6-25-2013

Detecting Subgroups in Children Diagnosed with Pervasive Developmental Disorder – Not Otherwise Specified

Laura A. Brennan
University of Connecticut - Storrs, laura.brennan@uconn.edu

Detecting Subgroups in Children Diagnosed with Pervasive Developmental Disorder – Not Otherwise Specified

Laura A. Brennan

B.A., College of the Holy Cross, 2004

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts at the

University of Connecticut

2013

APPROVAL PAGE

Master of Arts Thesis

Detecting Subgroups in Children Diagnosed with
Pervasive Developmental Disorder – Not Otherwise Specified

Presented by

Laura A. Brennan, B.A.

Major Advisor_____
Marianne L. Barton, Ph.D.

Associate Advisor_____
Chi-Ming Chen, Ph.D.

Associate Advisor_____
James Green, Ph.D.

University of Connecticut
2013

ACKNOWLEDGEMENTS

Committee

Marianne Barton, Ph.D., Chi-Ming Chen, Ph.D., James Green, Ph.D.

Early Detection Team

Deborah Fein, Ph.D., Thyde Dumont-Mathieu, MD, Sarah Hodgson, Ph.D., Jamie Kleinman, Ph.D. Katelin Carr, Lauren Herlihy, Kelley Knoch, Eva Troyb, Alex Hinnebusch,
Lauren Haisley, and Dasal Jashar

Undergraduate Research Assistants

Participating Families, Pediatricians, and Early Intervention Providers

TABLE OF CONTENTS

Abstract

Diagnosis of Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS) is assigned to children who exhibit some of the social and communicative impairments common to children with Pervasive Developmental Disorders (PDD) but fail to meet the detailed criteria of other PDDs. The lack of specific criteria for the diagnosis of PDD-NOS suggests a likely degree of heterogeneity within this population, yet there is little research exploring the similarities and differences between children with PDD-NOS. The current study utilized a hierarchical cluster analysis to detect subgroups within a sample of children with PDD-NOS that provided predictive information about diagnostic outcome at age 4. Results identified three clusters as best fitting the data. Cluster 1 demonstrated the fewest autism symptoms and highest cognitive scores of all clusters. 60% of Cluster 1 children no longer met criteria for a PDD at age 4. Cluster 2 demonstrated more social and communicative impairments and lower cognitive scores than Cluster 1, and the most repetitive behaviors of all three clusters. 89.5% of Cluster 2 children met criteria for Autistic disorder (AD) or PDD-NOS at age 4. Cluster 3 represented a small group of children difficult to diagnose at age two, as these children had the lowest cognitive scores and the most impaired social and communication skills, yet they did not demonstrate repetitive behaviors or interests. 80% of children from Cluster 3 were diagnosed with AD or PDD-NOS at age 4. These results raise questions regarding the increased importance of repetitive behaviors or interests for diagnosing ASD in the DSM-5.

Identifying Subgroups Within PDD-NOS

Pervasive Developmental Disorder (PDD) is a category of related disorders characterized by behavioral features across three domains: social reciprocity, communication, and restricted or stereotyped behaviors or interests (William Mandy, 2011; APA, 2000). The disorders within this category include Autistic Disorder, Rett's Disorder, Child Disintegrative Disorder, Asperger's Syndrome, and Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS; APA, 2000). In particular, Autistic Disorder, Asperger's Syndrome, and PDD-NOS are referred to as autism spectrum disorders (ASD; Bertrand et al., 2001). Symptoms from each of the three behavioral domains above are present within PDD in varying combinations and are sometimes described as falling along a continuum of severity, with more severe symptoms at one end of this spectrum and milder symptoms at the other (Buitelaar, Van der Gaag, Klin, & Volkmar, 1999; Walker et al., 2004).

The diagnosis of PDD-NOS was established in 1987 as a result of revisions made to the third version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III), which combined several diagnostic categories formally known as "nonautistic forms of PDD," such as "atypical autism," under one diagnostic label (Tidmarsh & Volkmar, 2003; APA, 1987). This change in diagnostic categorization expanded the range of symptoms considered to fall within the PDD spectrum, a change some identify as accounting for the dramatic increase in incidence rates of Autistic Disorder and PDD-NOS (Tidmarsh & Volkmar, 2003).

The current version of the DSM, the DSM-IV-TR, outlines specific diagnostic criteria for Autistic Disorder and Asperger's Syndrome. However, the diagnostic criteria for

PDD-NOS do not explicitly define the behaviors necessary for diagnosis. Rather, the diagnosis is assigned to children who exhibit a number of the social and communicative impairments common to children with PDD, but who fail to meet the more detailed criteria of other PDDs. A diagnosis of PDD-NOS is given when a child demonstrates a combination of symptoms, to include impairments in social interaction skills and *either* communication difficulties, *or* the presence of repetitive or stereotyped behaviors (APA, 2000).

As a result of its poorly defined criteria, PDD-NOS has been described as a potentially problematic, "catchall" diagnosis (Walker et al., 2004). The diagnosis has been criticized as constituting a "default diagnosis," lacking explicit operational definitions and having poor inter-rater reliability (Mandy, Charman, Gilmour, & Skuse, 2011; Prior et al., 1998; Walker et al., 2004). Despite these critiques, PDD-NOS remains a highly prevalent disorder; it is diagnosed at a rate 1.7 times that of Autistic Disorder (Chakrabarti & Fombonne, 2005).

The absence of more specific criteria for a diagnosis of PDD-NOS suggests a likely degree of heterogeneity within this population. There is, however, little research attempting to further "specify" the PDD-NOS diagnosis (Buitelaar et al., 1999). Instead, research has primarily described PDD-NOS in relation to other ASDs in order to examine whether each disorder presents unique and varied profiles or whether each diagnosis varies only by their position along a spectrum of symptom severity (Buitelaar et al., 1999; Paul et al., 2004; Walker et al., 2004).

Subgroups in the Literature

Attempting to validate subgroups under the umbrella of PDD was thought to be critical to articulating the etiology and trajectory of these disorders, as well as to developing effective treatment plans for children with PDDs (Stevens et al., 2000; Roux, Garreau, Barthelemy, & Hameury, 1994). Clarifying the characteristics of children within each subgroup under the PDD umbrella allows for a more comprehensive understanding of the distinct profiles and needs of children in each diagnostic category.

Research comparing PDD-NOS to other PDDs is extensive and reveals differing perspectives. Several studies propose that the PDD diagnoses represent a spectrum of symptom severity (Buitelaar et al., 1999; Fein et al., 1999; Prior et al., 1998; Stevens et al., 2000). This conceptualization of PDD argues that each disorder varies only by the severity of a child's autism related symptoms. Thus, under this interpretation, PDD-NOS does not differ qualitatively from other PDDs. A second perspective, however, suggests that the PDD-NOS profile varies distinctly from other PDDs, indicating that PDD-NOS may not fit neatly along the proposed continuum of symptom severity (Paul et al., 2004; Walker et al., 2004). One study articulated that children with PDD-NOS often demonstrate stronger cognitive and adaptive functioning than children with Autistic Disorder, have histories of language delays uncommon in Asperger's Syndrome, and exhibit repetitive and stereotyped behaviors less frequently than either children with Autistic Disorder or Asperger's (Walker et al., 2004).

In contrast to the number of studies examining the boundaries between PDD disorders as a whole, only two studies have looked within a sample of children diagnosed with PDD-NOS in order to detect subgroups and further define the characteristics of these

children. Darlene Walker and colleagues (2004) conducted a qualitative assessment of a small sample of children diagnosed with PDD-NOS (M=86.3 months, SD=38 months) and identified three groups that emerged from their data. The first group (n=11) demonstrated few repetitive and stereotyped behaviors, were described as cognitively "high functioning," and had a "transient or persistent language delay." The second group (n=5) exhibited numerous repetitive and stereotyped behaviors, yet had "good" current language skills. The authors hypothesized that this group might have met criteria for Asperger's Disorder, except for a mild language delay earlier in development. Finally, the third group (n=5) was characterized as being potentially "too young or too delayed" to effectively assess for repetitive or stereotyped behaviors. The authors also posited that these children might have presented with a late age of onset for Autistic Disorder (Walker et al., 2004).

A study conducted by William Mandy and colleagues (2011) looked at a sample of children diagnosed with PDD-NOS and grouped them according to DSM-IV-TR symptomotology. Their results indicated that 97% of children with PDD-NOS in their sample presented with a combination of social interaction and communication impairments, while only 3% presented with the combination of social interaction deficits and repetitive or stereotyped behaviors (Mandy et al., 2011). These findings suggest that a majority of children with PDD-NOS present with difficulties in communication skills in addition to deficits in reciprocal social interaction, but very few of these children will demonstrate repetitive and stereotyped behaviors. Both of these studies suggest that repetitive or stereotyped behaviors may appear later in development or might not be

consistently observed. This finding has important implications for the identification of ASDs in young children with less severe autism symptomotology.

Outcomes of Children with PDD-NOS

Several studies have shown that children with PDD-NOS are more likely than children with other PDD diagnoses to achieve "optimal outcomes" as they grew older. An "optimal outcome" refers to when children who were diagnosed with an ASD at an early age no longer demonstrate the symptoms required to receive an ASD diagnosis when revaluated later in development. A study by Berry and colleagues (2009) found that 17.1% of their sample of 35 children diagnosed with PDD-NOS at approximately 2 years of age no longer met criteria for an ASD by the time they were 4-years-old. This rate of achieving "optimal outcomes" was much greater than the rate of 6.8% (n=68) in children diagnosed with Autistic Disorder (Berry, 2009; Helt et al., 2008; Lord et al., 2006; Sutera et al., 2007). This same study looked at diagnostic outcomes for children diagnosed with PDD-NOS at age 2 when revaluated at age 4 and found several factors to be predictive of "optimal outcomes" at age 4 (Berry, 2009). These included better motor abilities early in development as reported by the parents, low symptom severity at initial diagnosis, presence of few repetitive behaviors, higher adaptive skills as measured by parent-report, and higher expressive language abilities on a developmental assessment measure (Berry, 2009). These findings suggest that there may be patterns of characteristics within PDD-NOS that might provide information about potential future outcomes.

Specific Aims

Identifying subgroups of children within the PDD-NOS population may enhance our ability to identify, understand, and provide services for these children. The literature examining subgroups within PDD more broadly suggests that defining these subgroups can be an important step in defining more explicitly the patterns of characteristics presented by each group of children (Stevens et al., 2000; Roux et al., 1994). The more refined our understanding of children with PDD-NOS becomes, the more precise our judgments can be in determining appropriate diagnoses, fine-tuning future research questions, and in developing and delivering the treatments best suited to the particular needs of children with PDD-NOS.

The current study sought to examine the characteristics of a sample of children diagnosed with PDD NOS at approximately 2 years of age. The study's specific aim was to identify more homogeneous and clinically meaningful subgroups within a sample of children diagnosed with PDD-NOS in the hope that those subgroups would have predictive validity for future diagnosis. This aim was addressed through: (a) utilization of a hierarchical cluster analysis to detect clusters in the current sample, (b) description of the characteristics within the subgroups detected by the cluster analysis, (c) determination of the predictive validity of subgroups by demonstrating differential outcomes based on the diagnosis received when the children were reevaluated at age 4, and (d) external validation of the subgroups using variables not included in the cluster analysis.

