



Published in final edited form as:

Psychol Assess. 2015 December ; 27(4): 1364–1378. doi:10.1037/pas0000116.

Convergent and Divergent Validity of the Mullen Scales of Early Learning in Young Children With and Without Autism Spectrum Disorder

Lauren B. Swineford¹, Whitney Guthrie², Audrey Thurm¹

¹National Institute of Mental Health, USA

²Autism Institute, Florida State University, USA

Abstract

The purpose of this study was to report on the construct, convergent, and divergent validity of the Mullen Scales of Early Learning (MSEL), a widely used test of development for young children. The sample included 399 children with a mean age of 3.38 years ($SD=1.14$) divided into a group of children with autism spectrum disorder (ASD) and a group comprised of children not on the autism spectrum, with and without developmental delays. The study used the MSEL and several other measures assessing constructs relevant to the age range, including developmental skills, autism symptoms, and psychopathology symptoms, across multiple methods of assessment. Multiple group confirmatory factor analyses revealed good overall fit and equal form of the MSEL one-factor model across the ASD and nonspectrum groups, supporting the construct validity of the MSEL. However, neither full nor partial invariance of factor loadings was established, due to the lower loadings in the ASD group, as compared to the nonspectrum group. Exploratory structural equation modeling revealed that other measures of developmental skills loaded together with the MSEL domain scores on a Developmental Functioning factor, supporting convergent validity of the MSEL. Divergent validity was supported by the lack of loading of MSEL domain scores on Autism Symptoms or Emotional/Disruptive Behavior Problems factors. Although factor structure and loadings varied across groups, convergent and divergent validity findings were similar in the ASD and nonspectrum samples. Together, these results demonstrate evidence for the construct, convergent, and divergent validity of the MSEL using powerful data analytic techniques.

Keywords

factor structure; Mullen Scales of Early Learning; validity; autism spectrum disorder

Introduction

Accurate measurement of developmental skills in young children is a critical component of assessment and diagnosis. Differential diagnosis of communication and neurodevelopmental disorders in young children relies upon valid measurement of developmental/cognitive

abilities, as diagnostic criteria require that disorders are not better accounted for by developmental delays (American Psychiatric Association, 2013). Understanding a child's developmental skills also provides important information for treatment planning and prognostic purposes. Given the rise in prevalence of developmental disorders such as autism spectrum disorder (ASD; Centers for Disease Control and Prevention, 2014) and the resulting rise in children referred for specialty testing, there is a clear need to examine the psychometric properties of developmental tests that are used in young children with and without delays and disabilities.

The Mullen Scales of Early Learning (MSEL; Mullen, 1995) is a measure of development that is used commonly in clinical and research settings to identify and quantify developmental delays, aid in differential diagnosis, and determine eligibility for and progress in interventions. The current version of the MSEL, derived from the combination of Infant MSEL (Mullen, 1989) and the Preschool MSEL (Mullen, 1992), measures aspects of nonverbal and verbal development in children birth through 68 months of age in five domains: gross motor, visual reception, fine motor, receptive language, and expressive language. A composite score of overall development, the Early Learning Composite (ELC), is derived from the T-scores of the domains, excluding gross motor.

Validity of the MSEL

While initial support for the concurrent and predictive validity of the original versions of the Infant and Preschool MSEL exists (see Mullen, 1995), scant research is available on the construct validity of the almost 20-year-old current version of the MSEL. Construct validity includes the degree to which the MSEL scores measure the construct they purport to assess (i.e., development), and how tightly the construct holds within the measure (in this case, how the MSEL domain scores relate to each other). Although the principal-axis factor analysis presented in the manual reports that the ELC provides an overall measure of general development approximating the *g* construct (Mullen, 1995), no other published factor analytic study has examined how well the MSEL domain scores fit into one composite score. Further, given that the MSEL has primarily been examined in children with typical development, it is unknown whether the ELC functions the same way in children with disorders such as ASD, where the distribution of scores is skewed and specific profiles emerge (e.g. verbal and nonverbal scores are often discrepant; Munson et al., 2008).

Convergent and divergent validity are also important aspects of construct validity. In the case of the MSEL, consideration should be given to the degree to which scores from other measures of development relate to MSEL scores (i.e., convergent validity) and conversely, the degree to which the MSEL does not measure dissimilar constructs such as autism symptoms and broader psychopathology (i.e., divergent validity). Convergent and divergent validity of the MSEL scores were examined in its validation study (Mullen, 1995) through correlations with the mental and psychomotor development indices from the Bayley Scales of Infant Development (BSID; Bayley, 1969). All MSEL domain scores, except gross motor, showed large correlations with the BSID mental developmental index, demonstrating convergent validity. Conversely, only the MSEL gross motor and expressive language scores showed large correlations with the BSID psychomotor index, demonstrating

divergent validity. However, it is noteworthy that significance values for the correlations were not reported and the sample had a substantially restricted age range (i.e., 6–15 months). Therefore, the findings from the validation study are not directly applicable for children above 15 months of age. Other support for convergent validity of the MSEL scores has been found with scores on the Preschool Language Scale (Zimmerman, Steiner, & Evatt, 1979) and the Peabody Fine Motor Scale (Folio & Fewell, 1983), which strongly correlated with the MSEL receptive and expressive language scores and the MSEL fine motor scores, respectively.

More recently, convergent validity of scores on the MSEL with the Differential Ability Scales (DAS; Elliot, 1990), was documented in children with ASD and other nonspectrum developmental disorders (Bishop et al., 2011). Neither nonverbal nor verbal scores were found to differ significantly between the MSEL (via ratio IQ scores, see Bishop et al., 2011 for calculation of ratio IQs) and the DAS (via IQ scores). The majority of children (69% and 73% for nonverbal and verbal IQ, respectively) fell in the same classification range (i.e., intellectual disability, borderline, average or above average) on both assessments, providing some evidence of convergent validity of the MSEL.

While these existing studies provide some support for the validity of the MSEL, the correlational methods employed in these studies do not allow for analysis of the validity of underlying constructs. Bivariate correlational methods do not account for the relationships among multiple variables or error terms, and cannot examine the underlying constructs which assessment tools are designed to measure. Further, in order to more comprehensively assess the constructs, there is a need for research exploring the validity of MSEL scores that includes a broader range of measures across different methods (i.e., standardized testing, clinical observation, and parent report/interview) expected to converge with the MSEL. There is also a need to examine the relationship between the MSEL and measures of ostensibly distinct constructs (i.e., social communication, repetitive behaviors, emotional problems, and disruptive behavioral problems) expected to diverge from the MSEL. Of particular importance is the degree to which the MSEL precisely measures development rather than confounding measurement of related constructs (e.g., social communication skills, attention) when used in children with disabilities such as ASD.

Measurement of developmental skills in ASD

The MSEL has been used extensively to describe developmental level and thereby estimate cognitive functioning of children with ASD. It has also been used to describe potential subtypes of ASD based on patterns of performance (Munson et al., 2008). While initial support exists for the clinical sensitivity of the MSEL for detecting general developmental delays, particular profiles of scores to distinguish children with neurodevelopmental disorders including cerebral palsy, epilepsy, and ASD have not been identified (Burns, King, & Spencer, 2013). Whether existing cognitive measures normed in typical populations should be used in atypical populations has been explored more generally and led some to question the validity of commonly used IQ measures in adults with ASD (Dawson, Soulieres, Gernsbacher, & Mottron, 2007). In particular, children with ASD have been shown to spend more time engaged in off-task behaviors and less time engaged with

the assessment during MSEL administrations than typical children, which may artificially deflate scores for children with ASD (Akshoomoff, 2006). Such score deflation may be due to impairments in social attention or secondary symptoms such as behavioral problems. MSEL scores may be confounded by these symptoms, though no existing study has examined these relations in ASD.

Conversely, studies that have examined the effect of factors associated with ASD symptoms (e.g., cognitive level, language, age) on the assessment of ASD symptoms have consistently found that measures of cognitive abilities influence the results of tests measuring ASD symptomatology. For example, lower scores on cognitive measures were associated with greater impairment as measured by the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005) in children with ASD (Hus, Bishop, Gotham, Huerta, & Lord, 2012). Similarly, scores from the Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R) are influenced by developmental level (Hus et al 2007; Gotham et al., 2009). While there appears to be evidence supporting the influence of cognitive level on ASD measurement (Gotham et al., 2007), the question of if and how ASD symptomatology influences measurement of cognitive ability remains unanswered. Therefore, divergent validity is of particular importance for an instrument such as the MSEL, given the concern about confounds of autism symptoms with scores on this measure. Thus, we specifically seek to compare MSEL psychometrics in ASD compared to nonspectrum samples.

Current Study

As indicated, limited research exists on the psychometric properties of the MSEL. Thus, our first aim was to explore aspects of construct validity of the MSEL by examining the degree to which the relationships among the MSEL domain scores conform to the MSEL Early Learning Composite (ELC), an overall measure of general development that approximates the *g* construct (AERA, APA, & NCME, 1999). To do this, we used confirmatory factor analysis (CFA) to determine overall model fit and the strength of MSEL domain score loadings onto a latent general development factor.

In order to determine whether the MSEL model fit is comparable in children with and without ASD, invariance of the measurement model across the ASD and nonspectrum groups was examined using a multiple group CFA. Measurement invariance, if confirmed across these groups, would indicate that the MSEL measures the same constructs in the same way in children with and without ASD (Kline, 2011). This is particularly important given that the MSEL was developed and validated in typically developing children, but is frequently used in children with ASD, whose symptoms may confound scores on this measure. Since there is concern about confounds of autism symptoms with scores on this measure, we specifically sought to compare MSEL psychometrics in ASD vs. nonspectrum samples.

Given that previous research raises questions about the degree to which MSEL scores are independent from constructs measured by other instruments used to assess behavior and development in young children, our second aim was to evaluate the convergent and divergent validity of the MSEL. Thus, we examined the relationship between test scores on

the MSEL and scores from measures also designed to assess development (i.e., convergent validity) and scores from measures intended to assess different constructs (i.e., divergent validity; AERA, APA, & NCME, 1999). Validity was tested using multiple methods of assessment (i.e., parent report, clinical observation, direct testing) and multiple constructs/traits that were included in exploratory structural equation models (ESEM). This novel approach to examining validity was chosen because ESEM allows for examination of possible relationships between scores on the MSEL and measures expected to converge and diverge with the MSEL whereas traditional correlational methods cannot examine underlying constructs which assessment tools are designed to measure. Although not yet commonly used, this more powerful, novel data analytic technique has been previously used to examine validity (e.g., Marsh, Lüdtke, Muthén, Asparouhov, Morin, Trautwein, & Nagengast, 2010; Rojas & Widiger, 2014; Van den Broeck, Hofmans, Cooremans, & Staels, 2014). Multiple group ESEMs examined measurement invariance across the ASD and nonspectrum groups in order to test whether patterns and strength of loadings were invariant across children with and without ASD and examine whether independence of MSEL scores is differentially affected across these two groups.

