The validity of psychiatric diagnoses: The case of 'specific' developmental disorders

The Validity of Psychiatric Diagnoses: The Case of 'Specific' Developmental Disorders

Murray J. Dyck,[1] Jan P. Piek[2] and Jeff Patrick[3]

[1] Professor of Clinical Psychology, School of Psychology, Griffith University, Gold Coast, Queensland, Australia, 4222

[2] Professor of Developmental Psychology, School of Psychology & Speech Pathology, Curtin University, Box U1987, Perth, Western Australia, 6845

[3] Associate Professor of Health Service Management, Faculty of Business, University of Tasmania, Level 11 King George V Building, Royal Prince Alfred Hospital, 83-117 Missenden Road, Camperdown, New South Wales, 2050

Running head: Diagnosis of Developmental Disorders

Correspondence to:

Murray J. Dyck, School of Psychology, Griffith University, Gold Coast, Queensland, Australia, 4222
Tel: 61 7 5552 8251
Fax: 61 7 5552 8291
Email: m.dyck@griffith.edu.au

Word Count: 6218

Abstract

We tested whether Developmental Coordination Disorder (DCD) and Mixed Receptive Expressive Language Disorder (RELD) are valid diagnoses by assessing whether they are separated from each other, from other childhood disorders, and from normality by natural boundaries termed zones of rarity. Standardized measures of intelligence, language, motor skills, social cognition, and executive functioning were administered to children with DCD (n=22), RELD (n=30), Autistic Disorder (n=30), Mental Retardation (n=24), Attention Deficit / Hyperactivity Disorder (n=53) and to a representative sample of children (n=449). Discriminant function scores were used to test whether there were zones of rarity between the DCD, RELD, and other groups. DCD and RELD were reliably distinguishable only from the Mental Retardation group. Cluster and latent class analyses both resulted in only two clusters or classes being identified, one consisting mainly of typical children and the other of children with a disorder. Fifty percent of children in the DCD group and 20% in the RELD group were clustered with typical children. There was no evidence of zones of rarity between disorders. Rather, with the exception of Mental Retardation, the results imply there are no natural boundaries between disorders or between disorders and normality.

Key words: psychiatric diagnosis, validity, developmental disorders, Developmental Coordination Disorder, Mixed Receptive-Expressive Language Disorder

1. Introduction

Planning for DSM-V has been strongly influenced by the fact that the syndromes defined in DSM-III and its successors have not been validated (Kupfer, First & Regier, 2002). Addressing the validity criteria proposed by Robins and Guze (1970) and Kendler (1980), Kupfer et al. summarised almost three decades of research:

> "despite many proposed candidates, not one laboratory marker has been found to be specific in identifying any of the DSM-defined syndromes. Epidemiologic and clinical studies have shown extremely high rates of comorbidities among the disorders, undermining the hypothesis that the syndromes represent distinct etiologies. Furthermore, epidemiologic studies have shown a high degree of short-term diagnostic instability for many disorders. With regard to treatment, lack of treatment specificity is the rule rather than the exception" (p. xviii).

Kendell and Jablensky (2003) questioned whether the criteria used to establish the validity of psychiatric diagnoses are appropriate. They argued that the Robins and Guze (1970) and Kendler (1980) criteria were based on the implicit and possibly unjustified assumption that psychiatric disorders are discrete, mutually discriminable, disease entities. Kendell and Jablensky suggested that "the possibility that disorders might merge into one another with no natural boundary in between ... was simply not considered" (p. 5). Kendell and Jablensky's critique does not imply that DSM-defined disorders are invalid. Rather, they state that "although there is a growing assumption, at least within the research community, that most currently recognized psychiatric disorders are not disease entities,

this belief has never been demonstrated, mainly because studies of the appropriate kind have rarely been mounted" (p. 7).

According to Kendell and Jablensky (2003), the key characteristic of valid disease entities is that they are separated from each other and from normality by natural boundaries. Natural boundaries are defined as *zones of rarity* (cf. Sneath, 1957) such that the characteristics or symptoms of one syndrome are discontinuous with normality and the symptoms of other syndromes. If zones of rarity cannot be identified, the implication is that there are no natural boundaries between syndromes and they "are merely arbitrary loci in a multidimensional space in which variation in both symptoms and etiology is more or less continuous" (Kendell & Jablensky, p. 7).

Based on available evidence, there is reason to doubt that there are zones of rarity between specific developmental disorders (Learning Disorders, Communication Disorders, Motor Skills Disorder) and between specific and pervasive developmental disorders (Autistic Disorder, Asperger's Disorder, Pervasive Developmental Disorder Not Otherwise Specified) or other disorders usually first evident in childhood. As Kupfer et al. (2002) noted about all DSM disorders, comorbidities among specific development disorders are high, as they would be between specific and pervasive developmental disorders if they were not precluded by hierarchical classification rules in DSM-IV. Indeed, comorbidity among developmental disorders is so high that a child who is diagnosed with any developmental disorder is more likely than not to also meet diagnostic criteria for one or more other developmental disorders (e.g., Cohen, Vallance, Barwick, Im, Menna et al., 2000; Frazier, Biederman, Bellordre, Garfield, Geller, et al., 2001; Hofvander, Delorme, Chaste, Nydén,

Wentz, et al., 2009; Kadesjo & Gillberg, 2001; Kaplan, Dewey, Crawford & Wilson, 2001; McGrath, Hutaff-Lee, Scott, Boada, Shriberg & Pennington, 2008; Miniscalco, Nygren, Hagberg, Kadesjo & Gillberg, 2006; Watemberg, Waiserberg, Zuk & Lerman-Sagie, 2007).

