# Comparison of the Factor Structure of the Child Behavior Checklist 1.5–5 between Children with ASD and Children with DD


Yi-Ling Cheng[a] Lai-Sang Iao[b] Chin-Chin Wu[c]*

[a]Department of Psychology, Kaohsiung Medical University, Kaohsiung, Taiwan

[b]Department of Psychology, Nottingham Trent University, Nottingham, United Kingdom

[c]Department of Psychology, Kaohsiung Medical University; Department of Medical Research, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan


Running Head: CBCL and ASD


**Keywords:** Autism Spectrum Disorder, Developmental Delay, CBCL1.5–5, measurement invariance



Chin-Chin Wu (Corresponding author), Kaohsiung Medical University;

Kaohsiung Medical University Hospital, e-mai:jinnchin@mail2000.com.tw

Conflict of interest: The authors declare that they have no conflict of interest.

# Highlights

- We tested measurement invariance to compare children with ASD and children with DD on each subscale of the CBCL 1.5-5.
- All of the subscales achieved basic level of invariance, suggesting similar factor structures across these two groups.
- Withdrawn, Aggressive behavior, and Sleep Problems did not achieve metric invariance, suggesting the relations between items and latent constructs are not similar across groups.
- Six out of seven subscales did not achieve scalar invariance, suggesting further group comparisons will not be ideal.

# Comparison of the Factor Structure of the Child Behavior Checklist 1.5–5 between Children with ASD and Children with DD

**Abstract**

**Background**: The Child Behavior Checklist 1.5–5 (CBCL 1.5–5) has been applied to identify emotional and behavioral problems on children with autism spectrum disorder (ASD). However, few studies explored whether the established factor model may be suitable for children with ASD and those with developmental delay (DD).

**Method**: To locate the potential sources of variations between these two groups, we tested measurement invariance multiple groups factor analysis.

**Results**: All subscales achieved the basic level of invariance (configural invariance). The findings suggested similar factor structures across these two groups. However, Withdrawn, Aggressive Behavior, and Sleep Problems did not achieve metric invariance. The findings suggested the relations between items and latent constructs are not similar across groups in these three scales.

**Conclusions**: Overall, there are different levels of invariances across subscales of the CBCL1.5–5. The attempt of using the CBCL1.5–5 to separate the profile of children with ASD and children with DD might be helpful, but only on particular aspects.

**Keywords:** autism spectrum disorder, developmental delay, CBCL1.5–5, measurement invariance

## 1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impaired social interaction and communication and repetitive behaviors or restricted interests (American Psychiatric Association [APA], 2013). Many individuals with ASD exhibit co-occurring emotional and behavioral problems (EBPs; Bauminger et al., 2010; Gau et al., 2010; Hou et al., 2018; Lindsey et al., 2020). Previous studies showed that EBPs were associated with cognitive ability (Falk et al., 2014; O'Brien & Pearson, 2004), autism symptoms (Jang et al., 2011; Lindsey et al., 2020), parenting stress and parents' mental health (Giovagnoli et al., 2015; Hou et al., 2018; Kim et al., 2016; Schiltz et al., 2018). Thus, it is important to measure EBPs of children with ASD. It is helpful for early intervention of children with ASD and their family.

Measuring EBPs could also be helpful for early diagnosis of ASD. In two decades, many studies examined the utilities of EBPs measures in detecting ASD in children. Compared to ASD-specific screening tools, an advantage of the EBPs measure relates to the fact that it assesses a wide range of EBPs rather than ASD in specific. Parents' responses are less likely to be biased depending on whether they believe their child have an ASD (Iao et al., 2020). Using the Developmental Social Disorders (DSD) scale of the Behavior Assessment System for Children-Second Edition (BASC-2), Bradstreet et al. (2017) recruited 224 toddlers and preschooler (age range: 24–63 months), including 117 children with ASD, 55 children with other diagnosis (e.g., developmental delay [DD]) and 52 children without diagnosis. Their findings showed that the sensitivity was .76 and specificity was .73 for distinguishing children with ASD from those without diagnosis. In addition, the sensitivity was .62 and specificity was .63 for distinguishing children with ASD from those with other diagnosis. Their findings indicated that the DSD scale could be used as a Level 1 ASD screening tool which is used to identify high risk cases in general populations.

Among the many used EBPs measures, the Child Behavior Checklist for Ages 1.5–5 (CBCL 1.5–5; Achenbach & Rescorla, 2000) is one of the most widely-used and well-studied in early childhood. The use of the CBCL 1.5–5 is well-studied in English-speaking populations and has been translated into different language versions (Ivanova et al., 2010). Several studies have shown that the CBCL 1.5–5 was a promising ASD screening tool in Italy (Muratori et al., 2011), Romania

(Predescu et al., 2013), Korea (Rescorla et al., 2015) and Germany (Limberg et al., 2017). In Taiwan, using the Withdrawn scale of the CBCL 1.5–5, Iao et al. (2020) recruited clinically referred toddlers and preschoolers (age range: 18–47 months), including 66 children with ASD and 68 children with DD. Their findings showed that the sensitivity was .74 and specificity was .77. Overall, their results showed that the Withdrawn scale can be used as a Level 2 ASD screening tool to identify ASD in high risk populations (i.e., children with DD).

The CBCL 1.5–5 was developed through a series of factor analyses including both exploratory factor analysis and confirmatory factor analysis (Achenbach & Rescorla, 2000). It was found that the best factor structure of the CBCL1.5–5 was a seven syndromes model. These seven syndromes are Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, and Aggressive Behavior. Accordingly, the first four subscales belong to the internalizing scale and the last two belong to the externalizing scale. The Sleep Problem scale, however, does not contribute to either the internalizing scale or the externalizing scale but it does contribute to the total score. The validity of the CBCL 1.5–5 was demonstrated over the years with USA samples (e.g., Achenbach & Rescorla, 2000; Basten et al., 2016; Kristensen et al., 2010). The invariance of CBCL 1.5–5 with other cultures has been validated. Studies showed factor invariance when assessing typically developing children's EBPs across different cultures (e.g., Ivanova et al., 2010; Rescorla et al., 2020; Tan et al., 2007). For example, Ivanova et al. (2010) validated the structures of CBCL with the data from 23 societies. Rescorla et al. (2020) also found similar results when examining the CBCL DSM-ASD scale with children across 24 societies. Because these societies included both Eastern and Western countries, their results indicated that the CBCL 1.5–5 can be an effective tool when assessing children's EBPs across a wide range of cultures.

