. p. 3-71.
- Mash EJ, Hunsley J. Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34(3):362–379. [PubMed: 16026210]
- Matson JL, LoVullo SV, Rivet TT, Boisjoli JA. Validity of the Autism Spectrum Disorder- Comorbid for Children (ASD-CC). *Research in Autism Spectrum Disorders*. 2009; 3:345–357.
- 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]
- Netemeyer, RG.; Bearden, WO.; Sharma, S. *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage Publications, Inc; 2003.
- Ooi YP, Rescorla L, Ang RP, Woo B, Fung DSS. Identification of autism spectrum disorders using the Child Behavior Checklist in Singapore. *Journal of Autism and Developmental Disorders*. 2010
- 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]
- Pandolfi, V.; Magyar, CI. Child behavior checklist in autism. In: Volkmar, FR., editor. *Encyclopedia of Autism Spectrum Disorders*. New York: Springer; in press
- Pandolfi V, Magyar CI, Dill CA. Confirmatory factor analysis of the Child Behavior Checklist 1.5–5 in a sample of children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009
- Raykov T. Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*. 1997; 21(2):173–184.

- Raykov T. Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*. 2001; 54:315–323. [PubMed: 11817096]
- Rescorla L. Cluster analytic identification of autistic preschoolers. *Journal of Autism and Developmental Disorders*. 1988; 18(4):475–492. [PubMed: 3215877]
- Reynolds, CR.; Kamphaus, RW. *Behavior assessment system for children- Second Edition*. Circle Pines, MN: AGS Publishing; 2004.
- Roid, GH. *Stanford-Binet Intelligence Scales- (5th Ed.)*. Itasca, IL: Riverside; 2003.
- Roid, GH.; Miller, LJ. *Leiter International Performance Scale- Revised*. Wood Dale, IL: Stoelting; 1997.
- Rutter, M.; Bailey, A.; Lord, C. *Social Communication Questionnaire*. Los Angeles: Western Psychological Services; 2003.
- Rutter, M.; LeCouteur, AL.; Lord, C. *Autism Diagnostic Interview- Revised*. Los Angeles: Western Psychological Services; 2003.
- Satorra, A.; Bentler, PM. Corrections to test statistics and standard errors in covariance structure analysis. In: Eye, Avon; Clogg, CC., editors. *Latent variable analysis: Applications to developmental research*. Thousand Oaks, CA: Sage; 1994. p. 399-419.
- Schopler, E.; Van Bourgondien, ME.; Wellman, GJ.; Love, SR. *Childhood Autism Rating Scale-2nd Edition*. Los Angeles: Western Psychological Services; 2010.
- 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]
- Sijtsma K. On the use, the misuse, and very limited usefulness of Cronbach's alpha. *Psychometrika*. 2009; 74(1):107–120. [PubMed: 20037639]
- Sparrow, SS.; Balla, DA.; Cicchetti, DV. *Vineland Adaptive Behavior Scales- Interview Edition*. Circle Pines, MN: American Guidance Service; 1984.
- Sparrow, SS.; Balla, DA.; Cicchetti, DV. *Vineland Adaptive Behavior Scales: Second Edition*. Circle Pines, MN: American Guidance Service; 2005.
- Steiger, JH.; Lind, JM. Statistically based tests for the number of common factors; Iowa City, IA. Paper presented at the meeting of the Psychometric Society; 1980.
- Volker MA, Lopata C, Smerbeck AM, Knoll VA, Thomeer ML, Toomey JA, Rodgers JD. BASC-2 PRS profiles for students with high functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2010; 40:188–199. [PubMed: 19705267]
- Wechsler, D. *Wechsler Adult Intelligence Scale- Third Edition*. San Antonio, TX: The Psychological Corporation; 1997.
- Wechsler, D. *Wechsler Intelligence Scale for Children- Fourth Edition*. San Antonio, TX: Harcourt Assessment, Inc; 2003.
- White SW, Oswald D, Ollendick T, Scahill L. Anxiety in children and adolescents with autism spectrum disorders. *Clinical Psychology Review*. 2009; 29:216–229. [PubMed: 19223098]
- Wirth RJ, Edwards MC. Item factor analysis: Current approaches and future directions. *Psychological Methods*. 2007; 12(1):58–79. [PubMed: 17402812]
- Yang-Wallentin F, Jöreskog KG, Luo H. Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling: A Multidisciplinary Journal*. 2010; 17(3): 392–423.
- Yu, CY. *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles: University of California; 2002. Unpublished doctoral dissertation