The following hypotheses were made concerning the outcomes of this analysis. First, given the common suggestion that PDD-NOS is a "catchall" diagnosis, we predicted that the characteristics of the subgroups detected by the cluster analyses would follow a

varied profile, meaning that the children in each subgroup would present with a profile that differed across multiple domains, rather than along a spectrum of symptom severity. Second, emphasis was placed on the predictive value of the clusters as a result of findings indicating that a higher percentage of children diagnosed with PDD-NOS at the age of 2 went on to attain an optimal outcome by the age of 4 when compared to children diagnosed with other ASDs. Patterns of behaviors in children with PDD-NOS at age 2 might provide important information about their potential developmental course. It was therefore hypothesized subgroup membership would be related to diagnostic outcome at age 4.

Methods

Participants

Participants were selected from a larger sample of children taking part in an ongoing study examining the effectiveness of a screening questionnaire designed to detect ASD symptoms in young children. These screening measures included the Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001) and a more recent, amended version, the M-CHAT-Revised (M-CHAT-R). Participants were enrolled in the study after receiving the screener either through a child's early intervention services, during pediatric well-child visits at 18 or 24 months of age, or by a caregiver's self-referral. A more detailed explanation of the Early Detection Study procedures can be found below.

Within the larger sample of children included in the Early Detection Study, 123 were diagnosed with PDD-NOS between the ages of 18-34 months. This subset of children

was considered for inclusion in the current study. Of the 123 children with PDD-NOS, data from 20 children were excluded from the analyses due to missing data. One child was determined to be an outlier and excluded due to the fact that he was the only child who received the Module 2 version of the Autism Diagnostic Observation Schedule (ADOS).

Participants for the current study therefore included 102 children diagnosed with PDD-NOS. The sample was 76% male (n = 78) and 24% female (n = 24). The mean age was 25.5 months of age (SD = 4.39). The majority of children were identified by their caregivers as White (n = 80, 78%), with fewer children identified as Hispanic/Latino (n=7, 7%), Black or African American (n = 5, 5%), Asian or Pacific Islander (n = 5, 5%), Biracial (n = 2, 2%), and "other" (n = 1, 1%). Data on race and ethnicity was not available for two children (2%).

Of the 102 children diagnosed with PDD-NOS, 71 (70%) received a re-evaluation between the ages of 48-64 months of age, as part of the Early Detection Study protocol. Thirteen (18%) of these 71 children were excluded from analyses due to missing data. As a result, 58 of the 71 children were included in our second series of analyses aimed at determining the predictive value of the clusters produced by the cluster analysis through looking at diagnostic outcome at age 4. This group was primarily male (n = 44, 76%), with 24% being female (n = 14). The mean age for this group was 51.1 (SD = 6.98) months of age. These children were mostly identified as White (n = 48, 83%), with 7% of the children being identified as Hispanic/Latino (n = 4), 5% as Black or African American (n = 3) and 3% as Asian or Pacific Islander (n = 2). Race/ethnicity data was not available for one child (2%).

Due to the exploratory nature of the hierarchical cluster analysis, the representative nature of the sample is critical to the generalizability of the findings (Hair & Black, 2004). The current sample is considered to be a close approximation to the current census data on the racial/ethnic breakdown and variation in socioeconomic status in the state of Connecticut and the United States, with over-sampling in low SES populations to increase participation in the study (United States Census Bureau, 2012). The gender ratio of 3.25:1 in children with ASD at age 2 and 3.14:1 at age 4 in the current sample (see Appendix A, Table A1) were slightly lower than the currently estimated gender prevalence ratio of 4.67:1 in children with ASD put forth by the Centers for Disease Control (CDC; Investigators, 2012)).

Procedures

Caregivers completed the M-CHAT or M-CHAT-R at their child's pediatrician's office, early intervention site, or home. The completed screener was then sent to the University of Connecticut Early Detection laboratory for scoring. If the child failed the M-CHAT or M-CHAT-R, caregivers were called to confirm items missed. Children who continued to fail the screener after the follow-up phone interview were invited for a developmental and diagnostic evaluation at the University of Connecticut free of charge. Transportation was provided to families unable to travel to the evaluation. Evaluations were conducted by a graduate student in the UConn Clinical Psychology Ph.D. program and by a licensed clinical psychologist. The families received the assessment results at the time of their appointment and were sent via post a comprehensive report summarizing testing results, along with recommendations, six to eight weeks following the evaluation.

After two years, the study invited participants who received an evaluation at approximately 2 years of age to return to UConn for a re-evaluation in order to assess the stability of the diagnosis indicated by their first evaluation.

During testing, the child and his or her caregiver(s) received a number of measures designed to assess the child's cognitive, language, and adaptive skill levels, as well as several ASD-specific measures, in order to gain a broad understanding of the child's development and to determine whether a diagnosis was appropriate. The parent-report measures for the current study included the Autism Diagnostic Interview, Revised (ADI-R) and the Vineland Adaptive Behavioral Scales. The children received the Mullen Scales of Early Learning (Mullen) and the Autism Diagnostic Observation Schedule (ADOS). The clinician completed the Childhood Autism Rating Scale (CARS), a measure of ASD symptom severity, using information gained from the caregiver interview, as well as their direct observations of the child.

The diagnosis of an ASD was assigned based upon the clinical judgment of experienced psychologists, using scores from the ADOS, ADI-R, CARS, and developmental and adaptive behavior measures and according to DSM-IV criteria for an ASD or PDD-NOS diagnosis. Assigning an ASD diagnosis on the basis of experienced clinical judgment is considered best practice and has been show to have high inter-rater reliability (Klin, Lang, Cicchetti, & Volkmar, 2000).

Measures

The current study analyzes data obtained from the measures described below. These measures have been used extensively in clinical practice and research in order to detect

and diagnose ASD in young children, and are considered to have strong psychometric properties (Kleinman et al., 2007; Lord et al., 2000; Mullen, 1994; Sparrow, Cicchetti, & Balla, 2005).

Modified Checklist for Autism in Toddlers (M-CHAT). The M-CHAT is the central measure in the Early Detection Study, as it serves as the study's sole enrollment criterion. The M-CHAT is a 23-item parent-report measure with 23 yes/no questions designed to detect ASD symptoms in young children (Robins, Fein, Barton & Green, 2001). This screening measure was adapted from the Checklist for Autism in Toddlers (CHAT; (Baron-Cohen, Allen, & Gillberg, 1992), in order to tailor the questionnaire to be appropriate for a parent-report format (Kleinman et al., 2007). Children who miss three or more of the 23 items on the M-CHAT are classified as having "failed" the screener and receive a scripted follow-up interview over the phone. If a child continues to fail the M-CHAT after the phone interview, the child is invited to receive a free developmental and diagnostic evaluation. Internal consistency was found to be sufficient for the complete screener and for six critical items (Cronbach's a values = .85 and .84, respectively) in a recent replication study of the M-CHAT (Kleinman et al., 2007).

Autism Diagnostic Observation Schedule (ADOS). The ADOS is a semi-structured, play-based interview that has been standardized for the purpose of diagnosing individuals with ASD (Lord et al., 2000). The ADOS assesses individual performance within four domains: Communication, Reciprocal Social Interaction, Play, and Repetitive Behaviors. The algorithm for scoring the ADOS follows this domain structure and cut-off scores for an ASD diagnosis have been established in the Communication and Reciprocal Social Interaction domains (Lord et al., 2000). Interrater reliability (mean

weighted kappas, $M\kappa_W$) was high for both Modules 1 and 2 ($M\kappa_W = .78$ and $M\kappa_W = .70$, respectively; Lord et al., 2000). Using the ADOS-Generic version algorithm, inter-rater agreement in assigning ASD vs. non-spectrum diagnoses was found to be 100% for Modules 1 and 3, 91% for Module 2, and 90% for Module 4 (Lord et al., 2000). Test-retest reliability indicated excellent stability for the Communication and Social domains, and good stability for the Stereotyped Interests and Restricted Interest domain (Lord et al., 2000).

Autism Diagnostic Interview (ADI). Both the original version of the ADI and a modified version, the ADI-Revised (ADI-R), were used to aid in the diagnosis of ASD in the current sample. The ADI and ADI-R are semi-structured interviews for parents of children with ASD that assesses autism symptomotology based on ICD-10 and DSM-IV (Lord, Rutter, & Le Couteur, 1994). Both measures are for use with children who have a mental age over 2 and have sound psychometric properties (Lord et al., 1994). Interrater reliability for the ADI and ADI-R communication and social domains was high (κ_W ranging from .64-.97 and κ_W ranging from .62-.89, respectively; Le Couteur et al., 1989; Lord et al., 1994), as were the interrater reliability results for the restricted and repetitive behaviors and interests of both versions ($M\kappa_W = .70$ and κ_W ranging from .55 to .87, respectively; Le Couteur et al., 1989; Lord et al., 1994).

Mullen Scales of Early Learning (Mullen). The Mullen Scales of Early Learning is a cognitive assessment standardized for use with children from birth to 68 months, which consists of five subdomains: Gross Motor, Visual Perception, and Fine Motor, as well as Receptive and Expressive Language (Mullen, 1994). Each subdomain score is assigned a t-score, as well as age equivalents and percentile rank for ease of interpretation. Internal

consistency for the measure is reported as being very satisfactory (.75 to .83) and the test re-test reliability remained high for both younger and older children (.84 and .76 respectively; Mullen, 1994). Children in the current sample completed this measure at both time points.

Vineland Adaptive Behavioral Scales – Interview Edition. The Vineland Adaptive Behavioral Scales (Vineland) is a parent-report measure designed to assess adaptive skills of children across four domains: Communication, Daily Living Skills, Socialization, and Motor Skills (Sparrow et al., 2005). Chronbach's alphas were computed for the domain scores and for the Adaptive Behavior Composite (ABC) score based on the internal-consistency reliabilities of the subdomains. All were found to be above .80 for the age ranges included in the Early Detection sample (Sparrow et al., 2005). Interclass correlations (ICC) indicate high test-retest reliability for each subdomain (ICC = .85 and higher) and high inter-interviewer reliability for the ABC score (ICC = .87) and Domain scores (ICC = .75). The use of an adaptive skill assessment when assigning a diagnosis of an ASD is recommended, as it can allow for better classification diagnostically and for more detailed treatment planning (Perry, Flanagan, Dunn Geier, & Freeman, 2009). The current study obtained scores on this measure for children at both time points.

Data Analytic Plan

Data analyses for the current study occurred in two phases, both of which utilized exploratory hierarchical cluster analyses to detect potential subgroups within a sample of 102 children diagnosed with PDD-NOS at age 2. Hierarchical methods are ideal for samples of this size in order to keep calculations feasible (Aldenderfer & Blashfield,

1984; Hair & Black, 2000). This procedure uses stepwise clustering methods to combine observations into subgroups using, in this case, agglomerative methods to assign observations to clusters (Hair & Black, 2000). Agglomerative methods place each observation into individual clusters initially and, through a stepwise process, merge the most similar clusters together to create a new cluster; this process continues until all clusters form a single group (Hair & Black, 2000).

While there are several agglomerative methods for creating these clusters, the current study utilized Ward's method (Ward, 1963) to detect clusters within the sample. Ward's method is a minimum variance procedure used for hierarchical cluster analyses that has been found to be preferable to other methods, such as the single-link method. This procedure joins two clusters based on their similarity to one another in order to decrease the variance within clusters. Similarity between two clusters is calculated by adding the sum of squares between the two clusters and dividing them by the sum of squares summed between all variables (Ward, 1963; Hair & Black, 2000). This method demonstrates a strong sensitivity to outliers and a tendency to suggest clusters that are similar in size (Milligan, 1980; Hair & Black, 2000). Each variable included in the analyses was plotted by observation to determine whether potential outliers existed. After examination of these results, one participant was excluded due to advanced language abilities, which required administration of Module 2 of the ADOS. As stated earlier in this section, because this measure differed significantly from Module 1 of the ADOS, this child was determined to be an outlier and excluded.