Method

Participants and Procedures

This research was conducted in a clinical research setting where children participated as part of a research protocol that included an evaluation to determine eligibility for autism research at [redacted for masked review]. This study was approved by and conducted in accordance with an Institutional Review Board at [redacted for masked review] and informed consent was obtained through participants' parents or guardians. The present sample included 399 children, with a mean age of 3.38 years ($SD=1.14$) grouped into categories of ASD ($n=184$) or nonspectrum ($n=215$).

A clinical diagnosis of ASD was the determining factor for group placement, as both groups included children with and without developmental delays as measured by the MSEL. Each of these groups (ASD and nonspectrum) was specifically recruited for; however, some children in the nonspectrum group (which included children recruited specifically for developmental delays or for typical development) were referred for concern about ASD (or vice versa). Clinical diagnosis of ASD or determination as nonspectrum was made by doctoral-level clinicians (i.e. clinical psychologists, speech and language pathologists) using DSM-IV diagnostic criteria, based on clinical judgment after administration of the ADI-R (Rutter, Le Couteur, & Lord, 2003) and the ADOS (Lord, Rutter, DiLavore, & Risi, 1999), described below.

Measures

The evaluation comprised a wide array of parent report, direct testing, and clinical observation measures, including measures of early development, language, adaptive behavior and symptomatology of autism or other emotional/behavioral problems. Test selection was based on the child's chronological age and abilities (e.g. whether a basal could be achieved on a specified test).

Mullen Scales of Early Learning.—The MSEL is a standardized test for children birth to 68 months that measures nonverbal and verbal developmental level through assessment of five domains: gross motor, visual reception, fine motor, receptive language, and expressive language. The test was standardized to 3 standard deviations from the mean. Thus, T-scores from each domain, which range from 20 to 80, were included in analyses, with the exception of the gross motor domain, which does not contribute to the Early Learning Composite (ELC). The MSEL has been shown to have good internal reliability and strong test-retest and interscorer reliability (Mullen, 1995).

Vineland Adaptive Behavior Scale, 2nd edition.—The Vineland Adaptive Behavior Scales, 2nd edition (VABS-II; Sparrow, Cicchetti, & Balla, 2005), is a standardized measure of adaptive behavior for children and adolescents from birth through 90 years of age. The VABS-II is administered as a semi-structured interview with caregivers. Standard scores, which span almost 5 standard deviations below the mean and 4 standard deviations above the mean, range from 20–160 for the four domains: Communication, Daily Living Skills, Socialization, and Motor Skills, were included in analyses. The VABS-II has been shown to have strong internal consistency, test-retest reliability, interinterviewer reliability, and validity for young children (Sparrow, Cicchetti, & Balla, 2005).

Autism Diagnostic Observation Schedule.—The ADOS (Lord, Rutter, DiLavore, & Risi, 1999) is a standardized assessment of autism symptoms through observation of communication, social interaction, and repetitive behaviors and restricted interests. Children received ADOS modules 1–3 based on their language level at the time of the screening evaluation. Children under 30 months of age without phrase speech received the ADOS-Toddler Module (Lord, Luyster, Gotham, & Guthrie, 2012). Diagnostic algorithms for all modules provided Social Affect and Restricted and Repetitive Behavior domain scores (Gotham, Risi, Pickles, & Lord, 2007), which were converted to Calibrated Severity Scores (CSS) that ranged from 0–10 (Hus, Gotham, & Lord, 2012), with a range of severity that included scores of 1–2 falling in the range of “minimal-to-no concern,” 3–4 “low,” 5–7 “moderate” and 8–10 “high.” CSS scores are not yet available for the ADOS-T; instead, Module 1 algorithms were calculated using ADOS-T items and Module 1 domain scores were converted to CSS scores. The ADOS has been shown to have strong internal consistency, interrater and test-retest reliability, and validity (Lord, Luyster, Gotham, & Guthrie, 2012; Lord, Rutter, DiLavore, & Risi, 1999).

Autism Diagnostic Interview-Revised.—The ADI-R (Rutter, Le Couteur, & Lord, 2003) is a standardized, semi-structured clinician-administered parent interview designed to measure autism symptoms in children and adults. Diagnostic algorithms provide Social, Communication, and Repetitive Behavior and Restricted Interest domain scores. The Communication domain is composed of separate verbal and nonverbal communication totals, with verbal communication totals only completed for children considered verbal. Since a significant portion of children were considered nonverbal, only the Communication nonverbal totals were included in the analyses. The ranges of scores for the Social, Nonverbal Communication, and Repetitive Behavior and Restricted Interest domains are 0–30, 0–14, and 0–12, respectively. Strong interrater reliability, test-retest reliability, and

diagnostic validity have been found for the ADI-R with higher estimates for Social and Communication domains compared to the Repetitive Behavior and Restricted Interest domain (Rutter, Le Couteur, & Lord, 2003).

Peabody Picture Vocabulary Test, 3rd edition.—The Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) is a standardized measure of receptive vocabulary for children (lower age range: 2 years, 6 months) and adults. Standard scores, which range from 20–160, were included in the analyses. The PPVT has been shown to have high internal consistency, test-retest reliability, and validity in young children (Dunn & Dunn, 1997).

Expressive One-Word Picture Vocabulary Test, 3rd edition.—The Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2000) is a standardized measure of expressive vocabulary for children (lower age range: 2 years) and adults. Standard scores, which range from 20–160, were included in the analyses. The EOWPVT has been shown to have high interrater and test-retest reliability and strong validity (Brownell, 2000).

Communication and Symbolic Behavior Scales – Caregiver Questionnaire.—The Communication and Symbolic Behavior Scales Caregiver Questionnaire (CSBS CQ; Wetherby & Prizant, 2002) is a parent completed measure of a child’s communication, social, comprehension and play skills. Standard scores are provided for children from 9 to 24 months of age in three areas: social, speech, and symbolic skills. Given that most children in this study were above 24 months of age (the oldest age at which normed standard scores are available), we used weighted raw scores for the social, speech, and symbolic subscales (ranges from 0–48, 0–40, and 0–51, respectively; Wetherby & Prizant, 2002). Strong test-retest reliability, concurrent validity, and predictive validity have been reported for the CSBS CQ (Wetherby, Allen, Cleary, Kublin, & Goldstein, 2002).

Social Communication Questionnaire.—The Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) is a parent completed measure of communication skills and social functioning designed as a screener for ASD. Total raw scores range from 0–39. The SCQ has been shown to have strong discriminate validity for older children and the authors recommend that the SCQ be used with caution for children under 4 years of age (Rutter, Bailey, & Lord, 2003). Given the age of participants in the current study, the SCQ was not used as a screener or diagnostic tool, but rather was collected for examination of its properties in relation to the other tools used to measure social communication in young children.

Repetitive Behavior Scale-Revised.—The Repetitive Behavior Scale (RBS-Revised; Bodfish, Symons, Parker, & Lewis, 2000) is a parent completed checklist of stereotyped, self-injurious, compulsive, and tic behaviors in children. A total raw score is obtained by adding all of the behaviors reported by the parents (range 0–129). Acceptable levels of reliability and validity have been reported for the RBS-R (Bodfish, Symons, Parker, & Lewis, 2000)

Child Behavior Checklist.—The CBCL 1.5–5 version (Achenbach & Rescorla, 2001) is a standardized parent completed form that measures early psychopathology symptoms

including behavioral and emotional problems in children age 1.5 to 5 years. The CBCL provides T-scores that range from 50 to 100 for seven syndrome scales: affect, anxiety, hyperactivity, oppositional defiant, emotionally reactive, sleep, and pervasive developmental problems. High test-retest reliability and acceptable validity levels have been reported for the CBCL (Achenbach & Rescorla, 2001).

Sample Characteristics

Descriptive statistics are presented in Table 1. Children in the ASD group scored significantly lower than children in the nonspectrum group on all domains of the MSEL compared to the nonspectrum group. A high percentage of children in the ASD group received below the floor (16%, 23%, 52%, 52%) on the MSEL Visual Reception, Fine Motor, Receptive Language and Expressive Language, respectively. As expected, when compared to children in the nonspectrum group, children in the ASD group had significantly higher scores on measures of autism and psychopathology symptoms except the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) oppositional defiant and anxiety scales. More specifically, the majority of children in the ASD group had ADOS Social Affect and Restricted and Repetitive Behavior CSS scores that fell in the moderate and high ranges compared to the nonspectrum group where the majority fell in the minimal-to-no evidence and low ranges.

Statistical Analyses

Construct validity.—General construct validity was analyzed using confirmatory factor analyses (CFAs) conducted using Mplus 7.11 software (Muthén & Muthén, 2012). Full Information Maximum Likelihood (FIML) estimation was used in order to handle missing data. FIML does not delete or impute data, but rather utilizes all available data to compute maximum likelihood estimates. This method has several advantages over other methods of handling missing data, including its production of less biased estimates and smaller standard errors compared to other methods (i.e., multiple imputation, listwise deletion; Enders & Bandalos, 2001). Goodness-of-fit statistics were considered for the overall model: Standardized Root Mean Square Residual (SRMR; criteria for good fit is .08; Hu & Bentler, 1999), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI; criteria for good fit is .95; Hu & Bentler, 1999). Root Mean Square Error of Approximation (RMSEA) values were reported because it is common practice. However, RMSEA values were not used to examine overall model fit, because this model had a small number of degrees of freedom and this fit statistic can be artificially large in this scenario (Kenny, Kaniskan, & McCoach, 2013). Strength of the loadings of the MSEL domain scores onto the latent factor measuring general development was examined, given that the ELC combines these domain scores and is used as a summary score to approximate g (Mullen, 1995).