$$SB\chi^2 = 11.21, df = 4, p = .024, CFI = .967, SRMR = .042$$

Figure 1. CFA Results for the two factor model

AD=Anxious/Depressed, WD=Withdrawn/Depressed, SC=Somatic Complaints, RB=Rule-Breaking Behavior, AB=Aggressive Behavior, Internal=Internalizing Domain, External=Externalizing Domain. Standardized factor 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. Factor correlation is presented between the latent factors.

Table 1

Participant Characteristics

Demographic Data	Percent ^a		
Gender: Male	82.8		
SES ^b			
Major Business/Professional	38.0		
Medium Business/Minor Professional	42.4		
Skilled Craftsman/Clerical/Sales	13.0		
Machine Operator/Semi-skilled	4.3		
Unskilled Laborers/Menial Service	2.2		
Educational Status ^c			
High School	13.8		
Partial College/Specialized Training	10.3		
College	75.9		
Race			
Asian	1.6		
Black/African-American	2.5		
White	95.9		
Ethnicity: Hispanic/Latino	0.8		
Percent with Down syndrome	23.8		
Adaptive Behavior Classification ^d			
Moderately High (115–129)	0.8		
Adequate (86–114)	6.6		
Moderately Low (71–85)	23.0		
Low (≤ 70)	69.7		
Mild (55–70)	45.9		
Moderate (40–54)	15.6		
Severe (25–39)	7.4		
Profound (≤ 24)	0.8		
<hr/>			
Cognitive Data	<i>M</i>	<i>SD</i>	<i>Range</i>
<hr/>			
Full Scale IQ			
WISC-IV (<i>n</i> = 48)	98.02	19.83	44–131
WAIS-III (<i>n</i> = 2)	100.00	-----	0
Stanford-Binet-5 (<i>n</i> = 30)	96.60	19.40	41–131
Leiter-R (<i>n</i> = 32)	53.31	26.15	36–125

Notes.

^a Percentages in some categories do not sum to 100 due to rounding error.

^b Hollingshead (1975) scale used for participants without Down Syndrome (*n* = 92).

^c Indicator of SES for families of participants with Down Syndrome (*n* = 29), Hollingshead data were not collected in that study.

^dCategories reported are based on Adaptive Behavior Composite standard score ranges consistent with Sparrow, Cicchetti, and Balla (2005).
Categorical data presented because 93 participants were administered the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) and 29 were administered the Vineland-II (Sparrow et al., 2005) in separate studies.

Table 2

CFA Results For Syndrome Scales

Model	$SB\chi^2$ ^a	df	RMSEA	CFI	Median Factor Loading	r_c ^b
Anxious/Depressed	78.51	65	.041	.995	.750	.94
Withdrawn/Depressed	34.59*	20	.078	.975	.655	.85
Somatic Complaints ^c	28.16	35	.000	1.00	.604	.88
Social Problems	77.98**	44	.080	.959	.607	.84
Thought Problems ^d	83.67*	35	.107	.900	.437	.76
Correlated disturbance ^e	68.14**	34	.091	.930	.340	.69
Attention Problems	62.33**	35	.080	.955	.519	.83
Rule-Breaking Behavior ^f	14.86	14	.023	.997	.612	.76
Aggressive Behavior ^g	207.74**	90	.104	.949	.658	.92
Correlated disturbance ^h	119.81*	88	.055	.986	.665	.90

Notes. N = 122. CFAs used Diagonally Weighted Least Squares estimation.

^a Satorra-Bentler chi-square.