Unlike other statistical procedures, cluster analyses are often considered exploratory because they do not meet standard assumptions of normality and there are multiple

methods for detecting clusters (Milligan & Hirtle, 2003). One consequence of this exploratory nature is that no clear parameters exist for determining the 'correct' number of clusters. Instead, selection of a cluster solution that best fits one's data is typically based on a combination of empirical judgments and practical or theoretical considerations (Hair & Black, 2000). The current study employed these dual criteria when evaluating the best fitting cluster structure for the data. First, the number of clusters was chosen based upon groupings depicted in the dendrogram--the tree diagram produced by the cluster analysis--and by examination of the scree plot produced by the hierarchical cluster analysis (Hair & Black, 2000). A dendrogram, or tree diagram, depicts the results of the hierarchical cluster analysis graphically by placing each observation in an individual cluster on the vertical axis and illustrating on the horizontal axis the agglomerative process of placing observations in a cluster and subsequently combining clusters (Hair & Black, 2000). The scree plot accompanying the dendrogram illustrates the joining of clusters, with each point on the scree line representing clusters combining and the spaces between each point representing the distances between the clusters at each step in the clustering process. When the distance between two points creates a sudden change in the direction of the scree plot (i.e., from a sharp downward slope to a more level slope), this is considered to be a natural cutting point for establishing the best fitting number of clusters (Dougherty, 2013).

Decisions about cluster numbers were also dependent upon theoretical and practical considerations, as is suggested by the literature (Hair & Black, 2000). Given that current research suggests the PDD-NOS population is heterogeneous and ill-defined, and that the intent of this study was to better understand the characteristics of the disorder in order to

guide diagnostic and treatment development, it was important to restrict the number of clusters to ensure the subgroups identified would be clinically relevant and applicable. It was decided that more than four clusters in a sample of 102 children would likely yield clusters with few observations in each and could potentially be less representative of the population as a whole. In a more practical sense, having more than four clusters was thought to be potentially cumbersome to those who might attempt to identify a child as a member of a particular cluster. Thus, the researcher did not consider cluster structures with more than four clusters.

The predictive value of the clusters produced by the hierarchical cluster analysis was also a key factor in determining the optimal cluster procedures. In order to establish 'outcome,' diagnostic data from 58 children included in the initial analyses, who received a re-evaluation at age 4, were examined. These participants were grouped according to whether they received either a diagnosis of Autistic Disorder or no longer met criteria for and an ASD or Developmental Delay (DD) diagnosis at age 4. Children who no longer met criteria for an ASD were considered by the experimenters to have achieved 'better' outcomes and children who received a diagnosis of Autistic Disorder were considered to have had 'poorer' outcomes. The examiners then calculated the specificity and sensitivity, as well as the positive and negative predictive values, of the clusters to determine whether cluster assignment at age 2 provided information about having 'better' or 'poorer' outcomes at age 4.

The cluster analyses were run using the software program, JMP® Version 9 (SAS Institute, Cary, NC).

Phase I. In the initial phase of the study, a hierarchical cluster analysis was conducted using Ward's method to detect clusters in the current sample. Variables included each individual item from the ADOS, Module 1, (29 total items, see Table A2) as well as each subdomain score from the Mullen Scales of Early Learning. Because the scales differed across assessment tools, scores from each measure included in the hierarchical cluster analysis were standardized to allow for comparison between measures (Hair & Black, 2000).

A three-cluster structure best fit the data (see Appendix B, Fig. B1 for dendrogram and scree plot). However, this cluster structure was found to have insufficient predictive performance due to poor specificity (0.68) and negative predictive value (0.42). Therefore, the examiners reevaluated the variables included in the analyses, as it became evident from these results that the quality of the variables included was more important for predicting outcome than quantity. Evidence supporting this conclusion can be found in the literature surrounding cluster analyses. Researchers suggest that selection of variables for cluster analyses must have theoretical and practical foundations (Hair & Black, 2000). More importantly, the literature suggests that only variables that describe the observations to be clustered and that directly pertain to the particular aims of the analyses should be included. Including variables that are irrelevant to either of these premises can mask the underlying cluster structure that exists, making identification of these clusters extremely difficult (Milligan & Hirtle, 2003; Hair & Black, 2000).

Phase II. In light of these findings, a second hierarchical cluster analysis was conducted, again using Ward's method, in which only variables that provided predictive information about participants at age 4 were included. The level of predictive

performance demonstrated by each variable was determined by plotting each item used in the original analyses against the reevaluation diagnosis of 'ASD' or 'No ASD/No DD' at age 4 (see Fig. B2 for an example of a plot used to determine predictive value). The examiners then selected the item from each subdomain from the ADOS Module 1 and the subdomain score from the Mullen Scales of Early Learning that best differentiated between groups at age 2 based on this diagnostic classification at age 4. The Visual Reception subdomain score from the Mullen and four items from the ADOS were selected, to include Item A6: Use of Other's Body to Communicate, B7: Requesting, C1: Functional Play with Objects, and D4: Unusually Repetitive Interests or Stereotyped Behaviors. This provided a list of variables collected during their first evaluation that offered the most information about the future diagnostic outcome for participants at age 4.

Evaluation of the clusters. Typical statistical procedures, such as using an analysis of variance (ANOVA) to demonstrate that significant differences exist between clusters by using the variables included in the cluster analysis, are not valid means of evaluating the cluster structures. Instead, what is referred to as 'external' validation procedures are suggested (Milligan & Hirtle, 2003; Hair & Black, 2000): External validity can be established by conducting ANOVAs that utilize variables not included in the hierarchical cluster analysis. In the current study, a one-way multivariate analysis of variance (MANOVA) was used to establish the external validity of the selected clusters by comparing the groups' standardized scores on each item of the Childhood Autism Rating Scale (CARS), an autism symptom severity measure, and subdomain scores from the Vineland Adaptive Behavior Scales. Due to missing CARS and Vineland data, three

persons were excluded from the external validity analyses; a total of 99 participants diagnosed with PDD-NOS at age 2 were included. Follow-up ANOVAs and appropriate post hoc tests were performed (i.e., if equal variances were not assumed, posthoc Games-Howell test was used; otherwise, post hoc Fisher's least significant difference [LSD] test was run). The alpha value was set at 0.05 for all statistical tests.

Results

Results from the Phase II hierarchical cluster analysis indicated that a three cluster structure best fit the data (see Fig. B3 for the dendrogram and scree plot produced by the analysis and corresponding cluster labels, Table A3 for demographic information on each cluster). Cluster 1 (n=26) scored the highest of the three clusters on the Visual Reception subdomain (M = 38.65, SD = 11.5) from the Mullen Scales of Early Learning. This cluster also exhibited the least social and communicative impairments and the fewest repetitive behaviors, as indicated by their scores on the item from each subdomain of the ADOS used in the analysis. Cluster 2 (n = 68) demonstrated lower scores than Cluster 1 on the Visual Reception subdomain (M = 30.95, SD =9), and presented with more social impairments and communication difficulties than Cluster 1 on the included ADOS items. Cluster 2 also had the most repetitive and stereotyped behaviors and interests of all three clusters. Profiles within the third cluster (Cluster 3, n = 8) were consistently varied and remained the most difficult to characterize. Cluster 3 received the lowest scores of all three clusters on the Mullen Visual Reception subdomain (M = 27.25, SD = 10.14) and remained the most impaired in areas of communication and social interaction on the

ADOS. Surprisingly, despite exhibiting greater impairment within these domains, children in Cluster 3 demonstrated fewer repetitive behaviors than those in Cluster 2.

Mullen Scales of Early Learning

The clusters demonstrated a consistent pattern across all subdomains of the Mullen (see Table A4; Figs. B4 and B5). In addition to the Visual Reception subdomain, which was the only score from this measure used in Phase II of the cluster analysis, Cluster 1 continued to have the highest scores of each cluster in the remaining subdomains (Fine Motor and Receptive and Expressive Language; M = 35.23, SD = 12.7, M = 31.85, SD = 10.45, and M = 31.62, SD = 8.26, respectively). Cluster 2's scores remained between Clusters 1 and 3 for each of these clusters, though its scores were more similar to Cluster 1 on the Fine Motor subdomain (M = 32.46, SD = 9.31) and closer to the lower scores found in Cluster 3 for the Receptive and Expressive Language subdomains (M = 22.72, SD = 6.38, M = 28.03, SD = 8.18, respectively), suggesting a more significant impairment in communication abilities in this cluster when compared to Cluster 1. Cluster 3 continued to receive the lowest scores across all remaining Mullen subdomains (M = 27.13, SD = 10.27, M = 20.75, SD = 2.12, M = 24.38, SD=4.96), which indicated the greatest cognitive impairment of all three clusters.

Autism Diagnostic Observation Schedule (ADOS)

ADOS A1, use of other's body to communicate. For the ADOS item examining a child's use of another person's body to communicate, results (see Fig. B6) indicated that Cluster 1 was the least likely to demonstrate this behavior, with 81% of the children in this cluster receiving a score of zero for "no use of another's body to communicate."

57% of children in Cluster 2 received scores indicating the children used another person's hand to lead them to or reach for an item they desired to a mild or moderate degree, as indicated by a score of one or two. Most striking about the current results for this ADOS item, was that 100% of Cluster 3 received a score of three, which is indicative of "little or no spontaneous communication," suggesting severe communicative impairments.

ADOS B7, requesting. Results for the ADOS B7 item assessing a child's ability to use joint attention, which consists of the child pointing to an object with his or her index finger, looking at the object, and then looking at the person to ensure they understand the communicative intent of the gesture, show that 65% of children in Cluster 1 were able to successfully request items using joint attention (see Fig. B7). The remaining 35% of the children in Cluster 1 received a mild score of 1, meaning these children used pointing to draw another's attention to an object, but their use of coordinated eye contact was not yet fluent enough for a score of zero. 84% of children in Cluster 2 demonstrated mild to moderate impairments in their ability to use joint attention to request objects and 63% of children in Cluster 3 demonstrated impairments in this skill area.

ADOS C1, functional play with objects. Cluster results for this item (see Fig. B8) investigating a child's ability to play appropriately and independently with a variety of toys indicated that the majority (92%) of children in Cluster 1 demonstrated unimpaired play skills, with 8% of the children in this cluster demonstrating mild impairment, as indicated by a score of one. Ninety-six percent of children in Cluster 2 demonstrated mild to moderate deficits this area, with only 4% of the children in this cluster being found to have no impairment in their play abilities. 88% of children in Cluster 3 received

a scores of one, two, or three, indicating mild to moderate impairment in this play domain.

ADOS D4, unusually repetitive interests or stereotyped behaviors. 85% of children in Cluster 1 did not demonstrate any repetitive or stereotyped behaviors during the administration of the ADOS (see Fig. B9). In contrast, 53% of the children in Cluster 2 received a score indicating that these behaviors were present to either a mild or moderate degree. Children in cluster 3 were more similar to children in Cluster 1 than on any previous item included in the cluster analysis, as 63% of children in Cluster 3 did not demonstrate any repetitive or stereotyped behaviors during the ADOS. When these behaviors were present in a child from Cluster 3, they received a milder score of one (37%).

