Testing the Construct Validity of Proposed Criteria for *DSM-5* Autism Spectrum Disorder

William P.L. Mandy, D.Clin.Psy., Tony Charman, Ph.D., David H. Skuse, M.D.

Objective: To use confirmatory factor analysis to test the construct validity of the proposed *DSM-5* symptom model of autism spectrum disorder (ASD), in comparison to alternative models, including that described in *DSM-IV-TR*. **Method:** Participants were 708 verbal children and young persons (mean age, 9.5 years) with mild to severe autistic difficulties. Autistic symptoms were measured using the Developmental, Dimensional and Diagnostic interview (3Di). The fit of the two-factor *DSM-5* model, which has a social communication and a restricted, repetitive behavior (RRB) factor, was compared with that of alternative models. In one half of the sample, properties of the *DSM-5* model were examined to investigate the validity of specific diagnostic criteria, informing the development of a better fitting *DSM-5* model. This was then cross-validated in the remaining "hold-out" half of the sample; and its stability was tested across groups defined by age, sex, and symptom severity. **Results:** The *DSM-5* model was superior to the three-factor *DSM-IV-TR* model. It was improved by the removal of items measuring "play and imagination" and "stereotyped and repetitive use of language." A scale measuring sensory abnormalities was added to the model, and loaded onto its RRB factor. This *DSM-5* model fit well in the hold-out sample; was stable across age and sex; and fit adequately in those with clinical and sub-threshold autistic presentations. **Conclusions:** Among higher-functioning individuals, ASD is a dyad, not a triad, with distinct social communication and repetitive behavior dimensions. As suggested in the proposed *DSM-5* criteria, sensory abnormalities are part of the RRB symptom cluster. J. Am. Acad. Child Adolesc. Psychiatry, 2012;51(1):41–50. **Key Words:** autism spectrum disorder, *DSM-5*, construct validity, confirmatory factor analysis, broader autism phenotype

Like the majority of neuropsychiatric disorders, autism is defined as a behavioral syndrome, such that a specific cluster of signs and symptoms are assumed to signify a latent disease entity. Current diagnostic criteria conceptualize the autism syndrome as a triad of observable impairments in social reciprocity, communication, and flexibility of thought and behavior.[1,2] This account of autism was based on clinical experience and judgement but has increasingly been subjected to rigorous, systematic testing by empirical means.[3,4] Such investigations have challenged our understanding of the autism construct and in doing so have provided an opportunity for a more empirically based conceptualisation to emerge. This process is especially relevant at the current time, given ongoing revision by the American Psychiatric Association and the World Health Organisation of their diagnostic manuals.

Draft *DSM-5* diagnostic criteria for "autism spectrum disorder" (ASD) propose several significant changes to the content and structure of the autistic syndrome.[5] Foremost among these is the suggested replacement of the autism triad with an autism dyad comprising social communication and restricted repetitive behavior (RRB) dimensions. Furthermore draft criteria propose the removal of impaired imagination from the list of core autistic symptoms[6]; the reassignment of stereotyped and repetitive use of language as a form of RRB, not a symptom of communication impairment; and the inclusion of hypo- or hyperreactivity to sensory stimuli as a core criterion

This article is discussed in an editorial by Dr. Bennett Leventhal on page 6.

within the RRB domain. In the current study we aimed to test the validity of these proposed changes to diagnostic criteria.

One approach to evaluating the construct validity of a diagnosis is confirmatory factor analysis (CFA), a hypothesis driven technique suitable for focused testing of specific symptom models. Recently, several groups have used this method to investigate the content and structure of autistic traits, and their findings have contributed to the development of an empirically-based model of ASD. When reviewed together, CFA studies using well validated measures of ASD symptoms are broadly supportive of the proposed *DSM-5* ASD construct. They are strongly suggestive that the triadic model proposed in *DSM-IV-TR* lacks validity,[7-11] as all but one study[12] found it to be inferior to alternative models. In support of *DSM-5* draft criteria, this in part reflects the fact that most of the symptoms of autism which *DSM-IV-TR* assumes to reflect two distinct social and communication dimensions are actually manifestations of a single social communication factor.[7-11] Furthermore, CFA provides strong evidence for a RRB factor that is separate from social communication impairments.[7-12]

Nevertheless there remain unanswered questions about the ASD construct and these are of relevance to imminent revisions of diagnostic criteria. There is uncertainty as to whether, as proposed for *DSM-5*, stereotyped and repetitive language loads onto a RRB, not a social communication, factor.[9,11] Similarly, there are mixed findings in the CFA literature with respect to the *DSM-5* proposition that play and imagination deficits be dropped as a symptom of autistic social communication impairment. Several analyses have suggested that these difficulties signify a distinct "make-believe and play" factor,[7,8] whereas others have yielded the contradictory finding that imagination difficulties are best conceptualized as a manifestation of disordered social communication.[11] Furthermore, the current CFA literature does not offer clear guidance as to whether sensory abnormalities should be understood as part of the RRB symptom cluster.

In addition, given the strong evidence that ASD is a dimensional disorder,[13] and the implicit recognition of this in the decision to subsume all pervasive developmental disorders under the "autism spectrum disorder" label in *DSM-5*, it will also be valuable to test whether models of autistic symptomatology apply beyond ASD to individuals with milder, sub-clinical difficulties characteristic of the broader autism phenotype (BAP).

