Research in Autism Spectrum Disorders 58 (2019) 19–29

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

journal homepage: www.elsevier.com/locate/rasd

Measurement invariance of the child behavior checklist in children with autism spectrum disorder with and without intellectual disability: Follow-up study

Kristen Dovgan[a,*], Micah O. Mazurek[b], John Hansen[c]

[a] Department of Psychology, Marist College, United States
[b] Curry School of Education, University of Virginia, 417 Emmet Street South, PO Box 400267, Charlottesville, VA, 22904-4260, United States
[c] Hansen Research Services LLC, United States

ARTICLE INFO

No. of reviews completed 2

Keywords:
Child behavior checklist
Autism spectrum disorder
Measurement invariance
Intellectual disability

ABSTRACT

Background: Children with autism spectrum disorder (ASD) have high rates of co-occurring emotional and behavioral problems. The Child Behavior Checklist (CBCL) has been used to assess emotional and behavioral functioning in many large-scale studies of children with ASD. However, the previously established factor model may account for symptom patterns differently in children with concurrent intellectual disability (ID).

Method: In a sample of children with ASD, a multi-group nested confirmatory factor analysis was used to compare participant groups with and without concurrent ID on each of the subscales of the CBCL.

Results: In almost all of the subscales, children with ASD and concurrent ID exhibit different baseline levels, measurement error, and overall predictive ability on the behavioral and emotional problems of the CBCL than children with ASD alone. In younger children with ASD, configural invariance was present in Emotional Reactivity, Anxious/Depressed, and Somatic Complaints, metric invariance was present in Sleep Problems, residual invariance was present in Withdrawn Behavior, and structural invariance was present in Attention Problems and Aggressive Behavior. In older children with ASD, configural invariance was present in Anxious/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, and Rule-Breaking; whereas, metric invariance was present in Withdrawn/Depressed and Aggressive Behavior.

Conclusions: In conclusion, when dealing with intellectually heterogeneous ASD samples, the item-level data of the CBCL should be used, rather than broad subscale-level data. These findings underscore a continued need for development and validation of measures that are appropriate for use in individuals with ASD across ages and intellectual functioning.

1. Introduction

Autism spectrum disorder (ASD) is characterized by difficulties with social communication and social interactions and by restricted and repetitive patterns of behavior and interests (American Psychiatric Association, 2013). Children with ASD also have very high rates of co-occurring emotional and behavioral problems (Simonoff et al., 2008). Therefore, it is important for both clinicians

* Corresponding author.
E-mail addresses: kristen.dovgan@marist.edu (K. Dovgan), mm5gt@virginia.edu (M.O. Mazurek), john@hrscat.com (J. Hansen).

https://doi.org/10.1016/j.rasd.2018.11.009
Received 25 April 2018; Received in revised form 6 November 2018; Accepted 13 November 2018
Available online 26 November 2018
1750-9467/

and researchers to routinely screen for potential emotional and behavioral symptoms across a range of areas using rating scales and behavior checklists (Lempp, de Lange, Radeloff, & Bachmann, 2012).

The Child Behavior Checklist (CBCL) (Achenbach & Rescorla, 2001) is a widely accepted broad-band rating scale for assessment of emotional and behavioral problems in children. The CBCL has been frequently used in research studies of children with ASD, including studies assessing behavioral and emotional problems (e.g., Gonzalez & Stern, 2016), measuring treatment outcomes, and verifying convergent and divergent validity of other measures (e.g., Kaat, Lecavalier, & Aman, 2014; Kim et al., 2016). The CBCL is also used to study comorbid psychiatric conditions in ASD, as seen in several literature reviews (e.g., Mannion & Leader, 2013; Matson & Cervantes, 2014).

The established factor structure of the CBCL has been validated in different samples using Confirmatory Factor Analyis (CFA) (Dedrick, Greenbaum, Friedman, Wetherington, & Knoff, 1997; Greenbaum & Dedrick, 1998). CFA can be used to confirm a priori hypotheses about the structure of data and to examine measurement invariance across different groups (Byrne, Shavelson, & Muthen, 1989). The CBCL syndrome scales were derived empirically through factor analysis using data from a large general population sample (Achenbach & Rescorla, 2001), and their construct validity has also been established in cross-cultural samples using CFA (De Groot, Koot, & Verhulst, 1994; Dedrick, Tan, & Marfo, 2008; Ivanova et al., 2010). However, the psychometric properties of the CBCL have only recently been examined in large samples of children with ASD.

Pandolfi and colleagues examined the factor structure of both the preschool version (Pandolfi, Magyar, & Dill, 2009) and the older version (Pandolfi, Magyar, & Dill, 2011) of the CBCL in samples of children with ASD. In the younger sample, the majority of the children had developmental delays, with average standard scores on both cognitive and adaptive measures falling within the intellectual disability (ID) range. The authors analyzed each syndrome scale separately to determine if there was evidence for a single latent factor. All children were examined as a single sample, despite differences in ID status. The authors concluded that each syndrome scale in the younger version (with the exception of Sleep Problems) was accounted for by a single factor (Pandolfi et al., 2009). In the older sample, IQ ranged from ID to superior functioning, and a minority (31.3%) of the sample met criteria for ID. Again, all of the children were examined as one sample. The results found support for the unidimensionality of each syndrome scale, with the exception of Thought Problems (Pandolfi et al., 2011).

In a more recent study, our team examined the factor structure of the CBCL in a larger sample of 623 children with ASD (Medeiros, Mazurek, & Kanne, 2017). In this initial study, we found that the established factor structure was the best fitting model for young children with ASD, but not for older children with ASD. The sample was representative of a wide range of intellectual functioning, with IQ ranging from 47 to 141. However, the study did not examine whether the CBLC performed differently among children with or without ID. This is an important consideration, given that almost one-third of the ASD population has co-occurring ID (Developmental Disabilities Monitoring Network, 2014), and children with co-morbid diagnoses of ASD and ID do not respond as positively to intervention as children with either diagnosis alone (American Psychiatric Association, 2013; Matson & Shoemaker, 2009).