There is also reason to doubt that there are zones of rarity between these disorders and normality. Specific developmental disorders are diagnosed when a child's measured performance in academic skills, language, or motor skills is substantially lower than the child's verbal and/or non-verbal intelligence. This means that a child's specific ability score, the child's intelligence score, and the discrepancy between these two ability scores all represent points on normal distributions of scores and the scores of children with a disorder are continuous with those of children with no disorder. This continuity makes it unlikely that there is a zone of rarity between disordered and normal children. Rather, just as Mental Retardation is partly defined by very low scores that represent the bottom tail of the normal distribution of intelligence test scores, children with specific developmental disorders have scores that place them at the bottom tail of the normal distributions of specific ability scores and/or at the bottom tail of the normal distribution of discrepancy scores.

According to Kendell (1989), discriminant function analyses are the most appropriate test to assess for zones of rarity separating syndromes from each other and from normality. The aim is to assess whether members of groups with and without disorders can be accurately classified, as would be expected of discrete entities, and also to see if the distribution of discriminant function scores is bimodal. It is the trough between the two

peaks in the distribution that represents the zone of rarity, the intermediate cases between the two syndromes or between a syndrome and normality.

A second approach to validating hypothetically discrete syndromes is cluster analysis. It was initially thought by Robins and Guze (1970) that cluster analysis would be useful in the first stage of the validation process, that is, in the initial description of syndromes. The procedure was expected to group individuals based on shared characteristics, the sets of correlated symptoms that would define different disorders, and these groups would then be shown to have other characteristics in common (like course and response to treatment) that would validate the grouping. A weakness of the approach was that effective clustering depends on prior knowledge of which symptoms need to be included in the analysis, something that can't be known until the validity of a syndrome has been demonstrated. Cluster analysis is more useful as a way of confirming that individuals who share a diagnosis are grouped with each other on the basis of the defining symptoms, and not with healthy people or people with another diagnosis.

A third approach is latent class analysis, a method for finding subtypes of related cases (referred to as latent classes) from multivariate ordinal or categorical data (Lazarsfeld & Henry, 1968). In most respects, latent class analysis is similar to cluster analysis, except that the assumption of conditional independence leads to more natural and meaningful groups (if any). Conditional independence simply means that within a latent class that corresponds to a distinct syndrome, the presence or absence of one symptom is viewed as unrelated to the presence or absence of all other symptoms.

In this study, five disorders were sampled for their value in assessing the validity of two specific developmental disorders: Developmental Coordination Disorder (DCD) and Mixed Receptive Expressive Language Disorder (RELD). DCD and RELD are defined by delays in acquiring either motor skills or receptive and expressive language skills relative to either general intelligence (DCD) or nonverbal intelligence (RELD). If these disorders represent discrete disease entities, there should be a zone of rarity separating children with these disorders from each other, from typically developing children, and from children with one of three other disorders. Attention Deficit/ Hyperactivity Disorder (ADHD) is partly defined by delays in the ability to sustain attention and/or to inhibit behavior, and the delays are inferred from behavior rather than from ability testing. Autistic Disorder (AD) is partly defined by qualitative impairments in social interaction and communication and by restricted repetitive and stereotyped behavior, interests, and activities. However, DSM-IV precludes the diagnosis of DCD and RELD when a pervasive developmental disorder is present, which implies that clinically significant problems with motor skills and language are common in autism. To the extent that motor or language problems are present, this implies that both DCD and RELD may be continuous with autism. Finally, Mental Retardation (MR) is partly defined by general intellectual functioning that is substantially below average. Both DCD and RELD may be diagnosed when MR is present if the delays in motor skills or language are disproportionate to general intellectual functioning.

2. Method

2.1 Participants

We studied 608 children aged 3 to 14 years (M=8.90 years, SD=3.11): a representative sample of 449 children and 159 children with a DSM-IV (American Psychiatric Association, 2000) diagnosis of disorder. Participants were recruited after the project had gained Institutional Ethics Committee approval.

The AD group comprised 30 children (23 boys) with a mean age of 8.47 years (SD=2.68) who were recruited through a state autism register (Glasson, 2002) and research networks. All had been diagnosed by appropriate health professionals. Diagnosis was confirmed by administering the Social Communication Questionnaire (M=26.35; SD=5.81) (Rutter, Bailey, & Lord, 2003), and children with a score of 15 or more were assessed with the Autism Diagnostic Interview – Revised (Lord, Rutter & Le Couteur, 1994). Children had shown abnormality in at least 1 domain prior to age 36 months (M=3.81; SD=1.19), obtained scores of 10 or more on social items (M=21,48; SD=4.87), 8 or more on communication items (M=15.84; SD=4.48), and 3 or more, except for 1 child who scored 1, on restrictive and repetitive behavior items (M=6.48; SD=2.61). Parents were asked to report comorbid disorders. In 16 cases, 1 to 8 additional disorders were reported: ADHD=10, MR=2, RELD=1, DCD=2, learning disorder=1, sleep disorder=3, depression=1, epilepsy=1 and cerebral palsy=1.

The MR group comprised 24 children (12 boys) with a mean age of 11.56 years (SD=2.45) recruited through a health department database. Selection was based on a diagnosis of mild mental retardation (based on standardized measurement of intelligence and adaptive functioning), no diagnosis of a genetic cause of the disorder, and no diagnosis of a pervasive developmental disorder. Parents were asked to report comorbid disorders. In

8 cases, 1 to 3 additional disorders were reported: ADHD=6, learning disorder=2, epilepsy=1 and undiagnosed physical delays=1.

Participants in the RELD group were 30 children (22 boys) with a mean age of 6.78 years (SD=1.85) recruited through Language Development Centres. Each child had been diagnosed by appropriate health professionals based on the results of standardized testing and significant impairments in social communication and academic functioning. Parents were asked to report comorbid disorders. In 7 cases, 1 additional disorder was reported: attention deficit disorder=3, learning disorder=1, and DCD=3.

Participants in the DCD group were 22 children (14 boys) with a mean age of 8.55 years (SD=2.07) recruited through special education and occupational therapist referrals. We confirmed that all children had coordination problems by testing them with the Movement Assessment Battery for Children (Henderson & Sugden, 1992); each child scored below the 5^{th} percentile. All children's motor coordination problems had significantly affected their academic achievement or activities of daily living. Parents were asked to report comorbid disorders. In 2 cases, ADHD was reported.