The validation of the factor invariances across different cultures has warranted the application of CBCL for other cultures. However, using CBCL in one culture setting to differentiate children's mental disorders has not been fully investigated. There are some studies explored the factor structures of CBCL between typically developing children and children with ASD. For example, Pandolfi et al. (2009) have found partial evidence suggesting that the factor structure of the CBCL

1.5–5 remained to be similar between typically developing children and children with ASD. Specifically, while they found that most items (92%) loaded significantly, several items appeared to have low loadings. For instance, one item (acts too young) in Withdrawn scale was not significant. The construct representations of CBCL can be different with the types of developmental problems. To the best of the authors' knowledge, there has been only one study investigating the factor structure (factor invariance) of the CBCL 1.5–5 by comparing children with ASD on different IQ levels (see Dovgan et al., 2019) but less was explored with other developmental problems. Some studies investigated children with ASD and without ASD using observed scores of the CBCL (Chericoni et al., 2021; Narzisi et al., 2013; Sikora et al, 2008). For example, Sikora et al. (2008) compared the screening performance of Gilliam Autism Rating Scale (GARS) and CBCL for ASD. They found that the Withdrawn scale and DSM-oriented scales: Pervasive Developmental Problem scale showed significant differences on these two groups with children aged 36-71 months. Similar findings were discovered with younger children. Narzisi et al. (2013) found that these two scales of CBCL had high sensitivity and specificity among children with ASD, children with other mental disorders, and typically developing children aged 18-36 months. Same as Chericoni et al. (2021) discovered that these two scales could serve as significant predictors to differentiate between children with and without ASD around 18 months. In fact, a meta-analysis conducted by Hampton and Strand (2015) confirmed the same finding. Among these studies, only a few have compared children with ASD and children with DD, and these studies focused on the use of observed scores. For instance, Predescu et al.'s (2013) study showed that only Pervasive Developmental Problem scale can differentiate between children with ASD and children with DD (sensitivity = 67.96, specificity =67.65). Another study also found that the Withdrawn scale could differentiate between children with ASD and children with DD (i.e., Iao et al., 2020). These studies showed initial promising results on the analysis using observed scores, but the differences on latent factor structures of the CBCL were not further investigated. Understanding the differences of latent factor structures between children with ASD and children with DD would help researchers identify the possible sources of discrepancy between these two neurodevelopmental disorders, separate the diagnostic profiles in clinic uses, enable parents and teachers recognizing the possible EBPs in early stages, and develop further diagnosis and interventions for children with ASD.

## 2. Methods

### *2.1. Participants*

Children were recruited through a teaching hospital in a city of south Taiwan. There were 378 children with developmental problems in the study. Children were diagnosed by a group of senior psychologists and psychiatrists based on their developmental histories, the current concerns of their parents, the results of tests that measure cognitive and adaptive functioning, the clinical observations of the child, and the results of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999). Based on their diagnoses, children were then divided into two groups. The first group was 192 (Female = 24) children that were being diagnosed with ASD. Their average age was 32.91 months (*SD* = 9.72). These children were diagnosed in accordance with the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; APA, 2013) criteria. The second group was 186 (Female = 59) children with DD. The diagnostic result was from a combined judgement that children failed to reach a total score of 85 in the MSEL (Mullen, 1995) and a T score of 35 on any cognitive scales (i.e., visual reception, fine motor, receptive language, and expressive language) of the MSEL. In addition, these children did not meet the DSM-5 criteria for ASD. Their average age was 31.25 months (*SD* = 10.59).

### *2.2. Procedure and Measures*

The children's parents were asked to complete the CBCL 1.5–5 (Achenbach & Rescorla, 2000) which is a standardized measure of behavioral and emotional problems in children around the world. The version we used here was the Chinese version that was translated from English version. The Chinese version has proper psychometric property. For example, the internal consistency is usually above .70 in various samples (see a review from Leung & Wong, 2003). The test-retest reliability of the CBCL 1.5–5 in a preschool sample in Taiwan were .52–.84 (Wu et al., 2012). The Chinese version has the same number of items and scoring system (0 = not true, 1 = somewhat or sometimes true, and 2 = very true or often true) as the English version of the CBCL 1.5–5. There were 99 items in the CBCL 1.5–5 and they can be divided into 7 subscales.

The MSEL (Mullen, 1995) was used to measure overall mental age of each participant. The MSEL is a standardized comprehensive developmental test that was designed for assessing preschool children aged 0 to 68 months. It consists of four cognitive scales: visual reception, fine motor, receptive language, and expressive language. The four cognitive scales yield T-scores, which have a mean of 50. The four subscale scores can be used to compute a composite score, which is an indicator of early learning and has a mean of 100. An overall mental age is computed for each child by averaging the age equivalents from these four cognitive scales.

The ADOS (Lord et al., 1999) is a semi-structured play-based and observational tool that consists of four modules, each of which is selected and administered based on child's age and expressive language. It is part of the gold-standard toolbox (along with Autism Diagnostic Interview-Revised, ADI-R) for ASD diagnosis because it serves as a standardized means by which communication, reciprocal social interaction, and stereotypic behaviors and restricted interests could be observed and scored. Each module provides an algorithm that entails cutoffs that can be used to assign examinee to one of the following three categories: autism, autism spectrum (i.e., pervasive developmental disorder-not otherwise specified; PDD-NOS), or non-ASD.

*2.3. Data Analysis Plan*

Descriptive statistics was conducted first to understand the characteristics of the sample. The reliability estimates with our current samples were also provided. While the CBCL 1.5–5 is a standardized measure with established reliability and validity, the original purpose of the test was not for assessing children with developmental problems. Even though the test has been used for measuring EBPs, we followed the advice of the standards for educational and psychological testing (American Educational Research Association [AERA] et al., 2014) to conduct our own reliability analysis. Therefore, we established our reliability estimates for each subscale from the current data. An analysis of reliability, i.e., Cronbach's alpha and greatest lower bound (glb) (Jackson & Agunwamba, 1977), on each subscale was conducted respectively. While Cronbach's alpha is being used prevalent across disciplines of psychology, it has been demonstrated that alpha can be misleading sometimes. Therefore, alternative reliability estimate: glb is provided here as well. Both alpha and glb are estimated values, but the calculation of glb gives us an interval about the location of true reliability. Specifically, the true reliability is located on a point between the value of glb and 1 (see the details in Sijtsma, 2009). The glb was generated from JASP 0.14.1 statistical software (Wagenmakers et al., 2018).

*2.4 Measurement Invariance*

Group comparisons in ASD research have been applying measurement invariance modeling to identify the level of group differences (Dovgan et al., 2019; Murray et al., 2014; Rescorla et al., 2019). In general, measurement invariance is used to examine whether the same construct was measured in different groups. During the process of the comparisons, this method also could be used to show the level of differences between groups, such as the comparison of ASD children with different IQ levels (Dovgan et al., 2019) and across different cultures (Rescorla et al., 2019). In addition, when understanding whether a diagnostic measurement assessed the same underlying constructs across different groups (e.g., children with ASD vs. typically developing children), measurement invariance is a very common method being used across ASD diagnostic scales to identify the sources of variations between two groups. For example, Murray et al. (2014) tested measurement invariance of the Autism Quotient Short Form in adult with and without ASD. Glod et al. (2017) also examined measurement invariance of the parent version of the Spence Children's Anxiety Scale-Parent version (SCAS-P).

