THE JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY

ACAMH

Journal of Child Psychology and Psychiatry 54:2 (2013), pp 206–215 doi:10.1111/j.1469-7610.2012.02588.x

Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach

Stelios Georgiades,[1] Peter Szatmari,[1] Michael Boyle,[1] Steven Hanna,[1] Eric Duku,[1] Lonnie Zwaigenbaum,[2] Susan Bryson,[3] Eric Fombonne,[4] Joanne Volden,[2] Pat Mirenda,[5] Isabel Smith,[3] Wendy Roberts,[6] Tracy Vaillancourt,[7] Charlotte Waddell,[8] Teresa Bennett,[1] Ann Thompson,[1] and Pathways in ASD Study Team

[1]Department of Psychiatry and Behavioural Neurosciences, McMaster University and Offord Centre for Child Studies, Hamilton, ON, Canada; [2]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada; [3]Departments of Pediatrics and Psychology, Dalhousie University and IWK Health Center, Halifax, NS, Canada; [4]Departments of Psychiatry and Pediatrics, McGill University, Montreal, QC, Canada; [5]Department of Educational and Counselling Psychology and Special Education, University of British Columbia, Vancouver, BC, Canada; [6]Department of Pediatrics and Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, University of Toronto, Toronto, ON, Canada; [7]Faculty of Education and School of Psychology, University of Ottawa, Ottawa, ON, Canada; [8]Faculty of Health Sciences, Simon Fraser University, Vancouver, BC, Canada

Background: Autism spectrum disorder (ASD) is characterized by notable phenotypic heterogeneity, which is often viewed as an obstacle to the study of its etiology, diagnosis, treatment, and prognosis. On the basis of empirical evidence, instead of three binary categories, the upcoming edition of the DSM 5 will use two dimensions – social communication deficits (SCD) and fixated interests and repetitive behaviors (FIRB) – for the ASD diagnostic criteria. Building on this proposed DSM 5 model, it would be useful to consider whether empirical data on the SCD and FIRB dimensions can be used within the novel methodological framework of Factor Mixture Modeling (FMM) to stratify children with ASD into more homogeneous subgroups. **Methods:** The study sample consisted of 391 newly diagnosed children (mean age 38.3 months; 330 males) with ASD. To derive subgroups, data from the Autism Diagnostic Interview-Revised indexing SCD and FIRB were used in FMM; FMM allows the examination of continuous dimensions and latent classes (i.e., categories) using both factor analysis (FA) and latent class analysis (LCA) as part of a single analytic framework. **Results:** Competing LCA, FA, and FMM models were fit to the data. On the basis of a set of goodness-of-fit criteria, a 'two-factor/three-class' factor mixture model provided the overall best fit to the data. This model describes ASD using three subgroups/classes (Class 1: 34%, Class 2: 10%, Class 3: 56% of the sample) based on differential severity gradients on the SCD and FIRB symptom dimensions. In addition to having different symptom severity levels, children from these subgroups were diagnosed at different ages and were functioning at different adaptive, language, and cognitive levels. **Conclusions:** Study findings suggest that the two symptom dimensions of SCD and FIRB proposed for the DSM 5 can be used in FMM to stratify children with ASD empirically into three relatively homogeneous subgroups. **Keywords:** Symptomatology, Autistic disorder, Classification, Diagnosis, DSM.

Evidence shows that there is notable heterogeneity in the phenotypic presentation of ASD, regarding both configuration and severity of behavioral symptoms (Geschwind, 2009; Wiggins, Robins, Adamson, Bakeman, & Henrich, 2011). To date, researchers have used different methodological approaches to investigate this heterogeneity. A number of studies have used factor analysis (FA) methods to examine the underlying structure of the ASD phenotype (Boomsma et al., 2008; Frazier, Youngstrom, Kubu, Sinclair, & Rezai, 2008; Georgiades et al., 2007; Georgiades et al., 2011; Kamp-Becker, Ghahreman, Smidt, & Remschmidt, 2009; Snow, Lecavalier, & Houts, 2009; Van Lang et al., 2006). As Snow et al. (2009) conclude, these studies have resulted in factor solutions that are not necessarily congruent with the three categorical domains of ASD as defined by the DSM-IV (i.e., social impairment, verbal/nonverbal communication impairment, and repetitive, restricted, stereotyped behaviors; American Psychiatric Association, 2000); rather, several of these studies suggest that ASD is best conceptualized using two symptom dimensions, namely social communication deficits (SCD) and fixated interests and repetitive behaviors (FIRB). This consistent finding has been incorporated in the proposed revisions of the ASD section of the upcoming DSM 5 (American Psychiatric Association, 2011).

In parallel, several studies have attempted to identify homogeneous subgroups of individuals with

Conflicts of interest statement: The authors report no conflicts of interest.

ASD using empirical methods. To date, cluster analytic studies have proposed anywhere from one-to-four clusters (or subgroups) for ASD that differ largely on symptom severity and intellectual abilities (see Wiggins et al., 2011). Ingram, Takahashi, and Miles (2008) used taxometric methods (Ruscio & Ruscio, 2004) to determine, which phenotypic domains would be most likely to divide a sample of ASD children into two discrete subgroups. Taxometric methods can test whether or not subjects in a given data set are best described in terms of two clusters or in terms of a single homogeneous population (Ruscio & Ruscio, 2004). Regarding ASD symptoms, results from the Ingram et al. study supported subgrouping participants based only on variation in social communication (i.e., high vs. low).

Munson et al. (2008) used latent class analysis (LCA) and taxometric methods to classify children with ASD. In this study, evidence for multiple subgroups was found using both methods and these subgroups differed in level of intellectual functioning and patterns of verbal versus nonverbal ability. The Munson et al. (2008) study suggests that within the ASD group, there are distinct subtypes of autism, which differ in severity of intellectual ability, patterns of cognitive strengths and weaknesses, and severity of autism symptoms.

