

The Structure of Autism Symptoms as Measured by the Autism Diagnostic Observation Schedule

Megan Norris · Luc Lecavalier · Michael C. Edwards

Published online: 20 August 2011
© Springer Science+Business Media, LLC 2011

Abstract The current study tested several competing models of the autism phenotype using data from modules 1 and 3 of the ADOS. Participants included individuals with ASDs aged 3–18 years ($N = 1,409$) from the AGRE database. Confirmatory factor analyses were performed on total samples and subsamples based on age and level of functioning. Three primary models were tested, including a one-factor model, the DSM-IV model, and the anticipated DSM-V model. Results indicated all models fit similarly. Module 1 ratings yielded better indices of fit across all models and higher inter-factor correlations than Model 3. Model fits were impacted by age and level of functioning. The lack of differentiation between models suggests that the structure of ASD symptoms is complex to measure statistically.

Keywords Confirmatory factor analysis · Autism · Autism Spectrum Disorders · Pervasive developmental disorders · Symptoms

Introduction

Much recent attention has been directed towards elucidating the structure of autistic symptoms. The heart of the matter is whether and how symptoms coalesce. This is important because such knowledge informs the use and development of instruments used to measure symptoms of Autism Spectrum Disorders (ASDs). Relatedly, these efforts also inform our conceptualization of diagnostic classification systems, an especially relevant point given the impending release of the DSM-V. A better understanding of the structure of ASD symptoms may also lead to a better understanding of the relationship between genes and behavior. In this respect, quantitative phenotypes have proven more useful than categorical designations (Szatmari et al. 2007a, b).

Factor analysis (FA) is one way researchers have examined symptom structure. A summary of such studies examining the structure of ASD symptoms can be found in Table 1. Thus far, results have been inconclusive. Studies largely support models composed of either two or three symptom domains. But there is variability in the composition as well as the number of factors identified. For instance, Frazier et al. (2008) submitted ADI-R data from 1,170 verbal individuals aged 2–46 years to FA and found support for a two-factor model, with a social/communication factor and a stereotyped behavior factor (which also included stereotyped language). A three-factor model also performed well, which separated difficulties in peer relationships and imaginative play from the social/communication factor (only containing two items). Likewise, using the ADI-R in a sample of 1,861 individuals with ASD aged 4–18 years, Snow et al. (2009) found that both two- and three-factor models were defensible.

Studies supporting a three-factor model have not always aligned with the three domains of the DSM-IV (e.g.,

M. Norris · L. Lecavalier (✉) · M. C. Edwards
Nisonger Center and Department of Psychology, The Ohio State University, Nisonger Center, 305 McCampbell Hall, 1581 Dodd Drive, Columbus, OH 43210, USA
e-mail: luc.lecavalier@osumc.edu

M. Norris
Strong Center for Developmental Disabilities and the Department of Pediatrics, The University of Rochester Medical Center, Rochester, NY, USA

Present Address:
M. Norris
Nationwide Children's Hospital Child Development Center, Columbus, OH, USA

Table 1 Summary of studies of ASD structure

Study	N	Instrument(s)	No. factors	Names	Notes (Methodology, sample details)
Boomsma et al. (2008)	263	ADI-R (12 subdomain scores)	3	Social/comm Stereotyped Speech/beh Impaired play skills ASD	CFA; verbal 4–25 year-olds with ASD
Constantino et al. (2004)	168	SRS (all items)	1		PCA; 245 parent ratings, 91 teacher ratings of 168 subjects aged 4–18 years with and without ASD
Frazier et al. (2008)	1,170	ADI-R 12 (subdomain scores)	2	Social/comm RRB	EFA, CFA; verbal 2–46 year-olds with ASD; also supported 3-factor model; AGRE database
Georgiades et al. (2007)	209	ADI-R (12 subdomain scores)	3	Social/comm Inflexible lang/beh RRB	PCA; verbal and nonverbal 2–40 year-olds with ASD; NVIQ range from 32 to 132 ($n = 183$)
Kamp-Beck et al. (Kamp-Becker et al. 2009)	140	ADI-R (47 items)	4	Social/comm Anxiety/compulsions Stereotyped beh Inadequate beh	PCA; verbal 6–24 years-olds with $FIQ \geq 70$; ASD and non-ASD individuals
		ADOS (24 items common between modules 3 and 4)	5	Social/comm Non/verbal beh Hyperactivity Stereotyped beh Interest and compulsions	Same methods and sample
Lecavalier et al. (2006)	226	ADI-R (algorithm items)	3	Social Communication RRB	EFA; verbal and nonverbal 5–17 year olds with ASD
Lecavalier et al. (2009)	730	ECI-4, CSI-4 (all items)	3	Social Communication RRB	CFA; included 3–12 year-olds with ASD
Mick (2005)	220	ADOS, Module 1 (algorithm items)	2	Social Communication RRB	PCA; 17–71 month-olds; ASD and non-ASD
	100	ADOS, Module 2 (algorithm items)	3	Social/comm Social language Joint attention	
Robertson et al. (1999)	51	PL-ADOS (social-communication algorithm items)	3	Joint attention Affective reciprocity Theory of mind	EFA; 3–6 year-olds with ASD

Table 1 continued

Study	N	Instrument(s)	No. factors	Names	Notes (Methodology, sample details)
Snow et al. (2009)	1,861	ADI-R (all items, algorithm items separately)	2	Social/comm RRB	EFA, CFA; 4–18 year-olds with ASD; also supported 3-factor DSM-IV model; AGRE database
Tadevosyan-Leyfer et al. (2003)	292	ADI, ADI-R (98 items common between both versions)	6	Spoken language Social intent Compulsions Dev milestones Savant skills Sensory aversions	PCA; 2–47 year-olds with ASD
van Lang et al. (2006)	255	ADI-R (12 subdomain scores)	3	Social/comm Stereotyped Speech/beh Impaired play skills	CFA; verbal and nonverbal 4–20 year-olds; ASD and non-ASD

Comm Communication, *Lang* Language, *Beh* Behavior, *Dev* Developmental

Georgiades et al. 2007; van Lang et al. 2006). The DSM-IV triad of symptom domains was largely based on clinical judgment, and has received limited empirical support. Additional models (e.g., a one-factor model) also have been proposed in the literature.