The first step of assessing the factor structure of CBCL 1.5–5 was to check the dimensionality of each subscale. One factor model was applied to each subscale for both groups respectively. Because each item in CBCL scale ranges from 0 to 2, the distribution is non-normal and specific estimators (maximum likelihood parameter estimates with standard errors, MLR or mean-and variance-adjusted weighted least squares estimator WLSMV) with robust estimation should be used. MLR and WLSMV are both estimators that can deal with non-normal data. They are often being compared and powers of estimations are slightly different in simulation studies (e.g., Li, 2016; Sass et al., 2014). Because item 93 in Somatic Complaints did not have response" 2" in DD group, and parallel responses between groups is necessary for the specification of ordinal items with WLSMV estimation, we then chose MLR estimator for measurement invariance. In addition, we considered Xia and Yang's (2019) suggestion that using diagonally weighted least squares (DWLS) estimators (e.g., WLSMV) with conventional cut off values might lead to the possibility of not detecting misfits. Because of these two reasons, we therefore used MLR to help us deal with this issue.

To examine the psychometric equivalence of the factor structure between two samples, measurement invariance is a standard method that is being applied across disciplines (Vandenberg & Lance, 2000). To identify the source of variations, measurement invariance progresses a series of steps on comparing and contrasting critical features on latent factor structures between groups. Because multiple groups are involved in this analysis, measurement invariance sometimes is also called multiple group comparisons of latent variable analysis (Sass, 2011). Studies that applied measurement invariance have different opinions on the level of invariances that needs to be achieved for the constructs to be considered as “measurement invariance” (Vandenberg & Lance, 2000). In general, in order to claim the construct is invariant across groups there are three major invariances that need to be established.

The first step is configural invariance. Configural invariance tests whether the overall factor structure is similar between samples. The second step is metric invariance. Metric invariance tests whether the factor loading is identical (being constrained to be the same in the model) between samples. The third step is scalar invariance. Scalar invariance tests whether the intercepts, as known as the initial status of the ability, can be identified to be the same (being constrained to be the same in the model) across groups (Cheung & Rensvold, 2002). Overall, measurement invariance is a set of latent variable model comparisons progressing from a relatively loose model to a relatively stricter model. The comparisons would stop when the comparison of the model fits suggested that the stricter model is a worse model than the relative loose ones. For example, if metric invariance is a worse model compared to the configural model, then scalar invariance would not be carried out as the next step. In a nutshell, measurement invariance investigates whether there are differences on the representations of the constructs across diverse samples and where these differences are. It is a particularly useful approach to examine the source of group differences on a measurement, if there is any. The analysis of one factor model and measurement invariance models were both conducted with R Lavaan package (Rosseel, 2012). Lavaan is a statistical package for structural equation modeling that was built under R environment. The functions of Lavaan are often being evaluated and compared with Mplus (Muthén & Muthén, 2015), which is a commercial and powerful statistical analysis tool for structural equation modeling. Because R software is free, whether the software is up to date is critical. As to the year of 2020, Lavaan still holds its validity in the evaluation of Structural Equation Modeling software (see a review from Svetina et al., 2020).

Fit indexes: There are several model fit indicators that can be used to identify the fit of the models. One type of model fit is absolute model fit that we use to identify whether one factor model is a good fit for each subscale. Absolute model fit is examining how the model fits with each group respectively. For that purpose, we use the value of the Root Mean Square Error of Approximation (RMSEA) and Comparative fit index (CFI) within the model. The value of RMSEA needs to be below .08 to be considered as a moderate fit, or below .05 to be a good fit. The value of CFI needs to be over .95 to be considered as good fits (Hu & Bentler, 1999; Iacobucci, 2010).

A second type of model fits is relative model fit. We used these indexes to compare model fits between measurement invariance models. For the relative model fits for comparing the fits between different models, we mainly used chi-square difference test first and the alternative fit indices, such as differences on CFI, standardized root mean square residual (SRMR), and RMSEA between models second (Chen, 2007; Cheung & Rensvold, 2002). Meade et al. (2008, p.590) suggested to consider, as such, when chi-square tests were significant, if alternative indices were not over the range with sample size greater than 200, it is still invariant models. This is because big sample size tends to lead to the significance of chi-square difference test (Raykov & Marcoulides, 2006). Because our sample size per group is under 200, we used the significant $p$ value ($\leq .05$) on chi-square different test as the primary criteria. If the chi-square test was significant, then the model comparison was ruled as non-invariant. If the chi-square test was NOT significant, then we check these three fit indices to make sure the invariance holds. If two out of three alternative fit indices (CFI, RMSEA, SRMR) are over the criteria, then we still considered it is non-invariant, even the chi-square was not significant. Furthermore, the recommendation of alternative fit indices values differs by total sample sizes in Chen (2007). Because our total sample size is over 300, according to Chen (2007), the significant *p* value (≤ .05) on chi-square test and/ or the changed value on CFI is ≥ -.010, RMSEA is ≥ .015, and SRMR ≥ .030 would suggest non-invariance. In addition, CFI ≥ -.010, RMSEA is ≥ .015, and SRMR ≥ .010 would suggest non-invariance for intercept or residual invariance (Chen, 2007).

In addition, a few papers have suggested the approach of doing partial invariance after full invariance was not able to achieve (Gregorich, 2006; Millsap & Kwok, 2004; Vandenberg & Lance, 2000). Partial invariance is an additional test to relax the problematic parameters and keep other parameters invariant to improve the model fit. Although the choice of parameters to be relaxed can be generated from the suggestion of modification index in most softwares, such a choice is data driven and could, sometimes, potentially result in type 1 errors as well (see Millsap & Kwok, 2004's discussion, p.94). As there are very few papers that had compared between ASD and DD currently, we do not have clear clues about these choices. Therefore, we chose to proceed with the full invariance models only.

## 3. Result

### *3.1 Descriptive Statistics*

The demographic comparisons of these two groups are showed in Table 1. Between children with ASD and children with DD, there were significant differences on mental age (DD group was older), parents' education (ASD group had longer years of education), and ADOS scores. However, after Bonferroni correction (.05/17 = .002, because we carried out multiple *t*-test comparisons on seven background variables, seven subscales, and three overall scales), only Father's year of education and ADOS scores remained significant (children with ASD were higher on both). Children with ASD also have more male participants than children with DD ($p$ <.001). However, independent sample *t* tests confirmed that there were no significant gender differences within either ASD or DD group on each subscale score.

Insert Table 1

The one factor model fit for each subscale for these two groups is presented on Table 2 and 3. All the model fits were mediocre under the critical values of fit indexes. This is expected, as we used stricter models (e.g., MLR) to fit our data. However, CFA models here were analyzed by group separately. We did not have the parallel item issues for this set of analysis. To help readers with the comparisons of previous studies, we provided both MLR estimations (table 2) and WLSMV estimations (table 3) here. These model fits had similar patterns as Pandolfi et al. (2009, they used WLSMV estimation) that Attention Problem fits best and Sleep Problems fit worst. Overall, the result indicated that most subscales were unidimensional and the items in each subscale mostly measured one latent trait.