To contribute to discussions of how ASD should best be conceptualized in *DSM-5*, in a sample of verbal, mostly high-functioning children with social communication difficulties, we tested three distinct symptom models using CFA: a three-factor *DSM-IV-TR* model; a one-factor model; and a two-factor model based on *DSM-5* draft criteria. ASD symptoms were measured using the Dimensional, Diagnostic and Developmental Interview (3Di).[14] For the *DSM-5* model, indices of model misspecification were inspected to see whether repetitive and stereotyped use of language is a form of RRB; and whether or not play and imagination difficulties should be considered a symptom of social communication difficulty. In addition we added a sensory abnormalities variable to the *DSM-5* model, to test the proposal that these behaviors should be considered a manifestation of RRB. As well as those with an ASD, participants included those with the BAP to enable evaluation of whether findings were applicable across the autism spectrum. Models were also tested for invariance with respect to age and sex.

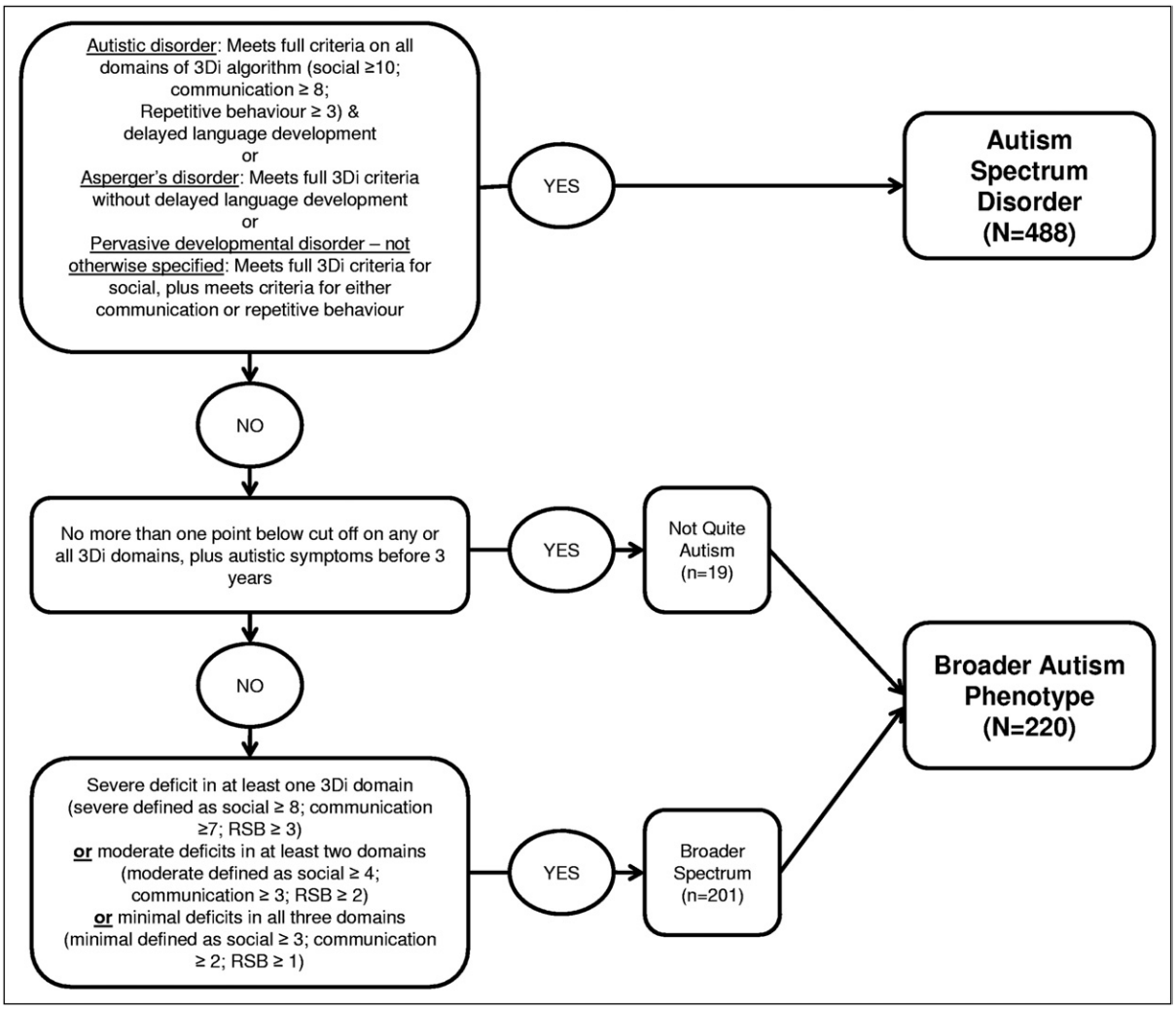
METHOD

Participants

Ethical approval for the current study was granted after review by the local hospital Research Ethics Committee. Participants (n = 708) had received a comprehensive assessment for an autism spectrum disorder (ASD) at a specialist service in London, UK. They were consecutive referrals between April 1999 and July 2009 who either received an ASD diagnosis, or met study criteria for having the broader autism phenotype (BAP). The diagnostic algorithm by which these inclusion criteria were implemented is described below and is depicted in Figure 1.