Results from studies may differ for several reasons, including the instrument used and type of statistical procedures employed. To date, most FA studies have used data from the ADI-R (8 out of 12 studies in Table 1). Perhaps the ADI-R has been used most often in this line of research because of its strong psychometric properties and the fact that it contains many items from each of the three DSM domains. However, use of other well-validated instruments can be informative. The ADOS is an ideal candidate for several reasons. It has been extensively validated (e.g., de Bildt et al. 2004; Mazefsky and Oswald 2006) and is based on clinician observation, thereby providing a different source of information. Recently, the ADOS scoring algorithms have been revised (Gotham et al. 2008; Gotham et al. 2007) to improve diagnostic validity. The revised algorithms also have been validated in an independent sample (Oosterling et al. 2010).

Methodological and procedural variations in the statistics can also impact FA results. For instance, while some authors use FA, others use PCA. Extraction methods and type of rotations also vary across studies. Some studies have analyzed data at the item level, while others have used subscales. Of course, results are dependent on the nature/content of items that are analyzed (i.e., the phenotype is limited to the behaviors that are analyzed).

The nature of the sample can also impact findings. As illustrated in Table 1, many previous studies have included a wide range of ages, verbal abilities, and/or levels of functioning. This complicates comparisons across studies, as research has shown that ASD symptoms change over time (Charman et al. 2005; Happé et al. 2006; Piven et al. 1996). Since ASD symptom patterns change with age, results may be ‘muddied’ in heterogeneous samples (e.g., if preschool-aged children and older children/adults are included in the same analyses). Other studies have shown that symptom severity and expression vary with level of functioning (e.g., Dawson et al. 2007; Munson et al. 2008; Szatmari et al. 2006). Examination of ASD subtypes also has indicated that variability among those with ASDs is perhaps best explained by differences in IQ (Witwer and Lecavalier 2008). In addition to providing further reason to revisit DSM criteria, these studies point to the importance of examining homogenous subgroups of those with ASDs.

The primary goal of the current study was to examine the structure of ASD symptoms as measured by the ADOS in a large sample of children and adolescents with ASD. In doing so, this study complemented previous research by using an observation-based measure of current behaviors. This study is unique in that it included all ADOS items

related to DSM domains and it examined several models of ASD symptoms within the same sample.

A secondary goal was to examine how different subgroups impacted model fit. Because previous studies have demonstrated that ASD symptoms may manifest differently as a function of IQ and age, the sample was split according to these variables. While the ADOS controls for level of functioning and age to some extent through appropriate module selection, there can still be significant variability in terms of age or level of functioning within a given module.

Methods

Participants

Participant data were collected from a large database available for research purposes, the Autism Genetic

Resource Exchange (AGRE; Geschwind et al. 2001). The majority of children in the AGRE database are from families in which at least two members have an ASD diagnosis (i.e., multiplex families). Participants with an ASD diagnosis are referred to AGRE; diagnoses are then confirmed via team consensus, based in part on both ADI-R and ADOS scores. Individuals are assigned to one of four diagnostic categories: Autism, Not Quite Autism (NQA, those within one point of domain autism cut-offs on the ADI-R), Broad Spectrum (more than one point away from domain cut-offs on the ADI-R, but still determined to fall on the spectrum), or Not Spectrum. Individuals were excluded if they were referred with an ASD but categorized as ‘Not Spectrum’ by AGRE researchers; all other individuals were included if they met spectrum cut-offs on the ADOS and had complete data.

Children and adolescents aged 3–18 years were included in analyses. Modules 1 and 3 were examined because

Table 2 Summary of module 1 participant characteristics

Module 1	Total (<i>N</i> = 720)	Under 6 (<i>n</i> = 290)	AB \leq 55 (<i>n</i> = 339)
Age years (M, SD)	7.2 (3.1)	4.5 (0.84)	7.9 (3.1)
Age range	3.02–17.81	3.02–6.0	3.1–17.81
Gender (males) <i>n</i> (%)	561 (78%)	224 (77%)	266 (79%)
Ethnicity	<i>n</i> = 605	<i>n</i> = 237	<i>n</i> = 288
Non-hispanic/latino	480 (79%)	184 (78%)	234 (81%)
Hispanic/latino	125 (21%)	53 (22%)	54 (19%)
Diagnostic category	<i>n</i> = 636	<i>n</i> = 252	<i>n</i> = 303
Autism <i>n</i> (%)	611 (96%)	238 (94%)	293 (97%)
NQA	11 (2%)	8 (3%)	6 (2%)
Broad spectrum	14 (2%)	6 (2%)	4 (1%)
ADOS classification			
Autism	665 (92%)	265 (91%)	314 (93%)
Spectrum	55 (8%)	25 (9%)	25 (7%)
ADOS domain scores (M, SD)			
Social	10.8 (2.2)	10.9 (2.4)	11.0 (2.1)
Range	4–14	4–14	4–14
Communication	6.2 (1.6)	6.2 (1.6)	6.3 (1.6)
Range	2–10	2–10	2–10
RRB	3.9 (1.5)	3.9 (1.6)	3.9 (1.5)
Range	0–6	0–6	0–6
ADOS revised domain scores (M, SD)			
Social affect	15.7 (3.2)	15.9 (3.4)	16.0 (3.1)
Range	5–20	5–20	5–20
RRB	5.3 (1.9)	5.1 (1.8)	5.2 (1.9)
Range	0–8	0–8	0–8
VABS (M, SD)	<i>n</i> = 542	<i>n</i> = 211	<i>n</i> = 339
AB composite	46 (15.8)	54.1 (12.9)	36.6 (11)
DLS SS	44.6 (18.5)	52.4 (14.3)	33.5 (13)
Raven NVIQ	<i>n</i> = 185	<i>n</i> = 46	<i>n</i> = 68
Mean, SD	89.4 (21.3)	102.8 (14.2)	86.1 (22.9)
Range	28–143	73–132	28–132