There are several reasons to expect a different factor structure for the CBCL in children with ASD who have concurrent ID. First, findings from previous research report consistently higher rates of comorbid symptoms in individuals with lower intellectual functioning compared to those with higher intellectual functioning (Kozlowski, Matson, Sipes, Hattier, & Bamburg, 2011; Matson, Cooper, Malone, & Moskow, 2008; Mayes, Calhoun, Murray, Ahuja, & Smith, 2011; Smith & Matson, 2010; Tureck, Matson, Cervantes, & Konst, 2014). Second, the co-occurrence of ASD and ID is associated with a greater rate of comorbid psychopathology than in those with either diagnosis alone. This has been found in populations of both children (e.g., Bradley & Isaacs, 2006; McClain, Hasty Mills, & Murphy, 2017) and adults (e.g., LoVullo & Matson, 2009). Third, the most common co-occurring disorders identified in children with ASD and ID include depression, bipolar disorder, schizophrenia, and anxiety (Matson, Hess, & Boisjoli, 2010), and the emotional and behavioral components of these disorders are measured by the CBCL (e.g., emotional reactivity, anxiety, depression, withdrawn behavior, aggression, social problems, thought problems, and conduct issues). Fourth, some research has found that parents may rate symptoms of ASD more severely when a child has concurrent ID. This was shown using the Social Communication Questionnaire (Vermeulen et al., 2017) and the Social Responsiveness Scale (Havdahl et al., 2015).

In summary, the literature shows that the CBCL has been used widely to examine internalizing and externalizing problems in children with ASD; however, the validity of inference from CBCL scores in this population has not been well examined. Researchers who compare groups of individuals on emotional and behavioral symptoms using the CBCL are assuming that the scores reflect the same latent traits on subscales across groups. If this assumption holds, the comparisons are valid, subgroups can be treated as one population, and the CBCL subscales can be meaningfully interpreted. However, the established factor models of the CBCL, or particular subscales, may not account adequately for the symptom patterns among children with ASD (Medeiros et al., 2017; Pandolfi et al., 2009, 2011). The goal of the current research is to examine if the presence of concurrent ID among children with ASD could explain the different symptom patterns seen on the CBCL. The purpose of this follow-up study is to analyze the measurement invariance of the CBCL across children with ASD who did or did not have concurrent ID.

Measurement invariance is the property of an instrument with scores that represent the same latent trait on the same scale across groups. When the relationship between observed item-level variables and the underlying constructs are the same across groups, measurement invariance is present, and we can theoretically use and interpret the measure similarly for both groups. However, without measurement invariance, scores may not reflect the same scale properties and conclusions about the overall factor structure or underlying constructs cannot be confidently applied to both groups. For example, children with ASD without concurrent ID might be driving the statistical relationship among the CBCL items, thus leading researchers to conclude that the factor structure is similar in children with ASD and TD children. Since children with ASD and concurrent ID exhibit different internalizing and externalizing problems than children with ASD alone, or typically developing children, then it is reasonable to suggest that the certain syndrome

scales of the CBCL may be less applicable in this particular subsample. Therefore, the use and interpretation of the CBCL in this population may be meaningfully different with regard to internalizing and externalizing symptom patterns depending on intellectual ability.

Assessing measurement invariance involves the testing of a series of increasingly restrictive hypotheses, each making more assumptions about the equality of the two groups. The most basic level of invariance is configural, wherein the same unobserved construct reflects the same collection of items across groups. This type of invariance implies that the CBCL subscales are comprised of the same items for both ASD and ASD/ID children. The next level is metric invariance, wherein the factor loadings (i.e., the correlation between items and a construct) are also equal across groups. This type of invariance implies that the CBCL subscales are comprised of the same items, and that those items have a similar influence on the subscale for both ASD and ASD/ID children. The next level is scalar invariance, wherein the item intercepts (i.e., the starting value of the scale) are also equal across groups. This implies that the CBCL subscales, comprised of the same similarly-weighted items, occur at the same rate across both ASD and ASD/ID children. For example, with scalar invariance of the emotionally reactive subscale, you could assume that emotionally reactive problems occur in general, in the same amount and rate across both ASD and ASD/ID children. The next level is residual invariance, wherein measurement error is also equal across groups. This implies that the amount of error in the CBCL subscale between the true value and the measured value is the same across ASD and ASD/ID children. The last and most-strict type of invariance is structural invariance, wherein the predictive ability is equal across groups. This implies that the CBCL subscale can be used to measure emotional and behavioral problems with the same amount of accuracy across children with ASD and ASD/ID.

To assess confidently emotional and behavioral problems in the ASD population using the CBCL, researchers and clinicians need to know the level of measurement invariance present for those with and without ID. In addition, different types of invariance would lead to different recommendations about using the individual items as opposed to the overall subscales on the CBCL. Because the type of measurement invariance may influence the clinical use of the CBCL in ASD populations it is therefore an important consideration in psychometric research. We hope that the results of this follow-up study will inform the use of the CBCL for diagnostic, research, or interventions purposes for children with ASD and concurrent ID.

2. Methods

2.1. Participants

Participants included 408 children with ASD (83.6% male) ranging in age from 2 to 17 years old ($M = 7.2$, $SD = 2.0$). This sample was slightly smaller than the original study, since we removed participants for whom Full Scale IQ scores were not available. As described in the methods for the previous study (Medeiros et al., 2017), participants had participated in autism research studies at one of two academic medical centers, and item-level CBCL data were available for secondary analysis. The sample was divided into those who received the younger version of the CBCL (26.5% of our sample; $n = 108$; average age = 4.24, $SD = .76$), and those who completed the older version of the CBCL (73.5%; $n = 300$; average age = 10.13, $SD = 3.22$).

2.2. Measures

2.2.1. Intelligence

Intelligence (IQ) was assessed by clinicians at each study site using a range of measures, including the Early Years Differential Ability Scales–Second Edition (30.6%) (Elliot, 2007), the School-Age Differential Ability Scales–Second Edition (30.4%) (Elliot, 2007), the Wechsler Abbreviated Scale of Intelligence – Second Edition (17.2%) (Wechsler, 2011), the Wechsler Intelligence Scale for Children – Fourth Edition (2.2%) (Wechsler, 2003), the Wechsler Preschool and Primary Scale of Intelligence – Third Edition (4.2%) (Wechsler, 2002), the Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition (.2%) (Wechsler, 2012), the Stanford Binet Scales of Intelligence - 5th Edition (11.8%) (Roid, 2003), and the Mullen Scales of Early Learning (MSEL, 3.4%) (Mullen, 1995). For those receiving the MSEL, the Early Learning Composite score was included as a measure of Full Scale IQ. In subsequent analyses, Full Scale IQ score was used to create a dichotomous variable indicating the presence (Full Scale IQ < 70) or absence (IQ ≥ 70) of Intellectual Disability.