Following a CFA using the entire sample, multiple group CFAs assessed measurement invariance across the ASD and nonspectrum groups, in order to examine whether MSEL measures the same construct across these groups. After examining the viability of the model separately in both groups, tests of equal form and factor loadings were considered in a sequential fashion. At each step, additional constraints were added to the model that increased equality across the groups. Comparisons were then made at each step to determine

whether the increased equality restrictions resulted in significantly poorer fit. If so, the less restrictive model (i.e., weaker invariance) was selected; if not, analyses proceeded to greater equality constraints across groups (i.e., stronger invariance). These comparisons were made using chi-square difference testing and change in CFI (i.e. increases of less than .01; Cheung & Rensvold, 2002; Chen, 2007). See Kline 2011 for detailed description of multiple group analyses to examine measurement invariance.

The first step was to test for configural invariance, or equal form invariance, which refers to equality of the structural configuration. If configural invariance held (i.e., good fit according to fit statistics), results indicated that the same constructs are measured in each group (i.e., structure of the model is the same across groups) and testing proceeded to construct-level metric invariance. If configural invariance did not hold, results indicated that the constructs are not equivalent across groups and testing did not proceed. In the second step, we tested for construct-level metric invariance, or equal factor loadings, which refers to equality of the unstandardized factor loadings onto the latent factor across groups. If construct-level metric invariance held (i.e., nonsignificant chi-square test and small increase in CFI [$<.01$]), results indicated that constructs are manifested the same way in each of the groups. If full construct-level metric invariance was not established, the less strict partial measurement invariance was tested, which holds some, but not all, factor loadings to be equal across groups. Even stricter forms of invariance (i.e., latent factor means, indicator means, variances) can also be tested, but were not examined in the present study given the known differences in level and variance in developmental functioning between children with and without ASD.

Convergent and divergent validity.—Convergent and divergent validity were tested using exploratory structural equation models (ESEM) that included variables from multiple methods of assessment and multiple constructs/traits within a latent model framework. ESEM combines features of CFA, exploratory factor analysis (EFA), and structural equation modeling (Asparouhov & Muthén, 2009; Marsh et al., 2009; Marsh et al., 2010). ESEM was chosen because it examines all possible relationships among scores on the MSEL and measures expected to converge and diverge with the MSEL and loadings of all indicators on all latent factors are freely estimated. In contrast, CFA typically places some constraints on loadings, with the overly restrictive independent clusters model (ICM) requiring all indicators to load onto only one factor with other loadings constrained to be zero. Similar to EFA, ESEM estimates all indicators and utilizes typical rotation methods. However in contrast to traditional EFA, ESEM provides standard errors, goodness-of-fit statistics, methods of statistical model comparison through significance and fit statistics, and multiple group invariance analyses. Of additional utility is ESEM's estimation of the relations between factors, which are less likely to be inflated than these parameters estimated within ICM-CFA models (Marsh et al., 2010). Within EFA specification (i.e., parameters estimated for loadings of all indicators on all factors) of ESEM, an oblique rotation was chosen given the lack of independence between the constructs of interest and the rarity of completely uncorrelated factors in social science research (Norris & Lecavalier, 2010). Given the lack of previous modeling of this type with the present constructs, Geomin rotation was chosen, as it is preferred when little is known about the true loading structure (Asparouhov & Muthén, 2009). Full Information Maximum Likelihood was chosen as the estimator, in order to

appropriately handle missing data (Muthén, Kaplan, & Hollis, 1987). See Table 1 for listing of all variables included in ESEM analyses.

Using ESEM, models with k and $k+1$ factors were directly compared using several different methods in order to determine the model with the optimal number of factors to retain. Up to four factors were examined given that measures of four primary constructs were included in the present study (see below). Chi-square difference testing and parallel analysis of eigenvalues (which index degree of information provided by each additional factor) were considered in choosing the number of factors to retain. Models were also compared on change in CFI (i.e. increases of less than .01) in order to determine whether improvement in model fit for the more complex model (i.e., more factors) was sufficiently large to reject the more parsimonious model (i.e., fewer factors).

Once the optimal number of factors for retention was identified, factor loadings were examined based on statistical significance rather than simply by a cutoff applied to the loading magnitude. Significance was calculated as the value of the parameter estimate divided by the standardized error of the loading, which follows the z -distribution and allows for hypothesis testing. Given the large number of tests for loading significance in the ESEM with a large number of variables, p was set at .001.

ESEM tested convergent validity through examination of whether the MSEL domain scores loaded significantly onto factor(s) composed of measures expected to index similar constructs (i.e., VABS, PPVT, EOWVT, CSBS Speech). Divergent validity of the MSEL was tested through examination of whether MSEL domains significantly loaded on factors composed of measures of ostensibly distinct constructs (i.e., autism symptoms and emotional/behavioral problems). It should be noted that although ESEM provides estimation of parameters for all variables on all factors, the parameters of interest for the present study were: loadings of MSEL domain scores on all factors and loadings of other measures together onto the same factor as the MSEL. While informative for other research aims, loadings of other measures on other factors (e.g., loading of autism symptom measures on the communication factor) were not pertinent to the present study. Such results are presented in tables but are not discussed.

Following an ESEM conducted using the entire sample, multiple group ESEMs were attempted in order to assess measurement invariance across the ASD and nonspectrum groups following the same procedure as the multiple group CFA described in the section detailing CFA procedures.

Results

Initial descriptive analyses indicated univariate normality of the MSEL domain scores (i.e., skew and kurtosis values within reasonable limits ($<|2|$ for skew and $<|7|$ for kurtosis; Curran, West, & Finch, 1996). Multivariate normality was also established for all variables (i.e., Mardia skewness, Mardia kurtosis, and Henze-Zirkler Tp values $> .05$), which is necessary for fitting CFAs and ESEMs using Maximum Likelihood estimation. A correlation matrix is presented in Table 2 to facilitate future meta-analyses.

Confirmatory Factor Analyses – Construct Validity

The MSEL CFA model, which specified that all four domain T-scores loaded onto a single latent factor approximating *g*, provided good fit according to CFI (.98) and SRMR (.02). TLI (.93) fell short of the .95 cutoff. As expected based on recommendations for use of RMSEA in low *df* models, RMSEA ([95% confidence interval]) = .20 [.14-.26]) did not indicate good fit. All MSEL domain scores demonstrated significant and strong loadings onto the latent factor (standardized loadings – Visual Reception: .89, Fine Motor: .84, Receptive Language: .92, Expressive Language: .89, all $p < .001$; See Figure 1). Thus, the MSEL CFA model supported the construct validity of this measure, as all of the MSEL domain scores were strong indicators of the latent factor approximating *g*.

Multiple group CFA.—Prior to conducting formal tests of invariance, viability of the model was first established separately in the ASD and nonspectrum groups. Within the ASD group, fit was good according to CFI (.95) and SRMR (.03), but not TLI (.86); indicator loadings were strong and significant (.69-.83). Within the nonspectrum group, fit was good according to CFI (.97) and SRMR (.02), but not TLI (.92); indicator loadings were also strong and significant (.80-.91). Results of invariance testing are summarized in Table 3.

Configural invariance/ Equal form.—Model fit remained good according to CFI (.97) and SRMR (.03), but not TLI (.89) when factor structure was constrained to be equal across the groups, demonstrating form invariance (see Table 3). Form invariance indicates equivalent form such that the factor structure (i.e. number of factors and pattern of loadings) is generally equivalent across groups.

Construct-level metric invariance/equivalent factor loadings.—Given that equivalent form was established, analysis advanced to tests of equivalence of the strength of factor loadings across groups. In this model, equality constraints were added such that indicator loadings were permitted to vary within but not across groups (i.e., equivalence between ASD and nonspectrum). Significantly reduced model fit was found, when compared to the previous model in which loadings were not constrained ($\chi^2_{diff}(3)=16.77$, $p < .001$; reduction in CFI=.02). Thus, factor loading equivalence could not be supported, as one or more parameters varied between the ASD and nonspectrum groups. Examination of factor loadings across groups indicated the standardized loadings were consistently lower (by .05-.11) in the ASD group. After failing to establish full construct-level metric invariance, partial measurement invariance was explored to determine whether freely estimating one or more loadings by group could improve model fit. However, partial measurement could not be established, as fit continued to be significantly worse when one or more loadings were freely estimated.

In sum, configural (but not metric) invariance was established, indicating that the MSEL domains were strongly related to the same, single latent factor (similar to ELC) in both the ASD and nonspectrum groups, but these loadings were significantly lower (though to a small degree) in children with ASD compared to the nonspectrum group. Results of the CFA models for the nonspectrum and ASD groups are presented in Figure 2.

Exploratory Structural Equation Modeling – Convergent and Divergent Validity

The first step in constructing the ESEM was to determine the number of classes to retain in the sample as a whole; models with one to four factors were compared. Three different methods were employed to guide this determination. A parallel analysis of eigenvalues preferred a two-factor solution. However, the comparative methods (chi-square difference testing and CFI comparisons) were significant for all comparisons of k and $k+1$ models (where k is the number of factors), indicating that the four-factor solution was superior to models with one, two, and three factors. Further, the four-factor solution optimized fit according to all goodness-of-fit statistics (CFI=.90, TLI=.86, SRMR=.05). No Heywood cases (i.e., communalities > 1.0), which can occur when too many factors are extracted, were observed in the four-factor model. Based on these findings, the four-factor solution was retained. See Table 4 for summary of model comparisons. As noted in Table 5, which summarizes factor loadings, factors were termed Autism Symptoms, Developmental Functioning, Communication, and Emotion/Behavior Problems.

Factor loadings in entire sample.—Scores from each MSEL domain demonstrated significant and strong loadings (.78-.93) on the Developmental Functioning factor. Scores from other measures of development (i.e., each VABS-II domain score, EOWPVT, and PPVT) also loaded significantly on this factor. None of the MSEL domain scores loaded significantly onto the Autism Symptoms or Emotion/Behavior Problems factors. The MSEL Expressive Language scores loaded significantly onto the Communication factor, although the magnitude was small.

However, before findings from analyses combining the ASD and nonspectrum groups should be interpreted with regard to validity, multiple group ESEMs should be conducted to determine whether the relationships between these constructs and variables are equal across these groups. Thus, testing of measurement invariance across the ASD and nonspectrum groups was initiated.