Under 6 under 6 years of age,
AB Adaptive behavior

Table 3 Summary of module 3 participant characteristics

Module 3	Total (N = 689)	Over 10 (n = 277)	AB ≤70 (n = 229)	AB ≥70 (n = 217)	IQ ≥100 (n = 399)
Age years (M, SD)	9.5 (2.7)	12.2 (1.7)	10.3 (2.8)	8.7 (2.4)	8.9 (2.3)
Age range	4.2–17.5	10.0–17.5	4.2–17.1	4.2–16.1	4.2–17.1
Gender (males) n (%)	549 (80%)	227 (82%)	186 (81%)	164 (76%)	323 (81%)
Ethnicity	n = 573	n = 224	n = 193	n = 185	n = 338
Non-hispanic/latino	460 (80%)	181 (81%)	153 (79%)	151 (82%)	277 (82%)
Hispanic/latino	113 (20%)	43 (19%)	40 (21%)	34 (18%)	61 (18%)
Diagnostic category	n = 599	n = 234	n = 206	n = 188	n = 345
Autism	542 (90%)	221 (94%)	188 (91%)	158 (84%)	302 (88%)
NQA	21 (4%)	6 (3%)	7 (3%)	11 (6%)	13 (4%)
Broad spectrum	36 (6%)	7 (3%)	11 (5%)	19 (10%)	30 (9%)
ADOS classification					
Autism	479 (70%)	201 (73%)	181 (79%)	121 (56%)	258 (65%)
Spectrum	177 (26%)	63 (23%)	37 (16%)	83 (38%)	123 (31%)
Not spectrum	33 (5%)	13 (5%)	11 (5%)	13 (6%)	18 (5%)
ADOS domain scores (M, SD)					
Social	8.2 (2.5)	8.3 (2.6)	8.6 (2.5)	7.7 (2.5)	8.0 (2.5)
Range	0–14	0–14	2–14	1–14	1–14
Communication	3.8 (1.6)	3.8 (1.6)	4.0 (1.6)	3.4 (1.5)	3.7 (1.5)
Range	0–8	0–8	0–7	0–8	0–8
RRB	1.8 (1.6)	1.8 (1.6)	2.1 (1.8)	1.5 (1.5)	1.7 (1.6)
Range	0–8	0–6	0–8	0–6	0–8
ADOS revised algorithms (M, SD)					
Social affect	10.2 (3.9)	10.5 (3.9)	10.9 (4.0)	9.2 (3.7)	9.9 (3.8)
Range	0–20	1–20	2–20	0–19	1–20
RRB	2.6 (1.7)	2.6 (1.7)	2.7 (1.8)	2.4 (1.7)	2.5 (1.7)
Range	0–8	0–8	0–8	0–7	0–8
VABS (M, SD)	n = 511	n = 213	n = 229	n = 217	n = 295
AB composite	69.1 (16.4)	63.6 (15.5)	55.2 (10.8)	83.2 (9.7)	72.1 (15.7)
DLS SS	67.9 (20.6)	64.9 (20.6)	50.7 (15.6)	85.4 (9.8)	70.5 (20.5)
Raven NVIQ	n = 617	n = 260	n = 202	n = 203	n = 399
M, SD	104.0 (17.3)	95.9 (17.0)	99.4 (18.0)	108.7 (15.8)	114.0 (10.6)
Range	43–140	43–131	43–140	58–140	100–140

Over 10 Over 10 years of age, AB Adaptive behavior

they are the most frequently used modules in children and adolescents. There were a total of 1,409 participants, with 720 individuals for Module 1 and 689 individuals for Module 3. Sample characteristics are reported in Tables 2 and 3.

Instruments

The *Autism Diagnostic Observation Schedule* (ADOS; Lord et al. 2000) is a semi-structured, standardized observation for the assessment of ASD. It consists of social “presses” which require a trained examiner to engage an individual in various social interactions. It consists of four modules; the appropriate module is selected based on the individual’s expressive language level and chronological age. Administration of

each module requires approximately 30–45 min. Most ADOS items are coded from zero (0, no evidence of abnormality) to three (3, markedly abnormal behavior). A code of seven (7) is used when the behavior is abnormal but in a different way than the protocol specifies and eight (8) is used when the rating is not applicable. For analyses, ratings of seven and eight were transformed to zeros. Module 1 consists of 10 activities and 29 ratings; 12 items are part of the scoring algorithm. Module 3 consists of 14 activities and 28 ratings; 11 items are part of the scoring algorithm.

The *Raven’s Progressive Matrices* (Raven et al. 2003) is a nonverbal measure of cognitive functioning. It consists of five sets of 12 items, with subsequent items increasing in difficulty. Estimated non-verbal age was used to select participants for the high-functioning subgroup in Module 3.