Insert Table 2 and Table 3

We conducted Cronbach's alpha and glb as the reliability estimates among the seven subscales of the CBCL 1.5–5. The Cronbach's alpha of each subscale was as follows: Emotionally Reactivity ($.73_{ASD}, .74_{DD}$), Anxious/Depressed ($.76_{ASD}, .76_{DD}$), Somatic Complaints ($.68_{ASD}, .57_{DD}$), Withdrawn ($.78_{ASD}, .73_{DD}$), Sleep Problems ($.76_{ASD}, .76_{DD}$), Attention Problems ($.59_{ASD}, .62_{DD}$), and Aggressive Behavior ($.87_{ASD}, .91_{DD}$). There were similar levels of internal consistencies between the two groups on most of the scales. However, some of the subscales have low alphas. For example, the Attention Problems on both groups showed lower alphas. The values of glb for each subscale was as follows: Emotionally Reactivity ($.82_{ASD}, .84_{DD}$), Anxious/Depressed ($.84_{ASD}, .85_{DD}$), Somatic Complaints ($.79_{ASD}, .75_{DD}$), Withdrawn ($.86_{ASD}, .78_{DD}$), Sleep Problems ($.84_{ASD}, .83_{DD}$), Attention Problems ($.66_{ASD}, .75_{DD}$), and Aggressive Behavior ($.94_{ASD}, .95_{DD}$). Attention Problems on both groups again showed lower values as well.

*3.2 Measurement Invariance*

The model comparisons are presented in Table 4. First of all, in terms of configural invariance, all of the subscales achieved this basic level of invariance. This suggested that each subscale has similar factor structures across the two groups. However, on the next step, only Emotionally Reactive, Anxious/Depressed, and Attention Problems achieved metric invariance, and Withdrawn, Aggressive Behavior, and Sleep Problems did not. This suggested that the relations between items and latent constructs were not similar across groups (Cheung & Rensvold, 2002) on Withdrawn, Aggressive Behavior, and Sleep Problems. Last, Somatic Complaints was the only scale that achieved scalar invariance across groups, but the model fit indexes: CFI:.69 and RMSEA:.110 were both under the critical values.

Insert Table 4

However, our sample sizes of each group in the current study were slightly on the low side for latent variable analysis. Previous studies (e.g., Dovgan et al., 2019) proposed that data simulation (e.g., bootstrapping) can deal with this issue so we used bootstrapping method on all subscales. We used Bollen-Stine bootstrapping method (Bollen & Stine, 1992) in Lavaan package (Rosseel, 2012) to compare the test statistics (i.e., Chi-square value) between the original sample and 100 resampling datasets. Specifically, we compared chi-square values for model fit index between these two samples. If none of the resampling dataset has a chi-square value exceeding the chi-square value of the original sample, this would suggest none of the resampling dataset has failed to fit and thus the original model fit is unbiased. We found that our result was unbiased (being significant below $p$ = .05) on Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, and Aggressive Behavior. However, there were some bias on Attention Problems (20% resamples had ill model fits).

Our measurement invariance analysis did not achieve scalar invariance in most subscales (six out of seven subscales). Based on the recommendations from previous reviews (e.g., Putnick &

Bornstein, 2016), further group comparisons would be biased and should not be carried out. The means and standard deviations of subscales for each group can be found on Table 5.

Insert Table 5

## 4. Discussion

The aim of the current study was to identify the source of variations when using the CBCL 1.5–5 to differentiate children with ASD from children with DD. For that purpose, we examined and found that the levels of invariance were varied across different subscales of the CBCL 1.5–5. In the one factor (unidimensional analysis of each subscale), we found that Attention Problems showed low reliabilities, low median factor loadings, and 20% bias in analysis in both of our ASD and DD group. One possible explanation is that there might be multiple factors in Attention Problems with ASD and DD. Low factor loading suggested that the items in this subscale might potentially not reflect this particular construct well. It is possible that it might measure more than one construct.

There are three subscales, Withdrawn, Aggressive Behavior, and Sleep Problems, that did not achieve metric invariance, suggesting that these three subscales may have similar factor structure but the relations between items and constructs varied between the samples. In terms of the variation on the Withdrawn scale, previous studies using other type of analyses have also found differences between children with ASD and children with DD. For example, Iao et al. (2020) found that there were score differences on Internalizing, Anxious/Depressed, Withdrawn, and Attention Problems subscales. Similarly, another study also found that Korean children with ASD scored significantly higher than typical children and children with DD (e.g., Rescorla et al., 2015) on the Withdrawn subscale. However, the results of the current study suggested that the scale representations of Withdrawn maybe different between children with ASD and children with DD. The application of Withdrawn might not be appropriate when screening children with subscale level.

Regarding to the variation on the Aggressive Behavior subscale, previous studies have showed that identifying certain Aggressive Behaviors can differentiate children with ASD from typically developing children (e.g., Mazefsky et al., 2011), especially when they are in school ages. Yet few studies explored whether this subscale can further differentiate between children with ASD and children with DD, especially when children are younger. Our finding showed that the Aggressive Behavior subscale can potentially have different scale representations between these two groups in the perspective of latent constructs. Thus, using Aggressive Behaviors for differentiating children with ASD from those with DD need to be cautious. In addition, the findings of this study suggested that the model fits of the Sleep Problems subscale on both groups were not ideal, the values were similar to what Pandolfi et al. (2009) found in their study. The variation on the Sleep Problems subscale also suggested that children with ASD and children with DD might have different sleep problems. In fact, previous studies have indicated that children with ASD, children with DD are significantly different on their sleep problems (see Reynolds et al., 2019). As mentioned earlier, the results of metric invariance model reflected the sensitivity on the equal factor loadings, particularly on how each item related to the latent constructs. The lack of metric invariance here on both subscales might suggest that even when in the condition that the total (raw) scores were the same, these total (raw) scores might be added up from different items in each group respectively. Overall, the results suggested that the ASD group and the DD group had variations on the Aggressive Behaviors and Sleep Problems subscale but the result of Sleep Problems might not be trustworthy as its CFA model fit was the worst.

Three subscales, Emotional Reactivity, Anxious/Depressed, and Attention Problems, have achieved metric invariance. This implied that the underlying relationships between items and constructs were the same across groups. In other words, these three constructs had similar components in both groups. However, these subscales did not achieve scalar invariance either. Our results demonstrated that group comparison of these subscale scores could potentially be problematic. For example, in terms of attention problems, while findings from previous studies were scarce, Dawson et al. (2002) found that young children with ASD (age 3 to 4 years old) did not perform differently on executive function tasks when compared to children with DD and to those with typical development. Their findings illustrated that children with ASD and those with DD may have similar attention problems, but it is worth noting that previous studies (e.g., Naizisi et al., 2013) also showed that children with ASD scored significantly higher on all subscales of the CBCL when compared to children with typical development. All evidence suggests that comparisons like these often lead to conflicting results. Perhaps it was because they were not comparable in the first place. We also found that Attention Problems subscale did show some biased results in our bootstrapping simulation when checking sampling bias. The finding indicated that the results with Attention Problems subscale may likely vary when different samples were used.