2.2.2. Adative functioning

The **Vineland Adaptive Behavior Scales – Second Edition (VABS-II)** survey interview form (Sparrow, Cicchetti, & Balla, 2005) was administered to parents/caregivers by clinicians or staff at each study site to assess adaptive functioning for 97.2% of the younger sample and 97.7% of the older sample. The VABS-II is a widely-used and psychometrically sound caregiver-report measure of adaptive functioning in the areas of Communication, Daily Living, and Socialization (Sparrow et al., 2005).

2.2.3. Behavioral and emotional functioning

The **Child Behavior Checklist (CBCL)** (Achenbach & Rescorla, 2001) is a parent-report questionnaire assessing children's behavioral and emotional functioning, which yields both broad-band and narrow-band scores across symptom domains. Response options for each item range from 0 (not true) to 2 (very true or often true). The CBCL has strong psychometric properties, including test-retest reliability, inter-rater agreement, and internal consistency (Achenbach, Dumenci, & Rescorla, 2003; Albores-Gallo et al., 2007; Frizzo, Pedrini, de Souza, Bandeira, & Borsa, 2014). The factor structure and validity of the empirically derived syndrome scales have been supported in both community and clinical samples (Achenbach & Rescorla, 2001; Achenbach et al., 2003; Hallerod

et al., 2010; Nakamura, Ebesutani, Bernstein, & Chorpita, 2009).

The CBCL contains two different age-based versions. For young children (Ages 1.5–5) there are 100 items and the following Syndrome Scales: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, and Aggressive Behavior. For older children (Ages 6–18) there are 113 items and the following Syndrome Scales: Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior. Some items had skewed rates of endorsement in the current study, wherein almost all parents responded, "less than average" for the item (e.g., Vomiting and Aches/Pains in the Somatic Complaints subscale; Alcohol, Runs Away, Sets Fires, Sex Problems, Thinks about Sex, Tobacco, Truancy, Drugs, and Vandalism from the Rule-Breaking subscale). These items were omitted, in line with other research modeling the CBCL in new populations (Ivanova et al., 2007).

2.3. Statistical analysis

All analyses were conducted using IBM SPSS 24 and AMOS 24.0.0 software programs. CFA was used to model latent factors associated with the associated subscale items. Parallel with our previous analyses of these data (Medeiros et al., 2017), we used the AMOS defaults, including maximum likelihood for estimation of model parameters. Substitution of missing values was necessary in order to produce meaningful results; thus missing values were replaced with the mode for a given CBCL item. This resulted in 26 substitutions (.12% of the data) in the younger sample and 119 substitutions (.26% of the data) in the older sample.

2.3.1. Multiple-group analyses

For the younger sample, we created seven individual path models for each of the CBCL subscales: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, and Aggressive Behavior. For the older sample, we created eight individual path models for subscales: Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior. Each model included the individual CBCL items for that subscale. We then designated the models to have two groups, ASD participants with and without ID. Because of the small sample size of our younger group when separated by ID status, the analyses and results for the younger age group should be considered highly exploratory.

To test for measurement invariance between groups, nested CFA models were specified with an increasing number of parameter constraints (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). Model specification starts with the *configural* model, the least constrained baseline model. Parameters in the *configural* model are freely estimated. However, in all CFA models, an indeterminate relationship between factor loadings and the scale of the latent factor requires resolving this metric identification problem by fixing one or the other. Commonly, the value of one factor loading is fixed to 1.0 (Joreskog & Sorbom, 1989). We follow this approach in line with our previous companion paper (Medeiros et al., 2017), and other CBCL research (e.g., Dedrick et al., 2008). Other approaches exist to address the indeterminacy problem, such as fixing the latent factor variance to 1.0 for both groups. And, because these approaches are all respecifications of each other, they will all result in the same level of practical and statistical fit (Reise, Widaman, & Pugh, 1993). For additional details on model identification see Long (1983) or Wu and Estabrook (2016). Following the baseline model is the first nested model, the *metric* model. For the *metric* model, the same identification constraints were used as the configural model, and item factor loadings are constrained to be equal across groups. The next model, *scalar*, adds constraints on item intercepts being equal across groups. In the *residual* model, factor loadings, intercepts, and error variances are set as equal across groups. Error variance is the measurement error associated with each observed variable. The last model, the *structural* model, constrains factor loadings, item intercepts, residual variances, and factor variances equal to one another across groups. Factor variance represents the overall error in prediction of the construct using the indicator variables, or the overall predictive ability of the subscale.

Each of the nested models is then compared with the preceding measurement invariance model. Difference are evaluated in model fit statistics and practical fit indicators, including the chi-square test, Tucker Lewis Index (TLI), Comparative Fit Index CFI, and Root Mean Square Error of Approximation (RMSEA). If there is a nonsignificant increase in chi-square when additional parameter constraints are imposed, then the hypothesis is tenable for measurement invariance at the level of the more constrained model (Reise et al., 1993). That is, we test the null hypothesis that the more constrained model is correct, under the assumption that the less constrained model is correct.

Based on past literature on nested CFA models and research conclusions (e.g., Cheung & Rensvold, 2002; Vandenberg & Lance, 2000), we generated recommendations for the use of the CBCL depending on the level of invariance upheld. Specifically, configural, metric, and scalar invariance would support the use of item-level CBCL data as the level of analysis when comparing groups; whereas, residual and structural invariance would support the use of broad, subscale-level CBCL data as the level of analysis when comparing groups.

2.3.2. Model fit and model comparison indices

Individual models were evaluated using goodness of fit statistics and fit indices, including the commonly used chi-square (χ^2) test and the above-mentioned non-statistical indicators: TLI, CFI, and the RMSEA. Nested models were evaluated using the difference of each chi-square (χ^2) value to determine if the more restricted nested model had a nonsignificant increase in chi-square, wherein nonsignificant increase indicated that the hypothesis of measurement invariance at the more restricted level was tenable. That is, if the models were statistically equivalent, a nonsignficant chi square value would be expected for that model comparison. Although chi-square difference tests are affected by sample size, in this study, the sample sizes were small enough to assume that trivial differences

Table 1
Descriptive statistics and statistical comparison of group characteristics by IQ level.