Participants in the ADHD group were 53 children (42 boys) with a mean age of 10.88 years (SD=2.07) recruited through the public education system; 27 had been diagnosed with the inattentive subtype and 26 with the combined subtype. We used the SWAN Rating Scale (Swanson et al., 2002), a parent rating scale, to confirm the diagnosis. Children with the inattentive subtype had scores <17 on the inattentive scale and scores >17 on the hyperactivity/impulsivity scale; children with the combined subtype had scores <17 on both scales. Parents were asked to report comorbid disorders. Learning (n=2) and

depressive disorders (n=4) were reported. All but 9 children were taking methylphenidate or dexamphetamine.

The typical children group comprised 449 children (220 males) with a mean age of 8.72 years (SD=2.30) recruited from schools/preschools in Perth, Australia. Schools were targeted based on their position on an index of average student achievement, i.e., they represented the distribution of academic achievement in Western Australia.

The groups differed significantly in both sex [χ^2 (5) = 29.75, $p < .001$] and age [F(5, 602) = 11.36, $p < .001$]. The sex differences reflect the fact that most developmental disorders are more prevalent in boys than girls, and our samples reflect the prevalence rates of the different disorders.

2.2 Measures

Intelligence was measured with 4 scales from the 3rd edition of the *Wechsler Intelligence Scale for Children* (WISC; Wechsler, 1992)—Vocabulary, Information, Block Design, and Picture Completion. These subtests represent the verbal comprehension and perceptual organization components of intelligence and provide a good estimate of full-scale IQ. Each test has excellent split-half and test-retest reliability, and both criterion and concurrent validity are well-established (Wechsler). Reliability of the tests in the samples of children with a developmental disorder and typical children, respectively, was α = .92 and .93 for Information, α = .94 and .93 for Vocabulary, α = .90 and .96 for Block Design, and α = .89 and .92 for Picture Completion (Dyck, Piek, Hay, Smith & Hallmayer, 2006).

Language ability was estimated with 4 scales from the 3rd edition of the *Clinical Evaluation of Language Fundamentals* (CELF; Semel, Wiig, & Secord, 1995)—Concepts and Directions, Word Classes, Recalling Sentences, and Formulated Sentences. The CELF has been standardized across a wide range of ages. The specific scales used are the ones that are administered to all age groups and sample receptive (Concepts and Directions, Word Classes) and expressive (Recalling Sentences, Formulated Sentences) language. These subscales have acceptable internal consistency (α = .54 to .91), test-retest reliability (.69 to .87), and concurrent validity [correlations with earlier versions of the test (r = .42 to .75) and with the Wechsler scales (r = .58 to .75) (Semel et al.)]. Reliability of the tests in disordered and typical samples, respectively, was α = .96 and .95 for Concepts and Directions, α = .93 and .95 for Word Classes, α = .95 and .96 for Recalling Sentences, and α = .96 and .96 for Formulating Sentences (Dyck et al., 2006).

Motor coordination was assessed with the *McCarron Assessment of Neuromuscular Development* (MAND; McCarron, 1997). The MAND comprises 10 tasks: 5 assess fine motor skills (Beads in a Box, Beads on a Rod, Nuts and Bolts, Finger Tapping, Rod on Slide) and 5 assess gross motor skills (Finger /Nose/Finger, Hand Strength, Heel to Toe Walking, Jumping, One Foot). These tasks have acceptable test-retest reliability (.67 to .98), criterion validity (e.g., prediction of work performance), and concurrent validity [correlations with the O'Connor Finger Dexterity Test (r = -.41 to -.62), simple reaction time (r = -.31 to -.58), finger tapping (r = .35 to .53), and choice reaction time (r = -.45 to -.62); McCarron]. Reliability of the tests in disordered and typical samples, respectively,

was α = .92 and .92 for Beads in a Box, α = .86 and .89 for Beads on a Rod, α = .89 and .95 for Nuts and Bolts, α = .76 and .70 for Finger Tapping, α = .70 and .64 for Rod on Slide, α = .93 and .92 for Finger /Nose/Finger, α = .68 and .91 for Hand Strength, α = .94 and .84 for Heel to Toe Walking, α = .17 and .18 for Jumping, and α = .82 and .86 for One Foot (Dyck et al., 2006).

Social Cognition. Social cognitive ability was estimated with 3 1st order and 1 2nd order theory of mind tasks, an advanced theory of mind task, and 6 subscales from the Emotion Recognition Scales (Dyck, Ferguson & Shochet, 2001; Dyck, Farrugia, Shochet & Holmes-Brown, 2004). First order *theory of mind* tasks are false belief tasks and included the "Sally Ann" (Baron-Cohen, Leslie & Frith, 1985), "Smarties" (Perner, Frith, Leslie & Leekam, 1989; Wimmer & Perner, 1983), and "Ella the Elephant" tasks (Harris, Johnson, Hutton, Andrews & Cooke, 1989). In each task, a child is asked whether a protagonist will act consistently with the protagonist's beliefs, known to be false, or consistently with what the test-taker knows to be the true state of the world. The 2nd order theory of mind task, the "John and Mary icecream story" (Perner & Wimmer, 1985), is identical except that a child must assess what the protagonist thinks that another person thinks. We treated these tasks as separate items on a 4-point theory of mind scale. The reliability of this scale in disordered and typical samples is relatively poor (α = .51 and .64; Dyck et al., 2006). The *Strange Stories Test* assesses the ability to provide context-appropriate mental state explanations for non-literal (irony, sarcasm, lies) statements (Happe, 1994). The test is internally consistent in disordered and typical samples (α = .85; Dyck et al., 2001).