The only subscale that supported strong invariance (scalar invariance) across groups was the Somatic Complaints subscale. This advocated that the Somatic Complaints subscale might be the only subscale that can be used universally on both children with ASD and children with DD. However, the item scores in both groups also implied that there might be flooring effect on this subscale with these young children of 1.5–5 years old. Thus, this result needs to be interpreted with caution as the Somatic Complaints subscale also had relatively low raw scores. Given the low verbal abilities of the children in both ASD and DD groups, they might be too young (in terms of developmental process of ASD and DD) to have enough vocabularies to articulate their feelings at this age. It is therefore possible that the items on this subscale did not exactly reflect children's feelings.

This study had a few limitations. One apparent limitation is the cultural difference between our sample and the samples in previous similar studies using measurement invariance to differentiate children's mental disorders. Compared to the western samples in previous studies (e.g., Dovgan et al., 2019), our sample was collected from an area in South Asia. Therefore, the cultural difference might limit the generalizability of our results. However, on the other hand, such a result also brings to the field a meaningful understanding of the diagnostic processes for children with ASD in different cultures.

Secondly, our sample sizes of both groups might be slightly small for latent variable analysis, though a combined sample was around 378 participants. Because of the small sample size, we were not able to run a seven factors model, which is each subscale serves as a latent factor. This approach would require a fairly large sample size if golden rule of sample size for Structural

Equation Modeling (SEM; Nunnally, 1967) was followed. Specifically, it would require 10 people per variable x 99 items (variables) =around 990 subjects. We choose a second-best approach that we tested the models on each subscale, respectively. We did not choose parceling because previous study (e.g., Dovgan et al., 2019) showed that some of these differences were on the item level analysis. Yet, the small sample size might reduce the power of the study and skew the parameters estimations, we dealt with this issue by conducting data simulation with a bootstrapping approach. The results showed that there were biased estimations on Attention Problems subscale. However, on the subscales that showed major differences (i.e., Withdrawn and Aggressive Behavior), the results showed unbiased estimations.

**5. Future Direction and Implications**

There are several possible future directions that can go from here. One possibility is testing our main findings on the Withdrawn and Aggressive Behavior subscales in different cultural or social contexts. Our study found that these two subscales can be the potential subscales to differentiate children with ASD and children with DD, but it is unclear whether it can be replicated in another culturally different sample (e.g., a Western sample). Another direction could be testing the same hypothesis with a longitudinal sample to identify if these differences changed over time when children grow older, for example, such as with the CBCL age 6–18 scale. Finally, a third possible direction could be using cognitive diagnostic models (Rupp et al., 2010) on the CBCL 1.5–5 with children with ASD and see if the same patterns of subscales can be found with other latent variable models. Such a psychometric model has been used recently to assess ASD knowledge scales (e.g., Harrison et al., 2017), and perhaps can be further explored with ASD screening tools.

The take home message that we can synthesize from these results is the use of CBCL1.5–5 for screening tools should be considered with caution, particularly on the subscales that have not achieved scalar invariance. Previous research also suggested that when researchers found variations of measurement invariance during the comparisons between children with ASD and other groups, item level data should be used instead of subscales (e.g., Dovgan et al., 2019). Therefore, the attempt to use only the items in Withdrawn and Aggressive Behavior subscale of the CBCL1.5–5 to separate the profile of children with ASD and children with DD might be helpful.

**Compliance with Ethical Standards**

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 2000 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the Ditmanson Medical Foundation Chia-Yi Christian Hospital Research Ethics Committee (CYCH-IRB102045, IRB2018084). Informed consent was obtained from all individual participants included in this study.

**Declaration of Competing Interest**

The authors declare that there are no conflicts of interest.

**Acknowledgements**

This study was supported by the Ministry of Science and Technology, Taiwan (MOST-103-2628-H-037-001-MY2, MOST-108-2410-H-037-001-SSS).

**Reference**

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. University of Vermont, Research Center for Children, Youth, and Families.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.

Basten, M., Tiemeier, H., Althoff, R. R., van de Schoot, R., Jaddoe, V. W., Hofman, A., Hudziak, J. J., Verhulst, C. F., & van der Ende, J. (2016). The stability of problem behavior across the preschool years: An empirical approach in the general population. *Journal of Abnormal Child Psychology, 44*(2), 393-404. https://doi.org/10.1007/s10802-015-9993-y

Bauminger, N., Solomon, M., & Rogers, S. J. (2010). Externalizing and internalizing behaviors in ASD. *Autism Research, 3*(3), 101–112. https://doi.org/10.1002/aur.131

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research, 21*(2), 205–229. https://doi.org/10.1177/0049124192021002004

Bradstreet, L. E., Juechter, J. I., Kamphaus, R. W., Kerns, C. M., & Robins, D. L. (2017). Using the BASC-2 parent rating scales to screen for autism spectrum disorder in toddlers and preschool-aged children. *Journal of Abnormal Child Psychology, 45*(2), 359–370. https://doi.org/10.1007/s10802-016-0167-3

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

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

Chericoni, N., Balboni, G., Costanzo, V., Mancini, A., Prosperi, M., Lasala, R., Tancredi, R., Scattoni, M. L., NIDA Network, Muratori, F., & Apicella, F. (2021). A combined study on the use of the Child Behavior Checklist 1½-5 for Identifying autism spectrum disorders at 18 months. *Journal of Autism and Developmental Disorders*. Advance online publication. https://doi.org/10.1007/s10803-020-04838-0

Dawson, G., Munson, J., Estes, A., Osterling, J., McPartland, J., Toth, K., Carver, L., & Abbott, R. (2002). Neurocognitive function and joint attention ability in young children with autism spectrum disorder versus developmental delay. *Child Development, 73*(2), 345–358. https://doi.org/10.1111/1467-8624.00411

Dovgan, K., Mazurek, M. O., & Hansen, J. (2019). Measurement invariance of the Child Behavior Checklist in children with autism spectrum disorder with and without intellectual disability: Follow-up study. *Research in Autism Spectrum Disorders, 58*, 19–29. https://doi.org/10.1016/j.rasd.2018.11.009

Falk, N. H., Norris, K., & Quinn, M. G. (2014). The factors predicting stress, anxiety and depression in the parents of children with autism. *Journal of Autism and Developmental Disorders, 44*(12), 3185–3203. https://doi.org/10.1007/s10803-014-2189-4