	1.5–5 Years Old			6–18 Years Old		
	ID (n = 33)	No ID (n = 75)	Test Statistic	ID (n = 74)	No ID (n = 226)	Test Statistic
Age	M = 4.12 SD = .86	M = 4.36 SD = .65	t (106) = −1.59	M = 10.07 SD = 3.28	M = 10.18 SD = 3.15	t (298) = −.267
% Male	75.8	90.7	$\chi^2(1) = 4.26$*	75.7	92.0	$\chi^2(1) = 14.13$**
IQ	R: 22–68 M = 55.09 S = 13.39	R: 70–130 M = 92.17 S = 13.78	t (106) = −12.99**	R: 20–69 M = 51.0 S = 13.08	R: 70–141 M = 96.58 S = 15.09	t (298) = −23.27**
Communication Skills	R: 42–100 M = 66.39 SD = 14.31	R: 57–118 M = 85.69 SD = 13.77	t (103) = −6.59**	R: 28–92 M = 64.73 SD = 9.82	R: 54–129 M = 79.03 SD = 12.02	t (290) = −9.26**
Daily Living Skills	R: 51–91 M = 69.82 SD = 10.57	R: 58–113 M = 82.6 SD = 12.43	t (103) = −5.12**	R: 28–101 M = 67.81 SD = 12.82	R: 48–134 M = 81.05 SD = 12.26	t (290) = −7.93**
Socialization Skills	R: 49–86 M = 67.61 SD = 9.91	R: 57–112 M = 75.6 SD = 11.65	t (103) = −3.41**	R: 40–87 M = 63.88 SD = 11.46	R: 42–112 M = 72.08 SD = 11.89	t (290) = −5.19**
Adaptive Composite	R: 48–83 M = 66.7 SD = 8.69	R: 60-107 M = 79.96 SD = 11.13	t (103) = −6.05**	R: 30–91 M = 64.31 SD = 10.16	R: 55–111 M = 75.47 SD = 10.01	t (290) = −8.26**

ID: intellectual disability (i.e., IQ < 70).
M = mean; SD = standard deviation; R = range.
Note: * $p < .05$; ** $p < .01$.

would not yield a nonsignificant result. As with individual models, differences in the three practical fit indices were included. More restricted models are considered favorable when differences in practical fit include: TLI ≤ .010, CFI ≤ .010, and a RMSEA ≤ .015 (Chen, 2007; Cheung & Rensvold, 2002). Additional details on non-statistical fit indices for measurement invariance models can be found in Reise et al. (1993).

3. Results

3.1. Participant characteristics

Table 1 displays the comparison of age, gender, IQ, and adaptive skills for children with and without ID in the younger and older age groups. Within each age group, children with and without ID varied significantly by gender, IQ, and adaptive skills.

3.2. Measurement invariance across populations

3.2.1. Exploratory results for the CBCL for ages 1.5–5

For young children, the baseline configural model was the best representation of emotional reactivity, anxious/depressed, and somatic complaints, wherein all parameters were free to vary. In regards to sleep problems, there was evidence to suggest that the remaining models fit the data worse than the metric model. A residual model was the best representation of withdrawn problems, and a structural model was the best representation for both attention problems and aggressive behavior. See Table 2 for the fit indices of both the baseline configural model and the differences tested among the subsequent nested measurement invariance models.

3.2.2. CBCL for ages 6–18

For older children with ASD, the configural baseline model was the best representation of anxious/depressed, somatic complaints, social problems, attention problems, and rule-breaking behavior. A metric model was the best representation of withdrawn/depressed, thought problems, and aggression. See Table 3 for the fit indices of both the baseline configural model and of the differences tested among the subsequent nested measurement invariance models.

4. Discussion

The purpose of this study was to test the measurement invariance of the CBCL in a sample of children with ASD with and without concurrent ID. Our sample size for younger children was quite small; therefore, our conclusions are highly exploratory and require future replication. For this sample, we found that younger children with ASD showed strong measurement invariance across those with and with ID for withdrawn problems, attention problems, and aggressive behavior. With residual invariance or higher, researchers might consider that these subscales are similar enough across the two groups to treat them as one population. Specifically, using the CBCL to measure attention problems and aggressive behavior in young children with ASD might be equally informative for those with and without ID. On the other hand, the measurement invariance for emotional reactivity, anxious/depressed, somatic

Table 2
Nested model comparisons of the CBCL subscales for children 1.5–5 across children with ASD who have or do not have concurrent intellectual disability.

Models	χ^2 ($\Delta\ \chi^2$)	df (Δ df)	p (Δ p)	TLI (Δ TLI)	CFI (Δ CFI)	RMSEA (Δ RMSEA)
Emotional Reactivity						
Configural*	119.528	54	.000	.623	.717	.107
Metric	(19.774)	(8)	(.011)	(.010)	(.051)	(−.001)
Scalar	(39.002)	(9)	(.000)	(.082)	(.129)	(−.011)
Residual	(31.530)	(9)	(.000)	(.035)	(.097)	(−.005)
Structural	(.009)	(1)	(.923)	(−.010)	(−.004)	(.002)
Anxious/Depressed						
Configural*	106.943	40	.000	.582	.702	.126
Metric	(35.462)	(7)	(.000)	(.089)	(.127)	(−.012)
Scalar	(19.650)	(8)	(.012)	(−.021)	(.052)	(.002)
Residual	(54.044)	(8)	(.000)	(.121)	(.205)	(−.015)
Structural	(4.767)	(1)	(.029)	(.005)	(.017)	(−.001)
Somatic Complaints						
Configural*	120.333	54	.000	.646	.735	.108
Metric	(47.833)	(8)	(.000)	(.139)	(.160)	(−.019)
Scalar	(4.00)	(9)	(.911)	(−.083)	(−.020)	(.011)
Residual	(54.746)	(9)	(.000)	(.119)	(.183)	(−.016)
Structural	(20.857)	(1)	(.000)	(.064)	(.079)	(−.007)
Withdrawn						
Configural	94.497	40	.000	.400	.571	.113
Metric	(11.121)	(7)	(.133)	(−.051)	(.032)	(.005)
Scalar	(9.879)	(8)	(.274)	(−.065)	(.015)	(.006)
Residual*	(8.40)	(8)	(.395)	(−.059)	(.003)	(.007)
Structural	(7.714)	(1)	(.005)	(.040)	(.053)	(−.005)
Attention Problems						
Configural	11.831	10	.297	.938	.969	.042
Metric	(2.761)	(4)	(.599)	(−.048)	(−.021)	(.022)
Scalar	(8.511)	(5)	(.130)	(.059)	(.060)	(−.025)
Residual	(2.606)	(5)	(.761)	(−.049)	(−.041)	(.019)
Structural*	(.522)	(1)	(.470)	(−.007)	(−.008)	(.004)
Aggressive Behavior						
Configural	522.026	304	.000	.770	.796	.082
Metric	(14.654)	(18)	(.686)	(−.016)	(−.003)	(.003)
Scalar	(29.475)	(19)	(.059)	(−.002)	(.010)	(.000)
Residual	(12.160)	(19)	(.879)	(−.017)	(−.006)	(.003)
Structural*	(.215)	(1)	(.643)	(−.001)	(−.001)	(.001)
Sleep Problems						
Configural	66.547	28	.000	.627	.752	.114
Metric*	(8.599)	(6)	(.197)	(−.045)	(.017)	(−.056)
Scalar	(15.764)	(7)	(.027)	(.002)	(.057)	(.000)
Residual	(10.518)	(7)	(.161)	(−.028)	(.022)	(.005)
Structural	(.535)	(1)	(.465)	(−.009)	(−.003)	(.001)