Multiple group ESEM.—As a first step, viability of the ESEM model was examined separately in each group. Substantial differences in model fit and factor loading patterns were observed between the ASD and nonspectrum groups, such that formal testing of form and loading equivalence testing of the whole group was not necessary. Simple examination of the model separately by group indicated that invariance across groups could not be established. Thus, loading patterns from the entire sample may not be accurate representations of convergent and divergent validity across children with and without ASD. As such, ESEMs are presented separately for children in the ASD and nonspectrum groups (see Table 5).

Nonspectrum group.—Within the nonspectrum group, fit of the four-factor model remained optimal compared to models with fewer factors (CFI=.90, TLI=.86, SRMR=.07). Patterns and magnitude of significant loadings supported convergent and divergent validity in the nonspectrum group. The MSEL domain scores were found to converge with scores from other measures of development across measurement methods, as these scores loaded together with all of the VABS domains and the PPVT on the Developmental Functioning

factor. The ADOS Social Affect CSS and CBCL Pervasive Developmental Problems scores also significantly loaded onto this factor. The MSEL Receptive and Expressive domain scores also loaded significantly into the Communication factor, along with VABS Communication, PPVT, EOWPVT, ADI-R Communication, and all CSBS composite scores, reflecting the relationship between the developmental functioning and communication skills constructs in children without ASD.

The MSEL domain scores did not significantly load onto the Autism Symptoms or Emotion/Behavior Problems factors, supporting divergent validity of the MSEL in the nonspectrum group. Significant factor loadings ($p < .001$) for the nonspectrum sample are presented in Figure 3.

ASD group.—Within the ASD group, the four-factor model showed poor overall fit (CFI=.76, TLI=.66, SRMR=.08), though solutions with fewer factors did not fit significantly better and were substantially less interpretable. Although patterns and magnitude of significant loadings differed in the ASD group, as compared to the nonspectrum group, results also supported convergent and divergent validity in the ASD group. The MSEL domain scores loaded significantly onto the Developmental Functioning factor, along with VABS Communication, PPVT, and EOWPVT scores, supporting the convergent of the MSEL in the ASD group. Further, the MSEL domain scores did not significantly load onto any of the other factors: Autism Symptoms, Communication, or Emotion/Behavior Problems, supporting the divergent validity of the MSEL in this group. The SCQ and ADOS Social Affect CSS also loaded onto the Developmental Functioning factor in the ASD group. However, the analyses here cannot determine whether this is due to measurement of developmental skills by the measures, or overlap between the constructs of ASD symptoms and developmental skills.

Differences in Loadings Between ASD and Nonspectrum groups

As described above, loadings were substantially different in the ASD and nonspectrum groups. In the ASD group, only the VABS-II Communication domain scores (but not the other VABS-II domains) converged with the MSEL domains on the Developmental Functioning factor, whereas scores for all VABS-II domains demonstrated significant, strong loadings on this factor in the nonspectrum group. The PPVT scores loaded significantly on the Developmental Functioning factor in both groups, but the EOWPVT scores loaded on this factor only in the ASD group. In addition, the MSEL Receptive Language and Expressive Language domain scores loaded significantly on the Communication factor in the nonspectrum group, but not the ASD group.

Discussion

The purpose of the current study was to report on aspects of the validity of MSEL scores, a widely used test of development for young children. Inclusion of scores from multiple methods of assessment in addition to the MSEL, and powerful, novel data analytic techniques allowed for examination of construct, convergent, and divergent validity of the MSEL in children with and without ASD.

Construct Validity

Results of confirmatory factor analyses examining general construct validity showed that each of the MSEL domain scores loaded onto a single, latent factor in the entire sample, and that this factor form held across the ASD and nonspectrum groups. Establishment of configural invariance across groups suggests that the same latent construct (i.e., overall development) is measured in both children with ASD and children classified as nonspectrum. These findings support the construct validity of the MSEL scores in both groups. In particular, use of the ELC, which is designed to approximate this latent construct, is supported.

Although the same factor structure fit comparably across groups, individual MSEL standardized loadings were weaker by .05-.11 in children with ASD compared to the nonspectrum group. However, these small differences do not necessarily reflect differences in utility of the MSEL across groups for two reasons. First, loadings were strong in both the ASD (.69-.83) and nonspectrum (.80-.91) groups. Second, loadings were also as strong as those found from other standardized cognitive tests (Bodin, Pardini, Burns, & Stevens, 2009; Devena, Gay, & Watkins, 2013). Factor loading differences between groups may have been in part due to different distributions of MSEL domain scores: in the ASD group scores were both lower and less variable than the nonspectrum group.

Convergent Validity

The results of the ESEM in the full sample demonstrated that most scores from measures of language (i.e., PPVT and EOWPVT) and adaptive behavior (i.e., VABS-II) loaded onto the same factor as the MSEL scores, supporting convergent validity by suggesting that these tests all index the same underlying construct. However, when examined separately, differences in patterns of significant loadings were observed between the ASD and nonspectrum groups. Importantly, the poor overall fit of the ESEM model in the ASD group indicates that loadings for the ASD group should be interpreted with caution.

In the nonspectrum group, the MSEL domain scores converged with the scores from measures hypothesized to assess aspects of development (the VABS-II domains and the PPVT) on the Developmental Functioning factor. In the ASD group, the VABS-II Communication domain (but not the other domains), PPVT, and EOWPVT converged with the MSEL domain scores on the Developmental Functioning factor. The recent finding that relationships between cognitive scores and adaptive scores in some domains may not go in the same direction for all children with ASD (Duncan & Bishop, 2013) may explain the lack of convergence with scores from the other VABS-II domains. Overall, support was provided for the convergent validity of the scores obtained on the MSEL as the MSEL converged with the VABS-II Communication domain and the PPVT across both groups.

The MSEL Receptive and Expressive Language domain scores also loaded onto the Communication factor in the nonspectrum group. This finding suggests that the scores from the MSEL Language domains tap a small, but statistically significant amount of variance in the broader construct of communication in children without ASD, as would be expected given the relatedness of the constructs of communication and developmental functioning.

However, this was not found in the ASD group. The significant heterogeneity of language abilities in children with ASD, in whom relationships between cognitive and language skills may differ compared to children without ASD, may explain the lack of relationship between the MSEL Language scores and the Communication factor for the ASD group.

Lastly, small loadings of ASD symptom measures were observed on the Developmental Functioning factor for the ASD and nonspectrum groups (i.e., ADOS Social Affect and SCQ in ASD group; ADOS Social Affect CSS and CBCL Pervasive Developmental Problems in nonspectrum group). The factor analytic models conducted cannot directly test directionality, which in this case would indicate whether these measures of autism symptoms partially measure developmental level or whether the inverse is true. However, the MSEL did not load into the Autism Symptoms factor, indicating its measurement of developmental skills does not measure the latent construct indexed by this factor. Thus, the current study, designed to test validity of the MSEL in young children with and without ASD, corroborates and improves upon current evidence for the MSEL's convergent validity. Multiple methods of measurement (e.g., direct testing and observation, parent report, and parent interview) as well as sophisticated statistical techniques failed to reveal a significant relationship between the MSEL scores and the latent autism symptoms construct but did reveal significant relationships between the MSEL and other measures of developmental functioning and communication.

Divergent Validity

Divergent validity was demonstrated by the loading of MSEL domain scores almost exclusively on the Developmental Functioning factor when examined in the entire sample and by group. The Visual Reception and Fine Motor domain scores did not significantly load onto the Autism Symptoms, Communication, or Emotional/Disruptive Behavioral Problems factors in either group, suggesting that measurement of these skills is independent of these latent constructs. The significant loadings of the MSEL Receptive and Expressive Language domain scores onto the Communication factor could be interpreted as evidence against the divergent validity of the MSEL. However, these loadings suggest that the MSEL measures additional aspect of communication not related to the developmental factor rather than a direct lack of divergence.

Limitations

Patterns of factor loadings relevant to examination of convergent and divergent validity of the MSEL differed in the ASD and nonspectrum groups. These differences may be generalizable to many children with ASD, but may also represent artifacts of the specific cohort of children in the current sample. Specifically, the MSEL scores were significantly lower and less variable in the ASD group than the nonspectrum group such that the current ASD sample contained a high percentage of children who were at the floor (score of 20; 28%, 36%, 16%, 8%) or below (16%, 23%, 52%, 52%) on the MSEL Visual Reception, Fine Motor, Receptive Language and Expressive Language, respectively. Though children with ASD often have comorbid intellectual disability, the current sample shows greater cognitive impairment than found in some other published samples of young children with ASD (e.g., Zwaigenbaum et al. 2012), although comparable to other community samples

(Barbaro & Dissanayake, 2012; Dawson et al. 2010). Similarly, the differing patterns of loadings for the PPVT and EOWPVT on the Developmental Functioning factor between the ASD and nonspectrum samples may have been due to the significantly lower scores and significantly smaller percentage (i.e. one third) of the ASD sample with observed scores on these vocabulary measures (although Maximum Likelihood was used to handle missing data in the present analyses). These group differences in patterns of relationships among measures combined with the poorer model fit in the ASD group (compared to the nonspectrum group) indicate that results and the loadings of the PPVT and EOWPVT in particular should be interpreted with caution.

Conclusions

The MSEL is designed to measure development in children from infancy to early school age. The current validity study included children covering almost the full age range for the MSEL (12–68 months), and provided support for validity of scores obtained on the MSEL across this developmental span. Subsequent steps in this line of research would be utilization of confirmatory procedures to further explore convergent and divergent validity of the MSEL. This study used a novel statistical method, ESEM, to explore relationships between scores from the MSEL and measures expected to converge and diverge with the MSEL. Future research could use these methods to explore the distinctness of measures on other constructs (e.g., RRBs) or construct validity of other tools (e.g., ADOS, CBCL).

Acknowledgements:

This research was supported by the Intramural Program of the National Institute of Mental Health. The views expressed in this paper do not necessarily represent the views of the NIMH, NIH, HHS, or the United States Government. Selected results were presented at the Society for Research on Child Development in Seattle, 2013. All authors have no conflict of interest to declare.

Dr. Swineford and Dr. Thurm are with Pediatrics and Developmental Neuroscience Branch, National Institute of Mental Health. Ms. Guthrie is with Florida State University.