Gau, S. S.-F., Chou, M.-C., Lee, J.-C., Wong, C.-C., Chou, W.-J., Chen, M.-F., Soong, W.-T., & Wu, Y.-Y. (2010). Behavioral problems and parenting style among Taiwanese children with autism and their siblings. *Psychiatry and Clinical Neurosciences, 64*(1), 70–78. https://doi.org/10.1111/j.1440-1819.2009.02034.x

Giovagnoli, G., Postorino, V., Fatta, L. M., Sanges, V., DePeppo, L., Vassena, L., Rose, P. D., Vicari, S., & Mazzone, L. (2015). Behavioral and emotional profile and parental stress in preschool children with autism spectrum disorder. *Research in Developmental Disabilities, 45–46*, 411–421. https:// doi.org/10.1016/j.ridd.2015.08.006

Glod, M., Creswell, C., Waite, P., Jamieson, R., McConachie, H., South, M. D., & Rodgers, J. (2017). Comparisons of the factor structure and measurement invariance of the Spence Children's Anxiety Scale–Parent version in children with autism spectrum disorder and typically developing anxious children. *Journal of Autism and Developmental Disorders*, *47*(12), 3834–3846. https:// doi.org/10.1007/s10803-017-3118-0

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11 Suppl 3), S78–S94. https:// doi.org/10.1097/01.mlr.0000245454.12228.8f

Harrison, A. J., Bradshaw, L. P., Naqvi, N. C., Paff, M. L., & Campbell, J. M. (2017). Development and psychometric evaluation of the Autism Stigma and Knowledge Questionnaire (ASK-Q). *Journal of Autism and Developmental Disorders*, *47*(10), 3281–3295. https://doi.org/10.1007/s10803-017-3242-x

Hou, Y.-M., Stewart, L., Iao, L.-S., & Wu, C.-C. (2018). Parenting stress and depressive symptoms in Taiwanese mothers of young children with autism spectrum disorder: Association with children's behavioural problems. *Journal of Applied Research in Intellectual Disabilities, 31*(6), 1113–1121. https://doi.org/10.1111/jar.12471

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hampton, J., & Strand, P. S. (2015). A review of level 2 parent-report instruments used to screen children aged 1.5-5 for autism: A meta-analytic update. *Journal of Autism and Developmental Disorders, 45*(8), 2519-2530. http://dx.doi.org/10.1007/s10803-015-2419-4

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*(1), 90-98. https://doi.org/10.1016/j.jcps.2009.09.003

Iao, L.-S., Yu, W.-H., & Wu, C.-C. (2020). Early screening for autism spectrum disorder in young children with developmental problems using the Chinese version of the Child Behavior Checklist. *Research in Autism Spectrum Disorders, 70*, 101477. https://doi.org/10.1016/j.rasd.2019.101477

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Harder, V. S., Ang, R. P., Bilenberg, N., Bjarnadottir, G., Capron, C., De Pauw, S. S. W., Dias, P., Dobrean, A., Doepfner, M., Duyme, M., Eapen, V., Erol, N., Esmaeili, E. M., Ezpeleta, L., Frigerio, A., Gonçalves, M. M., ... Verhulst, F. C. (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. https://doi.org/10.1016/j.jaac.2010.08.019

Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*(4), 567-578. https://doi.org/10.1007/BF02295979

Jang, J., Dixon, D. R., Tarbox, J., & Granpeesheh, D. (2011). Symptoms severity and challenging behavior in children with ASD. *Research in Autism Spectrum Disorders, 5*(3), 1028–1032. https://doi.org/10.1016/j.rasd.2010.11.008

Kim, I., Ekas, N. V., & Hock, R. (2016). Associations between child behavior problems, family management, and depressive symptoms for mothers of children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 26*, 80–90. https://doi.org/10.1016/j.rasd.2016.03.009

Kristensen, S., Henriksen, T. B., & Bilenberg, N. (2010). The Child Behavior Checklist for Ages 1.5–5 (CBCL/1½–5): Assessment and analysis of parent-and caregiver-reported problems in a population-based sample of Danish preschool children. *Nordic Journal of Psychiatry*, *64*(3), 203-209. https://doi.org/10.3109/08039480903456595

Leung, P. W., & Wong, M. M. (2003). Measures of child and adolescent psychopathology in Asia. *Psychological Assessment, 15*(3), 268–279. https://doi.org/10.1037/1040-3590.15.3.268

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936-949. https://doi.org/10.3758/s13428-015-0619-7

Limberg, K., Gruber, K., & Noterdaeme, M. (2017). The German version of the Child Behavior Checklist 1.5–5 to identify children with a risk of autism spectrum disorder. *Autism, 21*(3), 368–374. https://doi.org/10.1177/1362361316645932

Lindsey, R. A., Saltness, S. R., Lau, A. F., & Barry, T. D. (2020). A longitudinal examination of interactions between autism symptom severity and parenting behaviors in predicting change in child behavior problem. *Research in Autism Spectrum Disorders, 70*, 101469. https://doi.org/10.1016/j.rasd.2019.101469

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule (ADOS)*. Western Psychological Services.

Mazefsky, C. A., Anderson, R., Conner, C. M., & Minshew, N. (2011). Child Behavior Checklist scores for school-aged children with autism: Preliminary evidence of patterns suggesting the need for referral. *Journal of Psychopathology and Behavioral Assessment*, *33*(1), 31–37. https://doi.org/10.1007/s10862-010-9198-1

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, *9*(1), 93-115. https://doi.org/10.1037/1082-989X.9.1.93

Mullen, E. M. (1995). *Mullen Scales of Early Learning* (AGS Ed.). American Guidance Service.

Muratori, F., Narzisi, A., Tancredi, R., Cosenza, A., Calugi, S., Saviozzi, I., Santocchi, E., & Calderoni, S. (2011). The CBCL 1.5–5 and the identification of preschoolers with autism in Italy. *Epidemiology and Psychiatric Sciences, 20*(4), 329–338. https://doi.org/10.1017/S204579601100045X

Murray, A. L., Booth, T., McKenzie, K., Kuenssberg, R., & O'Donnell, M. (2014). Are autistic traits measured equivalently in individuals with and without an autism spectrum disorder? An invariance analysis of the Autism Spectrum Quotient Short Form. *Journal of Autism and Developmental Disorders*, *44*(1), 55–64. https://doi.org/10.1007/s10803-013-1851-6

Muthén, L. K. and Muthén, B. O. (2015). *Mplus user's guide* (7th ed.) Muthén & Muthén.

Narzisi, A., Calderoni, S., Maestro, S., Calugi, S., Mottes, E., & Muratori, F. (2013). Child Behavior Check List 1½–5 as a tool to identify toddlers with autism spectrum disorders: A case-control study. *Research in Developmental Disabilities, 34*(4), 1179–1189. https://doi.org/10.1016/j.ridd.2012.12.020

Nunnally, J. C. (1967). *Psychometric theory*. McGraw-Hill.