The Emotion Recognition Scales include 3 measures of emotion understanding ability. The *Emotion Vocabulary Test* measures the ability to define emotion words. The *Comprehension Test* measures the ability to understand the emotional consequences of exposure to an emotion-eliciting context. The *Unexpected Outcomes Test* measures the ability to apply reasoning skills and knowledge of the causes of emotions to explaining apparent incongruities between an emotion-eliciting context and the emotion elicited by the context. The internal consistency of these measures in disordered and typical samples, respectively, is: Emotion Vocabulary Test: α = .86 and 84; Comprehension Test: α = .78 and .79; Unexpected Outcomes Test: α = .64 and .77 (Dyck et al., 2001; Dyck, Farrugia et al., 2004).

The Emotion Recognition Scales also include 3 measures of emotion recognition ability. The *Fluid Emotions Test* (Dyck, Farrugia et al., 2004) measures the ability to recognize static and changed/changing facial expresions of emotion, including anger, contempt, disgust, fear, happiness, sadness, surprise or a neutral expression (Matsumoto & Ekman, 1995). Each item consists of 2 head and shoulders pictures of a person expressing an emotion. The test-taker is asked what emotion is being expressed and, after responding, the image is transformed to another person expressing a different emotion. Subjects identify, as quickly as they can, the second emotion. Speed of response is measured with a stop-watch. Two subscales were used: initial accuracy (initial emotions correct) and speed given accuracy, which is based on the speed of accurate post-morph responses. The internal consistency of the Accuracy and Speed Given Accuracy subscales were observed as α = .88

and α = .88, respectively, in samples of disordered children, and α = .65 and α = .84 in typical children (Dyck et al., 2001, 2006; Dyck, Farrugia et al., 2004).

The *Vocal Cues Test* (Dyck, Farrugia et al., 2004) measures the ability to recognize vocal intonations specific to 7 different emotions or a neutral expression. We used the "Unreal" scale in which emotions are expressed using non-semantic content: numerals, letters, nonsense syllables. The internal consistency of the test was α = .91 in disordered children and α = .85 in typical children (Dyck, Farrugia et al.).

Working memory was assessed with a Trailmaking/Updating Memory task, a simplification of a more complex task (Rabbit, 1997) designed to assess working memory and behavioral inhibition. In this task, the first 4 letters of the alphabet are designated as the target set, and within this target set, the actual target changes with successive presentations (from A to B to C to D to A). Participants indicate whether a letter, presented on screen, is part of the target set, and if it is, whether it is the current target. There are 2 trials of 120 stimulus presentations, of which 20 presentations are the target stimulus. For each presentation, a blue key is pressed if the stimulus is the target, otherwise a red key is pressed. We scored mean response time and variability and number correct from each of 2 trials of the test.

Response inhibition was assessed with a version of the Go/No-Go task used by Shue and Douglas (1992) to assess simple motor inhibition. In this task, letters are designated either as 'go' (respond) or 'no-go' (do not respond) stimuli, and are presented at 1-second intervals. When a go stimulus is presented, the child is asked to press a response key as

quickly as possible, and when a no-go stimulus is presented, no response is required. There were 2 trials of the task, each consisting of 120 stimuli (60 ‘go’ and 60 ‘no-go’). Responses to the ‘no-go’ stimulus were scored as commission errors, and failures to respond to the ‘go’ stimulus were scored as omission errors. In each trial, we scored the total number of commission and omission errors.

2.3 Procedure

Written informed parental consent and the agreement of participants was obtained. Participants with a disorder were individually assessed, usually at their schools, but sometimes at their home or at a university upon parental request. Testing followed a set order and was conducted in 3 sessions (2.5 hours, 2.5 hours, 1.25 hours) over 2 or 3 days. Parents of children with ADHD were asked not to administer medication on days their children were tested. Typical children were assessed individually at their schools (n=215) or at Project KIDS (n=234), a project at the University of Western Australia Child Study Centre in which research data are collected in school holidays. In Project KIDS, up to 12 children were scheduled for a day of activities that included 3 90-minute assessment sessions. The order of testing was uniform except for Project KIDS, where each child had his/her own schedule. If scheduled activities could not be completed, they were deferred to the end of the day when time was available to administer deferred tasks. Testing required 4.0 to 5.5 hours.

2.4 Data Transformations

To ensure that measures had the same scale, we used data from the representative sample to create standard scores (mean = 100, *sd* = 15) for each variable. These standard

scores were used to create composite scores, unweighted averages of standard scores on tests that defined the ability domain. The composite variables were: *perceptual organization* was the average of Block Design and Picture Completion (r = .39, p < .01, in typical children); *verbal comprehension* the average of Vocabulary and Information (r = .66, p < .01); *emotion recognition ability* was the average of Accuracy, Speed Given Accuracy, and Vocal Cues Test (r = .27 to .55, p < .01); *emotion understanding ability* was the average of Comprehension Test, Emotion Vocabulary Test, and Unexpected Outcomes Test (r = .22 to .35, p < .01); *theory of mind ability* was the average of the 5 theory of mind tasks (r = .13, p < .01); *receptive language ability* was the average of Concepts and Directions and Word Classes (r = .51, p < .01); *expressive language ability* the average of Formulating Sentences and Recalling Sentences (r = .52, p < .01); *fine motor coordination* was the average of fine motor tasks (r = .11 to .49, p < .01) and *gross motor coordination* the average of the five gross motor tasks (r = .10 to .34, p < .01; for hand strength, finger nose finger, r = .01, *ns*); *response inhibition* was the average of the two go/no go trials (r = .50, p < .01; *working memory accuracy* was the average of the two trials of the trailmaking task (r = .32, p < .01) and *working memory speed* (r = .52 to .78, p < .01) was the average response time and variability of the two trials of the trailmaking task.

Composite scores were restandardized by calculating age norms (based on the representative sample) for each composite score so that each had a mean of 100 and a standard deviation of 15. This procedure ensures that all composites have the same distribution in the population, but when used with low scoring clinical samples, it results in a larger range of scores than is obtained with conventional scoring. To ensure that very low

scores, and the associated increased range, cannot inflate observed correlations between measures, we set a minimum value of 20 for all composite scores.