References

- Achenbach T, & Rescorla L. (2001). *Manual for the Child Behavior Checklist/school-age forms & profiles*. Burlington, VT: University of Vermont.
- Akshoomoff N. (2006). Use of the Mullen Scales of Early Learning for the assessment of young children with autism spectrum disorders. *Child Neuropsychology*, 12(4–5), 269–277. [PubMed: 16911972]
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational & Psychological Testing (US). (1999). *Standards for educational and psychological testing*. American Educational Research Association.
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.
- Asparouhov T, & Muthén B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397–438.
- Barbaro J, & Dissanayake C. (2012). Developmental profiles of infants and toddlers with autism spectrum disorders identified prospectively in a community-based setting. *Journal of Autism and Developmental Disorders*, 42(9), 1939–1948. [PubMed: 22310906]
- Bayley N. (1969). *Bayley Scales of Infant Development*. San Antonio, TX: Psychological Corporation.

- Bishop SL, Guthrie W, Coffing M, & Lord C. (2011). Convergent validity of the Mullen Scales of Early Learning and the Differential Ability Scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities*, 116(5), 331–343. [PubMed: 21905802]
- Bodfish JW, Symons FJ, Parker DE, & Lewis MH (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders*, 30(3), 237–243. [PubMed: 11055459]
- Bodin D, Pardini DA, Burns TG, & Stevens AB (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology*, 15(5), 417–424. [PubMed: 19132580]
- Brownell R. (2000). *Expressive One-Word Picture Vocabulary Test*. Novato, California: Academic Therapy Publications.
- Burns TG, King TZ, & Spencer KS (2013). Mullen Scales of Early Learning: The utility in assessing children diagnosed with autism spectrum disorders, cerebral palsy, and epilepsy. *Applied Neuropsychology: Child*, 2(1), 33–42. [PubMed: 23427775]
- Centers for Disease Control and Prevention. (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, DC: 2002), 63, 1.
- Chen FF (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504.
- Cheung GW, & Rensvold RB (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.
- Constantino JN, & Gruber CP (2002). *The Social Responsiveness Scale*. Los Angeles: Western Psychological Services.
- Curran PJ, West SG, & Finch JF (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological methods*, 1(1), 16.
- Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, ... Varley J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics*, 125(1), e17–e23. [PubMed: 19948568]
- Dawson M, Soulières I, Gernsbacher MA, & Mottron L. (2007). The level and nature of autistic intelligence. *Psychological Science*, 18(8), 657–662. [PubMed: 17680932]
- Devena SE, Gay CE, & Watkins MW (2013). Confirmatory factor analysis of the WISC-IV in a hospital referral sample. *Journal of Psychoeducational Assessment*, 31(6), 591–599.
- Duncan AW, & Bishop SL (2013). Understanding the gap between cognitive abilities and daily living skills in adolescents with autism spectrum disorders with average intelligence. *Autism*, 1362361313510068.
- Dunn LM, & Dunn LM (1997). *Peabody Picture Vocabulary Test*. Circle Pines, MN: American Guidance Service. Inc. Publishing.
- Elliot C. (1990). *Differential Ability Scales*. San Antonio, TX: Psychological Corporation.
- Enders CK, & Bandalos DL (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8(3), 430–457.
- Folio MR, & Fewell R. (1983). *Peabody Developmental Motor Scales and Activity Cards* (Vol. 1). Chicago: Riverside Publishing Company.
- Gotham K, Pickles A, & Lord C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(5), 693–705. [PubMed: 19082876]
- Gotham K, Risi S, Pickles A, & Lord C. (2007). The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, 37(4), 613–627. [PubMed: 17180459]
- Hu LT, & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.

- Hus V, Bishop S, Gotham K, Huerta M, & Lord C. (2013). Factors influencing scores on the Social Responsiveness Scale. *Journal of Child Psychology and Psychiatry*, 54(2), 216–224. [PubMed: 22823182]
- Hus V, Gotham K, & Lord C. (2012). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*, 1–13. [PubMed: 21360019]
- Hus V, Pickles A, Cook EH Jr, Risi S, & Lord C. (2007). Using the Autism Diagnostic Interview—Revised to increase phenotypic homogeneity in genetic studies of autism. *Biological Psychiatry*, 61(4), 438–448. [PubMed: 17276746]
- Kenny DA, Kaniskan B, & McCoach DB (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 0049124114543236.
- Kline RB (2011). *Principles and practice of structural equation modeling*. Guilford press.
- Lord C, Luyster R, Gotham K, & Guthrie W. (2012). *Autism Diagnostic Observation Schedule – Toddler Module manual*. Los Angeles, CA: Western Psychological Services.
- Lord C, Rutter M, DiLavore PC, & Risi S. (1999). *Autism Diagnostic Observation Schedule-WPS (ADOS-WPS)*. Los Angeles, CA: Western Psychological Services.
- Marsh HW, Lüdtke O, Muthén B, Asparouhov T, Morin AJ, Trautwein U, & Nagengast B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22(3), 471. [PubMed: 20822261]
- Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJ, & Trautwein U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, 16(3), 439–476.
- Mullen EM (1989). *Infant Mullen Scales of Early Learning*. Cranston, RI: TOTAL Child, Incorporated.
- Mullen EM (1992). *Mullen Scales of Early Learning*. Cranston, RI: TOTAL Child, Incorporated.
- Mullen EM (1995). *Mullen Scales of Early Learning: AGS Circle Pines, MN*.
- Munson J, Dawson G, Sterling L, Beauchaine T, Zhou A, Koehler E, ... Estes A. (2008). Evidence for latent classes of IQ in young children with autism spectrum disorder. *American Journal on Mental Retardation*, 113(6), 439–452. [PubMed: 19127655]
- Muthén B, Kaplan D, & Hollis M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, 52(3), 431–462.
- Muthén B. & Muthén L. (2014). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Norris M, & Lecavalier L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. *Journal of Autism and Developmental Disorders*, 40(1), 8–20. [PubMed: 19609833]
- Rojas SL, & Widiger TA (2014). Convergent and Discriminant Validity of the Five Factor Form. *Assessment*, 21(2), 143–157. [PubMed: 24366956]
- Rutter M, Bailey A, & Lord C. (2003). *The Social Communication Questionnaire: Manual*. Los Angeles, CA: Western Psychological Services.
- Rutter M, Le Couteur A, & Lord C. (2003). *Autism Diagnostic Interview-Revised*. Los Angeles, CA: Western Psychological Services.
- Sparrow S, Cicchetti D, & Balla D. (2005). *Vineland Adaptive Behavior Scales (2nd ed.)* Circle Pines, MN: American Guidance Service.
- Van den Broeck W, Hofmans J, Cooremans S, & Staels E. (2014). Factorial validity and measurement invariance across intelligence levels and gender of the Overexcitabilities Questionnaire-II (OEQ-II). *Psychological Assessment*, 26(1), 55. [PubMed: 24079958]
- Wetherby AM, Allen L, Cleary J, Kublin K, & Goldstein H. (2002). Validity and reliability of the communication and symbolic behavior scales developmental profile with very young children. *Journal of Speech, Language, and Hearing Research*, 45, 1202–1218.
- Wetherby AM, & Prizant BM (2002). *Communication and Symbolic Behavior Scales: Developmental Profile*. Baltimore, MD: Paul H Brookes Publishing.
- Zimmerman IL, Steiner VG, & Pond RE (1979). *Preschool Language Scale*. Columbus, OH: Charles E. Merrill Publishing Company.

Zwaigenbaum L, Bryson SE, Szatmari P, Brian J, Smith IM, Roberts W, ... Roncadin C. (2012). Sex differences in children with autism spectrum disorder identified within a high-risk infant cohort. *Journal of Autism and Developmental Disorders*, 42(12), 2585–2596. [PubMed: 22453928]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