Data were collected by experienced psychiatrists and clinical psychologists as part of a multi-disciplinary clinical assessment for a high-functioning ASD in which diagnosis was based on parent report, supplemented by direct observation and information from school. Given the difficulties of differentiating Asperger's disorder and autistic disorder using *DSM* criteria, Szatmari guidelines[15] were used to distinguish these ASDs, depending on whether a delay in the onset of language was reported. Accordingly, for a diagnosis of autistic disorder impairments in each area of the autistic triad were required, as well as delayed development of onset of single word (>24 months) or

FIGURE 1 Assignment of affected status, including implementation of Autism Genetic Resource Exchange (AGRE) "not quite autism" and "broader spectrum" categories. RSB = repetitive and stereotyped behaviors.



phrase speech (>36 months). Asperger's disorder was diagnosed in the presence of the triad of impairments without the delayed development of either single-word or phrase speech. Pervasive developmental disorder—not otherwise specified (PDD-NOS) was diagnosed in accordance with *DSM-IV-TR* guidelines.[16] According to these rules, 488 members of the total sample met criteria for an ASD, with 189 (27% of the total sample) having autistic disorder, 135 (19%) having Asperger's disorder and 164 (23%) receiving a PDD-NOS diagnosis.

In the current study, we aimed to include individuals with the BAP, as well as those meeting criteria for an ASD. There is no universal definition of the BAP, and there are no agreed thresholds for inclusion in this group. To promote comparability and replicability between studies, we based our BAP inclusion criteria on the widely used, transparent, systematic definition of "broader spectrum" and "not quite autism" presentations provided by the Autism Genetic Resource Exchange (AGRE).[17] The 3Di PDD algorithm, which outputs scores equivalent to those obtained from the Autism Diagnostic Interview—Revised (ADI-R)[18] diagnostic algorithm, was used to implement these AGRE affected status categorizations, as described in Figure 1. Accordingly, 220 BAP participants were identified, of whom 201 (28% of the total sample) met criteria for AGRE "broader spectrum" difficulties, with the remaining 19 (3%) being classified as "not quite autism."

The mean age of participants was 9.5 years (range: 2.4-21.1 years) and the sample included 560 boys (79%) and 148 girls (21%). A measure of verbal IQ was available for 409 participants (57% of the sample) and performance IQ was available for 356 participants (50%). This reflects changes in clinic practice over time rather than any tendency to conduct IQ testing de-

TABLE 1 Characteristics of the Sample by Affected Status Category

	Autism Spectrum Disorder n = 488	Broader Autism Phenotype n = 220	Significance (*p* value)
Proportion male	81.6%	73.2%	.02
Mean age in years (SD)	9.60 (3.60)	9.37 (3.35)	.43
Mean Verbal IQ (SD)[a]	92.99 (19.01)	92.69 (20.91)	.89
Proportion verbal IQ below 70[a]	9.7%	15.5%	.10
Mean performance IQ[b]	93.84 (19.50)	96.15 (17.88)	.31
Proportion performance IQ below 70[b]	11.1%	9.6%	.67
Mean 3Di social	15.98 (3.46)	7.86 (2.39)	<.001
Mean 3Di communication	14.19 (3.47)	7.77 (3.05)	<.001
Mean 3Di repetitive and stereotyped behavior	4.38 (2.59)	1.67 (1.72)	<.001

Note: [a]*Verbal IQ available for N = 409.*
[b]*Performance IQ available for N = 356.*

pending upon a particular child's presenting difficulties. Mean VIQ was 92.9 (SD: 19.5, range: 40-153) and mean PIQ was 94.4 (SD: 19.1, range: 47-143). All participants had at least phrase speech and were in mainstream education, as these are criteria for accepting referrals at the clinic where this study took place. Table 1 shows participant characteristics for the ASD and BAP groups. These did not differ according to age or intelligence, but there were a higher proportion of males among those with an ASD diagnosis.

Measures

The Developmental, Dimensional and Diagnostic Interview (3Di). This validated, semi-structured, parent-report interview uses a hierarchical computerized algorithm to combine responses to 120 items, which cover both past and current behavior. Individual items are summed to generate 12 subscale scores, which represent the 12 individual *DSM-IV-TR* autistic symptom criteria. These are in turn summed to give three independent dimensional scores, one for each element of the *DSM-IV-TR* autism triad. The reciprocal social interaction scale of the 3Di is the sum of subscales measuring; (S1) nonverbal interaction; (S2) peer relationships; (S3) sharing; (S4) socio-emotional reciprocity. The communication scale comprises subscales measuring: (C1) nonverbal communication; (C2) conversational abilities; (C3) stereotyped and repetitive use of language; (C4) play and imagination. The repetitive behavior scale is made up of subscales measuring: (R1) unusual preoccupations; (R2) routines and rituals; (R3) stereotyped and repetitive motor behavior; (R4) persistent preoccupation with parts of objects. It is these subscales that were used in the CFA reported here. The 3Di algorithm is designed to generate autism triad scores that are equivalent to those derived from the ADI-R diagnostic algorithm, with identical ranges and thresholds for abnormality.