Note: Change values reflect nested model comparisons to the previous level of constraint. For example, in the Sleep Problems construct, assuming that the scalar model is correct, the change in chi-square to the residual model is 10.518.

* Denotes best fitting model.

complaints, and sleep problems did not support the use of the CBCL subscale scores. These subscales showed only configural or metric invariance, which supports the use of item-level data only when comparing children with and without ID. This means that children with and without ID have different baseline levels, different measurement error, and different levels of predictive ability for emotional reactivity, anxious/depressed, somatic complaints, and sleep problems. These findings support recommendations to use item-level data, rather than subscale-level data, for these constructs when examining samples of intellectually heterogeneous children with ASD.

Older children with ASD showed configural invariance for anxious/depressed, somatic complaints, social problems, thought problems, attention problems, and rule-breaking behavior. This means that these subscales only shared the same collection of items for those with and without ID. Withdrawn/depressed behavior and aggressive behavior showed metric invariance. This means that these subscales shared the same collection of items, and those items were equally important in the overall factor for both children with and without ID. The measurement invariance for the CBCL subscales in older children implies that researchers should use only item-level information, and not the total CBCL subscale scores, when assessing intellectually heterogeneous populations of children with ASD.

Table 3
Nested model comparisons of the CBCL subscales for children 6–18 across children with ASD who have or do not have concurrent intellectual disability.

Models	χ^2 ($\Delta\ \chi^2$)	df (Δ df)	p (Δ p)	TLI (Δ TLI)	CFI (Δ CFI)	RMSEA (Δ RMSEA)
Anxious/Depressed						
Configural*	366.642	130	.000	.701	.751	.078
Metric	(26.768)	(12)	(.008)	(−.008)	(.016)	(.001)
Scalar	(89.403)	(13)	(.000)	(.057)	(.081)	(−.007)
Residual	(226.715)	(13)	(.000)	(.182)	(.225)	(−.020)
Structural	(3.659)	(1)	(.056)	(−.001)	(.003)	(.000)
Withdrawn/Depressed						
Configural	87.639	40	.000	.800	.857	.063
Metric*	(3.242)	(7)	(.862)	(−.043)	(−.011)	(.007)
Scalar	(40.940)	(8)	(.000)	(.078)	(.099)	(−.012)
Residual	(33.663)	(8)	(.000)	(.039)	(.077)	(−.006)
Structural	(.964)	(1)	(.326)	(−.004)	(.000)	(.001)
Somatic Complaints						
Configural*	141.032	88	.000	.898	.919	.045
Metric	(44.747)	(10)	(.000)	(.049)	(.054)	(−.010)
Scalar	(26.955)	(11)	(.005)	(.009)	(.024)	(−.002)
Residual	(40.451)	(11)	(.000)	(.027)	(.045)	(−.004)
Structural	(1.708)	(1)	(.191)	(−.001)	(.001)	(.000)
Social Problems						
Configural*	188.819	88	.000	.785	.828	.062
Metric	(21.870)	(10)	(.016)	(.001)	(.010)	(.000)
Scalar	(78.161)	(11)	(.000)	(.094)	(.114)	(−.012)
Residual	(97.553)	(11)	(.000)	(.107)	(.148)	(−.012)
Structural	(.054)	(1)	(.816)	(−.005)	(−.002)	(.000)
Thought Problems						
Configural*	496.704	180	.000	.343	.437	.077
Metric	(24.126)	(14)	(.044)	(−.028)	(.018)	(.002)
Scalar	(35.511)	(15)	(.002)	(−.009)	(.037)	(.000)
Residual	(25.457)	(15)	(.044)	(−.024)	(.018)	(.002)
Structural	(.703)	(1)	(.402)	(−.003)	(.000)	(.000)
Attention Problems						
Configural*	193.875	70	.000	.764	.816	.077
Metric	(20.671)	(9)	(.014)	(−.007)	(.017)	(.001)
Scalar	(61.370)	(10)	(.000)	(.051)	(.076)	(−.008)
Residual	(6.281)	(10)	(.791)	(−.033)	(−.005)	(.005)
Structural	(1.211)	(1)	(.271)	(−.002)	(.000)	(.001)
Rule Breaking						
Configural*	175.089	48	.000	.713	.795	.106
Metric	(33.047)	(7)	(.000)	(.004)	(.039)	(−.001)
Scalar	(42.796)	(8)	(.000)	(.011)	(.053)	(−.002)
Residual	(67.828)	(8)	(.000)	(.042)	(.091)	(−.008)
Structural	(6.565)	(1)	(.010)	(.002)	(.008)	(.000)
Aggressive Behavior						
Configural	941.106	270	.000	.669	.708	.091
Metric*	(18.916)	(17)	(.333)	(−.019)	(.001)	(.002)
Scalar	(81.234)	(18)	(.000)	(.009)	(.028)	(−.001)
Residual	(57.432)	(18)	(.000)	(−.002)	(.017)	(.000)
Structural	(.988)	(1)	(.320)	(−.001)	(.000)	(.000)

Note: Change values reflect nested model comparisons of the current model to the previous level of constraint. For example, in the Aggressive Behavior construct, assuming that the scalar model is correct, the change in chi-square to the residual model is 57.432.

* Denotes best fitting model.

Overall, this study found that the measurement invariance of the CBCL across individuals with and without ID was not strong enough to warrant the use of subscale-level information in most constructs. These findings suggest that group comparisons among children with ASD are not appropriate for broad subscale-level CBCL scores. Table 4 presents the subscale measurement invariance and our recommendation for clinical use.