O'Brien, G., & Pearson, J. (2004). Autism and learning disability. *Autism, 8*(2), 125–140. https://doi.org/10.1177/1362361304042718

Pandolfi, V., Magyar, C. I., & Dill, C. A. (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. https://doi.org/10.1007/s10803-009-0716-5

Predescu, E., Sipos, R., Dobrean, A., & Miclutia, I. (2013). The discriminative power of the CBCL 1.5–5 between autism spectrum disorders and other psychiatric disorders. *Journal of Cognitive and Behavioral Psychotherapies, 13*(1), 75–87.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed.). Lawrence Erlbaum.

Rescorla, L. A., Adams, A., & Ivanova, M. Y. (2020). The CBCL/1½–5's DSM-ASD Scale: Confirmatory factor analyses across 24 societies. *Journal of Autism and Developmental Disorders*, *50*(9), 3326-3340. https://doi.org/10.1007/s10803-019-04189-5

Rescorla, L. A., Ghassabian, A., Ivanova, M. Y., Jaddoe, V. W. V., Verhulst, F. C., & Tiemeier, H. (2019). Structure, longitudinal invariance, and stability of the Child Behavior Checklist 1½–5's Diagnostic and Statistical Manual of Mental Disorders–Autism Spectrum Disorder scale: Findings from Generation R (Rotterdam). *Autism*, *23*(1), 223–235. https://doi.org/10.1177/1362361317736201

Rescorla, L. A., Kim, Y. A., & Oh, K. J. (2015). Screening for ASD with the Korean CBCL/1½–5. *Journal of Autism and Developmental Disorders, 45*(12), 4039–4050. https://doi.org/10.1007/s10803-014-2255-y

Reynolds, A. M., Soke, G. N., Sabourin, K. R., Hepburn, S., Katz, T., Wiggins, L. D., Schieve, L. A., & Levy, S. E. (2019). Sleep problems in 2-to 5-year-olds with autism spectrum disorder and other developmental delays. *Pediatrics*, *143*(3). https://doi.org/10.1542/peds.2018-0492

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rupp, A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications.* Guilford Press.

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*(4), 347–363. https://doi.org/10.1177/0734282911406661

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 167-180. https://doi.org/10.1080/10705511.2014.882658

Schiltz, H. K., McVey, A. J., Magnus, B., Dolan, B. K., Willar, K. S., Pleiss, S., Karst, J., Carson, A. M., Caiozzo, C., Vogt, E., & Van Hecke, A. V. (2018). Examining the links between challenging behaviors in youth with ASD and parental stress, mental health, and involvement: Applying an adaptation of the family stress model to families of youth with ASD. *Journal of Autism and Developmental Disorders, 48*(4), 1169–1180. https://doi.org/10.1007/s10803-017-3446-0

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107. https://doi.org/10.1007/S11336-008-9101-0

Sikora, D. M., Hall, T. A., Hartley, S. L., Gerrard-Morris, A. E., & Cagle, S. (2008). Does parent report of behavior differ across ADOS-G classifications: Analysis of scores from the CBCL and GARS. *Journal of Autism and Developmental Disorders, 38*(3), 440–448. https://doi.org/10.1007/s10803-007-0407-z

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using Mplus and the lavaan/sem tools packages. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Tan, T. X., Dedrick, R. F., & Marfo, K. (2007). Factor structure and clinical implications of Child Behavior Checklist/1.5–5 ratings in a sample of girls adopted from China. *Journal of Pediatric Psychology*, *32*(7), 807–818. https://doi.org/10.1093/jpepsy/jsm025

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

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E. J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wu, Y. T., Chen, W. J., Hsieh, W.-S., Chen, P.-C., Liao, H.-F., Su, Y.-N., & Jeng, S.-F. (2012). Maternal-reported behavioral and emotional problems in Taiwanese preschool children. *Research in Developmental Disabilities, 33*(3), 866–873. https://doi.org/10.1016/j.ridd.2011.11.018

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409-428. https://doi.org/10.3758/s13428-018-1055-2

Table 1

*Comparison of group demographic characteristics*

| | ASD (*n* = 192) | DD (*n* = 186) | *p* |
|---|---|---|---|
| CA (months) | | | |
| Mean (*SD*) | 32.91 (9.72) | 31.25 (10.59) | .112 |
| MA (months) | | | |
| Mean (*SD*) | 21.50 (10.72) | 24.40 (9.62) | .006 |
| RLAE (months) | | | |
| Mean (*SD*) | 19.54 (12.07) | 24.60 (10.42) | < .001 |
| ELAE (months) | | | |
| Mean (*SD*) | 17.06 (10.72) | 19.22 (9.91) | .044 |
| Reporter | | | |
| Mother: father | 180:12 | 172:14 | .624 |
| Parents' years of education | | | |
| Mean (*SD*): mother | 14.51 (2.35) | 13.87 (2.56) | .012 |
| Mean (*SD*): father | 14.56 (2.54) | 13.44 (2.74) | < .001 |
| ADOS total scores[a] | | | |
| Mean (*SD*): Module 1 | 17.14 (3.18) | 3.23 (1.76) | < .001 |
| Mean (*SD*): Module 2 | 15.77 (3.09) | 3.00 (1.86) | < .001 |
| Gender | | | |
| Male: female | 168:24 | 127:59 | < .001 |

*Note.* CA = chronological age; MA = mental age; RLAE = receptive language age equivalent; ELAE = expressive language age equivalent; ADOS = Autism Diagnostic Observation Schedule; ASD = autism spectrum disorder; DD = developmental delay.

[a]353 children (ASD:179, DD:174) were assessed with module 1 and 25 children (ASD:13, DD:12) were assessed with module 2

Table 2

*CFA results for subscales on ASD and DD groups (MLR estimation)*

| | $x^2$ | *df* | RMSEA | CFI | Median factor loading |
|---|---|---|---|---|---|
| ASD | | | | | |
| ER | 85.42 | 27 | .109 | .796 | .30 |
| AD | 51.40 | 20 | .096 | .898 | .33 |
| SC | 140.50 | 44 | .119 | .740 | .19 |
| WD | 52.67 | 20 | .095 | .901 | .44 |
| SP | 69.07 | 14 | .149 | .819 | .38 |
| AP | 1.88 | 5 | .000 | 1.000 | .30 |
| AB | 372.60 | 152 | .092 | .755 | .33 |
| DD | | | | | |
| ER | 79.73 | 27 | .111 | .810 | .22 |
| AD | 41.26 | 20 | .086 | .909 | .32 |
| SC | 124.28 | 44 | .113 | .685 | .12 |
| WD | 25.29 | 20 | .040 | .968 | .29 |
| SP | 62.75 | 14 | .143 | .839 | .38 |
| AP | 7.24 | 5 | .051 | .982 | .25 |
| AB | 286.80 | 152 | .073 | .875 | .38 |

*Note.* ASD = autism spectrum disorder; DD = developmental delay. ER = Emotionally Reactivity; AD = Anxious/Depressed; SC = Somatic Complaints; WD = Withdrawn; SP = Sleep Problems; AP = Attention Problems; AB = Aggressive Behavior.