More recently, Frazier et al. (2012) examined the structure of autism symptoms in a large sample of 14,744 children (8,911 ASD and 5,863 non-ASD; ages 2–18), included in a national registry, the Interactive Autism Network. After comparing different categorical, dimensional, and hybrid (i.e., combined categorical and dimensional) models, the authors concluded that a hybrid model that included both a category (ASD vs. non-ASD) and two symptom dimensions (SCD and FIRB as proposed in the DSM 5) was more parsimonious than all other models and replicated across measures and subsamples. Although the Frazier et al. (2012) study is informative in many ways, it is limited by the reliance on questionnaire data (i.e., the Social Responsiveness Scale and the Social Communication Questionnaire), and the wide age range of the sample (2–18 years) as the structure of the ASD phenotype might be different across age groups (i.e., early childhood vs. late adolescence). More importantly, the Frazier et al. study was based on a sample from an ASD registry of both ASD and non-ASD cases and therefore does not provide sufficient information on the phenotypic heterogeneity within the clinical ASD group alone. Interestingly, Frazier et al. (2012) noted that 'The two-factor/three-class FM model fit slightly better than all other models. However, the third class appears to overfit the symptom distribution by splitting ASD-affected youth according to extreme and less extreme groups across all SRS scales.' (page 32). On the basis of these results, further investigation of the distribution of symptom severity within the ASD group alone is warranted.

According to Rutter (2011), the complimentary use of categorical and dimensional classification has become the norm in most areas of medicine and the field of developmental psychopathology could also benefit from such an approach. A relatively new method called Factor Mixture Modeling (FMM) allows the examination of continuous dimensions and latent classes (i.e., categories) using both FA and LCA (LCA; Muthén, 2004) in a single analysis. FMM is based on the idea that complex phenotypes require complex measurement models. One of the novel aspects of FMM in relation to taxometric methods is that FMM goes beyond class detection and allows the specification of hypothesis-based multidimensional factor models within each class. Although taxometric methods have worked well to identify simple typologies (i.e., disorder is present vs. absent), FMM has been developed to identify the underlying structure of more complex data where there may be a combination of multiple dimensions and more than two categories. Therefore, for the study of complex phenotypes, FMM may be superior to taxometric methods both in terms of class detection and class assignment (Lubke & Tueller, 2010). To date, FMM has been used successfully in the study of one other child psychiatric disorder; attention-deficit/hyperactivity disorder (ADHD; Lubke et al., 2007). As far as we are aware, FMM has never been applied to a sample of newly diagnosed children with ASD.

Distinctions among ASD subtypes (i.e., autistic disorder, Asperger's disorder & pervasive developmental disorder not otherwise specified) have been found to be inconsistent over time, variable across sites, and often associated with severity of language deficits and intellectual impairment rather than a different manifestation of inherent ASD features, such as SCD and FIRB symptoms (American Psychiatric Association, 2011). Thus, the *Neurodevelopmental Disorders Work Group* for the upcoming version of the DSM 5 (anticipated release in 2013) is proposing a significant shift in the diagnostic conceptualization of ASD (American Psychiatric Association, 2011). Rather than representing ASD as multiple subtypes, ASD will be conceptualized as a single diagnostic category. Moreover, only two dimensions (instead of three categories) – social communication deficits (SCD) and fixated interests and repetitive behaviors (FIRB) – will be specified for the description of the ASD phenotype (American Psychiatric Association, 2011). Each individual with ASD will be dimensionally described with these two domains (SCD & FIRB) using a severity gradient based on the level of 'support required' by that individual (American Psychiatric Association, 2011 & Happe, 2011). To the best of our knowledge, it remains unclear how this gradient of ASD symptom severity will be defined empirically. Moreover, we are not aware of any (proposed) specific criteria on how to define informative subgroups/categories to compliment this dimensional approach.

Building on the DSM 5 model, it would be useful to consider whether empirical data on the SCD and FIRB symptom dimensions can be used within the novel methodological framework of FMM to stratify children with ASD into more homogeneous subgroups. Such a stratification could complement the dimensional approach as suggested by Rutter (2011) and provide the foundation for subgrouping children for genetic, imaging, outcome and response to treatment studies.

Methods

Participants

The study sample consisted of 391 newly diagnosed preschool children (mean age 38.3 months with SD of 8.7; 330 males) participating in a multisite longitudinal study (Pathways in ASD) examining the developmental trajectories of children with ASD (see Georgiades et al., 2011). All participants had a recent (i.e., within 4 months) clinical diagnosis of ASD, confirmed by the ADOS and the ADI-R, according to DSM-IV criteria (American Psychiatric Association, 2000). The sampling procedure was based on consecutive referrals within specified geographic regions across five Canadian provinces. The study was approved by the local Research Ethics Boards at all sites and all parents gave written informed consent for their children to participate.

Assessment measures

ASD symptom indicators: Autism diagnostic interview – revised (ADI-R; Rutter, LeCouteur, & Lord, 2003). The ADI-R is a standardized semi-structured interview used in the diagnosis of ASD. It is designed to be employed with a parent or caregiver who is familiar with the developmental history and current behavior of individuals over the age of 2 years. The ADI-R is scored using an algorithm that is organized in three domain scales; social, communication, and repetitive behaviors. Currently, there are two versions of the ADI-R algorithm; one for children of ages two to four and one for children aged four and above. As our sample comprised children aged 2–5, both algorithms were used in our study. To ensure comparability of scores across algorithm versions, 26 common algorithm items (scores of 3 recoded to 2) that apply to all children regardless of age or verbal abilities were selected for analyses. Algorithm items that were age-dependent or language-dependent were excluded. For a list of the 26 ADI-R algorithm items used in the current study see Table 1. There was no missing data on the ADI-R.