There are several potential reasons for the differences between ID and non-ID groups in regards to behavioral and emotional functioning. First, children with ASD and concurrent ID may exhibit different baseline levels of behavioral and emotional problems, which would explain the lack of scalar invariance in almost all of the CBCL subscales. This would be consistent with past research showing different levels of comborbid symptoms and psychopathologies in children with ID or ASD and ID (e.g., Kozlowski et al.,

Table 4
Measurement invariance of the CBCL subscales across those with and without ID.

Scale	Measurement Invariance	Recommendation
Younger Children 1.5–5		
Emotionally Reactive	Configural	Use item-level data
Anxious/Depressed	Configural	Use item-level data
Somatic Complaints	Configural	Use item-level data
Withdrawn	Residual	Use subscale-level data
Sleep Problems	Metric	Use item-level data
Attention Problems	Structural	Use subscales, with confidence
Aggressive Behavior	Structural	Use subscales, with confidence
Older Children 6-18		
Anxious/Depressed	Configural	Use item-level data
Withdrawn/Depressed	Metric	Use item-level data
Somatic Complaints	Configural	Use item-level data
Social Problems	Configural	Use item-level data
Thought Problems	Configural	Use item-level data
Attention Problems	Configural	Use item-level data
Rule-Breaking Behavior	Configural	Use item-level data
Aggressive Behavior	Metric	Use item-level data

2011; McClain et al., 2017; Tureck et al., 2014). Other research has shown that children with ASD without ID experience higher levels of anxiety (Mazurek & Kanne, 2010; Sukhodolsky et al., 2008) and depression (Mayes et al., 2011) than children with ASD and ID. Internalizing problems may be more common in this higher functioning group due to their adequate social cognition and desire for social interactions, which can cause psychological distress when combined with their social impairments (Bellini, 2004; Chamberlain, Kasari, & Rotheram-Fuller, 2007).

Second, the CBCL may be better at identifying behavioral and emotional problems in children with ASD who do not have concurrent ID, which would explain the lack of residual or structural invariance in the majority of the CBCL subscales. The CBLC subscales were developed using factor analysis of a large representative sample of "healthy" children in the general population (Achenbach & Rescorla, 2001). While the assessment may be useful for identifying children at risk for behavioral, emotional, or social problems, the tool was not designed to summarize the abilities of a heterogeneous sample that contains children with and without intellectual disabilities. This may lead to different levels of measurement error and overall predictive ability for the individual subscales in heterogeneous samples of children with ASD.

Third, parent respondents may be better able to describe the behavioral and emotional functioning of children without ID or they may rate the abilities of children with ASD and ID more harshly. This would be consistent with research on other assessments used in the ASD literature (Havdahl et al., 2015; Vermeulen et al., 2017). Difficulties with social communication may affect a child's ability to describe his/her own emotional states, thoughts, or internalizing symptoms. In addition, parents of children with lower intellectual functioning may have considerable difficulty interpreting their child's internal states, such as anxiety and depression, somatic complaints, and withdrawn or depressed behavior. However, we did not find a pattern of measurement variance specifically for subscales from the broad, internalizing disorder domain. Therefore, parents could just as easily have difficulty describing externalizing emotional and behavioral problems in their children with concurrent ID, such as emotional reactivity, sleep problems, social problems, and rule-breaking behavior. Future research should investigate the patterns of correspondence between intellectual ability and measurement invariance of the CBCL subscales and other parent-report assessments used in the ASD population.

Lastly, the differences between children with and without ID may become more pronounced over time, as a function of the different behavioral and emotional expectations placed on children and different parent expectations for their child's behavior. This may explain why there were a few subscales for younger children with structural measurement invariance, but no strong indicators of measurement invariance for older children. The groups may become more variant with age, as differences between those with and without ID become more apparent in every day social and emotional tasks. Longitudinal research is needed to examine the relationships among age, socio-emotional and behavioral problems, and parent expectations in this heterogeneous population of children with ASD.

5. Limitations and future research

Several items were not included in the subscale models because of limited response variation on these items. Constrained variance may reflect the varying set of behaviors exhibited by children with ASD, as Pandolfi et al. (2009, 2011) experienced the same issue with their samples of children with ASD. Another limitation was our sample size. Once the data were partitioned into subgroups of children with and without ID, our samples were quite small. This restricted an examination of the measurement invariance of the entire established factor structure. However, separating our examination by age group for the purposes of the CBCL versions was necessary, given the differences in number, type, and wording of items across age versions. In addition, the small sample size could also influence the restricted range of some responses, since the CBCL is a symptom count and many behaviors are (or should be) rare, creating a skewed item-level distribution. Studies with larger samples are needed to have a sufficient frequency of these rare

behaviors, and in particular, future studies with an adequate sample size are needed to replicate the exploratory findings of the younger age group.

Future research may address this limited sample size using Monte Carlo simulations to model distributional properties found in these data. Additionally, other methodologies such as Item Response Theory (IRT) may be used to determine group differences on individual items. Researchers commonly employ IRT's differential item functioning to determine if group membership affects the probability item endorsement for those with similar level of latent trait. Identifying these items may add to our understanding of CBCL subscale performance by ASD participants with and without ID. Finally, where measurement invariance is absent, factor scores, defined for each group, may be developed and scaled to existing CBCL scoring norms to preserve the utility and continuity of use by clinicians and practitioners.

6. Conclusions and implications

The CBCL is a widely-used measure of behavioral and emotional functioning in ASD research; however, its measurement properties are poorly understood when applied to this clinical population. Given that intellectual disability is a common comorbidity, affecting approximately 30% of the ASD population (Developmental Disabilities Monitoring Network, 2014), it is also important to understand whether the CBCL is equally applicable to those with and without ID. Of the three previous studies on the factor structure of the CBCL among children with ASD, none had examined the syndrome scales across subsamples of children with and without current ID. The current study suggests that the CBCL generally does not demonstrate measurement invariance, and most of the existing factor subscales may not be appropriate for use among children with ASD and concurrent ID. For clinicians and researchers using the CBCL among children with intellectually heterogeneous ASD, the current study provisionally recommends using only item-level data as opposed to subscale-level data, with the exception of attention problems and aggressive behavior in younger children. Using item-level CBCL data, rather than subscales, is also recommended for intellectually heterogeneous samples of older children with ASD. Further research is needed to establish a conclusive CBCL factor structure among children with ASD across levels of intellectual functioning.