3. Results

3.1 Between-Group Discrimination

We began with an omnibus discriminant function analysis based on the separate groups covariance matrices and using all variables, with missing values replaced by the respective group means, to assess the extent to which members of each group can be distinguished from each other group. The results (see Table 1, Analysis 1) show that 96.2% of typical children, 83.6% of all children, but only 47.7% of children with a disorder were correctly classified. Among children with a disorder, children with MR were most accurately classified (87.5%), followed by children with AD (66.7%) and RELD (63.3%). The most common error was for children with a disorder to be misclassified as typical, which occurred in 39.6% of cases, including 73.5% of children with ADHD. Only 12.5% of children with a disorder were misclassified as having some other disorder.

INSERT TABLE 1 ABOUT HERE

Follow-up analyses (see Table 1, Analyses 2 to 6) were conducted in which children in each diagnostic group were discriminated only from typical children. There was some improvement in classification for 2 groups: MR, where all children were accurately classified, and DCD, where 50% of children were accurately classified. Another follow-up analysis (see Table 1, Analysis 7) excluded typical children and reassessed the accuracy of classification among children with a disorder. The result showed that 88.6% of children with ADHD were correctly classified when typical children were excluded.

3.2 Zones of Rarity Between Groups

To maximize the chance of observing a zone of rarity between each group and each other group, analyses 2 to 6 were supplemented with additional ones comparing each diagnostic group with each other diagnostic group. The resulting scores on the discriminant functions were plotted as histograms to see if the distributions were bimodal. There was unambiguous evidence to support the validity of only one disorder, MR, for which the distributions of discriminant function scores were essentially non-overlapping with those of any other group. For the two specific developmental disorders, DCD and RELD, not only did each have bimodal distributions with the MR group, they also had bimodal distributions with the AD group and with each other. Both were continuous with the ADHD and TC groups. The discriminant function scores of the AD group were also continuous with those of the ADHD and TC groups, and those of the ADHD group were continuous with those of the TC group.

Observing zones of rarity between any two groups *except* the MR group depended on doing pairwise analyses. When discriminant function scores from analyses involving all groups are plotted, no bimodal distributions are observed between the AD, DCD, and RELD groups. Figure 1 presents the distribution of scores from the first and second discriminant functions in analysis 1, and shows that even where overlap between groups is minimal (e.g., between AD, MR, RELD), scores of different diagnostic groups occupy adjacent sectors of the 2-dimensional space, not distinctive clusters within that space. Similarly, the vectors that can be drawn through each sector through the group centroid of

typical children and the group centroid for each diagnostic group represent dimensions that are continuous with normality.

INSERT FIGURE 1 ABOUT HERE

3.3 Cluster Analysis

We took an empirical approach to clustering by using Schwartz's Bayesian criterion to specify how many clusters would be formed, with log-likelihood selected as the measure of distance. The procedure resulted in two clusters (see Table 2) that differ in ability level on all variables. The more able group, cluster 2, achieved at the population mean of 100 on each measure and the less able group obtained much lower scores on all measures except response inhibition. About 36% of children with a developmental disorder were in the higher scoring cluster. Only the diagnoses of MR (100% of cases), RELD (80%) and AD (73%) were mainly contained in the atypical (just over 8.5% of typical children) lower scoring cluster (see Table 3). With more than one third of children with a disorder misclassified as typical and no differentiation of children with a disorder, this analysis implies that there is no zone of rarity between disorders and normality or between different disorders.

INSERT TABLE 2 AND TABLE 3 ABOUT HERE

3.3 Latent Class Analysis

The analysis was conducted using LEM, a program for the analysis of categorical data (Vermunt, 1997). We dichotomized the 12 ability variables to distinguish cases more than one standard deviation below the mean from all the other cases. A pattern frequency matrix was created that noted the frequency of cases that fell into each possible symptom

profile. The number of symptom profiles is a function of the ordinal scale used (2 point scale in this case) and the number of variables (2^{12}=4,096 possible profiles). We then calculated all possible latent class solutions from a 2-class solution to an 11-class solution (k - 1). The relative merits of each solution were compared.

INSERT TABLE 4 ABOUT HERE

Model fit is generally assessed by comparing observed cross-classification frequencies to the expected frequencies predicted by the model (a likelihood ratio that is similar to a chi-square test of fit). In this case, because of the small sample size relative to the number of possible patterns, this usual test is not appropriate because the likelihood ratio does not resolve correctly when the pattern frequency matrix is sparse (Agresti & Yang, 1986). We instead used two parsimony indices, the Akaike Information Criterion and the Bayesian Information Criterion (von Davier, 1997). These indices adjust for the number of parameters in the model and the number of cases. The idea is that the model that indicates better parsimony is more likely to be correct. Both criteria indicated that the 2-latent class solution was the most parsimonious.

Table 4 suggests that the first latent class is a normally functioning class, and the second class is a low functioning one. The numbers in the Table are probabilities: for a child to be higher functioning on Perceptual Organization, there is only a 36% chance if the child is in the second latent class but a 94% chance if the child is in the first latent class. The same pattern is true for all variables. For the second latent class, a child is more likely to be low functioning on one or more of Perceptual Organization, Verbal Comprehension, Expressive Language, Receptive Language, Emotion Understanding, Theory of Mind, and

Gross Motor Coordination. On the remaining variables, a child in the second latent class is slightly more likely to be higher functioning than lower functioning, but the probability of being lower functioning remains relatively high.

4. Discussion

A DSM-IV definition of mental disorder is intended to "guide decisions regarding … the boundary between normality and pathology" (American Psychiatric Association, 2000, p. xxxi). For Kendell and Jablensky (2003), natural boundaries between valid disease entities are zones of rarity that separate them from each other and from normality. They pointed out that most research that has attempted to find natural boundaries between psychiatric disorders has failed, with "evidence suggesting that there may be no natural boundary between recognized mental disorder and normality or health" (p. 7). The major problem, however, is that there is very little research in this area, which means that few diagnostic categories have been identified as valid. Our aim was to assess whether there are zones of rarity separating two specific developmental disorders—DCD and RELD—from each other, from three other childhood disorders and from typical development. The results do not support a conclusion that either DCD or RELD is a valid disease entity based on the definitions provided by the DSM-IV. Rather, the results are consistent with the idea that these disorders are "loci in a multidimensional space in which variation in both symptoms and etiology is more or less continuous" (Kendell & Jablensky, p. 7).