The *Vineland Adaptive Behavior Scale* (VABS; Sparrow et al. 1984) is a semi-structured interview of adaptive behavior. Items are organized into four domains: Communication, Daily Living Skills, Socialization, and Motor Skills. An overall composite score also is obtained. Both the overall composite and the Daily Living Skills standard score were used to select participants for the low-functioning subgroups. This was because the composite score incorporates functioning in areas essential to the diagnosis of autism (i.e., socialization and communication) so a more independent domain was also used as criterion to ensure low-functioning status was not the result of severe autistic symptomology.

Procedure

Participant families for AGRE are found through various sources. At initial contact, a packet is given to families containing information about the program and a preliminary enrollment form. On the form, families identify the diagnosis they have received from their physician or specialist; Autism, PDD-NOS, and Asperger's Disorder are accepted. Then diagnoses are confirmed via team consensus, based in part on both ADI-R and ADOS scores. Raters are assessed periodically to ensure that they are research-reliable (Autism Genetic Resource Exchange 2009).

Statistical Analyses

Confirmatory Factor Analysis

Modules 1 and 3 were examined by CFA individually. Analyses were conducted on the entire module sample, then also on subgroups based age and level of functioning. CFA was performed using LISREL (Jöreskog and Sörbom 2004). Polychoric correlation matrices were used due to the ordinal nature of the data. Diagonally weighted least squares (DWLS) was used for estimation because it is preferred over other methods when data are categorical, and it requires smaller sample sizes than other methods (Wirth and Edwards 2007).

Measures of Fit

There are many different indices of fit, but little consensus regarding which measure (or combination of measures) is best to use for evaluating models (Bollen 1989). Often indices can disagree with each other. Indices in this study were chosen in order to follow recommendations which suggest selecting indices from different categories to reflect various criteria (Brown 2006; Garson 2009) and in order to maximize the ability to compare this study to previous work.

RMSEA was selected as the primary measure of model fit because it is one of the measures least impacted by

sample size and does not involve a null model (M. W. Browne, personal communication, June 1, 2010). It also has a confidence interval and graduated guidelines exist for its interpretation (rather than a single cut-off). Finally, it is commonly reported in the literature, allowing for direct comparison between previous studies and results found here. Additional measures considered were the Non-Normed Fit Index (NNFI), Closeness of Fit Index (CFI), Standardized Root Mean Square Residual (SRMR), and Goodness of Fit Index (GFI). Although recently GFI may have fallen out of favor as a preferred measure of fit (Garson 2009) it is frequently reported in previous studies, so was included here. Guidelines have been proposed to help researchers interpret goodness of fit indices, though there is some variability in recommendations. Generally accepted cut-offs were used (i.e., for the RMSEA, values less than 0.08 were considered acceptable; for the NNFI, CFI, and GFI, values greater than 0.95 were considered acceptable; and for the SRMR, values less than 0.10 were considered acceptable).

Model Specification

Three different models were tested within each module (total sample and subgroup). First, a one-factor model was specified, in which all items loaded onto a single factor. The second model was the three-domain DSM-IV model (as reported in the DSM-IV-TR; American Psychiatric Association 2000) that includes social impairments, communication, and repetitive behaviors and interest (see Fig. 1). Finally, a third model based on recently proposed changes to the DSM was specified (American Psychiatric Association 2010) called hereafter the DSM-V model (see Fig. 2). It consisted of one factor of social and communication items and one factor of restricted and repetitive behavior and language (RRB/L) items.

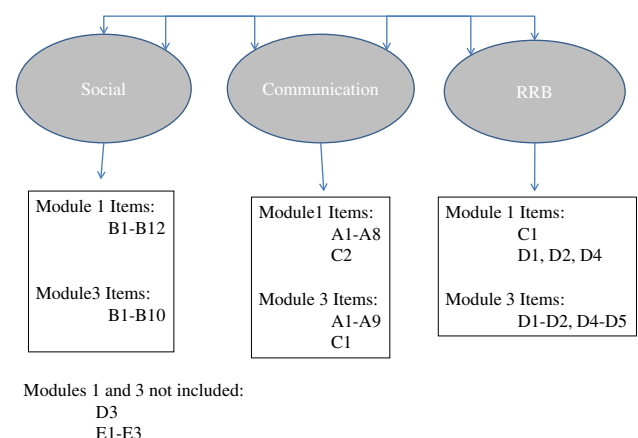


Fig. 1 DSM-IV model

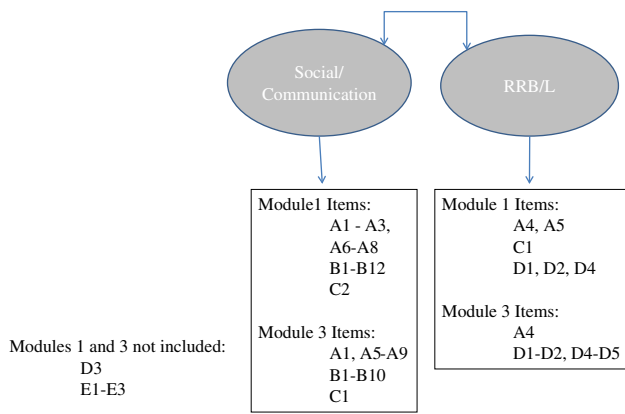


Fig. 2 Proposed DSM-V model

Subgroups

In order to examine the impact of subject characteristics, modules were divided into more homogenous subgroups by age and level of functioning (as determined by IQ or adaptive behavior scores). Decisions related to dividing modules 1 and 3 were made with two goals in mind: to create homogenous subgroups based on variables of interest, and to retain enough participants for stable factor solutions. As such, analyses were conducted on modules as a whole first and then by subgroup. A “youngest subgroup” in Module 1 consisted of those 6 years and younger and an “oldest subgroup” in Module 3 consisted of those 10 years and older. The subgroup analyses based on level of functioning were as follows: a “lowest-functioning subgroup” in Module 1 was based on VABS composite and DLS < 55; a “lowest-functioning subgroup” in Module 3 was based on VABS composite and DLS ≤ 70. There were two higher functioning groups in Module 3: a “highest-functioning subgroup” based on VABS composite and DLS > 70 and a “highest-functioning subgroup” based on Raven estimated NVIQ ≥ 100. This resulted in subgroups ranging from 217