External Validity

In order to explore whether the cluster structure (three clusters) detected by the hierarchical cluster analysis remained consistent when compared using variables not included in the original cluster analyses, a one-way MANOVA was conducted on all 15 CARS items, the CARS Total Score, and four Vineland subdomain scores. A trend toward differences was found among the three clusters on the dependent measures, Wilks's $\Lambda = .59$, $F(38,156) = 1.25$, $p = .17$, $\eta_p^2 = .23$. One-way ANOVAs on all dependent variables were conducted as follow-up tests to the MANOVA, and post hoc tests of the significant ANOVAs were further performed. ANOVAs and post hoc analyses revealed significant differences between clusters for seven items and for the

Total Score on the Childhood Autism Rating Scale (CARS). Clusters 1 and 2 differed significantly on the following items: I. Relating to People, II. Imitation, V. Object Use, VIII. Listening Response, XI. Verbal Communication, XII. Nonverbal Communication, and XV. General Impressions, as well as CARS total score. For each item, Cluster 1's CARS scores indicated the least severe autism symptom presentation, as they were significantly lower than Cluster 2's scores, which were suggestive of the most severe autism symptomotology. Cluster 3's scores, though not significantly different from either those of Clusters 1 or 2, fell consistently between Clusters 1 and 2's scores (see Table A5 for a summary of the external validity results).

No significant difference was found between clusters on the Vineland Adaptive Behavior Scales.

Diagnosis at Age 4 By Cluster

As would be expected, given that the variables included for the cluster analysis were selected on the basis of their predictive value, the clusters detected by the Phase II hierarchical cluster analysis were found to provide important information about outcome for children who received a reevaluation at age four. Fifteen of the children in Cluster 1 received a reevaluation (see Table A6 and Fig. B10). This cluster contained the greatest number of children who went on to no longer meet criteria for an ASD (n = 9, 60%), with the other six children (40%) remaining stable in their PDD-NOS diagnosis. In Cluster 2, 38 children were reevaluated at age 4. The majority of these children continued to meet criteria for PDD-NOS (n = 15, 39%) or went on to meet criteria for a diagnosis of

Autistic Disorder at age four (n = 19, 50%). Only 11% of the children (n = 4) in Cluster 2 did not demonstrate ASD symptoms at age 4. The majority of the five children in Cluster 3 who received a reevaluation went on to meet criteria for Autistic Disorder at age 4 (n = 3, 60%). One child (20%) continued to meet criteria for PDD-NOS and one child (20%) no longer met criteria for an ASD at age 4. Again, as expected given the variables used, the sensitivity and specificity values calculated for this three cluster structure was high. In Table A7, sensitivity for outcome at age 4 was 100%, while specificity was 83%. The positive and negative predictive values, as well as accuracy, were similarly high (95%, 100%, and 96%, respectively), indicating that our clusters demonstrated a highly accurate ability to predict age 4 diagnosis using scores at age 2.

Discussion

The current study used hierarchical clustering procedures to detect subgroups within a sample of children diagnosed with PDD-NOS in an attempt to clarify the characteristics of a diagnosis that has been portrayed as 'problematic' in the literature. The three clusters produced by these analyses are described in detail in the subsequent paragraphs.

Spectrum of Symptom Severity

In part, the results of the current study support the perspective that PDD-NOS is characterized by a spectrum of symptom severity (See Table A8). Clusters 1 and 2 appear to differ along a continuum. Cluster 1 represents the higher end of the spectrum, as this cluster consisted of children who received the highest scores on each subdomain of the Mullen Scales of Early Learning, demonstrated the least impairment on social and

communication skills, and exhibited the fewest repetitive behaviors and interests, as measured by the ADOS. These children also had the lowest total scores on the measure of autism symptom severity. Unsurprisingly, a majority of the children in Cluster 1 no longer met criteria for an ASD or remained stable in their PDD-NOS diagnosis when reevaluated at age four. Cluster 2 represents the lower end of the symptom severity spectrum, with these children receiving lower scores on the Mullen and demonstrating more impairment in social and communication skill areas than children in Cluster 1. Children in Cluster 2 also engaged in more restricted and repetitive behaviors or interests than children in either Clusters 1 or 3. Consistent with this profile, children in Cluster 2 either continued to meet diagnostic criteria for PDD-NOS or received a diagnosis of Autistic disorder when reevaluated at age four.

Evidence in the literature supports the current findings that link specific skill profiles in children to future outcome. A 2007 study suggested that children with higher cognitive scores and fewer early social interaction impairments demonstrate a greater ability to develop skills, such as receptive and expressive language, as well as play skills, over time (Ben-Itzchak & Zachor, 2007). Further, this study found that both cognitive levels and social-reciprocity skills were significantly correlated with outcome later in development (Ben-Itzchak & Zachor, 2007). Studies have also indicated that motor skills, symptom severity at age two, number of repetitive behaviors present, adaptive functioning, and expressive language skills are characteristic of children who no longer meet criteria for an ASD when reevaluated at age four (Berry, 2009).

In the current study, the children in Cluster 1 confirm earlier findings that children with PDD-NOS who receive higher scores on nonverbal problem solving measures,

demonstrate fewer social interaction impairments, and present with fewer repetitive behaviors and less severe autism symptomotology may be more likely to have 'better' outcomes later in development (i.e. no longer meet criteria for an ASD). Children in Cluster 2 demonstrated that lower cognitive scores combined with greater social impairment, more frequent repetitive behaviors, and more severe autism symptoms predict the retention of PDD-NOS diagnoses or the development of a more severe diagnosis of Autistic disorder by age four.

Finally, Clusters 1 and 2 may indicate that the PDD-NOS population is less heterogeneous than has been previously described in the literature. The implication of these findings may be that, regardless of the lack of explicitly defined criteria, there seems to be a somewhat consistent pattern of symptoms in a proportion of children diagnosed with PDD-NOS at age two, and that this pattern varies largely in terms of severity.

Varied Profile

The current findings also provide support, however, for the perspective that PDD-NOS represents a more varied profile and does not fit neatly within the spectrum of symptom severity. Cluster 3 demonstrates this varied profile (see Table A9). These children exhibit the most severe cognitive, social, and communicative impairments when compared to Clusters 1 and 2, yet they demonstrate far fewer repetitive and stereotyped behaviors than children in Cluster 2. These findings were not expected given that these children demonstrated deficits in the social and communication items from the ADOS. Also surprising was the fact that the children in Cluster 3 received lower scores than

children in Cluster 2 on the CARS, indicating milder autism symptom severity presentation at age two, despite more marked impairments in the cognitive, interpersonal, and communicative domains. Again, despite these more mild autism severity scores, it was found that a majority of the children in Cluster 3 who received a reevaluation at age four went on to develop Autistic disorder, suggesting that these children are more likely to have 'poorer' outcomes later in development.

Although Cluster 3 presents a varied profile in comparison to Clusters 1 and 2, the characteristics of the children found in Cluster 3 followed a consistent pattern. For example, the finding for children in Cluster 3 on Item A1 (Use of another's body to communicate), which indicate that these children made little or no spontaneous attempts to communicate, mirrors results on the receptive and expressive language subdomains of the Mullen, which showed that children in Cluster 3 had the lowest scores of all three clusters in their ability to understand language or to use language for communication with others.

Cluster 3 also demonstrated the greatest impairment in functional play skills, when compared to Clusters 1 and 2. Play skills have been found in the literature to be highly correlated with language, cognitive, and social development in young children (Bateson, 1955; Piaget, 1962; Vygotsky, 1978; Bates, 1979; Rapin, 1996). Toy play in particular is thought to be related to development of joint attention skills (Toth, Munson, N Meltzoff, & Dawson, 2006). In both high functioning and low functioning children with autism, the frequency with which they engage in toy play and the developmental level of this play have been found to be significantly lower than their non-autistic peers (Rapin, 1996). More important, evidence suggests that toy play in preschool aged children

diagnosed with autism has been found to be predictive of communication development over the next several years of development (Toth et al., 2006). Given the findings on the correlation between play skills and other important developmental areas, Cluster 3's profile of low cognitive scores and severe social and communicative impairments may lend further support to the interrelatedness of these developmental domains.

Cluster 3 represents a group of children with PDD-NOS who are potentially difficult to characterize and diagnose accurately at age two. As noted previously, prior research indicated that motor skills, severity scores, number of repetitive behaviors, and play skills, among others, are variables found to be predictive of developmental outcomes in children with ASD (Sutera et al., 2007; Berry, 2009). Lower functioning children with autism have been found in the literature to be more easily diagnosed at age two, especially when the children present with higher nonverbal than verbal scores (Rapin, 1996). Children in Cluster 3 exhibited this pattern of higher nonverbal problem solving scores than receptive and expressive language scores, and exhibited severe social and communication deficits. However, these children did not present with the repetitive behaviors required for an Autistic Disorder diagnosis, and their scores on the CARS were also less severe compared to children in Cluster 2. Despite their milder autism symptom presentation at age two, a greater proportion of children in Cluster 3 went on to have 'poorer' outcomes and met criteria for Autistic Disorder at age four. This finding has significant implications for the early identification of children with autism spectrum disorders.

Perhaps children in Cluster 3 presented with a late-onset form of Autistic disorder. One previous study indicated that by age three or four, children with late-onset autism do

not differ significantly in any diagnostic domain from children with early onset autism (Werner, Dawson, Munson, & Osterling, 2005). Though it is not possible in the present study to compare children with Cluster 3 profiles with other children with an Autism diagnosis at age four, it would be important to compare these profiles in order to determine whether children diagnosed with PDD NOS and the Cluster 3 profile appear to have a late onset form of the disorder.

It is also possible that children in Cluster 3 were not yet demonstrating the repetitive behaviors required for a diagnosis of Autistic Disorder. Research in repetitive behaviors has shown the number of repetitive behaviors exhibited by children at age four is often higher than was present in those children at age two (Moore & Goodson, 2003; Cox, Klein, Charman, Baird, Baron□Cohen, Swettenham, Drew, & Wheelwright, 1999; Stone, Lee, Ashford, Brissie, Hepburn, Coonrod, & Weiss, 1999). At age two ASD specific impairments in social and communication skills may be apparent on the ADOS and CARS, but symptoms in the restricted and repetitive behaviors and interests domain may not have developed yet.

External Validity

A MANOVA was utilized to determine whether differences between the clusters existed on variables not included in the original cluster analyses. Results indicated Clusters 1 and 2 differed significantly on seven items from the CARS, as well as the total score from this measure. Cluster 3 scores were not found to be significantly different from either cluster and remained consistently between Cluster 1 and 2 on each item and on the total score. The total scores for each cluster were above what has been found to be

the most accurate cut off score for PDD-NOS on the CARS, which is 25.5 for two-year olds (Chlebowski, Green, Barton, & Fein, 2010).