The basis of this conclusion is well illustrated by Figure 1, which shows the extent of continuity and overlap between typical children and children diagnosed with a disorder, and between children with different disorders. Although the group centroid for the DCD

group is close to the vector heading towards the AD group centroid (and distinctly away from that of the RELD group), there are children from the DCD group whose discriminant function scores place them near the RELD centroid and children from the RELD group whose scores place them near the DCD centroid. It is hard to construe such children as intermediate cases between otherwise distinct disorders when their ability scores are so nearly prototypical of some other disorder.

The cluster and latent class analyses also provided strong support for our conclusion. In neither analysis were children with one disorder distinguished from children with a different disorder. Rather, a higher functioning cluster or class was distinguished from a lower functioning cluster or class, and the lower functioning cluster included varying proportions of children with a disorder (from 37% of children with ADHD to 100% of children with MR) as well as a small proportion of typical children. In these analyses, there was no discontinuity between different disorders and clear continuity between the disorders and normality.

The only group that was separated from others by zones of rarity was mental retardation. This was clear from the discriminant function analyses in which there was a clear bimodal distribution of discriminant function scores in each contrast with another disorder and with typical children. This was also clear from the fact that no children with MR were clustered with the higher functioning group in the cluster analysis. This finding is useful for at least two reasons: first, because it is a vindication of the use of discriminant function and cluster analyses to assess zones of rarity, but more importantly and as will be

discussed below, it provides a clue as to why other disorders were not distinguishable from each other.

Because we do not have autism and ADHD symptom data from typical children and children in the DCD, RELD, and MR groups, we could not assess whether AD and ADHD are separated by zones of rarity from other groups. However, we think it unlikely that such zones of rarity exist. It is clear from our results that the ability profiles of children with AD or ADHD are continuous with those of typical children and children with some other disorder. Research using the Autism Diagnostic Interview and other symptom measures indicates that the defining symptoms of AD are also continuous with other disorders, especially on the language (Bishop & Norbury, 2002; Whitehouse, Barry & Bishop, 2008) and stereotyped movement dimensions (Gal, Dyck & Passmore, 2009). As a result, the manual of the Autism Diagnostic Interview (Rutter et al., 2003, p. 35) recognizes "the overlap between autism and severe developmental disorders, particularly those involving receptive language or pragmatic skills." Similarly, research on comorbidity indicates that symptoms of ADHD are very commonly found in children diagnosed with other disorders (Frazier et al., 2001; Hofvander et al., 2009; Kaplan et al., 2001; Miniscalco et al., 2006). In the absence of evidence that AD and ADHD are discrete disease entities, an appropriate default position might be to assume that they are not mutually discriminable disorders.

If DCD and RELD are not valid disease entities, an alternative way of conceptualizing developmental disorders needs to be adopted. Contemporary definitions of these disorders are based on an analogy with neurological syndromes that emerge following the loss of brain function caused by stroke, tumour, or other localized brain impairment. In

an early account of an adolescent's inability to learn to read, Morgan (1896, p. 1378) suggested that the case "is unique ... in that it follows upon no injury or illness, but is evidently congenital, and due most probably to defective development of that region of the brain, disease of which in adults produces practically the same symptoms—that is, the left angular gyrus." The idea was—and remains—that specific local defects in brain development are responsible for specific delays in the acquisition of reading, writing, arithmetic, expressive language, receptive language, or motor skills in the same way that different local brain impairments are responsible for the various agnosias, aphasias, and apraxias. This modular and static view of developmental disorders is challenged by dynamic models in which it is the reciprocal interaction of neural systems that underpins the acquisition of cognitive abilities (van der Maas, Dolan, Grasman, Wicherts, Huizenga et al., 2006) and in which any brain defects have cascading effects on multiple neural systems (Karmiloff-Smith, 2009; Thelen & Bates, 2003). Cascading effects imply that affected children will not have specific deficits or delays; rather, a broad range of cognitive functions will be more or less affected depending on which basic processes have been affected and how they have been affected.

Because of cascade effects, dynamic models of development predict that children with different brain defects will resemble each other, and will also resemble children with Mental Retardation, differing from them only in the severity and breadth of the deficits. Rather than having deficits in specific domains that are dissociated from general cognitive processes, any impairment will inhibit the development of specialized cognitive functions and cause associations, not dissociations, across domains (Dyck et al., 2006; Karmiloff-

Smith, 2009). For this reason, children with DCD not only have motor coordination problems, they also have poor receptive language and verbal working memory, and children with RELD not only have language problems, they also have poor social cognitive abilities, verbal working memory, and response inhibition (Wisdom, Dyck, Piek, Hay, & Hallmayer, 2007). Dynamic models of development predict what is depicted in Figure 1: whatever ability or other symptom dimensions are used to characterize children with developmental disorders, cascade effects ensure that “disorders ... merge into one another with no natural boundary in between” (Kendell & Jablensky, 2003, p. 5).

Neither DCD nor RELD can be considered valid or useful diagnoses if validity requires a zone of rarity between disorders and typical development. It is not possible on the basis of current knowledge to propose an alternative set of diagnostic constructs, other than to note that because underachievement in any domain is associated with behavioral disturbance (Dyck, Hay, Anderson, Smith, Piek et al., 2004), we do need to be attentive to delays in all domains. Later, when we understand how different patterns of delay in some domains are functionally related to delays in other domains, it may become possible to define diagnostic constructs that describe distinct patterns of disorder.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial interests that could be construed as a potential conflict of interest.