participants (Module 3, VABS Composite and DLS ≥ 70) to 399 participants (Module 3, Raven estimated NVIQ ≥ 100). It was necessary to divide groups in this manner due to missing data for many participants (e.g., the Raven was not administered to participants if they were too low-functioning, too young, or engaged in difficult behaviors which precluded test administration).

Results

Results are presented by module, with Module 1 first. Unless specified otherwise, results are presented for total sample analyses.

Module 1

Model Fit

Selected indices of fit for Module 1, all models, for the total sample and both subsamples are presented in Table 4.

Within the total Module 1 sample ($N = 720$), RMSEAs ranged from 0.056 (DSM-IV model) to 0.062 (the one-factor model). RMSEA confidence intervals overlapped, indicating that based on the RMSEA, no single model fit the data significantly better than other models. However, the DSM-IV model performed somewhat better than the other models. When other measures of model fit were considered, all three models had acceptable to good fits. In general, subsample fits were an improvement over the total sample fits. Fits tended to be best for each model within the lowest-functioning children.

Factor Loadings

Overall, the mean factor loadings for the total sample analyses were on the order of 0.50–0.60 for social and

Table 4 Module 1 indices of fit across models

	RMSEA (90% CI)	NNFI	CFI	SRMR	GFI
<i>One factor model</i>					
$N = 720$, total	0.062 (0.058–0.066)	0.96	0.96	0.088	0.96
$n = 290$, youngest	0.059 (0.052–0.066)	0.97	0.98	0.105	0.97
$n = 339$, lowest functioning	0.055 (0.049–0.062)	0.97	0.97	0.096	0.95
<i>DSM-IV model</i>					
$N = 720$, total	0.056 (0.052–0.06)	0.97	0.97	0.084	0.96
$n = 290$, youngest	0.057 (0.05–0.065)	0.96	0.98	0.102	0.97
$n = 339$, lowest functioning	0.050 (0.043–0.056)	0.97	0.98	0.093	0.96
<i>DSM-V model</i>					
$N = 720$, total	0.061 (0.057–0.065)	0.96	0.96	0.087	0.96
$n = 290$, youngest	0.059 (0.053–0.066)	0.97	0.97	0.105	0.97
$n = 339$, lowest functioning	0.054 (0.048–0.061)	0.97	0.97	0.096	0.95

communication factors and 0.35–0.4 for the RRB and RRB/L factors. For the one-factor model, the mean item loading was 0.50. For the DSM-IV model, the average loadings were: social = 0.59, communication = 0.51, and RRB = 0.37. For the DSM-V model, mean loadings were: social-communication = 0.54, and RRB/L = 0.41.

Association between Factors

The inter-factor correlations for each model were quite high. For the DSM-IV model the intercorrelations were: social and communication factors = 0.86, social and RRBs = 0.75, and communication and RRBs = 0.92. For the DSM-V model, correlation between factors was 0.89.

Module 3

Model Fit

Selected indices of fit for Module 3, all models, for the total sample and all subsamples are presented in Table 5.

Within the total Module 3 sample ($N = 689$) RMSEAs ranged from 0.076 (DSM-V model) to 0.083 (the one-factor model). As within Module 1, confidence intervals for RMSEAs overlapped between models, indicating similarities in model fit across models. Overall, the DSM-V model was slightly preferable as two indices fell within acceptable ranges.

The only subgroup analysis that yielded slightly better indices of fit than the total sample was with children over 10 years ($n = 277$). Analyses with the two highest-

functioning subgroups yielded poor fit, and the poorest fits for each model were found with the lowest-functioning subgroup ($n = 299$). As with the total sample, there was little difference between the three models in the subgroup analyses.

Factor Loadings

As in Module 1, the average factor loadings were approximately 0.50–0.60 for social and communication factors and 0.40 for the RRB and RRB/L factors. For the one-factor model, the mean loading was 0.47. For the DSM-IV model, the mean loadings were: social = 0.57, communication = 0.51, and RRB = 0.43. For the DSM-V model, mean loadings were: social-communication = 0.57 and RRB/L = 0.41.

Association between Factors

Inter-factor correlations for each model ranged from modest to strong. Inter-factor correlations for the DSM-IV model were: between social and communication factors = 0.93, between social and RRBs = 0.36, and between communication and RRBs = 0.33. The DSM-V model inter-factor correlation was: 0.48.