Research indicates that the 3Di has strong psychometric properties.[14,19] Interrater and test–retest reliability of the 3Di are high, yielding intraclass correlations greater than 0.86. Agreement between the 3Di and the ADI-R for threshold scores that comprise the ADI-R algorithm is also high: 86% for reciprocal social interaction, 100% for communication and 76% for repetitive and stereotyped behaviors. Agreement on caseness with clinician diagnosis is excellent (positive predictive power: 0.93, negative predictive power: 0.91).

For the current study, we also created a sensory abnormalities scale from five 3Di items concerning hypo- and hypersensitivity to sounds and textures. Cronbach's α for this scale was 0.8.

Measures of Intelligence. Data on IQ were collected as part of a routine clinical assessment, and as such a variety of measures were used. Instruments included the British Picture Vocabulary Scale,[20] the Wechsler Abbreviated Scale of Intelligence,[21] and the Wechsler Intelligence Scale for Children—Third[22] and Fourth[23] editions.

Statistical Analyses

Factor analysis is a technique for investigating the relationships between observed variables and their underlying latent constructs. In confirmatory factor analysis (CFA), these relationships are specified a priori, and then tested statistically. This involves the evaluation of models to see how well they fit the data from a given sample. In the current study, CFA was performed using AMOS 19.[24]

Initially, data from the 12 subscales of the 3Di diagnostic algorithm were used to test three alternative models:

1. One-factor model—all 12 3Di subscales loading onto one "autism" factor.
2. *DSM-IV* model—based on *DSM-IV-TR* and reflecting the existing structure of the 3Di algorithm with three factors of reciprocal social interaction (S1, S2,

S3, S4), communication (C1, C2, C3, C4) and repetitive behavior (R1,R2,R3,R4).

3. *DSM-5* model—based on draft criteria[5] with two factors, one for social communication deficits (S1, S2, S3, S4, C1, C2, C4) and one for restricted and repetitive behavior (C3, R1, R2, R3, R4). Note that in this model, sensory abnormalities were not included. This is because sensory abnormalities are not part of *DSM-IV-TR* criteria, and we wanted initially to test our *DSM-5* model against alternatives using exactly the same data set. Please see below for description of the subsequent addition of sensory abnormality data to a modified *DSM-5* model.

In CFA there is no single definitive indicator of model fit, so several indices of fit were used to assess model validity, chosen according to recommendations in the literature and to match those used in other CFA studies of autistic symptoms. These were the root mean square error of approximation (RMSEA); standardized root mean residual (SRMR); comparative fit index (CFI); the goodness of fit index (GFI); and the consistent version of Akaike's information criterion (CAIC). The RMSEA estimates a model's fit with the population covariance matrix. Values below 0.05 are considered good, and models with values above 0.1 should not be accepted.[25] The SRMR describes the average discrepancy between the observed and hypothesized correlation matrices, with values ranging from 0 to 1. Values below 0.08 indicate good fit. The CFI compares the hypothesized model to the interdependence model, in which all parameters are assumed to be zero. Values above 0.9 are considered acceptable, with a CFI above 0.95 indicating good fit.[26] The GFI is equivalent to r^2 in regression as it describes the proportion of sample variability explained by the model. Values close to 1 are desirable, with those in excess of 0.9 indicating acceptable fit. The CAIC reflects the extent to which parameter estimates will be valid in future samples, with smaller values relative to alternative models suggestive of better-fitting, more parsimonious models.

In addition to general measures of model fit, CFA yields information about how well individual variables fit within a model. Standardized residuals capture discrepancies between the sample covariance matrix and the restricted covariance matrix that is implied by a particular CFA model. As such they are an index of misspecification of individual variables, with values in excess of ±2 taken to indicate potentially problematic items in the current study. Modification indices are another source of information about item misspecification. These are estimates of how much a model would be improved (in terms of reduction in χ^2) if fixed parameters were to be freely estimated, and so can elucidate the nature of the relationship between misspecified items in a model. Standardized residuals and modification indices can be used to identify ways in which a model could be adapted (or "respecified") and improved in a post hoc procedure called "specification searching".[25]

Using specification searches to address a priori concerns, we aimed to derive a modified, better-fitting *DSM-5* model. We were particularly interested in understanding whether stereotyped and repetitive use of language (C3) loads onto the RRB factor; and whether or not play and imagination (C4) loads onto the social communication factor. To this end, once the *DSM-5* model had been estimated we inspected standardized residuals and modification indices for these variables to see whether they were well specified, and whether their removal or respecification would improve model fit. We also ran a *DSM-5* model that included sensory abnormalities as a manifestation of RRB, as is proposed in current draft criteria. For this model we were interested to see the overall model fit; the factor loading on RRB for sensory abnormality; and standardized residuals for the sensory abnormalities variable.

For the purpose of these model fitting procedures, the overall sample was randomly divided in two to create an "estimation sample" (n = 367) and a "hold-out sample" (n = 341). This enabled us to use specification searches to develop an improved *DSM-5* model in the estimation sample, and then to cross-validate the resultant model in the independent hold-out sample. This procedure reduced the risk of us deriving an over-fitted model that would not generalize beyond the dataset used in its development.