Assessments can serve many different functions in research, diagnostic, or clinical settings, and vary in their structure and specificity in coding and scoring (Fewell & Glick, 1993). It is common to extend tests to different populations, with the knowledge that the conclusions and generalizations should be limited to clinical or exploratory settings. In practice, the CBCL should be used cautiously in populations of individuals with ASD. Our findings indicate tentatively that the CBCL may be more appropriate for use among younger children with ASD than older children with ASD, and among those without comorbid ID. However, even among young children, only the attention problems and aggressive behavior subscales demonstrated adequate performance. This greatly limits the ability of researchers or clinicians to use or interpret the CBCL with confidence in cross-time, cross-clinic, or cross-study comparisons among children with ASD.

Our previous research (Medeiros et al., 2017) found that emotional and behavioral problems may manifest differently in children with ASD than among typically developing children, and syndrome scales may not yield useful information about the underlying construct of interest (e.g., anxiety, depression). For example, elevations on the CBCL Anxious/Depressed syndrome scale for a child with ASD may not represent true anxiety at a construct-level; instead, the anxiety scale could be elevated for other reasons associated with ASD (e.g., emotion-regulation problems in young children and social-emotional deficits in older children). The clinical utility of the CBCL lies in its ability to contribute to beneficial intervention development and beneficial treatment outcome (Nelson-Gray, 2004). However, if users cannot be confident that the measure is appropriately assessing intended constructs (i.e., symptom domains), and if the measure operates differently across subpopulations (i.e., age and IQ level), the utility of the CBCL is significantly diminished for use in assessing emotional and behavioral functioning in children with ASD. These findings underscore a continued need for development and validation of measures that are appropriate for use in individuals with ASD across ages and intellectual functioning.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment.* Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.

Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2003). DSM-Oriented and empirically based approaches to constructing scales from the same item pools. *Journal of Clinical Child and Adolescent Psychology, 32*, 328–340. https://doi.org/10.1207/S15374424JCCP3203_02.

Albores-Gallo, L., Lara-Muñoz, C., Esperón-Vargas, C., Zetina, J. A. C., Soriano, A. M. P., & Colin, G. V. (2007). Validity and reability of the CBCL/6-18. Includes DSM scales. *Actas Espanolas de Psiquiatria, 35*(6), 393–399.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (fifth edition). Washington, DC: American Psychiatric Association (DSM-5).