Discussion

This study examined the structure of ASD symptoms with items from the ADOS in a large group of children and

Table 5 Module 3 indices of fit across models

	RMSEA (90% CI)	NNFI	CFI	SRMR	GFI
<i>One factor model</i>					
$N = 689$, total	0.083 (0.079–0.087)	0.91	0.92	0.102	0.92
$n = 277$, oldest	0.079 (0.072–0.086)	0.93	0.93	0.111	0.94
$n = 229$, low functioning	0.148 (0.141–0.156)	0.71	0.74	0.174	0.89
$n = 217$, highest AB	0.086 (0.078–0.094)	0.91	0.92	0.122	0.92
$n = 399$, highest IQ	0.105 (0.103–0.107)	0.87	0.88	0.131	0.89
<i>DSM-IV model</i>					
$N = 689$, total	0.079 (0.074–0.083)	0.92	0.93	0.099	0.94
$n = 277$, oldest	0.076 (0.069–0.083)	0.93	0.94	0.108	0.94
$n = 229$, low functioning	0.144 (0.136–0.151)	0.73	0.76	0.167	0.90
$n = 217$, highest AB	0.081 (0.073–0.089)	0.92	0.93	0.119	0.92
$n = 399$, highest IQ	0.091 (0.086–0.097)	0.90	0.91	0.128	0.90
<i>DSM-V model</i>					
$N = 689$, total	0.076 (0.072–0.080)	0.93	0.93	0.096	0.94
$n = 277$, oldest	0.073 (0.066–0.080)	0.94	0.94	0.108	0.94
$n = 229$, low functioning	0.147 (0.14–0.154)	0.72	0.75	0.170	0.89
$n = 217$, highest AB	0.077 (0.069–0.085)	0.93	0.93	0.117	0.92
$n = 399$, highest IQ	0.092 (0.086–0.097)	0.90	0.90	0.128	0.90

adolescents with ASDs. Taken as a whole, the analyses did not support any single model over others. They also suggested that development may impact how symptoms coalesce. Indeed, data from Modules 1 and 3 yielded significantly different results, despite having most items in common. Overall, analyses in Module 1 yielded better fits across all models than analyses in Module 3. Indices of fit for Module 1 were acceptable, but suggested room for improvement. However, within Module 3, most indices were poor or less than acceptable. This pattern was more evident in subgroup analyses: Module 1 fits generally improved within subgroup analyses, whereas they generally worsened in Module 3. These results may help explain why so many plausible solutions are proposed in the literature: ultimately, there may not be much difference between the models, and fits may vary significantly with sample characteristics.

In Module 1, there was little differentiation between models. There is little reason to prefer one model over the others given that they all have some theoretical support. That being said, the DSM-IV model had slightly better indices of fit than other models, which echoes findings with the ADI-R and ECI/CSI (Lecavalier et al. 2006, 2009). Fit indices within subgroups generally did improve a little, but not to the extent that any model was clearly preferable.

There was no clear-cut winner between models in Module 3, but the DSM-V model did appear to perform marginally better than the others. This is encouraging as the DSM-V criteria have only recently been proposed and have yet to be studied. Although indices of fit in Module 3 were worse than in Module 1, they were comparable to several previous studies. For instance, Frazier et al. (2008) reported the following indices for the best models in their analyses: RMSEA = 0.07, CFI = 0.94, and NNFI = 0.92. Georgiades et al. (2007) found an RMSEA of 0.067, SRMR of 0.08, CFI of 0.92, and NNFI of 0.90 for their proposed model. Fit indices within subgroups were more variable than in Module 1, with model fit improving in the oldest subgroup but generally worsening in the other subgroups.

Sample composition is an important variable to consider in factor analytic studies. A possible explanation for the notable difference in fits across modules could be that Module 1 participants had more severe ASD symptoms than Module 3 participants, resulting in more clear-cut symptom relationships to latent factors. ADOS domain scores and autism diagnosis rates support this idea. Module 1 participants were generally lower functioning than those in Module 3. Comparatively, Module 3 may have been administered to individuals with a wider range of functioning, which may have impacted results.

Though analyses of the more homogenous subgroups did not yield expected results, they did provide evidence

that sample characteristics impact fit. Fit indices changed, sometimes drastically, when homogenous subgroups were analyzed. The poorest indices of fit of all analyses were found in the lowest-functioning subgroup of Module 3. This could be a reflection of Module 3 being less appropriate for lower-functioning individuals. The best fits were found in the lowest functioning subgroup of Module 1. This may indicate that autism symptoms, at least as they are currently defined by diagnostic instruments, are best measured in relatively young, lower-functioning individuals. This converges with previous work which has shown that DSM-IV criteria are most accurate in school-aged children with mild-to-moderate delays in functioning (see Lord and Corsello 2005).

Association between Factors

In Module 1, inter-factor correlations indicated a strong relationship between factors in all models (range from 0.75 to 0.89). Even the RRB and RRB/L correlated highly with other factors. This is in contrast to previous research indicating that the domains of ASD symptoms may be relatively distinct (e.g., Ronald et al. 2005; Ronald et al. 2006). However, the two studies by Ronald and colleagues were composed of community-based samples, not ASD-specific samples, which likely resulted in greater variability in scores. Indeed, other studies have reported a strong relationship between domains, including the RRB domain (e.g., Wing and Gould 1979; Lecavalier et al. 2009).

In Module 3, inter-factor correlations were still quite strong between social and communication factors, but much lower between RRB and RRB/L factors and other domains as compared to Module 1. Together, these findings may suggest that core ASD symptom domains become more distinct with development. These findings also converge with other studies (e.g., Charman et al. 2005) showing that the developmental trajectories of domains do indeed differ. The very high inter-factor correlations between social and communication factors across both modules provide some empirical support for collapsing these domains into one (as with the DSM-V model). However, the variability in inter-factor correlations for RRB and RRB/L factors across modules is less decisive. Research into the relationship between domains has been inconclusive, and further examination of this topic is necessary.