Finally, fit of the best-fitting *DSM-5* model was tested in subgroups defined according to sex, age, and symptom severity, to see if this model's structure and parameters were stable and generalizable. Three age groups were defined: early childhood (aged 2 to <7 years); middle childhood (aged 7 to <12 years); and late childhood/adolescence (aged 12 to <21 years). Two groups (ASD and BAP) were derived for symptom severity analyses. For each comparison, we initially inspected configural invariance (whether the same items loaded onto the same factors in each group), by examining the fit "free model" in which all parameters were allowed to vary across groups. Then we assessed construct-level metric invariance (whether factor loadings were similar in the different groups) by comparing the fit of the free model with that of a model in which all factor loadings were assumed to be the same across groups. A change in CFI greater than 0.01 was taken to indicate a lack of metric invariance between groups.[27] If we discovered noninvariance, we then systematically tested each item's factor loading for invariance using the procedure described by Byrne.[25]

RESULTS

Comparison of *DSM-5*, *DSM-IV-TR*, and One-Factor Models

Table 2 shows indices of fit for each model. The current *DSM-IV-TR* model did not meet criteria for acceptable fit. The one-factor model performed especially badly, scoring worst on each index of fit.

TABLE 2 Fit Indices for Confirmatory Factor Analysis Models Tested Against Whole Sample (N = 708)

	χ^2	DF	SRMR	GFI	CFI	RMSEA (90% CI)	CAIC
DSM-IV	405.6	51	0.067	0.910	0.892	0.099 (0.090–0.108)	609.870
One-factor model	568.6	54	0.074	0.856	0.844	0.116 (0.101–0.125)	750.140
DSM-5	285.4	53	0.049	0.935	0.929	0.079 (0.070–0.088)	474.416
Modified *DSM-5*	124.1	34	0.037	0.966	0.964	0.061 (0.05–0.073)	282.865
DSM-5 with sensory abnormalities	132.6	43	0.035	0.967	0.967	0.054 (0.044–0.065)	305.943

Note: CAIC = consistent version of Akaike's information criterion (lower values suggest better models); CFI = comparative fit index (>0.9 suggests adequate fit); CI = confidence interval; GFI = goodness of fit index (>0.9 suggests adequate fit); RMSEA = root mean square error of approximation (<0.1 suggests adequate fit); SRMR = standardized root mean residual (<0.08 suggests adequate fit).

Only the *DSM-5* model had acceptable values for all the indices of fit. For each index, its value was better than those of the other models.

Modifying the *DSM-5* Model

In the estimation sample, we examined the standardized residuals of the *DSM-5* model to evaluate whether stereotyped and repetitive use of language (C3) loads onto the RRB factor; and whether play and imagination (C4) loads onto the social communication factor. C3 appeared to be a problematic item, as its standardized residual covariance was outside the acceptable range (greater than ±2) for two other communication subscales (C1 and C4). C4 (play and imagination) also appeared to be misspecified, as it was highly correlated with C2 (conversational abilities), as indicated by their high standardized residual covariance (2.79) and the fact that the largest reduction of the model χ^2 (19.18) would be achieved by allowing their error terms to co-vary.

Given the possibility that stereotyped and repetitive language may map onto both the RRB and SC domains, we ran a model that allowed C3 to load onto both these factors. This resulted in no improvement in model fit (ΔCFI = 0.001). On the basis of indices of model misspecification, and our a priori concerns about stereotyped and repetitive use of language and play and imagination criteria in *DSM-5*, we decided to remove C3 and C4 to see how this improved our *DSM-5* model. In the estimation sample the fit indices of the resultant 'modified *DSM-5* model' were suggestive of good fit (CFI = 0.97, RMSEA = 0.06), whereas those of the initial *DSM-5* model implied only adequate fit (CFI = 0.93, RMSEA = 0.08).

Next, we added the 3Di sensory abnormalities variable to the modified *DSM-5* model, as depicted in Figure 2. The fit indices of this "*DSM-5* sensory model" were indicative of good fit in the estimation sample (CFI = 0.98, RMSEA = 0.05). The loading of the sensory abnormalities variable on to RRB was 0.58, and all standardized residuals for this variable were comfortably in the range of −2 to 2, suggesting that it was well specified as a manifestation of the RRB factor.

We then tested how well this *DSM-5* sensory model fit in the hold-out sample. Fit indices of a free model allowing all parameters to vary between the estimation and hold-out groups suggested configural invariance (CFI = 0.97, RMSEA = 0.04). Metric invariance was also demonstrated, since a model assuming equality of all factor loadings in both groups did not fit worse than in the free model (ΔCFI = 0.001). Given this evidence for the generalisability of the *DSM-5* sensory model to the hold-out sample, all subsequent analyses used the whole sample. In Table 2, we present fit statistics for the modified *DSM-5* model and *DSM-5* sensory model tested against the whole sample, to allow comparison with other models.

Tests of *DSM-5* Model Invariance According to Age, Sex, and Autism Symptom Severity

CFAs examining model invariance with respect to age, sex, and autism spectrum severity were performed for the *DSM-5* sensory model. Model fit indices for this model in the different sex, age, and autism spectrum severity groups are shown in Table 3. Fit statistics for the free model estimated across age (CFI = 0.97, RMSEA = 0.03), sex (CFI = 0.96, RMSEA = 0.04) and symptom severity (CFI = 0.94, RMSEA = 0.03) suggest that it has configural invariance (i.e., that its structure

FIGURE 2 *DSM-5* sensory model. Note: RRB = restricted, repetitive behavior; SC = social communication.

and pattern of item loadings are stable) in relation to these variables.