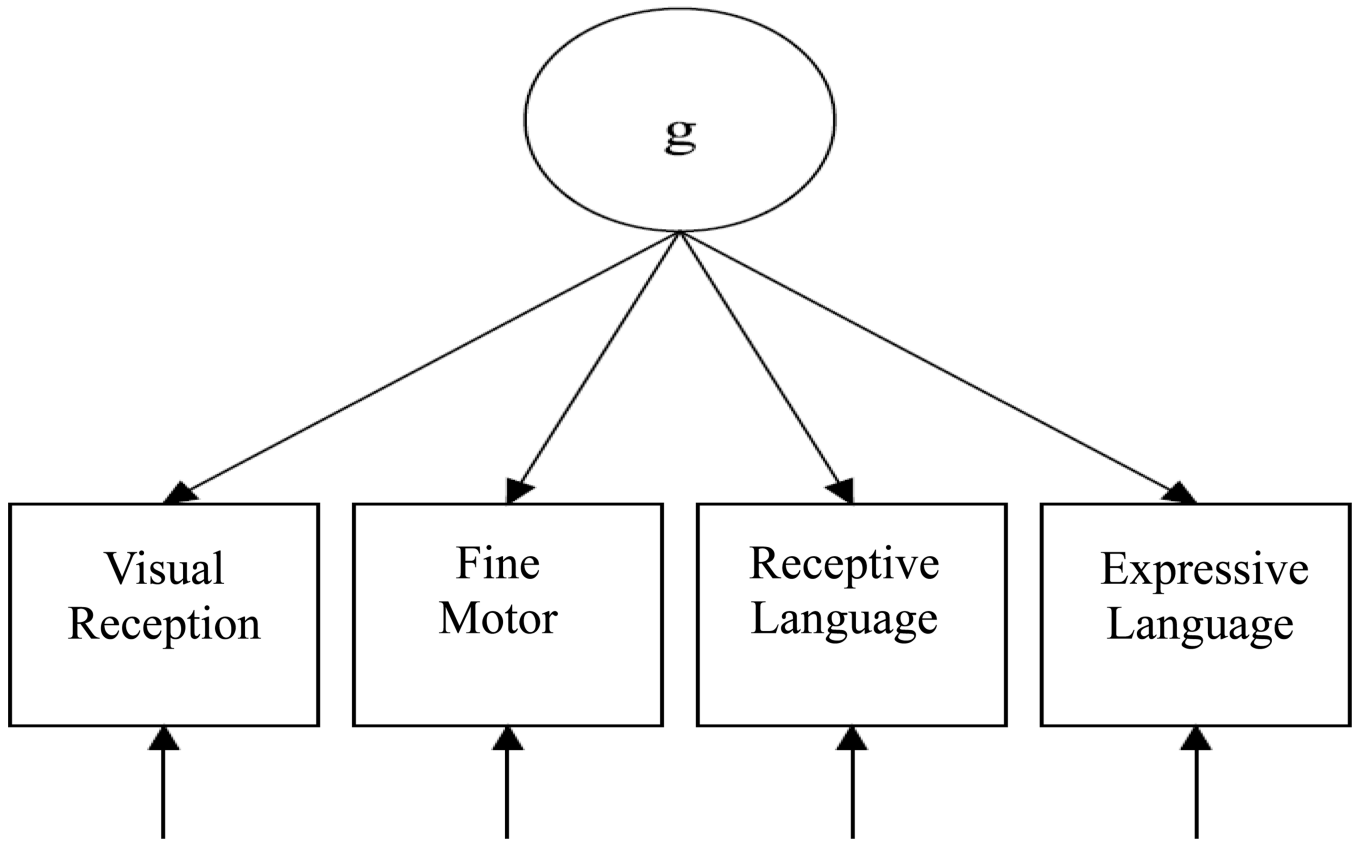
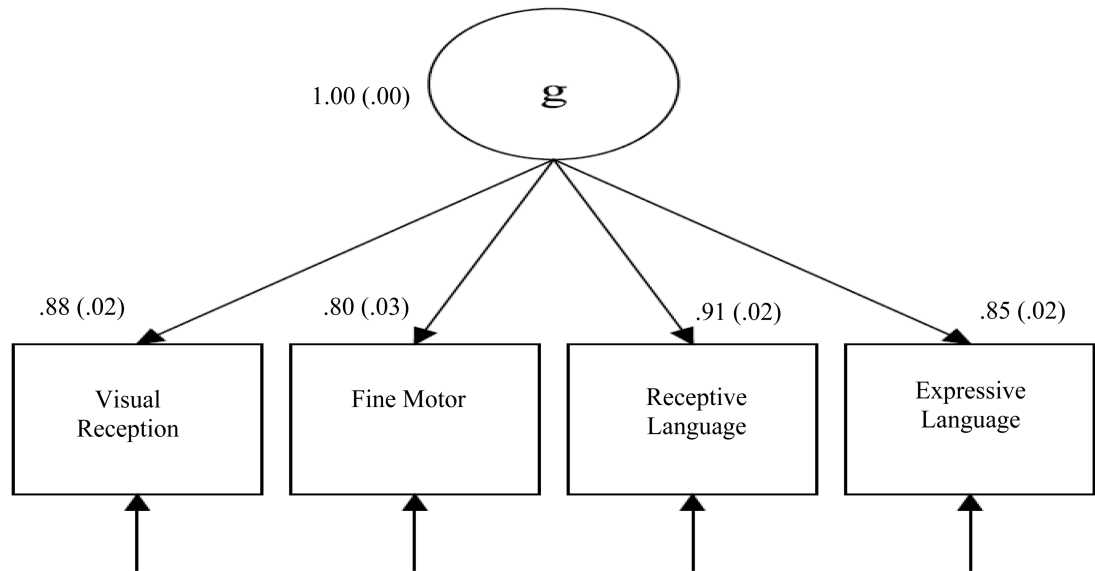


Figure 1.
Simplified diagram of CFA model for entire sample

Panel A. Nonspectrum Sample



Panel B. ASD Sample

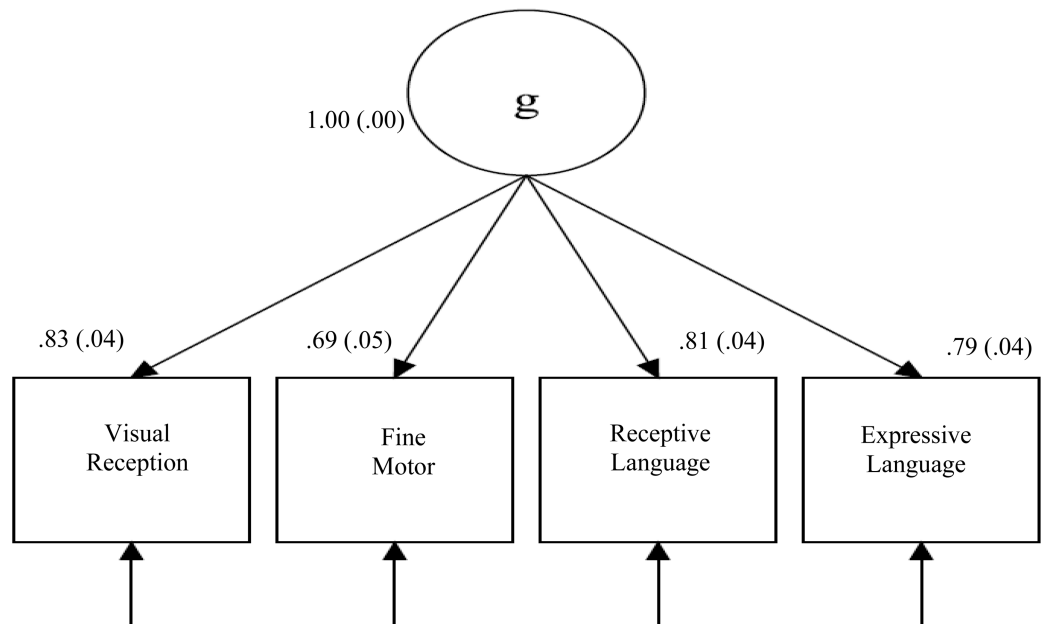


Figure 2.
Path diagram of CFA models for nonspectrum and ASD samples

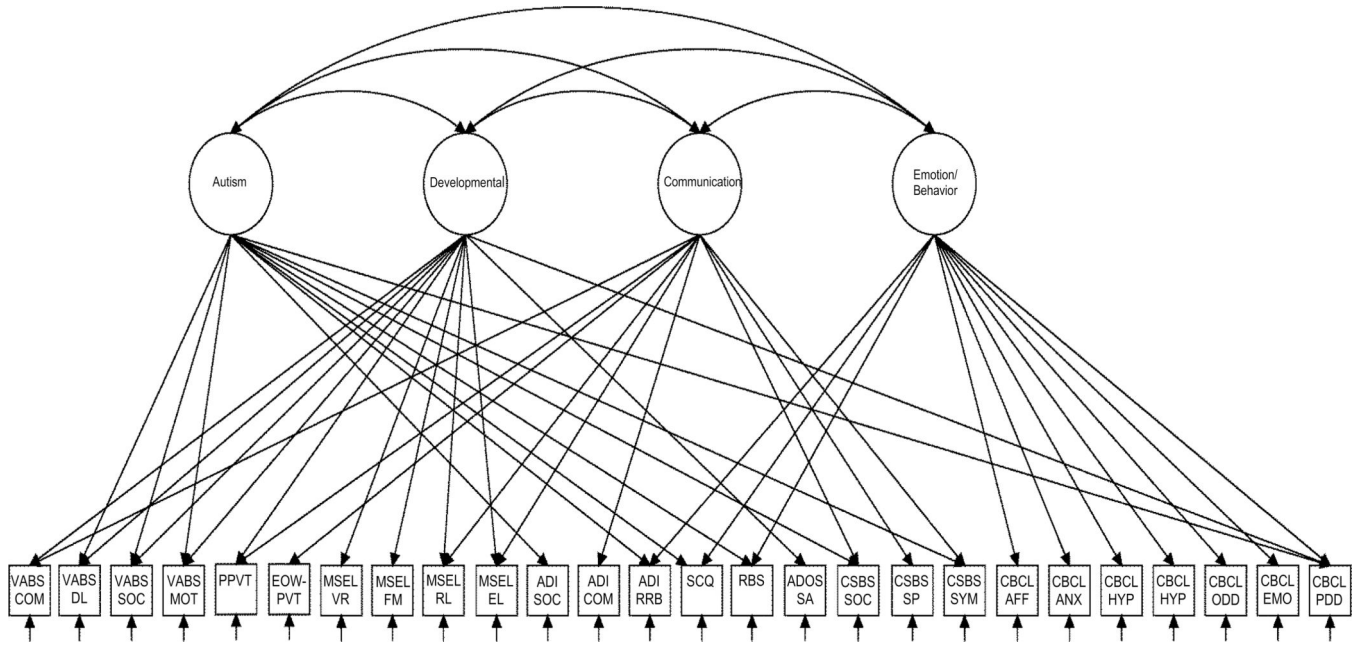


Figure 3. ESEM for nonspectrum group Note: The ESEM conducted estimated the loading of all observed variables on all latent factors. However, for the purpose of clarity, only those paths that represent loadings significant at $p < .001$ are shown in this simplified path diagram. Parameter estimates can be found in Table 5. The Autism Diagnostic Observation Schedule Repetitive Behavior Calibrated Severity Score is not included in the figure as it did not load significantly onto any of the factors. VABS = Vineland Adaptive Behavior Scales; COM = Communication Standard Score; DL = Daily Living Standard Score; SOC = Socialization Standard Score; MOT = Motor Skills Standard Score; PPVT = Peabody Picture Vocabulary Test Standard Score; EOWPVT = Expressive One Word Picture Vocabulary Test Standard Score; MSEL = Mullen Scales of Early Learning; VR = Visual Reception T-Score; FM = Fine Motor T-Score; RL = Receptive Language T-Score; EL = Expressive Language T-Score; ADI = Autism Diagnostic Interview; Social = Social Interaction Domain; Comm = Communication Domain; RRB = Restricted and Repetitive Behavior Domain; SCQ = Social Communication Questionnaire Total Score; RBS-R = Repetitive Behavior Scales-Revised Total Score; ADOS = Autism Diagnostic Observation Schedule; SA CSS = Social Affect Calibrated Severity Score; CSBS CQ = Communication and Symbolic Behavior Scales Caregiver Questionnaire; Social = Social Total Score; Speech = Speech Total Score; Symbolic = Symbolic Total Score; CBCL = Child Behavior Checklist; Affect = Affect Problems T-score; Anxiety = Anxiety Problems T-score; Hyper = Hyperactivity Problems T-score; OD = Oppositional Defiant Problems T-score; ER = Emotionally Reactive T-score; Sleep = Sleep Problems T-score; PDD = Sleep Problems T-score