Acknowledgements

This research was supported by grants from the National Health and Medical Research Council and the Research Centre for Applied Psychology, Curtin University of Technology. The funders had no role in the design or conduct of the research, the analysis of results, the preparation of the manuscript, or the decision to publish. We wish to thank the participating children and parents who made this study possible.

References

Agresti, A. & Yang, M. (1986). An empirical investigation of some effects of sparseness in contingency tables. *Computational and Statistical Data Analysis*, 5, 9-21.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: American Psychiatric Association.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition*, *21*, 37-46.

Bishop, D., & Norbury (2002). Exploring the borderlands of autistic disorder and specific language impairment: A study using standardised diagnostic instruments. *Journal of Child Psychology & Psychiatry, 43*, 917-929.

Cohen, N., Vallance, D., Barwick, M., Im, N., Menna, R., Horodezky, N., & Isaacson, L. (2000). The interface between ADHD and language impairment: An examination of language, achievement, and cognitive processing. *Journal of Child Psychology & Psychiatry*, *41*, 353-362.

Dyck, M., Farrugia, C., Shochet, I., & Holmes-Brown, M. (2004). Emotion recognition/understanding ability in hearing or vision-impaired children: Do sounds, sights, or words make the difference? *Journal of Child Psychology & Psychiatry, 45*, 789-800.

Dyck, M., Ferguson, K., & Shochet, I. (2001). Do autism spectrum disorders differ from each other and from non-spectrum disorders on emotion recognition tests? *European Child & Adolescent Psychiatry, 10*, 105-116.

Dyck, M., Hay, D., Anderson, M., Smith, L., Piek, J., & Hallmayer, J. (2004). Is the discrepancy criterion for defining developmental disorders valid? *Journal of Child Psychology & Psychiatry, 45*, 979-995.

Dyck, M., Piek, J., Hay, D., Smith, L., & Hallmayer, J. (2006). Are abilities abnormally interdependent in children with autism? *Journal of Clinical Child & Adolescent Psychology, 35*, 20-33.

Frazier, J. A., Biederman, J., Bellordre, C. A., Garfield, S. B., Geller, D. A., Coffey, B. J., & Faraone, S. V. (2001). Should the diagnosis of Attention-Deficit / Hyperactivity Disorder be consider in children with Pervasive Developmental Disorder? *Journal of Attention Disorders, 4*, 203-211.

Gal, E., Dyck, M., & Passmore, A. (2009). The relationship between stereotyped movements and self-injurious behaviour in children with developmental or sensory disabilities. *Research in Developmental Disabilities, 30*, 342-352.

Glasson, E. (2002). The Western Australian register for autism spectrum disorders. *Journal of Pediatrics & Child Health, 38*, 321.

Happe, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism & Developmental Disorders, 24*, 129-154.

Harris, P., Johnson, C., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition and Emotion, 3*, 379-400.

Henderson, S. E., & Sugden, D. A. (1992). *Movement assessment battery for children*. UK: The Psychological Corporation.

Hofvander, B., Delorme, R., Chaste, P., Nydén, A., Wentz, E., Ståhlberg, O., Herbrecht, E., Stopin, A., Anckarsäter, H., Gillberg, C., Råstam, M., & Leboyer, M. (2009). Psychiatric and psychosocial problems in adults with normal-intelligence autism spectrum disorders. *BMC Psychiatry*, *9*, 35-43.

Kadesjo, B. & Gillberg, C. (2001). The comorbidity of ADHD in the general population of Swedish school-age children. *Journal of Child Psychology & Psychiatry, 42*, 487-492.

Kaplan, B. J., Dewey, D. M., Crawford, S. G., & Wilson, B. N. (2001). The term *comorbidity* is of questionable value with reference to developmental disorders: Data and theory. *Journal of Learning Disabilities, 34*, 555-565.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology*, *45*, 56-63.

Kendell, R. E. (1989). Clinical validity. *Psychological Medicine*, *19*, 45-55.

Kendell, R. & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, *160*, 4-12.

Kendler, K. S. (1980). The nosologic validity of paranoia (simple delusional disorder): A review. *Archives of General Psychiatry*, *37*, 699-706.

Kupfer, D. J., First, M. B., & Regier, D. A. (eds.) (2002). *A research agenda for DSM-V*. Washington, DC: American Psychiatric Association.

Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview – Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659-685.

Matsumoto, D., & Ekman, P. (1995). *Japanese And Caucasian Facial Expressions Of Emotion (JACFEE) And Neutral Faces (JACNeuF).* San Francisco State University, San Francisco.

McCarron, L. (1997). *McCarron assessment of neuromuscular development: fine and gross motor abilities* (revised). Texas: McCarron-Dial Systems, Inc.

McGrath, L., Hutaff-Lee, C., Scott, A., Boada, R., Shriberg, L., & Pennington, B. (2008). Children with comorbid speech sound disorder and specific language impairment are at increased risk for Attention Deficit / Hyperactivity Disorder. *Journal of Abnormal Child Psychology*, *36*, 151-163.

Miniscalco, C., Nygren, G., Hagberg, B., Kadesjo, B., & Gillberg, C. (2006). Neuropsychiatric and neurodevelopmental outcome of children at age 6 or 7 years who had screened positive for language problems at 30 months. *Developmental Medicine & Child Neurology*, *48*, 361-366.

Morgan, W. P. (1896). A case of congenital word blindness. *The British Medical Journal*, *November 7*, 1378.

Perner, J., Frith, U., Leslie, A., & Leekam, S. (1989). Exploration of the autistic child's theory of mind: Knowledge belief and communication. *Child Development,60*, 689-700.

Perner, J. & Wimmer, H. (1985). "John thinks that Mary thinks that…" Attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437-471.

Rabbit, P. (1997). *Methodology of frontal and executive function*. East Sussex, UK: Psychology Press.

Robins, E. & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its applications to schizophrenia. *American Journal of Psychiatry, 126*, 983-987.

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire: Current Version*. Los Angeles, CA: Western Psychological Services.