Next we checked the invariance of factor loadings (metric invariance). Restricting the model in the three different age groups and by sex did not significantly decrease model fit compared with that in the unrestricted model (ΔCFI = .001). When comparing ASD and BAP groups, the restricted *DSM-5* model performed less well than an unrestricted model (ΔCFI = 0.041). Systematic investigation of individual items identified S3 (sharing), S4 (socio-emotional reciprocity), and C1 (nonverbal communication) as having different sized factor loadings in ASD and BAP groups. All had weaker loadings onto the social communication factor in the BAP compared with the ASD group.

DISCUSSION

In a sample of verbal children with social communication difficulties, we used confirmatory factor analysis to assess the construct validity of "autism spectrum disorder" (ASD) as it is described in *DSM-5* draft criteria. Specifically, we investigated whether ASD has two dimensions of social communication and restricted, repetitive behavior (RRB); whether repetitive and stereotyped use of language should be considered a symptom of RRB; whether impaired play and imagination should be considered an autistic social communication symptom; and whether sensory abnormalities are part of the RRB symptom cluster. Furthermore, we tested whether our *DSM-5* model was consistent for male and female participants and throughout childhood and adolescence, and whether it applied to the broader autism phenotype (BAP) as well as to those with more severe autistic difficulties.

In common with the majority of CFA studies using ADI-R measured symptoms,[7-11] a three-factor model emulating *DSM-IV-TR* was not supported in our analyses of 3Di data. We also confirmed previous findings that autism is not well conceptualized as a single dimension.[9,11,12] Instead, all indices of fit pointed to the superiority of a two-factor model of autistic symptomatology, comprising social communication and RRB dimensions. Thus the current study offers an endorsement of the dyadic model proposed in *DSM-5* draft criteria, and shows that previous findings that informed their development generalize beyond ADI-R measured symptoms. Our

TABLE 3 *DSM-5* Model Goodness of Fit Statistics for Groups of Participants (Degrees of Freedom for All Models = 43)

Symptom severity	n	χ^2	SRMR	GFI	CFI	RMSEA (90% CI)	CAIC
Autism Spectrum Disorder	488	98.740	0.050	0.964	0.929	0.052 (0.038–0.065)	264.117
Broader autism Phenotype	220	58.469	0.052	0.953	0.955	0.041 (0.000–0.065)	205.523
Age							
Early childhood	190	80.315	0.055	0.924	0.946	0.067 (0.044–0.090)	122.237
Middle childhood	337	79.675	0.038	0.959	0.975	0.051 (0.033–0.068)	236.400
Late childhood and adolescence	181	49.765	0.046	0.953	0.989	0.030 (0.000–0.061)	192.331
Sex							
Male	560	110.244	0.031	0.965	0.967	0.053 (0.041–0.065)	278.745
Female	148	80.038	0.058	0.920	0.946	0.077 (0.050–0.102)	217.974

Note: CAIC = consistent version of Akaike's information criterion (lower values suggest better models); CFI = comparative fit index (>0.9 suggests adequate fit); CI = confidence interval; GFI = goodness of fit index (>0.9 suggests adequate fit); RMSEA = root mean square error of approximation (<0.1 suggests adequate fit); SRMR = standardized root mean residual (<0.08 suggests adequate fit).

analyses are of parent-reported symptoms, but it is notable that a similar two-factor model has been validated using direct observational data from the Autism Diagnostic Observation Schedule and multi-factor item-response theory.[28]

We investigated the validity of specific proposed changes to diagnostic criteria through the planned examination of indices of misspecification ("specification searches") for individual 3Di subscales within the *DSM-5* model. Evidence for the proposal that stereotyped, repetitive language (subscale C3 of the 3Di algorithm) be considered a type of RRB, not a symptom of communication impairment, was equivocal. Although the *DSM-5* model that treated C3 as a type of RRB was adequate, and C3 did not cross-load onto the social communication factor, inspection of standardized residuals suggested that C3 covaried strongly with items measuring social communication. Furthermore its removal improved model fit. This phenomenon has been reported elsewhere[11] and suggests that questions measuring stereotyped, repetitive language may capture both social communication and RRB difficulties. We do not suggest that stereotyped speech be removed from the list of RRB symptoms; rather, we highlight the need, when designing instruments, to implement *DSM-5* criteria, to carefully select items that focus on stereotyped language but not stereotyped language used specifically within the context of social interaction. Furthermore, there may be items capturing other aspects of stereotyped language, such as a reliance on scripted phrases in conversation, that do load onto the SC factor.

We also used planned specification searches to examine whether deficits in play and imagination should be considered a core symptom of autistic social communication impairment. It has been removed from current drafts of *DSM-5*, although shared imaginative play is mentioned as an example of a joint activity with the context of a typical relationship. The subscale measuring play and imagination (C4) was found to be problematic within the *DSM-5* model tested, reflecting its covariation with another subscale measuring conversation abilities (C2). This finding is compatible with the notion embodied in draft *DSM-5* criteria that difficulties with play and imagination are relevant to an ASD diagnosis, but that the key issue is the extent to which they can occur socially, in the context of reciprocal exchange.