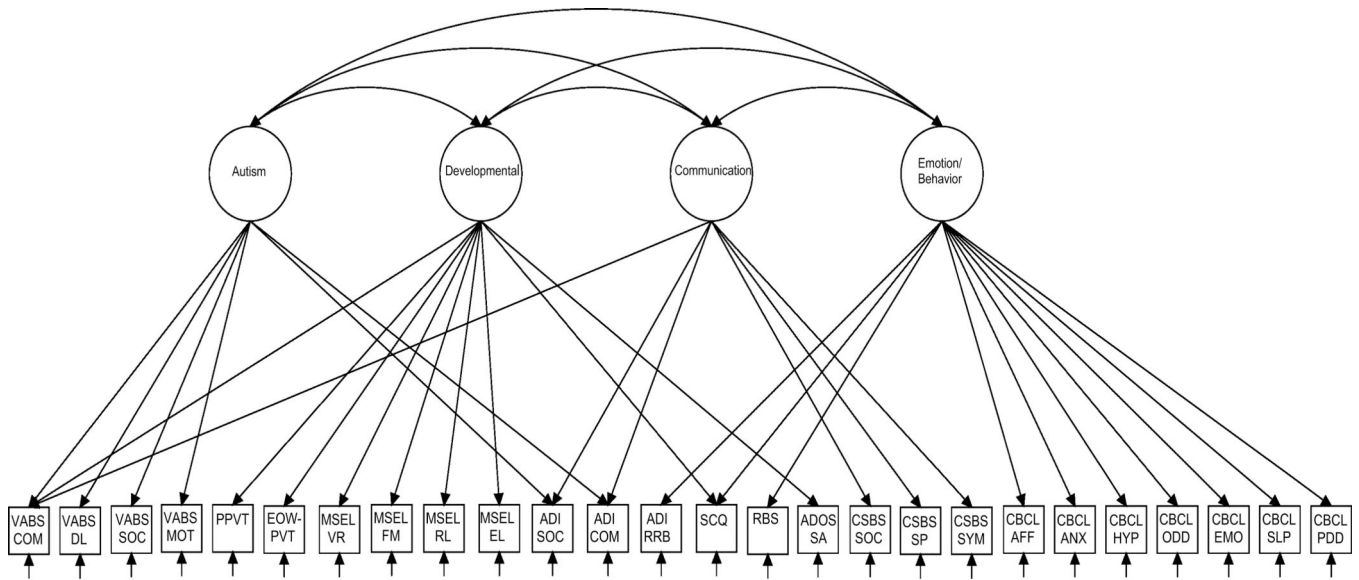


Figure 4.

ESEM for ASD sample Note: The ESEM conducted estimated the loading of all observed variables on all latent factors. However, for the purpose of clarity, only those paths that represent loadings significant at $p < .001$ are shown in this simplified path diagram.

Parameter estimates can be found in Table 5. The Autism Diagnostic Observation Schedule Repetitive Behavior Calibrated Severity Score is not included in the figure as it did not load significantly onto any of the factors. VABS = Vineland Adaptive Behavior Scales; COM = Communication Standard Score; DL = Daily Living Standard Score; SOC = Socialization Standard Score; MOT = Motor Skills Standard Score; PPVT = Peabody Picture Vocabulary Test Standard Score; EOWPVT = Expressive One Word Picture Vocabulary Test Standard Score; MSEL = Mullen Scales of Early Learning; VR = Visual Reception T-Score; FM = Fine Motor T-Score; RL = Receptive Language T-Score; EL = Expressive Language T-Score; ADI = Autism Diagnostic Interview; Social = Social Interaction Domain; Comm = Communication Domain; RRB = Restricted and Repetitive Behavior Domain; SCQ = Social Communication Questionnaire Total Score; RBS-R = Repetitive Behavior Scales-Revised Total Score; ADOS = Autism Diagnostic Observation Schedule; SA CSS = Social Affect Calibrated Severity Score; CSBS CQ = Communication and Symbolic Behavior Scales Caregiver Questionnaire; Social = Social Total Score; Speech = Speech Total Score; Symbolic = Symbolic Total Score; CBCL = Child Behavior Checklist; Affect = Affect Problems T-score; Anxiety = Anxiety Problems T-score; Hyper = Hyperactivity Problems T-score; OD = Oppositional Defiant Problems T-score; ER = Emotionally Reactive T-score; Sleep = Sleep Problems T-score; PDD = Sleep Problems T-score

Table 1.

Descriptive Statistics for all Variables Included in Exploratory Structural Equation Model and Child Age.

	ASD			Nonspectrum		
	<i>n</i>	<i>Mean</i>	<i>SD</i>	<i>n</i>	<i>Mean</i>	<i>SD</i>
Age in Years	184	3.60	1.05	215	3.20	1.18
Vineland Adaptive Behavior Scales						
Communication Standard Score	184	72.18	13.73	215	91.70	16.75
Daily Living Standard Score	184	71.85	10.96	215	89.07	14.59
Socialization Standard Score	184	70.11	8.58	215	90.51	13.12
Motor Skills Standard Score	184	79.53	10.36	215	90.13	14.62
Peabody Picture Vocabulary Test Standard Score	56	78.27	22.45	109	95.80	21.30
Expressive One Word Picture Vocabulary Test Standard Score	58	79.47	20.18	124	94.44	21.68
Mullen Scales of Early Learning						
Visual Reception T-Score	176	29.24	11.26	208	45.12	15.71
Fine Motor T-Score	176	25.09	7.85	208	40.99	16.14
Receptive Language T-Score	176	24.78	9.26	208	41.41	14.95
Expressive Language T-Score	176	25.00	7.75	208	40.24	14.02
Autism Diagnostic Interview						
Social Interaction Domain Total Score	173	17.16	5.00	118	7.07	5.88
Communication Domain Total Score	173	10.14	3.00	118	4.78	3.67
Restricted and Repetitive Behavior Domain Total Score	173	4.40	1.91	118	2.32	2.02
Social Communication Questionnaire Total Score	154	19.41	5.56	164	9.16	6.65
Repetitive Behavior Scales-Revised Total Score	144	21.97	13.95	156	10.29	12.20
Autism Diagnostic Observation Schedule						
Social Affect CSS Score	176	6.45	1.75	168	1.91	1.13
Restricted and Repetitive Behavior CSS Score	176	8.14	1.76	168	4.28	2.64
Communication and Symbolic Behavior Scales Caregiver Questionnaire						
Social Total Score	119	27.31	8.09	133	38.62	8.24
Speech Total Score	119	20.50	13.21	133	29.80	11.28
Symbolic Total Score	119	27.39	13.03	133	38.91	11.47
Child Behavior Checklist						
Affect Problems T-score	137	61.76	8.96	152	55.74	7.29
Anxiety Problems T-score	137	55.64	8.24	152	54.14	7.28
Hyperactivity Problems T-score	137	60.07	8.04	152	55.37	6.54
Oppositional Defiant Problems T-score	137	56.34	8.08	152	55.14	7.65
Emotionally Reactive T-score	137	57.55	8.69	152	55.01	7.41
Sleep Problems T-score	137	59.23	10.55	152	55.36	7.31
Pervasive Developmental Problems T-score	137	71.88	8.77	152	60.10	10.59

Table 2.

Correlation matrix among measures for the ASD and nonspectrum groups

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
VABS																		
1. COM		.79*	.80*	.71*	.73*	.67*	.75*	.66*	.81*	.78*	-.47*	-.50*	-.19	-.55*	-.32*	-.39*	-.39*	.64*
2. DL	.61*		.82*	.77*	.60*	.47*	.70*	.69*	.71*	.64*	-.48*	-.50*	-.27	-.60*	-.34*	-.36*	-.37*	.61*
3. SOC	.60*	.76*		.73*	.59*	.51*	.66*	.65*	.68*	.64*	-.64*	-.55*	-.41*	-.80*	-.45*	-.41*	-.42*	.63*
4. MOT	.51*	.67*	.54*		.56*	.48*	.65*	.69*	.62*	.53*	-.27	-.40*	-.19	-.52*	-.34*	-.27*	-.28*	.51*
5. PPVT	.53*	.15	.06	.11		.85*	.68*	.61*	.74*	.73*	-.30	-.23	-.08	-.48*	-.40*	-.32	-.30	.59*
6. EOWPVT	.57*	.12	.03	.20	.77*		.59*	.46*	.67*	.66*	-.30	-.35	-.16	-.46*	-.37*	-.28	-.28	.53*
MSEL																		
7. VR	.52*	.29*	.24	.37*	.57*	.66*		.77*	.79*	.72*	-.29	-.37*	-.16	-.43*	-.29*	-.41*	-.40*	.49*
8. FM	.47*	.26	.23	.46*	.48*	.62*	.66*		.71*	.65*	-.28	-.35*	-.19	-.48*	-.29*	-.45*	-.43*	.41*
9. RL	.57*	.18	.20	.18	.64*	.71*	.65*	.52*		.80*	-.46*	-.47*	-.17	-.54*	-.38*	-.45*	-.44*	.60*
1. EL	.60*	.33*	.32*	.27*	.58*	.51*	.64*	.49*	.70*		-.32*	-.31	-.05	-.39*	-.27	-.43*	-.34*	.48*
ADI																		
11. Social	-.44*	-.51*	-.67*	-.30*	.04	.06	-.11	-.14	-.23	-.24		.74*	.42*	.71*	.37*	.17	.16	-.81*
12. Comm	-.57*	-.43*	-.50*	-.32*	.03	-.20	.19	.18	-.30*	-.33*	.56*		.29	.62*	.25	.26	.24	-.81*
13. RRB	.04	-.17	-.30*	-.14	.17	.05	-.01	.03	.03	.11	.31*	.07		.59*	.59*	.19	.23	-.33
14. SCQ	-.23	-.29*	-.39*	-.16	.18	.19	.04	.02	-.01	.03	.53*	.43*	.46*		.70*	.36*	.38*	-.84*
15. RBS-R	-.09	-.21	-.30*	-.07	.17	.11	.09	.10	.13	.07	.32*	.12	.47*	.56*		.21	.24	-.53*
ADOS																		
16. SA CSS	-.25*	-.04	.01	-.04	-.26	-.40	-.39*	-.32*	-.43*	-.27*	0.11	0.12	-.10	-.14	-.16		.49*	-.26
17. RRB CSS	-.09	.01	.01	.05	-.37	-.26	-.29*	-.20	-.31*	-.24	-.01	0.18	.09	.09	.01	.17		-.19
CSBS CQ																		
18. Social	.67*	.46*	.51*	.44*	-.06	.21	.26	.27	.31	.27	-.53*	-.64*	-.15	-.59*	-.19	-.12	-.09	
19. Speech	.75*	.41*	.30	.32*	.33	.53	.36*	.25	.34*	.58*	-.28	-.56*	.13	-.26	-.06	-.13	-.23	.65*
20. Symbolic	.71*	.31	.22	.31	.22	.48	.33*	.30	.38*	.37*	-.30	-.58*	.03	-.40*	-.11	-.17	-.21	.76*
CBCL																		
21. Affect	-.17	-.25	-.31*	-.24	.12	.16	.03	.03	.05	-.02	.34*	.19	.24	.36*	.50*	-.16	-.05	-.26
22. Anxiety	.11	-.01	-.08	-.05	.25	.16	.13	.19	.27	.18	.08	.03	.33*	.22	.51*	-.25	-.06	.08
23. Hyper	-.24	-.33*	-.35*	-.18	.05	-.14	-.02	-.11	.01	.01	.34*	.23	.27	.33*	.53*	-.13	-.07	-.17