Table 3

*CFA results for subscales on ASD and DD groups (WLSMV estimation)*

| | $x^2$ | *df* | RMSEA | CFI | Median factor loading |
|---|---|---|---|---|---|
| ASD | | | | | |
| ER | 76.42 | 27 | .098 | .907 | .52 |
| AD | 64.70 | 20 | .108 | .933 | .71 |
| SC | 63.43 | 44 | .048 | .967 | .78 |
| WD | 51.78 | 20 | .091 | .954 | .67 |
| SP | 71.46 | 14 | .147 | .904 | .70 |
| AP | 2.18 | 5 | .000 | 1.000 | .52 |
| AB | 357.78 | 152 | .084 | .883 | .64 |
| | | | | | |
| DD | | | | | |
| ER | 52.22 | 27 | .071 | .957 | .67 |
| AD | 49.49 | 20 | .089 | .945 | .71 |
| SC | 85.13 | 44 | .071 | .907 | .64 |
| WD | 27.70 | 20 | .046 | .978 | .61 |
| SP | 58.48 | 14 | .131 | .928 | .74 |
| AP | 7.18 | 5 | .049 | .994 | .41 |
| AB | 304.28 | 152 | .074 | .940 | .72 |

*Note.* ASD = autism spectrum disorder; DD = developmental delay. ER = Emotionally Reactivity; AD = Anxious/Depressed; SC = Somatic Complaints; WD = Withdrawn; SP = Sleep Problems; AP = Attention Problems; AB = Aggressive Behavior.

Table 4

*Model comparison of CBCL subscales (MLR estimation)*

| | $X^2$ ($\Delta X^2$) | df ($\Delta$df) | *p* value of $\Delta X^2$ test | RMSEA (Δ RMSEA) | SRMR (Δ SRMR) | CFI (Δ CFI) |
|---|---|---|---|---|---|---|
| ER | | | | | | |
| Configural | 164.8 | 54 | | .110 | .073 | .803 |
| **Configural vs.Metric*** | 166.8(2) | 62(8) | .45 | .103(-.007) | .085(.012) | .804(.001) |
| Metric vs. Scalar | 188(21.2) | 70(8) | .00 | .102(-.001) | .090(.005) | **.783 (-.021)** |
| AD | | | | | | |
| Configural | 91.9 | 40 | | .091 | .059 | .903 |
| **Configural vs.Metric*** | 95.2(3.3) | 47(7) | .72 | .082(-.009) | .068(.009) | .908(.005) |
| Metric vs. Scalar | 114.4(19.2) | 54(7) | .00 | .085(.003) | .073(.005) | **.888(-.020)** |
| SC | | | | | | |
| Configural | 264.4 | 88 | | .116 | .081 | .717 |
| Configural vs.Metric | 254.8(-9.6) | 98(10) | .16 | .114(-.002) | .097(.016) | **.696(-.021)** |
| **Metric vs. Scalar*** | 274.3(11.93) | 108(10) | .14 | .110(-.004) | .099(.002) | .691(-.007) |
| WD | | | | | | |
| **Configural*** | 77.1 | 40 | | .074 | .050 | .925 |
| Configural vs.Metric | 96.2(19.00) | 47(7) | **.01** | .078(.004) | .075(.025) | **.900(-.025)** |
| Metric vs. Scalar | NA | NA | NA | NA | NA | NA |
| AP | | | | | | |
| Configural | 9.2 | 10 | | .00 | .027 | 1.000 |
| **Configural vs.Metric*** | 16.0(6.8) | 14(4) | .14 | **.028(.028)** | .045(.018) | .991(-.009) |
| Metric vs. Scalar | 43.3(27.3) | 18(4) | **.00** | .087(.059) | **.073(.028)** | **.890(-.101)** |
| AB | | | | | | |
| **Configural*** | 275.7 | 88 | | .112 | .075 | .899 |
| Configural vs.Metric | 293.8(18.1) | (10) | **.05** | .103(-.009) | .085(.010) | **.799(-.10)** |
| Metric vs. Scalar | NA | NA | NA | NA | NA | NA |
| SP | | | | | | |
| **Configural*** | 131.8 | 28 | | .146 | .076 | .829 |
| Configural vs.Metric | 157.7(25.9) | 34(6) | **.00** | .145(-.001) | .091(.015) | **.797(-.032)** |
| Metric vs. Scalar | NA | NA | NA | NA | NA | NA |
| Chen (2007) Non invariance Alternative fits criteria | | | | Δ RMSEA ≥ .015 | (Δ SRMR) ≥ .030 | (Δ CFI) ≥ -.010 |

*Note.* ER = Emotionally Reactivity; AD = Anxious/Depressed; SC = Somatic Complaints; WD = Withdrawn; SP = Sleep Problems; AP = Attention Problems; AB = Aggressive Behavior.
*indicated the best model fit. Numbers in ( ) are the differences.
We used the significant *p* value (≤ .05) on chi-square different test as the primary criteria. So, if the chi-square test was significant, then the model comparison was ruled as non-invariant. If the chi-square test was not significant, then we check these three fit indices. If two out of three values (CFI, RMSEA, SRMR) are over the criteria, then we still considered it is non-invariant. (see our method section for more details)
"NA" means that scalar invariance was not carried out because the last (metric) invariance was not achieved.

Table 5

*Means and Standard deviations for CBCL raw scores*

| | ASD (*n* = 192) | DD (*n* = 186) |
|---|---|---|
| | Mean (*SD*) | Mean (*SD*) |
| ER | 4.90 (3.24) | 4.31 (3.08) |
| AD | 4.62 (3.05) | 4.02 (2.87) |
| SC | 2.84 (2.72) | 2.77 (2.31) |
| WD | 6.65 (3.52) | 3.48 (2.75) |
| SP | 3.89 (3.01) | 3.83 (2.90) |
| AP | 4.60 (2.09) | 4.52 (2.04) |
| AB | 14.32 (6.93) | 13.93 (7.98) |
| IN | 19.01 (9.54) | 14.59 (8.82) |
| EN | 18.93 (8.13) | 18.08 (9.28) |
| Overall | 60.30 (24.96) | 53.26 (26.33) |

*Note.* ER = Emotionally Reactive; AD = Anxious/Depressed; SC = Somatic Complaints; WD = Withdrawn; SP = Sleep Problems; AP = Attention Problems; AB = Aggressive Behavior; IN = Internalizing; EN = Externalizing; Overall = Total Problems.

**Declaration of Competing Interest**

The authors declare that there are no conflicts of interest.

**Author Contributions**: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors. All authors wrote first draft of the manuscript and discussed on revision of the manuscript. All authors read and approved the final manuscript.