DSM-5

The currently proposed criteria for ASD in the Diagnostic and Statistical Manual, 5th edition (DSM-5) includes several significant changes to the existing criteria found in the DSM-IV-TR. First, the DSM-5 collapses Autistic Disorder, Asperger's Syndrome, and PDD-NOS diagnoses into one, Autism Spectrum Disorder diagnosis (Frazier, Youngstrom, Speer, Embacher, Law, Constantino, Findling, Hardan, & Eng, 2012). Second, the DSM-5 symptom domains have been reduced to two symptom clusters (A. Social Communication & B. Restricted, Repetitive Behaviors; RRB), rather than three (Social Interaction, Communication, and Restricted, Repetitive, and Stereotyped Behaviors; (Mandy, Charman, & Skuse, 2012). In order to receive a diagnosis of ASD, a child must demonstrate symptoms from both symptom domains A (Social Communication) and B (Restricted, Repetitive Behaviors). In order to meet criteria for the symptoms described in Criteria A, a child must have demonstrated one symptom in all three of the symptom subdomains, which includes A1 (Social-Emotional Reciprocity), A2 (Nonverbal Communication), and A3 (Relationships). In order to meet criteria for the symptom cluster defined by Criteria B, a child must have demonstrated one symptom from at least two symptom subdomains. These subdomains included B1 (Stereotyped or Repetitive Speech, Motor Mannerisms, or Use of Objects), B2 (Excessive Adherence to Routines or Ritualized Speech), B3 (Restricted, Fixated Interests), and B4 (Hyper-or Hypo-reactivity to Sensory Input or Unusual Sensory Interests).

Two of the Clusters detected in the current study, Clusters 1 and 3, did not demonstrate consistent repetitive and stereotyped behaviors at age two. These findings are consistent with other studies investigating the characteristics of PDD-NOS samples. Walker et al.'s (2004) study found that 50% of their sample demonstrated only mild or transient repetitive behaviors or interests, while Mandy et al. (2011) found that 97% of their sample did not demonstrate these behaviors. This data may suggest that our current model for understanding PDD in young children is inaccurate. Children who do not present with repetitive behaviors at age 2 may not truly have a PDD. If PDD is defined as a more severe disorder, it can be expected that positive results in children diagnosed with PDD will decrease, even when these children receive quality intervention. However, if we define PDD in more broad terms, as has occurred since the revisions to the DSM in 1987 and 1994 when PDD-NOS Asperger's Disorder were included as diagnoses, the number of children diagnosed with PDD will likely increase, but it can also be expected that children diagnosed with PDD will demonstrate more positive outcomes (APA, 1987; APA, 1994; Tidmarsh & Volkmar, 2003).

Without the repetitive and stereotyped behaviors or interests, it is unclear whether young children would meet criteria for an ASD diagnosis under the currently proposed DSM-5 criteria. It is therefore imperative that future research continue to understand the trajectories of children who do not present with consistent repetitive and stereotyped behaviors at two years of age in order to ensure that children are not prevented from accessing the autism specific early intervention services needed. Limiting access to such services when autism symptoms are present to either a mild or moderate degree, as found

in Clusters 1 and 3, would likely have significant impacts on children's outcome at age four, though further research will be necessary to support this claim.

Limitations

Several limitations should be considered when interpreting the results of the current study. First, of the 102 children who received a diagnosis of PDD-NOS at age two, only 58 were reevaluated at the approximate age of four, as a result of caregivers being unable to contact, having relocated, or refusing the evaluation. Therefore, the results relating to outcome were based on a subset of children included in the original cluster analyses. However, the percentages of children from each cluster that received a reevaluation were roughly equal, indicating that there was almost equal access to data on outcome for each cluster (See Table 3).

The number of children in Cluster 3 was extremely small, thus limiting our ability to draw generalizable conclusions from this data. However, it should be noted that during the Phase II cluster analyses, when the number of clusters was expanded to include four total clusters or contracted to include only two clusters, Cluster 3 remained a distinct group while the configurations of Clusters 1 and 2 changed. It was therefore determined that Cluster 3 represented a discrete cluster with characteristics that differed notably from the other possible clusters.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Beverly Hills; London: Sage.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders: DSM-III*. American Psychiatric Publishing, Inc.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV*. American Psychiatric Publishing, Inc.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. American Psychiatric Publishing, Inc.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry*, *161*(6), 839–843. doi:10.1192/bjp.161.6.839

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.

Bateson, G. (1955). A theory of play and fantasy. *Psychiatric research reports*, *2*(39), 39-51.

Berry, L. N. (2009). *Early treatments associated with optimal outcome in children with autism spectrum disorders.* (Unpublished doctoral dissertation) University of Connecticut, Storrs, Connecticut.

Ben-Itzchak, E., & Zachor, D. A. (2007). The effects of intellectual functioning and autism severity on outcome of early behavioral intervention for children with autism. *Research in Developmental Disabilities*, *28*(3), 287–303. doi:10.1016/j.ridd.2006.03.002

Bertrand, J., Mars, A., Boyle, C., Bove, F., Yeargin-Allsopp, M., & Decoufle, P. (2001). Prevalence of Autism in a United States Population: The Brick Township, New Jersey, Investigation. *PEDIATRICS*, *108*(5), 1155–1161. doi:10.1542/peds.108.5.1155

Buitelaar, J. K., Van der Gaag, R., Klin, A., & Volkmar, F. (1999). Exploring the Boundaries of PervasiveDevelopmental Disorder Not Otherwise Specified: Analyses of Data from the DSM-IV Autistic Disorder Field Trial. *Journal of Autism and Developmental Disorders*, *29*(1), 33–43.

Chakrabarti, S., & Fombonne, E. (2005). Pervasive Developmental Disorders in Preschool Children: Confirmation of High Prevalence. *The American Journal of Psychiatry,* 162(6), 1133–1141.

Chlebowski, C., Green, J. A., Barton, M. L., & Fein, D. (2010). Using the Childhood Autism Rating Scale to Diagnose Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *40*(7), 787–799. doi:10.1007/s10803-009-0926-x

Cox, A., Klein, K., Charman, T., Baird, G., Baron□Cohen, S., Swettenham, J., Drew, A., & Wheelwright, S. (1999). Autism spectrum disorders at 20 and 42 months of age: Stability of clinical and ADI□R diagnosis. *Journal of Child Psychology and Psychiatry*, *40*(5), 719-732.

Dougherty, G. (2013). (Agglomerative) Hierarchical Clustering *Pattern Recognition and Classification* (pp. 150-154): Springer New York.

Fein, D., Stevens, M., Dunn, M., Waterhouse, L., Allen, D., Rapin, I., & Feinstein, C. (1999). Subtypes of Pervasive Developmental Disorder: Clinical Characteristics. *Child Neuropsychology*, *5*(1), 1–24.

Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., Findling, R.L., Hardan, A.Y., & Eng, C. (2012). Validation of Proposed DSM-5 Criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(1), 28-40.

Helt, M., Kelley, E., Kinsbourne, M., Pandey, J., Boorstein, H., Herbert, M., & Fein, D. (2008). Can Children with Autism Recover? If So, How *Neuropsychology Review*, *18*(4), 339–366. doi:10.1007/s11065-008-9075-9

Investigators, A. A. D. D. M. N. S. Y. 2. P. (2012). *Prevalence of Autism Spectrum Disorders - Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008* (Vol. 61, pp. 1–24). Surveillance Summaries.

Kleinman, J. M., Robins, D. L., Ventola, P. E., Pandey, J., Boorstein, H. C., Esser, E. L., Wilson, L. B., et al. (2007). The Modified Checklist for Autism in Toddlers: A Follow-up Study Investigating the Early Detection of Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *38*(5), 827–839. doi:10.1007/s10803-007-0450-9

Klin, A., Lang, J., Cicchetti, D. V., & Volkmar, F. R. (2000). Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV autism field trial. *Journal of Autism and Developmental Disorders*, *30*(2).

Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holdgrafer, M., & McLennan, J. (1989). Autism Diagnostic Interview: A Standardized Investigator-

Based Instrument. *Journal of Autism and Developmental Disorders*, *19*(3), 363–387.

Lecavalier, L., Gadow, K. D., DeVincent, C. J., Houts, C., & Edwards, M. C. (2009). Deconstructing the PDD clinical phenotype: internal validity of the DSM-IV. *Journal of Child Psychology and Psychiatry*, *50*(10), 1246–1254. doi:10.1111/j.1469-7610.2009.02104.x

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism From 2 to 9 Years of Age. *Arch Gen Psychiatry*, *63*, 694–701.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., et al. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3).

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*(5), 659–685.

Mandy, W. P. L., Charman, T., & Skuse, D. H. (2012). Testing the Construct Validity of Proposed Criteria for DSM-5 Autism Spectrum Disorder. *JAAC*, *51*(1), 41–50. Elsevier Inc. doi:10.1016/j.jaac.2011.10.013

Mandy, W., Charman, T., Gilmour, J., & Skuse, D. (2011). Toward specifying pervasive developmental disorder-not otherwise specified. *Autism Research*, *4*(2), 121–131. doi:10.1002/aur.178

Milligan, G. W. (1980). An Examination of the Effect of Six Types of Error Perturbation On Fifteen Clustering Algorithms, *Psychometrika*, *45*(3), 325–342.

Moore, V., & Goodson, S. (2003). How well does early diagnosis of autism stand the test of time? Follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. *Autism*, *7*(1), 47-63.

Mullen, E. M. (1994). *The Mullen Scales of Early Development*. Circle Pines, MN: American Guidance Services.

Paul, R., Miles, S., Cicchetti, D., Sparrow, S., Klin, A., Volkmar, F., Coflin, M., et al. (2004). Adaptive Behavior in Autism and Pervasive Developmental Disorder-Not Otherwise Specified: Microanalysis of Scores on the Vineland Adaptive Behavior Scales. *Journal of Autism and Developmental Disorders*, *34*(2), 223–228.

Perry, A., Flanagan, H. E., Dunn Geier, J., & Freeman, N. L. (2009). Brief Report: The Vineland Adaptive Behavior Scales in Young Children with Autism Spectrum

Disorders at Different Cognitive Levels. *Journal of Autism and Developmental Disorders*, *39*(7), 1066–1078. doi:10.1007/s10803-009-0704-9

Piaget, J. (1962). *Play, dreams and imitation* (Vol. 24). New York: Norton.

Prior, M., Eisenmajer, R., Leekam, S., Wing, L., Gould, J., Ong, B., & Dowe, D. (1998). Are There Subgroups within the Autistic Spectrum? A Cluster Analysis of a Group of Children with Autistic Spectrum Disorder. *Journal of Child Psychology and Psychiatry*, *39*(6), 893–902.

Rapin, I. (Ed.). (1996). *Preschool children with inadequate communication* (Vol. 139). Mac Keith Press.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*(2), 131–143.

Stevens, M. C., Fein, D. A., Dunn, M., Allen, D., Waterhouse, L. H., Feinstein, C., & Rapin, I. (2000). Subgroups of Children With Autism by Cluster Analysis: A Longitudinal Examination. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(3), 346–352.

Stone, W. L., Lee, E. B., Ashford, L., Brissie, J., Hepburn, S. L., Coonrod, E. E., & Weiss, B. H. (1999). Can autism be diagnosed accurately in children under 3 years? *Journal of Child Psychology and Psychiatry*, *40*(2), 219-226.

Sutera, S., Pandey, J., Esser, E. L., Rosenthal, M. A., Wilson, L. B., Barton, M., Green, J., et al. (2007). Predictors of Optimal Outcome in Toddlers Diagnosed with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders.*, *37*(1), 98–107. doi:10.1007/s10803-006-0340-6

Sparrow, S. S., & Cicchetti, D. V., & Balla, D.A., (2005). The Vineland Adaptive Behavior Scales-II, Second Edition. Circle Pines, MN: American Guidance Service, Inc. 109-117.

Tidmarsh, L., & Volkmar, F. R. (2003). In Review. *Canadian Journal of Psychiatry*, *48*, 517-525.