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
24. OD	.10	-.03	-.09	.04	.25	.06	.09	.13	.18	.10	.12	.03	.23	.08	.46*	-.27	-.12	.08
25. ER	-.05	-.13	-.17	-.16	.26	.26	.11	.13	.16	.05	.20	.10	.28	.32*	.59*	-.24	-.09	-.10
26. Sleep	.09	-.06	-.15	.01	.26	.27	.04	.09	.14	.01	.22	.13	.16	.24	.37*	-.11	-.12	-.06
27. PDD	-.18	-.24	-.36*	-.08	-.01	-.08	.07	.12	-.06	.01	.46*	.28	.34*	.50*	.52*	-.10	-.18	-.26

Note. Nonspectrum is above the diagonal; ASD is below the diagonal.

* p<.001. VABS = Vineland Adaptive Behavior Scales; COM = Communication Standard Score; DL = Daily Living Standard Score; SOC = Socialization Standard Score; MOT = Motor Skills Standard Score; PPVT = Peabody Picture Vocabulary Test Standard Score; EOWPVT = Expressive One Word Picture Vocabulary Test Standard Score; MSEL = Mullen Scales of Early Learning; VR = Visual Reception T-Score; FM = Fine Motor T-Score; RL = Receptive Language T-Score; EL = Expressive Language T-Score; ADI = Autism Diagnostic Interview; Social = Social Interaction Domain; Comm = Communication Domain; RRB = Restricted and Repetitive Behavior Domain; SCQ = Social Communication Questionnaire Total Score; RBS-R = Repetitive Behavior Scales-Revised Total Score; ADOS = Autism Diagnostic Observation Schedule; SA CSS = Social Affect Calibrated Severity Score; RRB CSS = Restricted and Repetitive Behavior Calibrated Severity Score; CSBS CQ = Communication and Symbolic Behavior Scales Caregiver Questionnaire; Social = Social Total Score; Speech = Speech Total Score; Symbolic = Symbolic Total Score; CBCL = Child Behavior Checklist; Affect = Affect Problems T-score; Anxiety = Anxiety Problems T-score; Hyper = Hyperactivity Problems T-score; OD = Oppositional Defiant Problems T-score; ER = Emotionally Reactive T-score; Sleep = Sleep Problems T-score; PDD = Sleep Problems T-score

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Results of Multiple Group CFA Testing Measurement Invariance in ASD and Nonspectrum Groups

	χ^2	df	χ^2_{diff}	df	CFI	CFI	TLI	SRMR	RMSEA (90% CI)
<i>Single Group Solutions</i>									
ASD	17.40	2	--	--	.95	--	.86	.03	.21 (.13-.31)
Nonspectrum	20.28	2	--	--	.97	--	.91	.02	.21 (.13-.30)
<i>Measurement Invariance</i>									
Equal form	37.68	4	--	--	.97	--	.89	.03	.21 (.15-.27)
Equal factor loadings	54.45	7	16.77*	3	.95	.00	.92	.10	.19 (.14-.24)

Note. χ^2 = Chi-Square; df = degrees of freedom; χ^2_{diff} = Chi-square difference testing between previous model

* p<.05; SRMR = Standardized Root Mean Residual; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; CI = confidence interval.

Table 4.

ESEM Model Comparison

Number of Factors	Eigenvalue	p for χ^2_{diff}	SRMR	CFI	TLI	RMSEA (90% CI)
1	13.65	<.001	.13	.63	.60	.15 (.15-.16)
2	3.65	<.001	.07	.77	.73	.13 (.12-.13)
3	1.41	<.001	.07	.85	.81	.11 (.10-.12)
4	1.30	<.001	.05	.90	.86	.09 (.09-.10)
5	.95	<.001	.04	.94	.91	.07 (.07-.08)
6	.88	<.001	.03	.96	.93	.06 (.06-.07)

Note. χ^2 = Chi-Square; df = degrees of freedom; χ^2_{diff} = Chi-square difference testing between k and $k-1$ when k is the number of factors; CFI = Comparative fit index; TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Square Error of Approximation; CI = confidence interval.

Table 5.

ESEM Factor Loadings for All Samples

	Factor 1			Factor 2			Factor 3			Factor 4		
	Autism Symptoms			Developmental Functioning			Communication			Emotion/Behavior Problems		
	Entire Sample	Non-spectrum	ASD	Entire Sample	Non-spectrum	ASD	Entire Sample	Non-spectrum	ASD	Entire	Non-Spectrum	ASD
VABS Communication	-.18*	-.12	.39*	.62*	.64*	.52*	.32*	.36*	.39*	-.06	-.03	.01
VABS Daily Living	-.39*	-.34*	.74*	.55*	.73*	.27	-.01	-.02	-.01	-.05	.08	-.00
VABS Socialization	-.86*	-.41*	.90*	.42*	.57*	.13	-.03	.04	.02	-.06	-.11	.00
VABS Motor Skills Standard Score	-.14	-.26*	.56*	.63*	.71*	.35	-.00	-.03	-.06	-.11	.10	-.07
PPVT	.03	.00	-.09	.60*	.37*	.68*	.51*	.71*	.27	-.04	-.00	.06
EOWPVT	.06	.04	-.11	.53*	.20	.78*	.61*	.85*	.25	-.01	-.03	-.04
MSEL Visual Reception	.08	.07	.05	.93*	.87*	.88*	.18	.09	-.09	.01	-.01	-.11
MSEL Fine Motor	-.10	-.07	.10	.85*	.88*	.74*	-.13	-.09	-.07	.05	.07	-.03
MSEL Receptive Language	-.10	-.02	.04	.78*	.69*	.72*	.12	.31*	.10	.04	-.09	.07
MSEL Expressive Language	-.00	.13	.14	.79*	.67*	.75*	.20*	.41*	.06	-.01	-.05	.04
ADI Social	.92*	.72*	-.64*	-.02	.06	.11	-.05	-.28	-.29*	-.07	.09	.01
ADI Communication	.79*	.63*	-.37*	-.05	-.03	-.00	-.24*	-.35*	-.57*	-.11	-.05	-.07
ADI RRB	.72*	.43*	-.28	-.04	-.10	-.17	.24*	.20	-.07	.17	.40*	.36*
SCQ Total Score	.88*	.68*	-.31	.05	-.07	.29*	-.03	-.08	-.44*	.13	.32*	.27*
RBS-R Total Score	.52	.42*	-.16	-.02	.01	.06	.05	.01	-.03	.47*	.50*	.71*
ADOS Social Affect CSS	.59*	.11	.05	-.25*	-.33*	-.37*	-.04	-.09	-.04	-.21*	.08	-.16
ADOS RRB CSS	.48*	.11	.14	-.30*	-.30	-.24	-.01	-.10	-.15	-.11	.13	-.02
CSBS CQ Social	-.70*	-.61*	.33	-.01	.02	.03	.41*	.42*	.72*	-.04	-.14	-.03
CSBS CQ Speech	-.13	-.13	.04	.16	.04	.31	.78*	.87*	.67*	-.01	.14	.05
CSBS CQ Symbolic	-.40*	-.45*	-.12	.00	-.02	.18	.74*	.74*	.92*	.01	.16	-.05
CBCL Affect Problems	.23	.04	-.11	.20	.05	-.01	-.18	-.16	-.10	.61*	.70*	.63*
CBCL Anxiety Problems	-.02*	.07	.13	.04	.12	-.02	.00	.05	.13	.80*	.79*	.84*

	Factor 1			Factor 2			Factor 3			Factor 4		
	Autism Symptoms			Developmental Functioning			Communication			Emotion/Behavior Problems		
	Entire Sample	Non-spectrum	ASD	Entire Sample	Non-spectrum	ASD	Entire Sample	Non-spectrum	ASD	Entire	Non-Spectrum	ASD
CBCL Hyperactivity Problems	.12	-.17	-.22	-.25*	-.34	-.10	.02	-.02	-.01	.42*	.40*	.51*
CBCL Oppositional Defiant Problems	-.14	-.24	.03	-.11	-.02	-.05	.03	-.07	.20	.66*	.74*	.63*
CBCL Emotionally Reactive	-.03	-.01	.05	-.06	-.05	-.01	-.04	.09	-.01	.86*	.85*	.83*
CBCL Sleep Problems	.12	.01	.04	.08	.10	.01	-.15	-.07	-.09	.54*	.63*	.56*
CBCL Pervasive Developmental Problems	.53*	.37*	-.20	-.16	-.20*	.05	.03	.04	-.14	.43*	.61*	.54*

Note.

* $p < .001$. VABS = Vineland Adaptive Behavior Scales, Standard Scores; PPVT = Peabody Picture Vocabulary Test Standard Score EOWPVT = Expressive One Word Picture Vocabulary Test Standard Score; MSEL = Mullen Scales of Early Learning T-Scores; ADI = Autism Diagnostic Interview Domain Scores; RRB = Restricted and Repetitive Behavior; SCQ = Social Communication Questionnaire Total Score; RBS-R = Repetitive Behavior Scales-Revised Total Score; ADOS = Autism Diagnostic Observation Schedule; CSS = Calibrated Severity Score; CSBS CQ = Communication and Symbolic Behavior Scales Caregiver Questionnaire Total Scores; CBCL = Child Behavior Checklist T Scores