When a sensory abnormalities subscale was added to the *DSM-5* model, there was no evidence of a reduction in model fit. Inspection of factor loadings and standardized residuals showed that this subscale, based on items about hypo- and hyper-sensitivity to sound and textures, loads comfortably onto the RRB factor. This is consistent with findings elsewhere that sensory abnormalities are more strongly correlated with RRB than with autistic social and communication difficulties.[29] Sensory difficulties are common in ASD,[30] represent a risk factor for reduced functional adaption,[31] and are a prominent feature of the phenomenology of ASD.[32] Despite this, they have received little attention from researchers, who have tended to focus on autistic social communication impairment. The

inclusion of sensory difficulties among core ASD diagnostic criteria may now stimulate research into their etiology and treatment.

Previous CFA findings have shown that models of autistic symptoms can vary according to participant characteristics.[12] In our mostly high-functioning sample, the *DSM-5* sensory model showed stability according to age and sex. In groups defined according to these variables, its basic factor structure and pattern of item loadings were stable (configural invariance) and the size of factor loadings were similar (metric invariance). As our sample included individuals from right across the autism spectrum, including individuals with the BAP, we were able to test the invariance of our best-fitting *DSM-5* sensory model according to symptom severity. The model fit well for both the ASD and BAP participants, demonstrating configural invariance across the range of symptom severity. However there was evidence that loadings onto the SC factor for subscales measuring sharing (S3), socio-emotional reciprocity (S4), and non-verbal communication (C1) were lower in the BAP than in ASD. One interpretation of this finding is that these sub-scales, which were initially designed to measure difficulties of clinical severity, are somewhat better indicators of social communication difficulties in ASD than in individuals with sub-threshold presentations. It will be useful in future to develop and test items specifically designed to capture the more subtle social difficulties characteristic of the BAP.

This study used a large, heterogeneous sample to test the construct validity and factor invariance of ASD using a well-validated measurement instrument. A *DSM-5* model was developed in an estimation sample and cross-validated in a hold-out sample, to test its generalisability. Nevertheless, our findings must be considered in the light of the following limitations. Firstly, in line with other clinic-based[33] and epidemiological[34] samples, our participants were mainly high-functioning. There is some evidence that the factor structure of the autistic phenotype can vary with IQ such that better model fit is found in more intellectually able groups.[12] We acknowledge that our findings may not generalize to lower functioning groups. In fact there is a possibility that high- and low- functioning ASD have distinct factor structures, because those CFA studies that included a large proportion of ASD individuals with comorbid intellectual disability have tended not to endorse a two-factor *DSM-5*-type model,[7,8,10] because they report, in addition to social communication and RRB, an "imagination" factor. It will be valuable in future to formally test for invariance of the *DSM-5* model right across the range of intellectual ability in ASD.

Second, because we used subscales from the 3Di validated diagnostic algorithm, which was based on *DSM-IV* and ICD-10 criteria, we were able to test a model that approximates, but that is not identical to, the model proposed in *DSM-5* draft criteria. This is because each 3Di subscale does not exactly correspond to a single proposed *DSM-5* criterion: specific *DSM-5* algorithms that have been validated and applied in sufficient numbers will be required for a full test of the construct validity of ASD.

Our findings suggest that the autism phenotype is inadequately described by *DSM-IV-TR* criteria, and that the proposed autism dyad of *DSM-5* has greater validity, with the core impairments of ASD being manifestations of separable social communication and RRB dimensions. In addition, we present support for the notion that sensory abnormalities are best conceptualized as an aspect of RRB. Evidence is emerging that it is "time to give up on a single explanation for autism,"[3] and that distinct aetiologies may underpin different dimensions of autistic impairment.[4] On this basis, it will be valuable to investigate whether the social communication and RRB dimensions we describe are associated with distinct endophenotypes and genotypes.[35] &

Accepted October 21, 2011.

Dr. Skuse is with the Institute of Child Health, University College London (UCL), London UK. Dr. Mandy is with The Research Department of Clinical, Health and Educational Psychology, University College London, London, UK. Dr. Charman is with the Centre for Research in Autism and Education, Institute of Education, London, UK.

The authors thank Dr. Pasco Fearon for his valuable expert advice on several analyses in the current study.

Disclosure: Dr. Skuse is a stockholder in IxDx Ltd., which owns exclusive rights to 3Di software and to the dissemination of 3Di technology and intellectual property. Dr. Charman receives grant or research support from the Medical Research Council, the European Science Foundation, Autistica, the Autism Education Trust, and the Department of Education. He receives royalties from Guilford Press and Sage. Dr. Mandy reports no biomedical financial interests or potential conflicts of interest.