Toth, K., Munson, J., N Meltzoff, A., & Dawson, G. (2006). Early Predictors of Communication Development in Young Children with Autism Spectrum Disorder: Joint Attention, Imitation, and Toy Play. *Journal of Autism and Developmental Disorders*, *36*(8), 993–1005. doi:10.1007/s10803-006-0137-7

United States Census Bureau. (2012). U.S. Census Bureau: State and County QuickFacts, Connecticut. Retrieved from http://quickfacts.census.gov/qfd/states/09000.html

Vygotsky L. L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard university press.

Walker, D. R., Thompson, A., Zwaigenbaum, L., Goldberg, J., Bryson, S. E., Mahoney, W. J., Strawbridge, C. P., et al. (2004). Specifying PDD-NOS: A Comparison of PDD-NOS, Asperger Syndrome, and Autism. *Journal of The American Academy of Child and Adolescent Psychiatry*, *43*(2), 172–180.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Werner, E., Dawson, G., Munson, J., & Osterling, J. (2005). Variation in Early Developmental Course in Autism and its Relation with Behavioral Outcome at 3–4 Years of Age. *Journal of Autism and Developmental Disorders.*, *35*(3), 337–350. doi:10.1007/s10803-005-3301-6

Appendix A
Tables

Table A1
Sample Demographics

Sample	N	Mean Age in Months (SD)	Gender (Ratio)	Race/Ethnicity
Age 2	102	25.5 (4.39)	Males = 78 Females = 24 (3.25:1)	White (n=80) Hispanic/Latino (n=7) Asian or Pacific Islander (n=5) Black or African American (n=5) Biracial (n=2) Other (n=1) Missing (n=2)
Age 4	58	51.1 (6.98)	Males = 44 Females = 14 (3.14:1)	White (n=48) Hispanic/Latino (n=4) Black or African American (n=3) Asian or Pacific Islander (n=2) Missing (n=1)

Table A2
Autism Diagnostic Observation Scale (ADOS) Items by Subdomain

Item #	A Language & Communication	B Reciprocal Social Interaction	C Play	D Stereotyped Behaviors & Restricted Interests.	E Other Abnormal Behaviors
1	Overall Level of Non-Echoed Language.	Unusual Eye Contact.	Functional Play With Objects.*	Unusual Sensory Interest in Play Material/Person.	Overactivity.
2	Frequency of Vocalizations Directed Toward Others.	Responsive Social Smile.	Imagination/ Creativity.	Hand and Finger and Other Complex Mannerisms.	Tantrums, Aggression, Negative or Disruptive Behavior.
3	Intonation of Vocalizations or Verbalizations.	Facial Expressions Directed to Others.		Self-Injurious Behavior.	Anxiety.
4	Immediate Echolalia.	Integration of Gaze & Other Behaviors During Social Overtures.		Unusually Repetitive Interests or Stereotyped Behaviors.*	
5	Stereotyped/ Idiosyncratic Use of Words or Phrases.	Shared Enjoyment in Interaction.			
6	Use of Other's Body to Communicate.*	Response to Name.			
7	Pointing.	Requesting.*			
8	Gestures.	Giving.			
9		Showing.			
10		Spontaneous Initiation of Joint Attention.			
11		Response to Joint Attention.			
12		Quality of Social Overtures.			

** Indicates item included in Phase II of cluster analyses*

Table A3
Demographic Information of Clusters

Demographic Information	Cluster 1	Cluster 2	Cluster 3
n	26	68	8
Mean Age in Months (SD)	25.6 (4.71)	26.2 (4.44)	23.71 (3.31)
Gender (Ratio)	Males = 17 Females = 9 (1.89:1)	Males = 55 Females = 13 (4.23:1)	Males = 6 1 Females = 2 (3:1)
Race/Ethnicity	White (n=20) Hispanic/Latino (n=2) Asian or Pacific Islander (n=0) Black or African American (n=1) Biracial (n=1) Other (n=0) Missing (n=2)	White (n=54) Hispanic/Latino (n=4) Asian or Pacific Islander (n=5) Black or African American (n=4) Biracial (n=0) Other (n=1) Missing (n=0)	White (n=6) Hispanic/Latino (n=1) Asian or Pacific Islander (n=0) Black or African American (n=0) Biracial (n=1) Other (n=0) Missing (n=0)
Received Reevaluation at Age 4	15 (58%)	38 (55%)	5 (63%)

Table A4
Average Mullen Scales of Early Learning (Mullen) T Scores by Cluster

	Mean Mullen Subdomain Scores (SD)			
Cluster	Visual Processing*	Fine Motor	Receptive Language	Expressive Language
Cluster 1 (n=26)	38.65 (11.5)	35.23 (12.7)	31.85 (10.45)	31.62 (8.26)
Cluster 2 (n=68)	30.95 (9)	21.083 (3.53)	22.67 (6.59)	29.65 (8.38)
Cluster 3 (n=8)	27.25 (10.14)	27.13 (10.27)	20.75 (2.12)	24.38 (4.96)

** Indicates item included in Phase II of cluster analyses*

Table A5
External Validity Results: Post Hoc Analyses,
Significant Childhood Autism Rating Scale (CARS) Items

CARS Item	Cluster 1 Mean (n=25)	Cluster 2 Mean (n=66)	Cluster 3 Mean (n=8)	Follow-up ANOVA and p value	Post hoc test and p value
I. Relating to People	1.940[2]	2.333[1]	2.188	.015	LSD, .004
II. Imitation	1.740[2]	2.220[1]	2.125	.007	LSD, .002
V. Object Use	1.580[2]	1.924[1]	1.625	.005	LSD, .002
VIII. Listening Response	1.760[2]	2.114[1]	2.000	.032	LSD, .009
XI. Verbal Communication	2.360[2]	2.674[1]	2.313	.024	LSD, .017
XII. Nonverbal Communication	1.980[2]	2.326[1]	2.188	.012	Games-Howell, .009
XV. General Impressions	1.820[2]	2.136[1]	1.875	.011	LSD, .005
CARS Total Score	26.420[2]	29.455[1]	27.438	.002	LSD, .001

Note: Superscripts indicate significant differences between clusters.

Table A6
Diagnoses at Age Four by Cluster

	Diagnoses				
Cluster	Autistic Disorder (AD)	PDD-NOS	No Longer Meets Criteria for ASD	Developmental Delay (DD)	Developmental Language Delay (DLD)
Cluster 1 (n=15)	0	6 (40%)	5 (33.3%)	3 (20%)	1 (6.7%)
Cluster 2 (n=38)	19 (50%)	15 (39.5%)	1 (2.6%)	2 (5.3%)	1 (2.6%)
Cluster 3 (n=5)	3 (60%)	1 (20%)	0	1 (20%)	0

	ASD vs. No ASD Diagnosis	
Cluster	ASD	No ASD
Cluster 1 (n=15)	6 (40%)	9 (60%)
Cluster 2 (n=38)	34 (89.5%)	4 (10.5%)
Cluster 3 (n=5)	4 (80%)	1 (20%)

Table A7
Phase II Cluster Analysis Results: Sensitivity/Specificity Estimates for Diagnosis at Age Four

Cluster Analysis (Test)	Diagnosis at Re-Evaluation (Gold Standard)		
	Autistic Disorder	No ASD/No DD	Total
Autistic Disorder (1)	21	1	22
No ASD/No DD (2)	0	5	5
Total	21	6	27

Sensitivity	1
Specificity	0.8333
Positive Predictive Value	0.9545
Negative Predictive Value	1
Accuracy	0.963

Table A8
Spectrum of Symptom Severity: Clusters One and Two

Dimension	Cluster 1 (n=26)	Cluster 2 (n=68)
Cognitive Functioning	Highest	Lower
Social & Communication Deficits	Lowest	Higher
Repetitive Behaviors	Low	Highest
CARS Total Score	Lowest	Highest
Age 4 Diagnosis	Most Likely No-ASD/No-DD	PDD-NOS or AD

Table A9
Varied Profile: Cluster Three

autism	Cluster 3 (n=8)
Cognitive Functioning	Lowest
Social & Communication Deficits	Highest
Repetitive Behaviors	Low
CARS Total Score	Mild
Age 4 Diagnosis	Most Likely AD

Appendix B
Figures

Figure B1
Methods: Phase I Cluster Analysis Dendrogram and Scree Plot

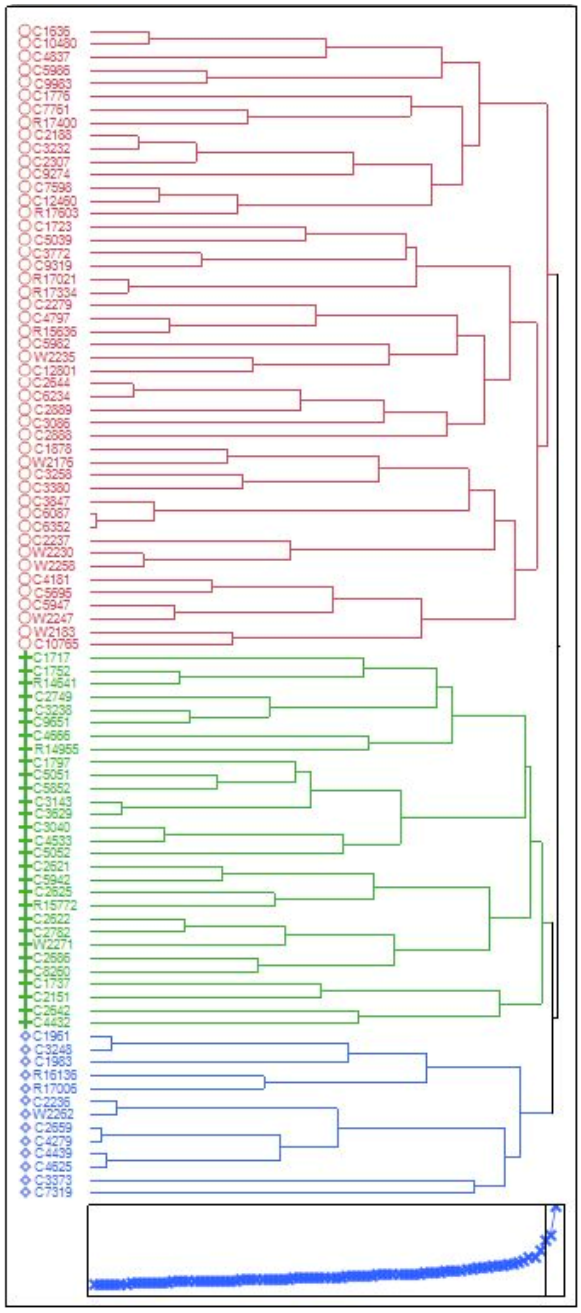


Figure B2
Autism Diagnostic Observation Schedule (ADOS) Scores in Subdomain A, Language and Communication, by Diagnosis at Age Four

Example of a plot used to determine the item from each subdomain of the ADOS that best differentiated between groups who were reevaluated at age four. For subdomain A, item A1 best differentiated between participants who went on to have 'worse' outcomes, demonstrated by their receiving a diagnosis of Autistic Disorder, and those who went on to have 'better' outcomes, as demonstrated by their no longer meeting criteria for an Autism Spectrum Disorder (ASD).