Table 1 The two-factor structure of the 26 ADI-R algorithm items indexing ASD symptoms ($N = 391$)

	No.	ADI-R diagnostic algorithm items (item no.)	SCD factor	FIRB factor
SCD	1	Direct gaze (item 50)	.405	.238
	2	Social smiling (item 51)	.509	.147
	3	Range of social expressions used to communicate (item 57)	.535	.131
	4	Interest in children (item 62)	.650	.020
	5	Response to approaches of other children (item 63)	.617	.031
	6	Showing and directing attention (item 52)	.656	.147
	7	Offering to share (item 53)	.546	.001
	8	Seeking to share enjoyment with others (item 54)	.608	.140
	9	Use of other's body to communicate (item 31)	.283	.145
	10	Offering comfort (item 55)	.607	.133
	11	Quality of social overtures (item 56)	.636	.100
	12	Inappropriate facial expressions (item 58)[a]	.063	.536
	13	Appropriateness of social responses (item 59)	.616	.025
	14	Pointing to express interest (item 42)	.626	.009
	15	Nodding (item 43)	.507	.097
	16	Head shaking (item 44)	.547	−.047
	17	Conventional/instrumental gestures (item 45)	.647	.045
	18	Spontaneous imitation of actions (item 47)	.583	.179
	19	Imaginative play (item 48)	.521	.199
	20	Imitative social play (item 61)	.507	−.006
FIRB	21	Unusual preoccupations (item 67)	.047	.414
	22	Compulsions/rituals (item 70)	.018	.348
	23	Hand and finger mannerisms (item 77)	−.007	.547
	24	Other complex mannerisms or stereotyped body movements (item 78)	.088	.595
	25	Repetitive use of objects or interest in parts of objects (item 69)	.203	.529
	26	Unusual sensory interests (item 71)	.100	.619

ADI-R, Autism Diagnostic Interview-Revised; In the factor mixture modeling (FMM) analysis, the first 20 items were 'forced' to load on the Social Communication Deficits (SCD) factor; the remaining six items loaded on the Fixated Interests and Repetitive Behaviors (FIRB) factor.
[a]Item 58 (inappropriate facial expressions) was the only item that did not load as expected; however, for the FMM analysis it was 'forced' to load on the SCD factor for practical reasons.

Class correlates: Autism diagnostic observation schedule (ADOS; Lord et al., 2000). The ADOS uses standardized activities and 'presses' to elicit communication, social interaction, imaginative use of play materials, and repetitive behaviors, allowing the examiner to observe the occurrence/nonoccurrence and severity of behaviors important to the diagnosis of ASD. The ADOS consists of four modules, each of which is appropriate for individuals with differing language levels. A calibrated total severity metric that accounts for differences in age and module is used in this study (Gotham, Pickles, & Lord, 2009).

Vineland adaptive behavior scales, second edition (VABS II; Sparrow, Cicchetti, & Balla, 2005). The VABS II assesses child adaptive behavior in the communication, socialization, daily living skills and motor domains, and expresses overall functioning in the 'Adaptive Behavior Composite' (ABC) score (used in current analyses). The VABS II is administered to a parent or caregiver using a semi-structured interview format.

Merrill-palmer-revised scales of development (MP-R; Roid & Sampers, 2004). This is an individually administered measure of intellectual ability that is appropriate for children aged 2–78 months. The 'Developmental Index standard score' (used in current analyses) comprises cognitive, receptive language, and fine motor scales.

Preschool language scale – fourth edition (PLS-4; Zimmerman, et al., 2002). The PLS-4 is a language test used to identify children with language disorder between birth and 83 months or for older children (such as children with ASD) who function developmentally within this range. The 'Total Language Score' is used in this study.

Statistical analyses

Factor mixture modeling Factor Mixture Modeling allows the simultaneous examination of continuous dimensions and latent classes (or categories, or subgroups) using both FA and LCA (LCA; Muthén, 2004). FMM is a general framework extending FA and LCA by combining the two as submodels into a single general model (Lubke & Muthén, 2005). Unlike taxometric methods that can only test for dichotomous classes derived using data on only one dimension at a time, FMM permits the specification (i.e., hypothesis) of a multidimensional factor model for each class (Lubke & Tueller, 2010; Lubke et al., 2007). In FMM, individuals are stratified into discrete classes, but within each class, continuous latent factors account for potential differences in the severity of the disorder (Walton, Ormel, & Krueger, 2011). Specific FMMs can be compared and evaluated using well-established indices of goodness-of-fit (Lubke & Muthén, 2005). In this study, FMMs were applied to identify more homogeneous subgroups (or classes) of ASD using data indexing the SCD and FIRB severity dimensions of ASD within each class.

The 26 ADI-R algorithm items measuring ASD symptoms were subjected to a Principal Component Analysis (PCA) with Varimax (i.e., orthogonal) rotation to derive the most parsimonious model containing uncorrelated factors. Results indicated that compared to the one, three, and four-factor solutions, the two-factor solution (explaining 32% of the variance; see Table 1) was the most parsimonious solution in terms of both the *Scree Plot* criterion and a clear pattern of item loadings. Moreover, the specific two-factor solution was selected for subsequent FMM analysis because of its conceptual interpretability and its consistency with established results of numerous factor-analytic studies in the literature as well as the current DSM 5 proposal for the structure of the ASD symptom phenotype (American Psychiatric Association, 2011).

On the basis of previous studies that have proposed the existence of one-to-four ASD subgroups, a total of four competing FMMs were tested using 26 ADI-R indicators. Models 2f1c (two factors, one class), 2f2c, 2f3c, and 2f4c are FMMs with one, two, three, and four