Correspondence to Dr. William Mandy, Research Department of Clinical, Educational and Health Psychology, University College London, UK, WC1N 6BT; E-mail: w.mandy@ucl.ac.uk

0890-8567/$36.00/©2012 American Academy of Child and Adolescent Psychiatry

DOI: 10.1016/j.jaac.2011.10.013

REFERENCES

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, text revision. Washington, DC: American Psychiatric Association; 2000.
2. World Health Organisation. The International Classification of Diseases and Disorders—10: Diagnostic Criteria for Research. Geneva: World Health Organization, 1993.
3. Happe F, Ronald A, Plomin R. Time to give up on a single explanation for autism. Nat Neurosci. 2006;9:1218-1220.
4. Mandy WP, Skuse DH. Research review: what is the association between the social-communication element of autism and repetitive interests, behaviours and activities? J Child Psychol Psychiatry. 2008;49:795-808.
5. American Psychiatric Association. Proposed Revision. Available at: http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=94. Accessed July 6, 2011.
6. Wing L, Gould J, Gillberg C. Autism spectrum disorders in the DSM-V: better or worse than the DSM-IV? Res Dev Disabil. 2011;32:768-773.
7. Boomsma A, Van Lang ND, De Jonge MV, De Bildt AA, Van EH, Minderaa RB. A new symptom model for autism cross-validated in an independent sample. J Child Psychol Psychiatry. 2008;49:809-816.
8. Van Lang ND, Boomsma A, Sytema S, *et al.* Structural equation analysis of a hypothesised symptom model in the autism spectrum. J Child Psychol Psychiatry. 2006;47:37-44.
9. Frazier TW, Youngstrom EA, Kubu CS, Sinclair L, Rezai A. Exploratory and confirmatory factor analysis of the Autism Diagnostic Interview—revised. J Autism Dev Disord. 2008;38:474-480.
10. Georgiades S, Szatmari P, Zwaigenbaum L, *et al.* Structure of the autism symptom phenotype: a proposed multidimensional model. J Am Acad Child Adolesc Psychiatry. 2007;46:188-196.
11. Snow AV, Lecavalier L, Houts C. The structure of the Autism Diagnostic Interview—Revised: diagnostic and phenotypic implications. J Child Psychol Psychiatry. 2009;50:734-742.
12. Lecavalier L, Gadow KD, Devincent CJ, Houts C, Edwards MC. Deconstructing the PDD clinical phenotype: internal validity of the DSM-IV. J Child Psychol Psychiatry. 2009;50:1246-1254.
13. Constantino JN, Todd RD. Autistic traits in the general population: a twin study. Arch Gen Psychiatry. 2003;60:524-530.
14. Skuse D, Warrington R, Bishop D, *et al.* The Developmental, Dimensional and Diagnostic Interview (3di): a novel computerized assessment for autism spectrum disorders. J Am Acad Child Adolesc Psychiatry. 2004;43:548-558.
15. Szatmari P. The classification of autism, Asperger's syndrome, and pervasive developmental disorder. Can J Psychiatry. 2000;45:731-738.
16. Mandy W, Charman T, Gilmour J, Skuse D. Toward specifying pervasive developmental disorder—not otherwise specified. Autism Res. 2011;4:121-31.
17. Autism Genetic Research Exchange. AGRE Affected Status Categories. Available at: http://research.agre.org/agrecatalog/algorithm.cfm. Accessed July 6, 2011.
18. Lord C, Rutter M, Le CA. Autism Diagnostic Interview—Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord. 1994;24:659-685.
19. Santosh PJ, Mandy WP, Puura K, Kaartinen M, Warrington R, Skuse DH. The construction and validation of a short form of the Developmental, Diagnostic and Dimensional Interview. Eur Child Adolesc Psychiatry. 2009;18:521-524.
20. Dunn LM, Dunn DM, Whetton C, Burley J. British Picture Vocabulary Scale: Second Edition. NFER Nelson; 1997.
21. Wechsler D. Manual for the Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: Psychological Corporation, 1999.
22. Wechsler D, Golombok J, Rust S. Manual for the Wechsler Intelligence Scale for Children—Third UK Edition. London, UK: Psychological Corporation, 1991.
23. Wechsler D. Manual for the Wechsler Intelligence Scale for Children—Fourth UK Edition. San Antonio, TX: Psychological Corporation, 2003.
24. Arbuckle J. IBM SPSS AMOS 19 User's Guide. Chicago, IL: SPSS Inc.; 2010.
25. Byrne BM. Structural Equation Modelling with AMOS. 2nd ed. New York, NY: Routledge, 2010.
26. Hu L-T and Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equat Model. 1999:6;1-55.
27. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equat Model. 2002;9:233-255.
28. Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. J Autism Dev Disord. 2007;37:613-627.
29. Chen YH, Rodgers J, McConachie H. Restricted and repetitive behaviours, sensory processing and cognitive style in children with autism spectrum disorders. J Autism Dev Disord. 2009;39:635-642.
30. Ben-Sasson A, Hen L, Fluss R, Cermak SA, Engel-Yeger B, Gal E. A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. J Autism Dev Disord. 2009;39:1-11.
31. Lane AE, Young RL, Baker AE, Angley MT. Sensory processing subtypes in autism: association with adaptive behavior. J Autism Dev Disord. 2010;40:112-122.
32. Grandin T. Thinking in pictures and other reports from my life with autism. London: Bloomsbury, 2006.
33. Fischbach G, Lord C. The Simons Simplex Collection: a resource for the identification of autism genetic risk factors. Neuron. 2010;68:192-195.
34. Kim Y, Leventhal BL, Koh Y, *et al.* Prevalence of autism spectrum disorders in a total population sample. Am J Psychiatry. 2010;168:904-912.
35. Charman T, Jones CRG, Pickles A, *et al.* Defining the cognitive phenotype of autism. Brain Res. 2011;1380:10-21.