Considering both modules, one important point is that overall, the newly proposed DSM-V criteria performed as well as the other models. This is one of the first studies to explicitly evaluate this new model and to provide some evidence supporting it. The original conceptualization of autism discussed two main symptom domains: social/communication impairments and RRBs (Kanner 1943).

This study provides evidence that the social and communication domains may be so closely related that they represent one factor (as in Kanner's work and the recent DSM-V conceptualization). However, as discussed above, inter-factor correlations across modules for the DSM-IV RRB and DSM-V RRB/L factors were more variable. In summary, this indicates the DSM-V model is promising, but the RRB/L domain may need further examination.

The primary method by which models were compared in this study was by examining different indices of fit, most notably the RMSEA. It is important to note that the guidelines for interpretation of these measures are arbitrary. Researchers need to consider several measures of fit as well as clinical and theoretical meaningfulness of solutions. It is important to note that even unacceptable fits can represent progress when compared to previous results (Bollen 1989). It is also important to remember that not all indices of fit yield similar conclusions. For instance, there is less variability in NNFI values across modules and subsamples compared to RMSEA values. Furthermore, it is not clear how indices of fit translate into clinical significance. A model with an RMSEA of 0.03 will not necessarily be more clinically meaningful than a model with an RMSEA of 0.06. Research into this topic is lacking.

Limitations

Almost all participants in this study came from multiplex families. Because of this, results may not generalize to individuals from simplex families. Additionally, all individuals had a significant level of ASD symptoms. This can lead to decreased variability in the data. Some researchers have argued that studying autistic symptoms among individuals with ASD may artificially inflate associations (e.g., Happé et al. 2006). Further examination of models within community samples or family members with ASD symptoms without a full-blown ASD might increase variability in scores and provide valuable information in understanding the structure of ASD symptoms. Additionally, because the ADOS controls somewhat for language level, age, and level of functioning, even total sample analyses were (theoretically) being performed on somewhat homogenous samples. This may have led to less variability in scores, which could have resulted in less differentiation between models.

This was a sample of convenience. The nature of the AGRE database dictates that not every child is rated on each instrument, due to age, level of functioning, or behavior difficulties. For this reason, some information (e.g., NVIQ scores) was not available for some participants. Additionally, creating more homogenous subgroups to examine the impact of age and level of functioning was difficult due to missing information. It was impossible to

use both IQ and adaptive behavior to create subgroups for level of functioning, and if subgroups had been independent of each other, they would not have been large enough for CFA to be performed. For these reasons results of subgroup analyses, especially those based on level of functioning, should be interpreted with caution.

Future Directions

Elucidation of the structure of ASD symptoms will necessitate different research strategies and data sources. Continued work with large samples and homogenous subgroups offers a number of advantages. If the trend for large research databases continues, such studies may become possible sooner rather than later. Used in isolation, factor analysis is unlikely to yield a completely satisfying answer to the question: "what is the structure of ASD symptoms?" With the variety of etiologies related to the variable phenotypes, this question is obviously a complicated one. Genetic and family studies suggest that a single underlying etiology is unlikely to explain the observed behavioral heterogeneity. For example, some studies have demonstrated that over half the genes which contribute to variation in one symptom domain are independent from those responsible for variation in the other domains (Ronald et al. 2005; Ronald et al. 2006). Ultimately, a growing body of literature is indicating that many different phenotypes make up ASDs, each associated with different symptom profiles, and each possibly with distinct etiologies (Abrahams and Geschwind 2008).

References

- Abrahams, B. S., & Geschwind, D. H. (2008). Advances in autism genetics: On the threshold of a new neurobiology. *Nature Reviews*, *9*, 341–355.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders—text revised* (4th ed.). Washington, DC: APA.
- American Psychiatric Association. (2010). DSM-V development: Autistic disorder. <http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=94>. Retrieved March 5, 2010.
- Autism Genetic Resource Exchange (AGRE). (2009). The AGRE program. <http://www.agre.org/program/intro.cfm?do=program>. Retrieved July 7, 2009.
- Bollen, K. A. (1989). *Structural equations with latent variables*. USA: John Wiley and Sons, Inc.
- Boomsma, A., van Lang, N. D. J., de Jonge, M. V., de Bildt, A. A., van Engeland, H., & Minderaa, R. B. (2008). A new symptom model for autism cross validated in an independent sample. *Journal of Child Psychology and Psychiatry*, *49*, 809–816.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. NY: The Guilford Press.
- Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism