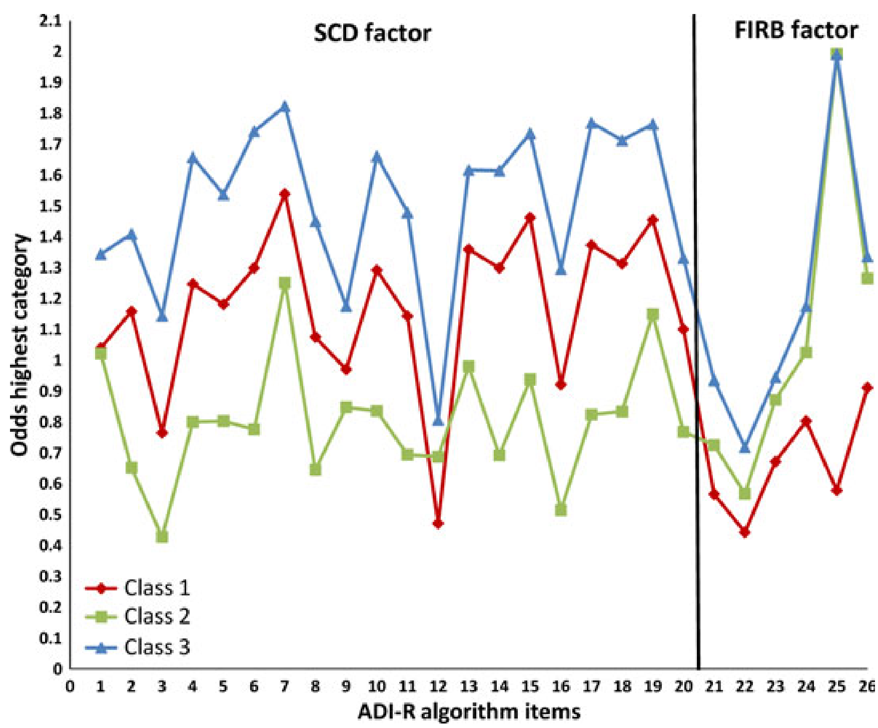

Figure 1 Factor mixture model of the ASD symptom phenotype with 'two factors/three classes' (N = 391) The horizontal axis represents the 26 algorithm items from the ADI-R. Items 1–20 load on the SCD factor and items 21–26 load on the FIRB factor. The vertical axis represents the probability of scoring in the highest response category/class for each item in proportion to scoring in any of the other categories/classes for 391 children with ASD in the 'two factor/three class' factor mixture model. ADI-R, Autism Diagnostic Interview-Revised; SCD, Social Communication Deficits; FIRB, Fixated Interested and Repetitive Behaviors.

classes, all with two factors. Specifically, the FMMs have two factors/dimensions (SCD & FIRB) with 20 of the ADI-R indicators forced to load only on the SCD factor and the remaining six indicators forced to load only on the FIRB factor (see Table 1 & Figure 1 below).

To confirm whether FMMs are a better overall fit to the data than structural models proposed in previous studies, five LCA models (with one-to-five classes) were evaluated in relation to the four FMMs described above. Finally, to confirm that the two-factor model (shown in Table 1) used in the FMM analysis had a comparable fit to the data as other previously proposed solutions, five FA models (with one-to-five factors) were also tested.

The fit of all competing models to the data was tested simultaneously using established goodness-of-fit criteria, such as the *Akaike Information Criterion (AIC)*, the *Bayesian Information Criterion (BIC)*, and the *Sample Size Adjusted BIC*. In general, lower values of AIC and BIC indicate a better model fit to the data (Lubke et al., 2007). Large simulation FMM studies have shown that the specific goodness-of-fit criteria can help researchers determine, which model correctly depicts the data at hand (Lubke & Spies, 2008). All models were run using *MPlus Version 5.0* statistical software (Muthen & Muthen, 2007).

Characterization of classes For the best fitting model, *factor scores* and *class assignment* were calculated for each individual child. Factor scores were calculated using the observed means of items that load on each factor; class assignment was implemented using modal assignment by placing subjects in the class with the highest posterior class probability (Lubke & Tueller, 2010). These scores were then used in post hoc analyses to describe the derived classes using other child phenotypic indicators believed to be important for the characterization of ASD (American Psychiatric Association, 2011; Volkmar, State, & Klin, 2009).

Specifically, derived classes were described in relation to the child's age at diagnosis, adaptive functioning (indexed by the VABS II composite score), cognitive abilities (indexed by the M-P-R standard score), and language abilities (indexed by the PLS-4). To describe class profiles in terms of ASD symptoms, class mean scores were also compared on: (a) the two derived symptom factors, SCD and FIRB (b) the original ADI-R algorithm domain scales (i.e., social, nonverbal communication, and repetitive behaviors); and (c) the ADOS severity metric. To better represent class variability on ASD symptom severity, a two-dimensional (convex hull) plot of SCD by FIRB for the derived classes was created. For these analyses, effect sizes (ES) were estimated by computing class mean differences taking two classes at a time, divided by the overall standard deviation of all three classes combined. An effect size of 0.2–0.3 could be interpreted to be a 'small' effect, around 0.5, a 'medium' effect and 0.8 to infinity, a 'large' effect (Cohen, 1992).

Cross-tabulation (chi-square analysis) was used to compare the proportion of children across classes by gender; for all other comparisons, one-way analysis of variance (ANOVA) was used. These analyses were conducted using the SPSS Inc. (2011) statistical software.

Results

Factor mixture modeling

A direct statistical comparison of all competing models showed that the 'two-factor/three-class' FMM provided the best fit to the data and was clearly superior across all models (FA, LCA, and FMM) based on all goodness-of-fit criteria – the AIC, BIC, and adjusted BIC (see Table 2). The specific FMM was estimated using a relatively small number of parameters, suggesting parsimony in the description

Table 2 ASD structural symptom model comparisons, fit indices, and class proportions (N = 391)

	Number of classes (c) or factors (f)	Log Likelihood	Number of free parameters	AIC	BIC	Adjusted BIC	Class percentages
LCA models	1c	−11,124.957	52	22,353.915	22,560.288	22,395.294	100%
	2c	−10,326.232	79	20,810.464	21,123.992	20,873.329	32%, 68%
	3c	−10,135.720	106	20,483.440	20,904.123	20,567.790	31%, 22%, 47%
	4c	−10,049.148	133	20,364.297	20,892.135	20,470.132	20%, 14%, 41%, 25%
	5c	−10,011.462	160	20,342.925	20,977.918	20,470.246	11%, 19%, 23%, 22%, 25%
FA models	1f	−10,212.711	78	20,581.423	20,890.982	20,643.492	
	2f[a]	−10,138.238	103	20,482.476	20,891.253	20,564.439	
	3f	−10,057.198	127	20,368.395	20,872.421	20,469.456	
	4f	−10,015.669	150	20,331.337	20,926.644	20,450.701	
	5f	No convergence					
FMMs	2f1c	−10,187.397	78	20,530.795	20,840.354	20,592.864	100%
	2f2c	−10,143.391	81	20,448.782	20,770.247	20,513.238	70%, 30%
	2f3c	−10,073.055	84	20,314.109	20,647.480	20,380.953	34%, 10%, 56%
	2f4c	No convergence					

ASD, Autism Spectrum Disorder; LCA, latent class analysis; FA, factor analysis; FMM, factor mixture model; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion. The best fitting model from a direct comparison of all models (FA, LCA, and FMM) and across all goodness-of-fit criteria is presented in bold font.
[a]The specific two-factor model is based on the principal component analysis depicted in Table 1.