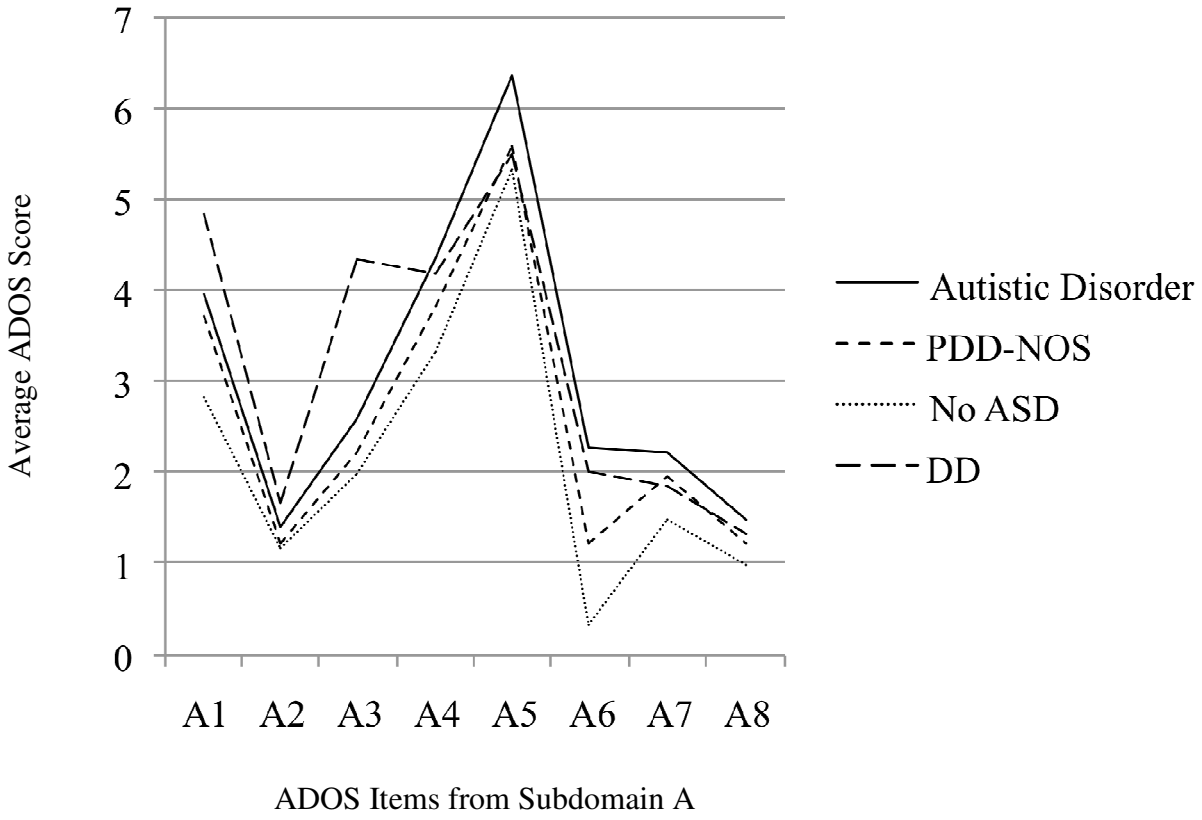


Figure B3
Results: Phase II Cluster Analysis Dendrogram and Scree Plot

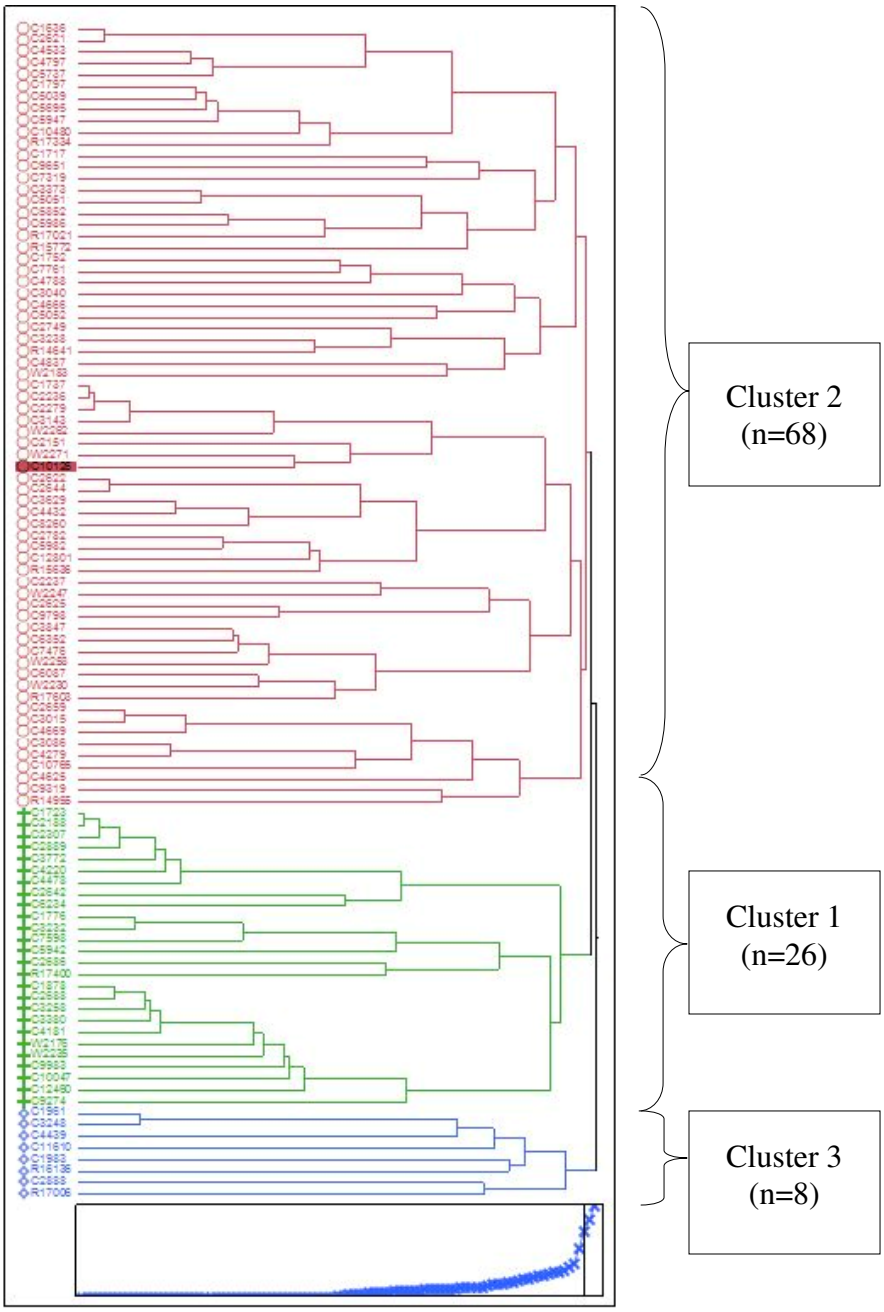
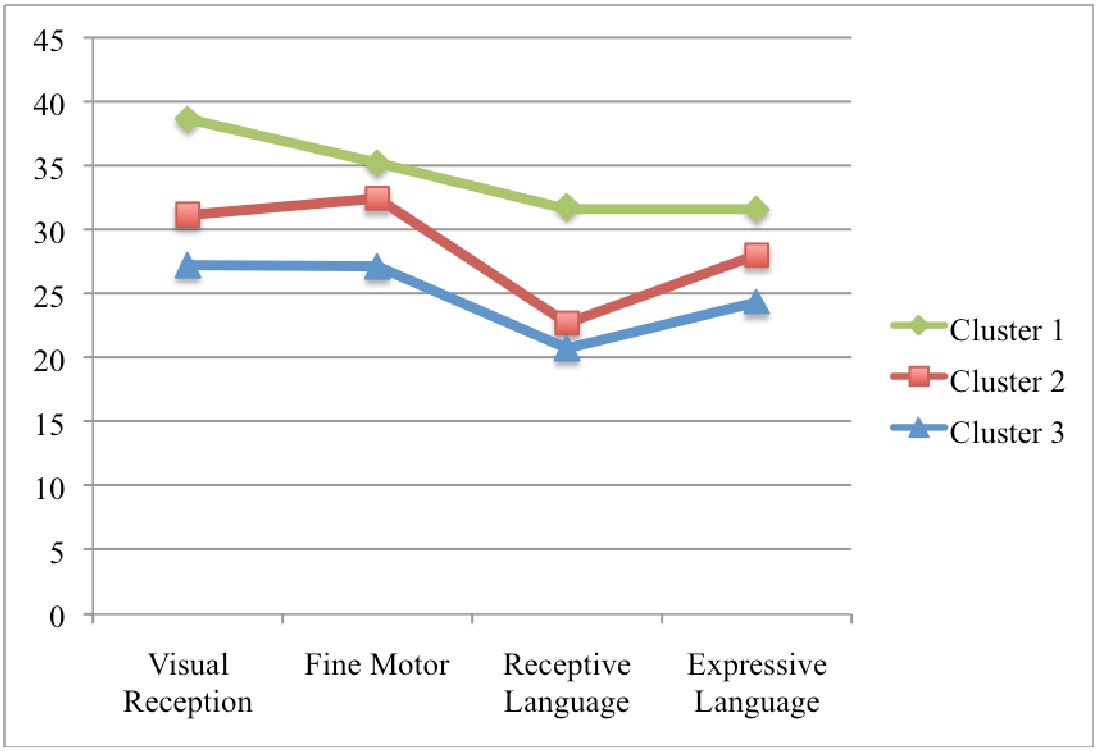