- at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry*, 45, 500–513.
- Constantino, J. N., Gruber, C. P., Davis, S., Hayes, S., Passanante, N., & Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry*, 45, 719–726.
- Dawson, G., Estes, A., Munson, J., Schellenberg, G., Bernier, R., & Abbott, R. (2007). Quantitative assessment of autism symptom-related traits in probands and parents: Broader phenotype autism symptom scale. *Journal of Autism and Developmental Disorders*, 37, 523–536.
- de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., et al. (2004). Interrelationship between the autism diagnostic observation schedule-generic (ADOS-G), autism diagnostic interview-revised (ADI-R), and the diagnostic and statistical manual of mental disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders*, 34, 129–137.
- Frazier, T. W., Youngstrom, E. A., Kubu, C. S., Sinclair, L., & Rezaei, A. (2008). Exploratory and confirmatory factor analysis of the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders*, 38, 474–480.
- Garson, G. D. (2009). Structural equation modeling. <http://faculty.chass.ncsu.edu/garson/PA765/structur.htm>. Retrieved February 2, 2010.
- Georgiades, S., Szatmari, P., Zwaigenbaum, L., Duku, E., Bryson, S., Roberts, W., et al. (2007). Structure of the autism symptom phenotype: A proposed multidimensional model. *Journal of the American Academy of Child and Adolescent Psychiatry*, 46, 188–196.
- Geschwind, D. H., Sowiński, J., Lord, C., Iversen, P., Shestack, J., Jones, P., et al. (2001). The autism genetic resource exchange: A resource for the study of autism and related neuropsychiatric conditions. *American Journal of Human Genetics*, 69, 463–466.
- 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, 613–627.
- Gotham, K., Risi, S., Dawson, G., Tager-Flusberg, H., Joseph, R., Carter, A., et al. (2008). A replication of the autism diagnostic observation schedule (ADOS) revised algorithms. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 642–651.
- Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, 9, 1218–1220.
- Jöreskog, K. G., & Sörbom, D. (2004). *LISREL 8.7 for windows [computer software]*. Lincolnwood, IL: Scientific Software International, Inc.
- Kamp-Becker, I., Ghahreman, M., Smidt, J., & Remschmidt, H. (2009). Dimensional structure of the autism phenotype: Relations between early development and current presentation. *Journal of Autism and Developmental Disorders*, 39, 557–571.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250.
- Lecavalier, L., Aman, M. G., Scahill, L., McDougle, C. J., McCracken, J. T., Vitiello, B., et al. (2006). Validity of the autism diagnostic interview-revised. *American Journal on Mental Retardation*, 111, 199–215.
- Lecavalier, L., Gadow, K. D., DeVincent, C. J., Houts, C., & Edwards, M. C. (2009). Deconstructing the PDD clinical phenotype: Internal validity of the DSM-IV. *The Journal of Child Psychology and Psychiatry*, 50, 1246–1254.
- Lord, C., & Corsello, C. (2005). Diagnostic instrument in autistic spectrum disorders. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders, vol.2: Assessments, interventions, and policies* (pp. 730–771). Hoboken: John Wiley and Sons, Inc.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30, 205–223.
- Mazefsky, C. A., & Oswald, D. P. (2006). The discriminative ability and diagnostic utility of the ADOS-G, ADI-R, and GARS for children in a clinical setting. *Autism*, 10, 533–549.
- Mick, K. A. (2005). Diagnosing autism: Comparing the childhood autism rating scale (cars) and the autism diagnostic observation schedule (ADOS). (Doctoral dissertation, Wichita State University, 2005). *Dissertation Abstracts International*. 67(6068).
- Munson, J., Dawson, G., Sterling, L., Beauchaine, T., Zhou, A., Koehler, E., et al. (2008). Evidence for latent classes of IQ in young children with autism spectrum disorder. *American Journal on Mental Retardation*, 113, 439–452.
- Oosterling, I., Roos, S., de Bildt, A., Rommelse, N., de Jonge, M., Visser, J., et al. (2010). Improved diagnostic validity of the ADOS revised algorithms: A replication study in an independent sample. *Journal of Autism and Developmental Disorders*, 40, 689–703.
- Piven, J., Harper, J., Palmer, P., & Arndt, S. (1996). Course of behavior change in autism: A retrospective study of high-IQ adolescent and adults. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35, 523–529.
- Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's progressive matrices and vocabulary scales*. San Antonio, TX: Harcourt Assessment.
- Robertson, J. M., Tanguay, P. E., L'Ecuyer, S., Sims, A., & Waltrip, C. (1999). Domains of social communication handicap in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 738–745.
- Ronald, A., Happé, F., & Plomin, R. (2005). The genetic relationship between individual differences in social and nonsocial behaviours characteristic of autism. *Developmental Science*, 8, 444–458.
- Ronald, A., Happé, F., Bolton, P., Butcher, L. M., Price, T. S., Wheelwright, S., et al. (2006). Genetic heterogeneity between the three components of the autism spectrum: A twin study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 691–699.
- Snow, A., Lecavalier, L., & Houts, C. (2009). Structure of the autism diagnostic interview-revised: Diagnostic and phenotypic implications. *Journal of Child Psychology and Psychiatry*, 50, 734–742.
- Sparrow, S. S., Balla, D., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales (Survey ed.)*. Circle Pines, MN: American Guidance Service.
- Szatmari, P., Georgiades, S., Bryson, S., Zwaigenbaum, L., Roberts, W., Mahoney, W., et al. (2006). Investigating the structure of the restricted, repetitive behaviors and interests domain of autism. *Journal of Child Psychology and Psychiatry*, 47, 582–590.
- Szatmari, P., Maziade, M., Zwaigenbaum, L., Mérette, C., Roy, M., Joobor, R., et al. (2007a). Informative phenotypes for genetic studies of psychiatric disorders. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)*, 144B, 581–588.
- Szatmari, P., White, J., & Merikangas, K. R. (2007b). The use of genetic epidemiology to guide classification in child and adolescent psychopathology. *International Review of Psychiatry*, 19, 483–496.
- Tadevosyan-Leyfer, O., Dowd, M., Mankoski, R., Winklosky, M. A., Putnam, S., McGrath, L., et al. (2003). A principal components analysis of the autism diagnostic interview—revised. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 864–872.

- van Lang, N. D. J., Boomsma, A., Sytema, S., de Bildt, A. A., Kraijer, D. W., Ketelaars, C., et al. (2006). Structural equation analysis of a hypothesized symptom model in the autism spectrum. *Journal of Child Psychology and Psychiatry*, *47*, 37–44.
- Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, *9*, 11–29.
- Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*, 58–79.
- Witwer, A. N., & Lecavalier, L. (2008). Examining the validity of autism spectrum disorder subtypes. *Journal of Autism and Developmental Disorders*, *38*, 1611–1624.