Table 3 Means, standard deviations, and effect sizes for the three classes of children with ASD on variables of interest

	Mean (*SD*)			Effect Size		
	Class 1	Class 2	Class 3	C1 versus C2	C1 versus C3	C2 versus C3
SCD dimension	1.18 (0.36)[a]	0.72 (0.19)[a]	1.51 (0.26)[a]	1.18	−0.86	−2.04
FIRB dimension	0.67 (0.37)	1.08 (0.34)[b]	1.18 (0.38)[b]	−0.93	−1.17	−0.24
ADI-R social domain scale	15.27 (5.05)[a]	10.59 (3.18)[a]	19.56 (3.94)[a]	0.90	−0.83	−1.73
ADI-R communication (nonverbal) domain scale	11.30 (3.17)[a]	9.05 (3.60)[a]	12.81 (3.07)[a]	0.67	−0.45	−1.12
ADI-R repetitive behaviors domain scale	3.74 (2.04)	5.78 (1.97)[b]	5.88 (2.12)[b]	−0.88	−0.93	−0.04
ADOS severity metric	7.23 (1.80)[c]	7.65 (1.69)	7.79 (1.63)	−0.24	−0.33	−0.08
VABS II adaptive behavior	75.57 (9.60)[d]	79.49 (10.14)[d]	70.30 (9.424)	−0.39	0.52	0.91
M-P-R developmental index score	60.94 (26.75)[d]	64.97 (25.38)[d]	52.03 (24.34)	−0.16	0.35	0.50
PLS-4 total language score	68.70 (20.44)[d]	72.57 (21.69)[d]	61.76 (17.11)	−0.20	0.36	0.57

ASD, Autism Spectrum Disorder; ADI-R, Autism Diagnostic Interview-Revised; SCD, Social Communication Deficit; FIRB, Fixated Interests and Repetitive Behavior; VABS II, Vineland Adaptive Behavior Scales, Second Edition; M-P-R, Merrill-Palmer-Revised Scales of Development; PLS-4, Preschool Language Scale – Fourth Edition.
Class 1: 34% of sample; Class 2: 10% of sample; Class 3: 56% of sample.
[a]All three classes are significantly different from each other ($p < 0.05$).
[b]Each of these classes is significantly different from Class 1 ($p < 0.05$).
[c]This class is significantly different than the other two Classes 2 and 3 ($p < 0.05$).
[d]Each of these classes is significantly different from Class 3 ($p < 0.05$).
Effect size C1 versus C2 = [mean (C1) – mean (C2)]/overall *SD*, Effect size C1 versus C3 = [mean (C1) – mean (C3)]/overall *SD*, Effect size C2 versus C3 = [mean (C2) – mean (C3)]/overall *SD*.

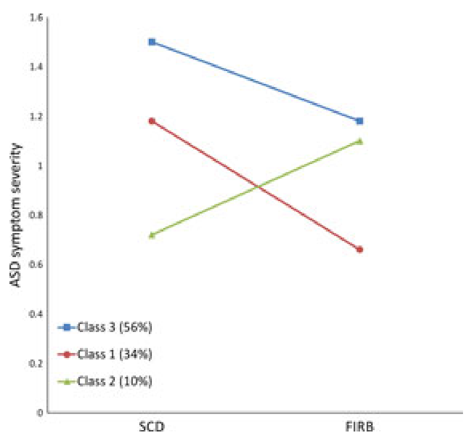

Figure 2 Class profiles using mean scores of SCD and FIRB symptom dimensions ($N = 391$) Notes: SCD, Social Communication Deficits; FIRB, Fixated Interested and Repetitive Behaviors; Class 1: 34% of sample; Class 2: 10% of sample; Class 3: 56% of sample.

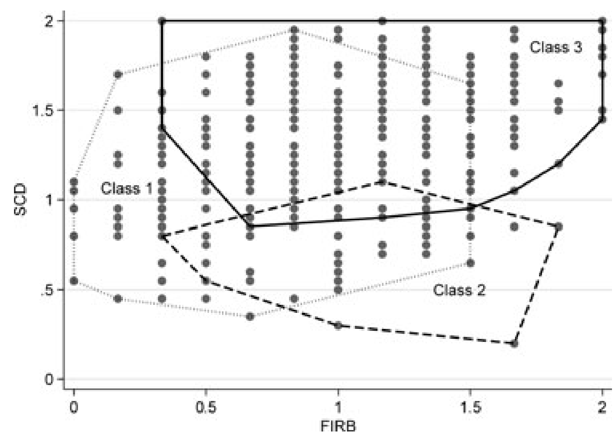

Figure 3 SCD by FIRB two-dimensional (convex hull) plot for the three derived ASD classes ($N = 391$) Notes: The horizontal axis represents scores on the Fixated Interested and Repetitive Behaviors (FIRB) symptom dimension; The vertical axis represents scores on the Social Communication Deficits (SCD) symptom dimension; Class 1 (34% of sample); Class 2 (10% of sample); Class 3 (56% of sample).

of the underlying phenotypic structure of ASD. According to this FMM, ASD can be described in this sample using data on the two independent severity dimensions of SCD and FIRB to stratify children into three relatively homogeneous classes (Class 1: 34%, Class 2: 10% and Class 3: 56% of the sample).

Characterization of classes

Table 3 presents the ANOVA results and effect sizes (ES) for comparing classes on variables of interest (i.e., class correlates). In terms of ASD symptoms, on average, children assigned to *Class 1* (34% of the sample) score moderately high on social communication impairments (indexed by the SCD factor and the original ADI-R social and communication domains) and have the lowest scores of repetitive behaviors (indexed by the FIRB factor and the original ADI-R behaviors domain). Children assigned to *Class 2* (10% of the sample) have a reverse profile with the lowest scores on social communication impairments and moderately high scores of repetitive behaviors. Children assigned to *Class 3* (56% of the sample) have the highest scores on both social communication impairments and repetitive behaviors. The estimated effect sizes for class differences on ASD symptoms above were large (see Table 3). The between and within class variability on ASD symptoms is also shown in the two-dimensional (convex hull) plot (Figure 3). It must be noted that class differences on the ADOS severity metric ranged from small to moderate.

Children assigned to *Class 2* were diagnosed at a later age on average (mean age = 43.99; *SD* = 9.18 months; $p < 0.01$) compared to children assigned to *Class 1* (mean age = 38.42; *SD* = 8.77 months) and *Class* 3 (mean age = 37.31; *SD* = 8.1 months).