Figure B4
Average Scores on Mullen Scales of Early Learning by Cluster



** Line graph for visual purposes only*

Figure B5
Average Scores on Mullen Scales of Early Learning by Cluster

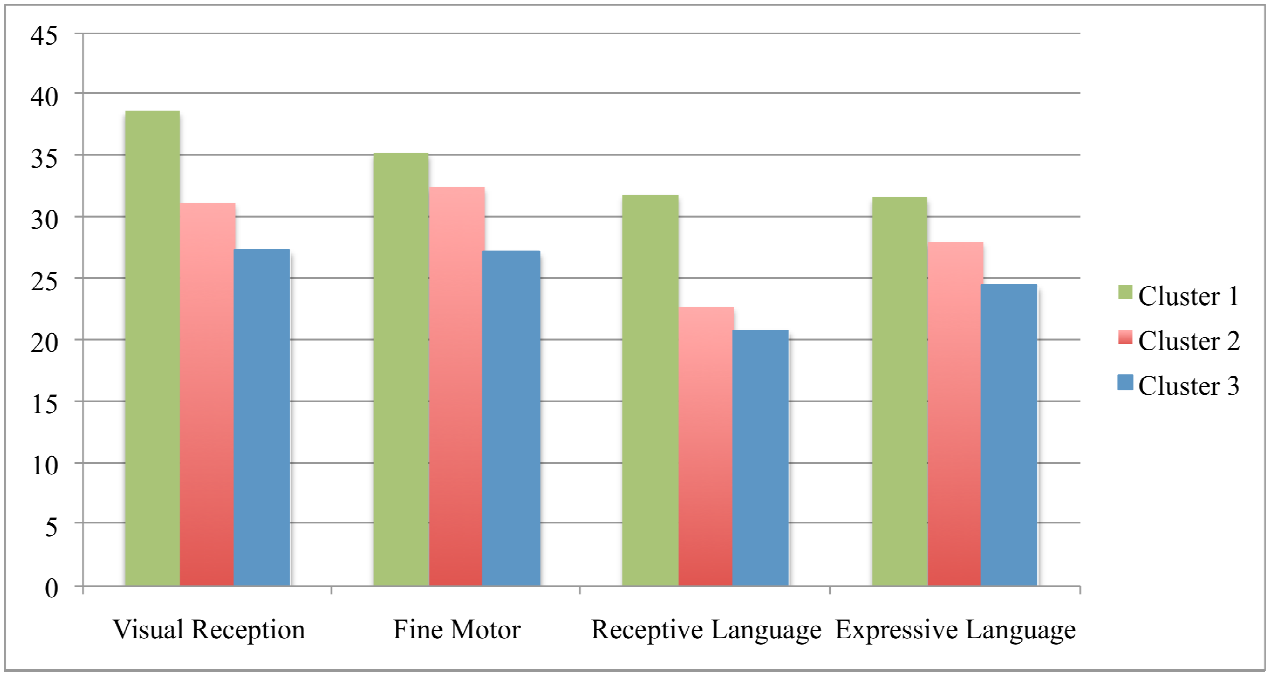
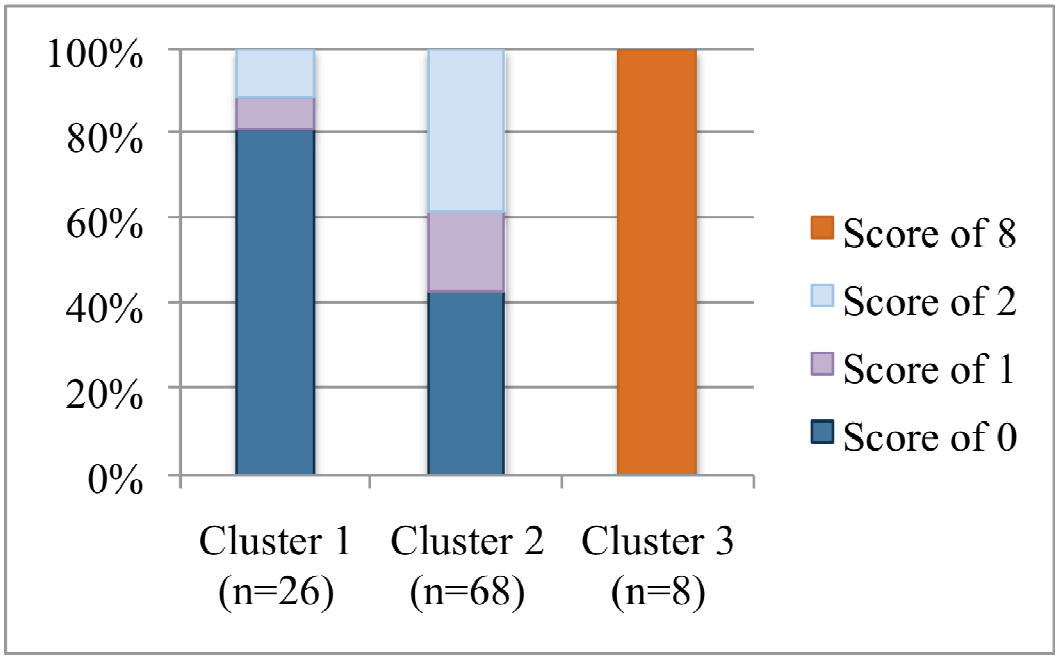
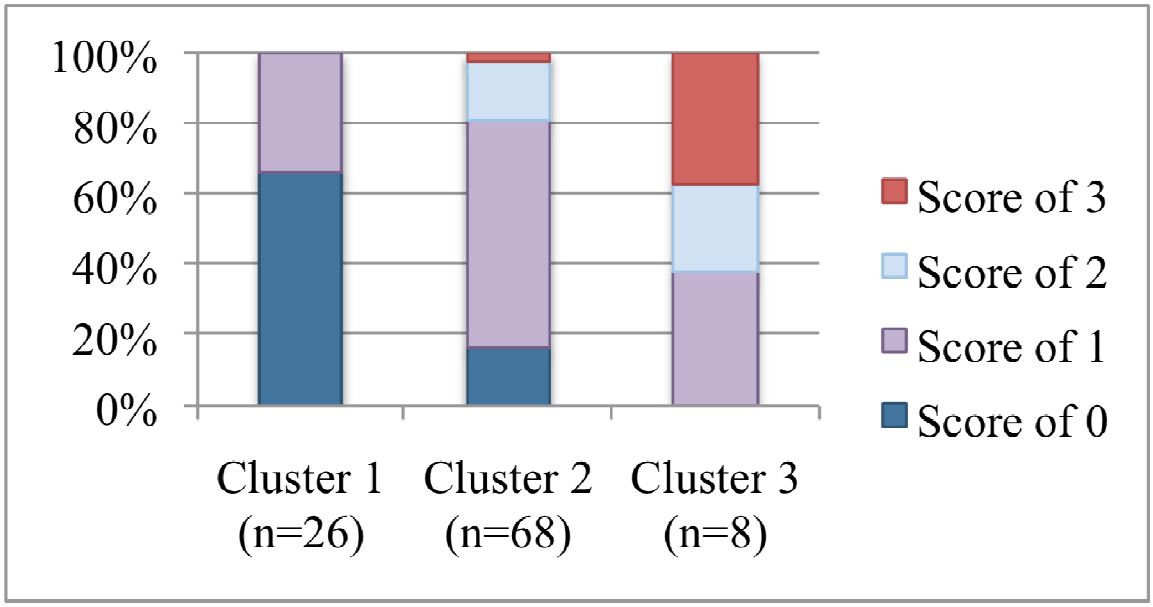


Figure B6
Autism Diagnostic Observation Scale (ADOS) Item A6 Scores by Cluster



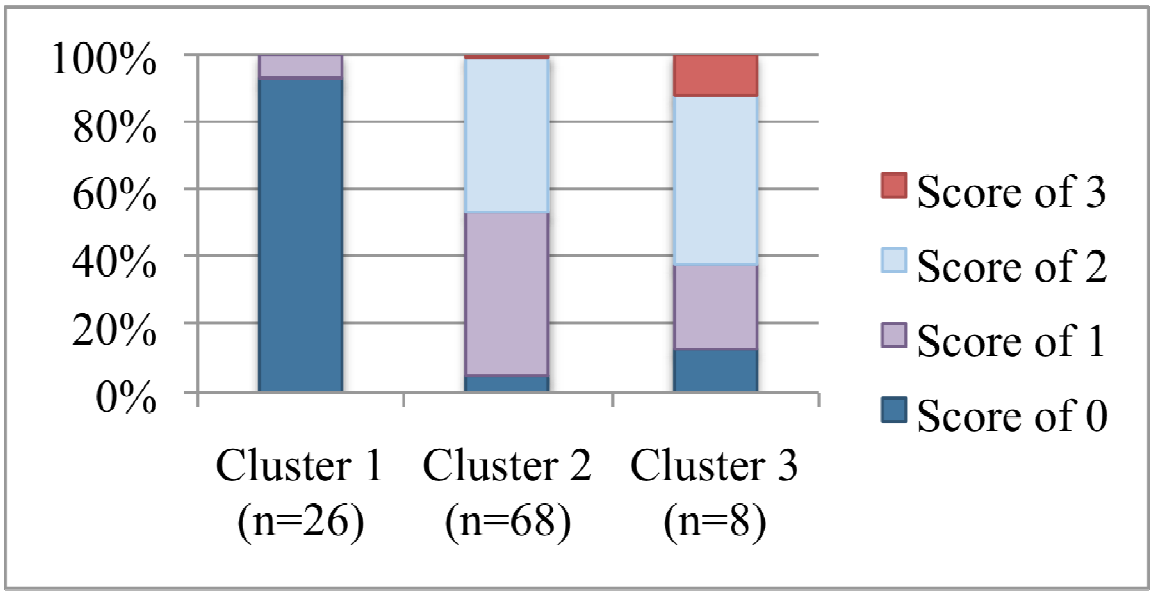
0 = No use of another's body to communicate
1 = Takes another person's hand and leads him/her places without coordinated gaze
2 = Placement of another person's hand or other body part on object
8 = Little or no spontaneous communication

Figure B7
Autism Diagnostic Observation Scale (ADOS) Item B7 Scores by Cluster



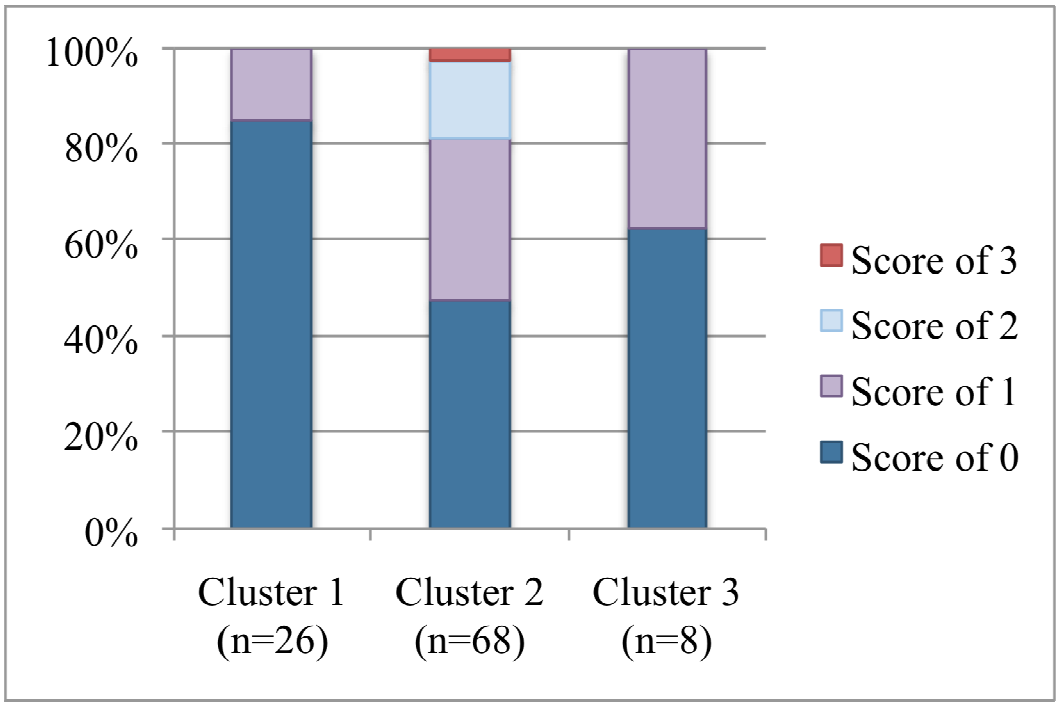
0 = Points with index finger using coordinated gaze to object and person
1 = Using pointing to reference objects, without sufficient flexibility or frequency for '0'
2 = Points to objects when close to or touching object, no coordinated eye gaze or vocalization
3 = Does not point to objects in any way

Figure B8
Autism Diagnostic Observation Scale (ADOS) Item C1 Scores by Cluster



0 = Spontaneously and appropriately plays with variety of toys
1 = Some spontaneous functional play with cause-and-effect toys with at least 1 miniature
2 = Plays appropriately with cause-and-effect toys only, and/or pushing car
3 = No play with toys or only stereotyped play

Figure B9
Autism Diagnostic Observation Scale (ADOS) Item D4 Scores by Cluster



0 = No repetitive or stereotyped behaviors during the ADOS evaluation
1 = An interest or behavior that is repetitive or stereotyped to an unusual degree
2 = Repetitive or stereotyped interests and/or behaviors are minority of child's interests or behaviors
3 =Repetitive or stereotyped interests and/or behaviors form majority of child's interests

Figure B10
Diagnosis at Age Four by Cluster