^b Scale reliability.

^c Sans Item 56g vomiting, throwing up.

^d Sans Items 18 deliberately harms self or attempts suicide, 40 hears sound or voices that aren't there, 46 nervous movements or twitching, 70 sees things that aren't there, 85 strange ideas.

^e Correlated errors for Items 59 plays with own sex parts and 60 plays with own sex parts too much.

^f Sans Items 2 drinks alcohol without parents' approval, 72 sets fires, 73 sexual problems, 81 steals at home, 82 steals outside the home, 96 thinks about sex too much, 99 smokes, chews, or sniffs tobacco, 101 truancy, skips school, 105 uses drugs for nonmedical purposes, 106 vandalism.

^g Sans Items 37 gets in many fights, 86 stubborn, sullen, or irritable and 88 sulks a lot.

^h Correlated errors for items 20 (destroys his/her own things) and 21 (destroys things belonging to his/her family or others); and items 22 (disobedient at home) and 23 (disobedient at school).

* $p < .05$

** $p < .005$

Table 3
 Mean Raw Score Differences on CBCL Scales Between Groups With and Without a Co-occurring EBD

Scale	ASD+EBD ^a		ASD only ^b		Percentage of ASD+EBD above ASD sample mean	
	M	SD	M	SD	M	g ^c
Anxious/Depressed	7.85**	5.34	4.50	3.86	.67	74.9
Withdrawn/Depressed	5.10*	3.34	3.36	2.72	.55	70.9
Somatic Complaints	3.02*	2.70	1.61	2.18	.55	70.9
Social Problems	6.97	3.25	5.61	3.54	.41	65.9
Thought Problems	8.05*	4.53	5.61	3.61	.57	71.6
Attention Problems	9.95	4.06	9.07	4.29	.21	58.3
Rule Breaking	2.73	2.34	2.25	2.29	.21	58.3
Aggressive Behavior	8.14	6.77	5.96	5.54	.34	63.3
Internalizing	15.94**	9.12	9.46	6.87	.76	77.6
Externalizing	10.87	8.48	8.21	7.19	.33	62.9
Total Problems	57.59**	25.09	41.86	20.32	.66	74.5

^a n = 63

^b n = 28

^c Hedges' g (see Hedges & Olkin, 1985)

* p < .05

** p < .01

Table 4
 CBCL Syndrome and Broadband Scales With Best Diagnostic Accuracy for the Most Commonly Observed Disorders in the Sample

Disorder	AUC (95% CI) ^e	p	Sensitivity	Specificity	PPV ^b	NPV ^c	Cut Score ^d
Depressive Disorder ^e							
Anxious/Depressed	.857 (.768, .947)	<.001	.867	.645	.325	.961	6.5
Withdrawn/Depressed	.816 (.695, .937)	<.001	.867	.632	.317	.960	4.5
Internalizing Domain	.909 (.847, .971)	<.001	.933	.776	.452	.983	15.5
Anxiety Disorder ^f							
Anxious/Depressed	.724 (.621, .828)	<.001	.914	.446	.508	.893	3.5
Internalizing Domain	.715 (.609, .820)	.001	.914	.446	.508	.893	8.5
ADHD							
Attention Problems	.723 (.619, .828)	<.001	.875	.424	.452	.862	7.5
Total Problems	.677 (.560, .794)	.006	.906	.254	.397	.833	35.5
ODD							
Aggressive Behavior	.836 (.722, .951)	.006	1.00	.612	.154	1.00	6.5
Externalizing Domain	.823 (.696, .949)	.009	1.00	.565	.140	1.00	8.5
Total Problems	.773 (.594, .951)	.026	.833	.741	.185	.984	58.5

Notes. N= 91.

^a Area under the curve and 95% confidence interval.

^b Positive predictive value.

^c Negative predictive value.

^d Empirically-derived cut score.

^e Included Major Depressive Disorder and Depressive Disorder- NOS

^f Included Specific Phobia, Obsessive-Compulsive Disorder, and Generalized Anxiety Disorder