In terms of overall functioning (i.e., developmental level, language abilities, and adaptive behavior), on average, children in *Class 2* had the highest scores followed by children in *Class 1*; children assigned to *Class 3* had the lowest scores in relation to children from the other two classes ($p < 0.01$). From Table 3, we can see that the largest effect sizes were seen for differences between Classes 2 and 3.

Finally, there were no differences in distribution by gender (i.e., the proportion of males and females) across the three classes, although the small proportion of females in the sample (16%) may limit our ability to detect statistically significant differences in this distribution across classes.

Discussion

This study used the novel method of FMM to stratify children with ASD into empirically derived subgroups based on their severity levels on the two diagnostic symptom domains of SCD and FIRB proposed for the DSM 5. Our findings confirm those from previous studies suggesting notable heterogeneity in the phenotypic presentation within the ASD spectrum even at this young age (see Munson et al., 2008). Our data suggest that there is evidence of three relatively homogeneous ASD subgroups or classes (Class 1: 34%, Class 2: 10%, Class 3: 56% of the sample) that lie on two spectra (i.e., SCD & FIRB) of ASD symptoms. Although the three subgroups/classes could be described using a total ASD severity gradient, this gradient does not follow the same pattern for both the SCD and FIRB symptom dimensions. Specifically, for the SCD dimension, Class 3 has the highest mean score followed by Class 1 and then Class 2; for the FIRB dimension, Class 3 has the highest mean score followed by Class 2 and then Class 1 (see Figure 2). The fact that the class severity gradient pattern differs across dimensions speaks to the importance of treating SCD and FIRB as independent spectra that together make up the overall compound ASD phenotype. These data support the idea that the two ASD symptom domains of SCD and FIRB may potentially arise from largely independent (although possibly overlapping) underlying risk factors (Mandy & Skuse, 2008).

Statistically significant differences (see large effect sizes in Table 3) in ASD symptom severity as well as notable differences in profiles related to child functioning provide suggestive support for the potential utility of the three ASD subgroups proposed herein. However, a closer inspection of the between and within class variability suggests that the three subgroups have overlapping distributions of both ASD symptoms (see Figure 3) and overall level of functioning (see Table 3). Hence, even if the three subgroups are more homogeneous in relation to a single ASD spectrum, we still observe wide variability/heterogeneity within each subgroup. Therefore, until these subgroups are tested in genetic, imaging, outcome, and treatment studies, it would be premature to claim that their statistically different profiles are clinically meaningful and/or useful.

Children in Class 2 were diagnosed (on average) at a later age than children from the other two classes. Although this finding is intriguing, it cannot be taken as evidence for later onset of ASD in this group because the age of diagnosis is directly connected to the age a child gets referred, as well as to the time a child spends on a wait list for a diagnostic assessment. This finding could be due to the more 'subtle' presentation of ASD-related symptoms (i.e., low impairment on SCD) in children assigned to Class 2.

Data presented in this exploratory study do not offer definite answers to the complex issue of ASD heterogeneity; however, our empirical findings could be used to generate specific hypotheses related to the utility of the three derived ASD subgroups. For example, it is possible that children from the different ASD classes might follow different developmental trajectories, which could be helpful in determining prognosis. Moreover, one could hypothesize that children from different classes would have a differential response to treatment (see Szatmari, 2011). Such research findings could offer clinicians flexible and practical solutions that allow for the utilization of dimensional symptom severity data that can be converted into categorical classes (e.g., mild, moderate, severe). This way clinicians will be able to reliably communicate the information to patients and colleagues and apply prespecified inclusion/exclusion criteria for treatment purposes (Kamphuis & Noordhof, 2009; Kraemer, 2007). Furthermore, the empirical organization of children into more homogenous ASD classes could yield informative phenotypes for stratifying children in genetic studies and in studies in search of biological markers of ASD (Liu et al., 2011; Szatmari et al., 2007).

By exploring substantial data on the two symptom dimensions proposed for the DSM 5 (i.e., SCD and FIRB), we were able to derive more homogeneous classes of newly diagnosed children with ASD based on severity levels. It is important to note that we chose not to refer to these ASD groups as 'subtypes', something that would suggest (a priori) 'qualitative' differences in etiology, diagnosis, and prognosis (Witwer & Lecavalier, 2008). Rather, we chose to use the terms 'classes' or 'subgroups' that simply refer to empirical (i.e., data-driven), potentially informative groupings of children, in this case with similar scores on ASD symptom severity dimensions and other related phenotypes.

Limitations

The present study is of an exploratory nature and has several limitations, which call for a cautious interpretation of findings. First, FMM analyses were based on parent-report data from the ADI-R and are subject to potential reporting biases. Although data on the direct observation ADOS were available, they were not used in FMM analysis because it was not possible to overcome the complex measurement issues arising from the administration of different ADOS modules to different children. In addition, the ADOS severity metric that can account for differences in modules was not used in the models tested in this study. This metric is comprised of a total score and does not provide separate scores on the SCD and FIRB symptom dimensions – the core indicators in our study (the lack of notable class differences on the ADOS severity metric supports our decision at the beginning of the study not to use it in the FMM). Second, this study uses cross-sectional data from a sample of newly diagnosed children in a limited age range; thus, the specific findings cannot be generalized to older children, and must be interpreted within the context of the diagnostic process. Third, to ensure comparability of scores across ADI-R algorithm versions, items that were age-dependent or language-dependent were excluded. As a result, it is possible that important phenotypic information related to language and/or age was missed in this analysis. For example, the exclusion of language-dependent items from the analyses prohibits the exploration of an additional language-related dimension, a construct known to be important in the clinical characterization of ASD. Fourth, the fact that the FMM method artificially imposes a common factor structure in each class, does not allow for the examination of potentially different factor structures within different classes/subgroups. For example, one could hypothesize that the phenotypic structure of ASD might have a different appearance among higher functioning children from Class 2 compared with lower functioning children from Class 3. Fifth, the FMMs with more than three classes did not converge, preventing any test for the potential superiority of more complex models with four or more classes. In fact, in the LCA tests, the four-class model was a better fitting model than the three-class model (the same was true for the three-vs. two-factor solutions; see Table 2). This could be perceived as a limitation of the data; for example, variability in symptom presentation may be restricted by the narrow age range in our sample (i.e., ages 2–5 years), something that may in turn hinder the ability of the FMM procedure to identify structural models that are more complex than the '2f/3c' solution. Sixth, the use of ordinal (0– 2) algorithm items from the ADI-R might have had an effect on the estimation of model parameters in FMM; preferably, normally distributed indicators derived from continuous measures of ASD symptoms should be used in future research. Perhaps, the biggest limitation of this study is the absence of a construct validity criterion against which the utility of the proposed ASD model can be tested. However, as longitudinal and genetic data on this sample become available, we plan to evaluate the ability of this model to predict specific developmental trajectories, response to treatment, and genetic markers of ASD.