Semel, E., Wiig, E., & Secord, W. (1995). *Clinical evaluation of language fundamentals (3rd ed.): Technical manual*. San Antonio, TX: The Psychological Corporation.

Shue, K., & Douglas, V. (1992). Attention deficit hyperactivity disorder and the frontal lobe syndrome. *Brain and Cognition, 20,* 104-124.

Sneath, P. H. A. (1957). Some thoughts on bacterial classification. *Journal of General Microbiology, 17*, 184-200.

Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., Clevenger, W., Wasdell, M., & McLeary, R. (2002). *Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN rating scales*. Retrieved April 30, 2004, from www.adhd.net.

Thelen, E., & Bates, E. (2003). Connectionism and dynamic systems: Are they really different? *Developmental Science, 6,* 378-391.

van der Maas, H., Dolan, C., Grasman, R., Wicherts, J., Huizenga, H., & Raijmakers, M. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842-861.

Vermunt, J. K. (1997). LEM: A general program for the analysis of categorical data. Department of Methodology and Statistics, Tilburg University. http://www.tilburguniversity. edu/nl/over-tilburg-university/schools/socialsciences/organisatie/departementen/mto/ software2.html

von Davier, M. (1997). Bootstrapping goodness-of-fit statistics for sparse categorical data: Results of a Monte Carlo study. *Methods of Psychological Research, 2*, 29-48.

Watemberg, N., Waiserberg, N., Zuk, L., & Lerman-Sagie, T. (2007). Developmental coordination disorder in children with attention deficit hyperactivity disorder and physical therapy intervention. *Developmental Medicine and Child Neurology, 49*, 920-925.

Wechsler, D. (1992). *Wechsler intelligence scale for children – Third edition: Manual* (Australian Adaptation). San Antonio, TX: The Psychological Corporation.

Whitehouse, A. J. O., Barry, J. G., & Bishop, D. V. M. (2008). Further defining the language impairment in autism: Is there a specific language impairment subtype? *Journal of Communication Disorders*, *41*, 319-336.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in children's understanding of deception. *Cognition*, *13*, 103-128.

Wisdom, S., Dyck, M., Piek, J., Hay, D., & Hallmayer, J. (2007). Can autism, language and coordination disorders be differentiated based on characteristic ability

profiles? *European Child & Adolescent Psychiatry*, *16*, 178-186.

Table 1

Classification summaries

		Predicted					
Analysis 1	Observed	TC	AD	MR	RELD	DCD	ADHD
	Typical	**432**	5	0	3	0	9
	AD	6	**20**	1	0	1	2
	MR	0	2	**21**	0	1	0
	RELD	9	1	1	**19**	0	0
	DCD	9	2	1	1	**7**	2
	ADHD	39	0	0	4	1	**9**
Analysis 2	Typical	**445**	4				
	AD	11	**19**				
Analysis 3	Typical	**448**		1			
	MR	0		**24**			
Analysis 4	Typical	**444**			5		
	RELD	11			**19**		
Analysis 5	Typical	**445**				4	

	DCD	11				**11**	
Analysis 6	Typical	**440**					9
	ADHD	44					**9**
Analysis 7	AD		**21**	1	0	2	6
	MR		2	**22**	0	0	0
	RELD		2	1	**21**	0	6
	DCD		3	1	3	**5**	10
	ADHD		0	0	4	2	**47**

Abbreviations: TC=Typical Children, AD=Autistic Disorder, MR=Mental Retardation, RELD=Receptive-Expressive Language Disorder, DCD=Developmental Coordination Disorder, ADHD =Attention Deficit Hyperactivity Disorder

Table 2

Mean standardized ability scores by cluster.

	Cluster	
	1	2
Emotion Recognition	79.07	101.28
Emotion Understanding	73.85	101.30
Theory of Mind	71.35	101.35
Response Inhibition	94.47	100.77
Working Memory Accuracy	90.61	101.35
Working Memory Speed	80.63	99.78
Perceptual Organization	72.63	101.60
Verbal Comprehension	68.73	101.49
Expressive Language	68.55	101.93
Receptive Language	66.73	101.37
Fine Motor Skills	75.59	100.48
Gross Motor Skills	69.07	99.49

Table 3

Correspondence Between Clusters and Samples

	Cluster	
Diagnosis	1	2
Typical	39	410
AD	22	8
MR	24	0
RELD	24	6
DCD	11	11
ADHD	20	33
Total	140	468

Abbreviations: TC=Typical Children, AD=Autistic Disorder, MR=Mental Retardation, RELD=Mixed Receptive Expressive Language Disorder, DCD=Developmental Coordination Disorder, ADHD =Attention Deficit Hyperactivity Disorder

Table 4

Manifest probabilities of functioning within each latent class

		Latent Class 1	Latent Class 2
Perceptual Organization	Low functioning	0.053	0.638
	Higher functioning	0.947	0.362
Verbal Comprehension	Low functioning	0.021	0.741
	Higher functioning	0.979	0.259
Expressive Language	Low functioning	0.042	0.741
	Higher functioning	0.958	0.259
Receptive Language	Low functioning	0.035	0.534
	Higher functioning	0.965	0.466
Emotion Recognition	Low functioning	0.074	0.467
	Higher functioning	0.926	0.533
Emotion Understanding	Low functioning	0.056	0.641
	Higher functioning	0.944	0.359
Theory of Mind	Low functioning	0.045	0.652
	Higher functioning	0.955	0.348
Fine Motor Coordination	Low functioning	0.045	0.488
	Higher functioning	0.955	0.512
Gross Motor Coordination	Low functioning	0.055	0.565
	Higher functioning	0.946	0.435

Working Memory Speed	Low functioning	0.071	0.434
	Higher functioning	0.929	0.566
Working Memory Accuracy	Low functioning	0.075	0.387
	Higher functioning	0.925	0.613
Response Inhibition	Low functioning	0.098	0.326
	Higher functioning	0.902	0.674

Note: Low functioning = Less than 1 standard deviation below the mean. Higher functioning refers to every other case.