Bellini, S. (2004). Social skill deficits and anxiety in high-functioning adolescents with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 19*(2), 78–86.
Bradley, E. A., & Isaacs, B. J. (2006). Inattention, hyperactivity, and impulsivity in teenagers with intellectual disabilities, with and without autism. *The Canadian Journal of Psychiatry, 51*, 598–606.
Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466.
Chamberlain, B., Kasari, C., & Rotheram-Fuller, E. (2007). Involvement or isolation? The social networks of children with autism in regular classrooms. *Journal of Autism and Developmental Disorders, 37*(2), 230–242.
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504.
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.
De Groot, A., Koot, H. M., & Verhulst, F. C. (1994). Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychological Assessment, 6*(3), 225–230.
Dedrick, R. F., Greenbaum, P. E., Friedman, R. M., Wetherington, C. M., & Knoff, H. M. (1997). Testing the structure of the Child Behavior Checklist/4-18 using confirmatory factor analysis. *Educational and Psychological Measurement, 57*(2), 306–313. https://doi.org/10.1177/0013164497057002009.
Dedrick, R. F., Tan, T. X., & Marfo, K. (2008). Factor structure of the Child Behavior Checklist/6-18 in a sample of girls adopted from China. *Psychological Assessment, 20*(1), 70–75.
Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, & Centers for Disease Control and Prevention (CDC) (2014). *Prevalence of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010. Morbidity and Mortality Weekly Report.SUrveillance Summaries*(Washington, D.C.: 2002), 63(2)1–21.
Elliot, C. D. (2007). *Differential ability scales* (second edition). San Antonio, TX: PsychCorp Harcourt Assessment Inc.
Fewell, R. R., & Glick, M. P. (1993). Observing play: An appropriate process for learning and play assessment. *Infants and Young Children, 5*(4), 35–43.
Frizzo, G. B., Pedrini, J. R., de Souza, D. S., Bandeira, D. R., & Borsa, J. C. (2014). Reliability of child behavior checklist and teacher's report form in a sample of brazilian children. *Universitas Psychologica, 14*(1), 149–156. https://doi.org/10.11144/Javeriana.upsy14-1.rcbc.
Gonzalez, M., & Stern, K. (2016). Co-occurring behavioral difficulties in children with severe feeding problems: A descriptive study. *Research in Developmental Disabilities, 58*, 45–54.
Greenbaum, P. E., & Dedrick, R. F. (1998). Hierarchical confirmatory factor analysis of the Child Behavior Checklist/4-18. *Psychological Assessment, 10*(2), 149–155.
Hallerod, S. L., Larson, T., Ståhlberg, O., Carlström, E., Gillberg, C., Anckarsäter, H., et al. (2010). The Autism- Tics, AD/HD and other comorbidities (A-TAC) telephone interview: Convergence with the Child Behavior Checklist (CBCL). *Nordic Journal of Psychoatry, 64*(3), 218–224.
Havdahl, K. A., Hus Bal, V., Huerta, M., Pickles, A., Oyen, A. S., Stoltenberg, C., et al. (2015). Multidimensional influences on autism symptom measures: Implications for use in etiological research. *Journal of the American Academy of Child and Adolescent Psychiatry, 55*, 1054–1063.
Ivanova, M. Y., Achenbach, T. M., Resorla, L. A., Dumenci, L., Almqvist, F., Bilenverg, N., et al. (2007). The generalizability of the Youth Self-Report syndrome structure in 23 societies. *Journal of Consulting and Clinical Psychology, 75*, 729–738.
Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Harder, V. S., Ang, R. P., Bilenberg, N., et al. (2010). Preschool psychopathology reported by parents in 23 societies: Testing the seven-syndrome model of the child behavior checklist for ages 1.5-5. *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(12), 1215–1224.
Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.
Kaat, A. J., Lecavalier, L., & Aman, M. G. (2014). Validity of the Aberrant Behavior Checklist in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 44*, 1103–1116.
Kim, S. H., Joseph, R. M., Frazier, J. A., O'Shea, T. M., Chawarska, K., Allred, E. N., et al. (2016). Predictive validity of the Modified Checklist for Autism in Toddlers (M-CHAT) born very preterm. *The Journal of Pediatrics, 178*, 101–107.
Kozlowski, A. M., Matson, J. L., Sipes, M., Hattier, M. A., & Bamburg, J. W. (2011). The relationship between psychopathology symptom clusters and the presence of comorbid psychopathology in individuals with severe and profound intellectual disability. *Research in Developmental Disabilities, 32*, 1610–1614.
Lempp, T., de Lange, D., Radeloff, D., & Bachmann, C. (2012). The clinical examination of children, adolescents and their families. In J. M. Rey (Ed.). *IACAPAP e-textbook of child and adolescent mental health*. Geneva: International Association for Child and Adolescent Psychiatry and Allied Professions.
Long, J. S. (1983). *Confirmatory factor analysis.* Beverly Hills, CA: Sage.
LoVullo, S. V., & Matson, J. L. (2009). Comorbid psychopathology in adults with autism spectrum disorders and intellectual disabilities. *Research in Developmental Disabilities, 30*, 1288–1296.
Mannion, A., & Leader, G. (2013). Comorbidity in autism spectrum disorder: A literature review. *Research in Autism Spectrum Disorders, 7*(12), 1595–1616.
Matson, J. L., & Cervantes, P. E. (2014). Commonly studied comorbid psychopathologies among persons with autism spectrum disorder. *Research in Developmental Disabilities, 35*(5), 952–962.
Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabities, 30*(1107-), 1114.
Matson, J. L., Cooper, C., Malone, C. J., & Moskow, S. L. (2008). The relationship of self-injurious behavior and other maladaptive behaviors among individuals with severe and profound intellectual disability. *Research in Developental Disabilities, 29*, 141–148.
Matson, J. L., Hess, J., & Boisjoli, J. A. (2010). Comorbid psychopathology in infants and toddlers with autism and pervasive developmental disorders-not otherwise specified (PDD-NOS). *Research in Autism Spectrum Disorders, 4*, 300–304.
Mayes, S. D., Calhoun, S. L., Murray, M. J., Ahuja, M., & Smith, L. A. (2011). Anxiety, depression and irritability in children with autism relative to other neuropsychiatric disorders and typical development. *Research in Autism Spectrum Disorders, 5*, 474–485.
Mazurek, M. O., & Kanne, S. (2010). Friendship and internalizing symptoms among children and adolescents with ASD. *Journal of Autism and Developmental Disorders, 40*(12), 1512–1520.
McClain, M. B., Hasty Mills, A. M., & Murphy, L. E. (2017). Inattention and hyperactivity/impulsivity among children with attention-deficit/hyperactivity-disorder, autism spectrum disorder, and intellectual disability. *Research in Developmental Disabilities, 70*, 175–184.
Medeiros, K., Mazurek, M. O., & Kanne, S. (2017). Investigating the factor structure of the CBCL in a large sample of children with ASD. *Research in Autism Spectrum Disorders, 40*, 24–40. https://doi.org/10.1016/j.rasd.2017.06.001.
Mullen, E. M. (1995). *Mullen scales of early learningCircle pines* (ags ed.). MN: American Guidance Service Inc.
Nakamura, B. J., Ebesutani, C., Bernstein, A., & Chorpita, B. F. (2009). A psychometric analysis of the child behavior checklist DSM-oriented scales. *Journal of Psychopathology and Behavioral Assessment, 31*, 178–189. https://doi.org/10.1007/s10862-008-9119-8.
Nelson-Gray, R. (2004). Treatment utility of psychological assessment. *Psychological Assessment, 15*, 521–531.
Pandolfi, V., Magyar, C., & Dill, C. (2009). 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, 39*(7), 986–995.
Pandolfi, V., Magyar, C. I., & Dill, C. A. (2011). An initial psychometric evaluation of the CBCL 6-18 in a sample of youth with autism spectrum disorders. *Research in Autism Spectrum Disorders, 6*, 96–108. https://doi.org/10.1016/j.rasd.2011.03.009.
Reise, P. S., Widaman, K., & Pugh, H. R. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552–566. https://doi.org/10.1037/0033-2909.114.3.552.
Roid, G. H. (2003). *Stanford-Binet intelligence scalesTechnical manual* (fifth edition). Itasca, IL: Riverside Publishing.
Simonoff, E. M., Pickles, A. P., Charman, T. P., Chandler, S. P., Loucas, T. O. M. P., & Baird, G. F. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*(8), 921–929.
Smith, K. R. M., & Matson, J. L. (2010). Psychopathology: Differences among adults with intellectual disability, and comorbid autism spectrum disorders and epilepsy.

Research in Developmental Disabilities, 31, 743–749.
Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior Scales-II.* Circle Pines, MN: American Guidance Service.
Sukhodolsky, D., Scahill, L., Gadow, K., Arnold, L., Aman, M., McDougle, C., et al. (2008). Parent-rated anxiety symptoms in children with pervasive developmental disorders: Frequency and association with core autism symptoms and cognitive functioning. *Journal of Abnormal Child Psychology, 36*(1), 117–128.
Tureck, K., Matson, J. L., Cervantes, P. E., & Konst, M. (2014). An examination of the relationship between autism spectrum disorder, intellectual functioning and comorbid symptoms in children. *Research in Developmental Disabilities, 35*, 1766–1772.
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002.
Vermeulen, K., Egger, J. I. M., Janzing, J. G. E., van Dongen, L., van Bokhoven, H., Kleefstra, T., et al. (2017). The context of symptom measures: Interpretation and clinical diagnosis of autism spectrum disorders in intellectual disabilities. *Journal of the American Academy of Child and Adolescent Psychiatry, 56*, 618–619.
Wechsler, D. (2002). *The wechsler preschool and primary scale of intelligence* (third edition). San Antonio, TX: The Psychological Corporation.
Wechsler, D. (2003). *Wechsler intelligence scale for children–Fourth edition (WISC-IV).* San Antonio, TX: NCS Pearson.
Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence–Second edition (WASI-II).* San Antonio, TX: NCS Pearson.
Wechsler, D. (2012). *Wechsler preschool and primary scale of intelligence* (4th ed.). Bloomington, MN: Pearson.
Wu, H., & Estabrook, C. R. (2016). Identification of CFA models of different levels of invariance for ordered categorical outcomes. *Psychometrika, 81*(4), 1014–1045.