Conclusion

Heterogeneity within the autism spectrum is, perhaps, the biggest obstacle to research and translation of research into clinical practice (Newsschaffer, Fallin, & Nora, 2002). Abandoning the 'single entity' approach to autism is a necessary step to overcome that obstacle (Happé, Ronald, & Plomin, 2006). Moreover, as Rutter (2011) notes, although empirical findings indicate that most mental disorders operate in a dimensional manner, it is still useful to continue using categories (in a complementary way), as they can be quite informative for clinical practice as well as for stratification purposes in clinical, intervention, biological, and genetic research.

We propose, herein, a factor mixture model that uses dimensional severity scores on the SCD and FIRB symptom spectra to stratify children with ASD into three relatively homogeneous subgroups. Children from these subgroups have different severity levels of ASD symptoms, are diagnosed at different ages and function at different adaptive, language, and cognitive levels. However, as noted by Szatmari (2011), rather than focusing on assigning labels to these three ASD subgroups, we should focus instead on identifying markers that capture diversity – in developmental trajectories, in responses to treatment, and in the genetic heterogeneity inherent in ASD.

We believe that the proposed 'two-factor/three-class' ASD model could inform the ongoing work of the DSM 5 revisions (American Psychiatric Association, 2011). For example, as noted by Lord and Jones (2012), even if the two proposed symptom dimensions (SCD & FIRB) can be informative for the characterization of children with ASD, quantitative measures are needed to accurately map these dimensions. The observed (in the current study) between and within class variability on ASD symptoms as well as the differential, but overlapping class distributions of child functioning indicators must be taken into consideration when defining 'severity levels' and 'clinical specifiers' for the revised ASD criteria in the DSM 5 (American Psychiatric Association, 2011). Ideally, carefully designed DSM 5 field trials will incorporate these observations and test the relevant hypotheses empirically. In the mean time, findings from the current study serve as a renewal of our quest for understanding the complex issue of ASD phenotypic heterogeneity, and thus contribute to the study of the etiology, diagnosis, treatment and prognosis of ASD.

Acknowledgements

This study was supported by the Canadian Institutes of Health Research, Autism Speaks, the Government of British Columbia, the Alberta Innovates Health Solutions, and the Sinneave Family Foundation. Stelios Georgiades is supported by an Autism Research Training (ART) fellowship by the Canadian Institutes of Health Research. The authors thank all the families who participated in the *Pathways in ASD* study. The authors also acknowledge the members of the *Pathways in ASD Study Team*. These members had equal contribution to the study and are listed here alphabetically: Liliana Abruzzese, Megan Alexander, Susan Bauld, Ainsley Boudreau, Colin Andrew Campbell, Mike Chalupka, Lorna Colli, Melanie Couture, Bev DaSilva, Vikram Dua, Miriam Elfert, Lara El-Khatib, Lindsay Fleming, Kristin Fossum, Nancy Garon, Shareen Holly, Stephanie Jull, Karen Kalynchuk, Kathryne MacLeod, Preetinder Narang, Julianne Noseworthy, Irene O'Connor, Kaori Ohashi, Sarah Peacock, Teri Phillips, Sara Quirke, Katie Rinald, Jennifer Saracino, Cathryn Schroeder, Cody Shepherd, Rebecca Simon, Mandy Steiman, Richard Stock, Benjamin Taylor, Lee Tidmarsh, Larry Tuff, Kathryn Vaillancourt, Stephen Wellington, Isabelle Yun, and Li Hong Zhong.

Correspondence

Peter Szatmari, Director, Offord Centre for Child Studies, McMaster University, Chedoke Site, Patterson Building, 1200 Main St. W, Hamilton, ON, L8N 3Z5, Canada; Email: szatmar@mcmaster.ca

Key points

- Autism Spectrum Disorder (ASD) is characterized by notable phenotypic heterogeneity, which is often viewed as an obstacle to the study of its etiology, diagnosis, treatment, and prognosis.
- This study used the novel method of Factor Mixture Modeling (FMM) that allows for the integration of both categories and dimensions to stratify children with ASD into relatively more homogeneous subgroups.
- Results showed that children with ASD can be classified into three subgroups based on their severity on the symptom dimensions of social communication deficits (SCD) and fixated interests and repetitive behaviors (FIRB). Children within these subgroups were diagnosed at different ages and were functioning at different adaptive, language, and cognitive levels.
- Clinically, it is possible that children from these subgroups might follow different developmental trajectories and/or or have a differential response to treatment.
- Study findings can inform the ongoing work on the DSM 5 revisions for ASD.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn., Text revision). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2011). Report of the DSM-V Neurodevelopmental Disorders Work Group. Available from: http://www.dsm5.org [last accessed 6 June 2011].

Boomsma, A., Van Lang., N.D., 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.

Cohen, J. (1992). "A power primer". *Psychological Bulletin*, *112*, 155–159.

Frazier, T. W., Youngstrom, E. A., Kubu, C. S., Sinclair, L., & Rezai, A. (2008). Exploratory and confirmatory factor analysis of the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders*, *38*, 474–480.

Frazier, T.W., Youngstrom, E.A., Speer, L., Embacher, R., Law, P., Constantino, J., … & Eng, C. (2012). Validation of proposed DSM-5 criteria for autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*, 28–40.

Georgiades, S., Szatmari, P., Duku, E., Zwaigenbaum, L., Bryson, S., Roberts, W., … & Pathways in ASD Study Team. (2011). Phenotypic overlap between core diagnostic features and emotional/behavioral problems in preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*, 1321–1329.

Georgiades, S., Szatmari, P., Zwaigenbaum, L., Duku, E., Bryson, S., Roberts, W., … & Mahoney, W. (2007). Structure of the autism symptom phenotype: A proposed multidimensional model. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*, 188–196.

Geschwind, D. (2009). Advances in autism. *Annual Review of Medicine*, *60*, 367–380.

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*, 693–705.

Happe, F. (2011). Criteria, categories, and continua: Autism and related disorders in DSM-5. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*, 540–542.

Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience*, *9*, 1218–1220.

Ingram, G. D., Takahashi, T. N., & Miles, J. H. (2008). Defining autism subgroups: A taxometric solution. *Journal of Autism and Developmental Disorders*, *38*, 950–960.

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.

Kamphuis, J.H., & Noordhof, A. (2009). On categorical diagnoses in DSM-V: Cutting dimensions at useful points? *Psychological Assessment*, *21*, 294–301.

Kraemer, H. C. (2007). DSM categories and dimensions in clinical and research contexts. *International Journal of Methods in Psychiatric Research*, *16*(Suppl. 1), S8–S15.

Liu, X., Georgiades, S., Duku, E., Thompson, A., Devlin, B., Cook, E.H., … & Szatmari, P. (2011). Identification of genetic loci underlying the phenotypic constructs of autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*, 687–696.

Lord, C., & Jones, R. M. (2012). Annual research review: Rethinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *53*, 490–509.

Lord, C., Risi, S., Lambrecht, L., Cook, E., Leventhal, B., DiLavore, P., … & Rutter, M. (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.

Lubke, G.H., & Muthén, B.O. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21–39.

Lubke, G.H., Muthén, B.O, Moilanen, I.K, McGough, J.J., Loo, S.K., Swanson, J. M., … & Smalley, S.L. (2007). Subtypes vs. severity differences in attention deficit hyperactivity disorder in the Northern Finnish Birth Cohort (NFBC). *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*, 1584–1593.

Lubke, G.H., & Spies, J. (2008). Choosing a 'correct' factor mixture model: Power, limitations, and graphical data exploration. In G. R. Hancock, & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models*. Charlotte, NC: Information Age Publishing, 343–362.

Lubke, G., & Tueller, S. (2010). Latent class detection and class assignment: A comparison of the MAXEIG taxometric procedure and factor mixture modeling approaches. *Structural Equation Modeling*, *17*, 605–628.

Mandy, W.P.L., & Skuse, D.H. (2008). Research review: What is the association between the social-communication element of autism and repetitive interests, behaviours and activities. *Journal of Child Psychology and Psychiatry*, *49*, 795–708.

Munson, J., Dawson, G., Sterling, L., Beauchaine, T., Zhou, A., Koehler, E., … & Abbott, R. (2008). Evidence for latent classes of IQ in young children with autism spectrum disorder. *American Journal on Mental Retardation*, *113*, 427–438.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data; In D Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Sage: Thousand Oaks.

Muthen, L.K., & Muthen, B.O. (2007). *Mplus Version 5.0 User's Guide*. Los Angeles: Muthen & Muthen.

Newsschaffer, C. J., Fallin, D., & Nora, L. L. (2002). Heritable and nonheritable risk factors for autism spectrum disorders. *Epidemiologic Reviews*, *24*, 137–153.

Roid, G., & Sampers, J. (2004). *Merrill-Palmer-Revised Scales of Development*. Wood Dale, IL: Stoelting Co.

Ruscio, J., & Ruscio, A. M. (2004). Clarifying boundary issues in psychopathology: The role of taxometrics in a comprehensive program of structural research. *Journal of Abnormal Psychology*, *113*, 24–38.

Rutter, M. (2011). Research review: Child psychiatric diagnosis and classification: Concepts, findings, challenges and potential. *Journal of Child Psychology and Psychiatry*, *52*, 647–660.

Rutter, M., LeCouteur, A., & Lord, C. (2003). *ADI-R: The Autism Diagnostic Interview-Revised*. Los Angeles, CA: Western Psychological Services.

Snow, A. V., 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., Cicchetti, D.V., & Balla, D.A. (2005). *Vineland Adaptive Behavior Scales: Second Edition (Vineland II), Survey Interview Form/Caregiver Rating Form*. Livonia, MN: Pearson Assessments.

SPSS Inc. (2011). *SPSS for Windows, Rel. 19*. Chicago.

Szatmari, P. (2011). New recommendations on autism spectrum disorder. *BMJ;*, *342*, d2456.

Szatmari, P., Maziade, M., Zwaigenbaum, L., Merette, C., Roy, M-A., Joober, R., & Palmour, R. (2007). Informative phenotypes for genetic studies of psychiatric disorders. *American Journal of Medical Genetics Part B*, *144B*, 581–588.

Van Lang, N.D.J, Boomsma, A., Sytema, S., De Bildt, A.A., Kraijer, DW., Ketelaars, C. & Minderaa, R. B. (2006). Structural equation analysis of a hypothesised symptom model in the autism spectrum. *The Journal of Child Psychology and Psychiatry.*, *47*, 37–44.

Volkmar, F.R., State, M., & Klin, A. (2009). utism and autism spectrum disorders: Diagnostic issues for the coming decade. *Journal of Child Psychiatry and Psychology.*, *50*, 108–115.

Walton, K. E., Ormel, J., & Krueger, R. F. (2011). The dimensional nature of externalizing behaviors in adolescence: Evidence from a direct comparison of categorical, dimensional, and hybrid models. *Journal of Abnormal Child Psychology*, *39*, 553–561.

Wiggins, L. D., Robins, D. L., Adamson, L. B., Bakeman, R., & Henrich, C (2011). Support for a dimensional view of autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*, *42*, 191–200.

Witwer, A. N., & Lecavalier, L. (2008). Examining the validity of autism spectrum disorder subtypes. *Journal of Autism and Developmental Disorders*, *38*, 1611–1624.

Zimmerman, I. L., Steiner, V.G., & Pond, R.E. (2002). *Preschool language scale (4th edn)*. San Antonio, TX: The Psychological Corporation.

Accepted for publication: 6 June 2012
Published online: 1 August 2012