Evaluation of the Pervasive Developmental Disorder Behavior Inventory

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Kristy Lynne Hall, B.A.

Graduate Program in Psychology

The Ohio State University

2010

Master's Examination Committee:

James A. Mulick, Advisor
Luc Lecavalier
Michael Aman
Eric Butter
Abstract

This study assessed the psychometric properties of the PDD Behavior Inventory (PDDBI) as a diagnostic instrument for autism spectrum disorders. Individually matched samples of children and adolescents referred for evaluation of possible autism spectrum disorders (ASDs) were rated by their parents on the instrument. The sample consisted of 84 children and adolescents between the ages of 37 and 141 months (42 each in an ASD group and a non-ASD group). Optimal sensitivity and specificity were achieved using a cutoff score of 45 on the Autism Composite T-score. Independent samples T-tests indicated that groups differed significantly on the Autism Composite T-score and on one of the 14 subscale/composite scores. Analyses were also conducted on individually matched IQ-based subgroups of participants (one group consisting of children and adolescents with nonverbal IQs ≤ 70, and the other consisting of children and adolescents with nonverbal IQs > 70). The cutoff score of 32 that was recently proposed by the developers of the instrument was also evaluated in the entire sample and in the IQ-based subgroups. Results indicated that the instrument performed relatively poorly overall, with slightly better results obtained when used in the subsample of children and adolescents with nonverbal IQs ≤ 70.
This document is dedicated to my husband and my parents.
Acknowledgments

I wish to thank my advisor, Jim Mulick, and my committee member, Eric Butter, for their guidance in the data collection and design of the project.

I would also like to thank Luc Lecavalier, my committee member, for his guidance, wisdom, generosity with his time, and unending patience. His advice helped to greatly increase my knowledge of statistical analysis, autism spectrum disorders, assessment, and research design.
Vita

2002……………………………………..B.A., Psychology, Boston University

2003-2006……………………………….Research Coordinator, Research Units
on Pediatric Psychopharmacology and Psychosocial Intervention (RUPP-PI)

2006- 2007……………………………….Enrichment Fellow, Psychology

The Ohio State University

2007- present……………………………….Graduate Teaching Associate

The Ohio State University

Publications

Aman MG, McDougle CJ, Seahill L, Handen B, Arnold LE, Johnson C, Stigler
KA, Bearss K, Butter E, Swiezy NB, Sukhodolsky DD, Ramadan Y, Pozdol
SL, Nikolov R, Lecavalier L, Kohn AE, Koenig K, Hollway JA, Korzekwa
S, Vitiello B, Wagner A; the Research Units on Pediatric Psychopharmacology
Pervasive Developmental Disorders and Serious Behavior Problems: Results
From a Randomized Clinical Trial. Journal of the American Academy of Child
and Adolescent Psychiatry, 48, 1143-1154

Clark, DL, Arnold, LE, Crowl, L, Bozzolo, H, Peruggia, M, Ramadan, Y, Bornstein,
R, Hollway, J, Thompson, S, Pappas, K, Shelton, S, Bozzolo, D, Hall, K, Cook,
A (2008). Vestibular Stimulation for ADHD: Randomized Controlled Trial of
Comprehensive Motion Apparatus. Journal of Attention Disorders, 11, 599-
611.


Fields of Study

Major Field: Psychology
Table of Contents

Abstract .............................................................................................................................................. ii

Dedication ........................................................................................................................................ iii

Acknowledgments .......................................................................................................................... iv

Vita .................................................................................................................................................. v

List of Tables ................................................................................................................................... viii

List of Figures ................................................................................................................................. ix

Chapter 1: Introduction .................................................................................................................. 1

Chapter 2: Method ......................................................................................................................... 12

Chapter 3: Results ......................................................................................................................... 20

Chapter 4: Discussion .................................................................................................................... 25

Appendix: Tables and Figures ....................................................................................................... 32

References ...................................................................................................................................... 40
## List of Tables

Table 1. Characteristics of Sample................................................................. 33

Table 2. $T$-test Results on PDDBI Subscale and Composite Scores...................... 34

Table 3. PDDBI Autism Composite T-score Sensitivity, Specificity, PPV, NPV, and Efficiency ................................................................. 35

Table 4. IQ $\leq 70$ subgroup: Selected PDDBI Autism Composite T-score Sensitivity, Specificity, PPV, NPV, and Efficiency............................................ 36

Table 5. IQ $> 70$ subgroup: Selected PDDBI Autism Composite T-score Sensitivity, Specificity, PPV, NPV, and Efficiency............................................ 37

Table 6. Coefficient Alpha for PDDBI Subscales and Composites.......................... 38
List of Figures

Figure 1. Histograms of Autism Composite T-scores for ASD ($n = 42$) and non-ASD ($n = 42$) groups……………………………………………………………… ……...39
Chapter 1: Introduction

The pervasive developmental disorders (PDDs) are a set of behaviorally defined developmental disorders first manifested in childhood and characterized by impairment in social interaction and communication and the presence of restricted, repetitive, and stereotyped behavioral patterns, interests, and activities. The PDDs include Rett’s Disorder, Childhood Disintegrative Disorder, and three disorders known collectively as the Autism Spectrum Disorders (ASDs). The ASDs include Autistic Disorder (AD), Asperger’s Disorder, and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS; American Psychiatric Association, 2000).

ASDs share a number of similar features, but differ in number and severity of specific symptoms. Diagnostic criteria for AD include the presence of at least six total symptoms from the following three categories, including (a) at least two specific impairments in social interaction; (b) at least one specific impairment in communication; and (c) at least one restricted, repetitive, and/or stereotyped pattern of behavior, interests or activities. The onset of these symptoms must have been evident before the age of three years. Diagnostic criteria for Asperger’s disorder are very similar to AD, except that they do not include clinically significant delays in language or communication. To meet criteria for a diagnosis of PDD-NOS, the individual must have impairments in social interaction and...
either impairments in communication or the presence of stereotyped behaviors, interests, or activities. However, the individual must not meet full criteria for a more specific PDD in order to be given the diagnosis of PDD-NOS (American Psychiatric Association, 2000).

The median prevalence rate for AD has reportedly risen from 4.4 in 10,000 in 1966 to the current rate, estimated to be 10 in 10,000 (Fombonne, 2003). Males are 4.3 times as likely to be diagnosed with autism as females, and intellectual disability (ID) accompanies autism in approximately 70% of cases (Fombonne, 2003). Current prevalence estimates for all ASDs combined are approximately 60 per 10,000 (Fombonne, 2003), which is the equivalent of 1 in 167 individuals, or 0.6%.

Although prevalence estimates for ASDs (estimates of the proportion of individuals in a given population who, at a given point in time, have an ASD) have risen markedly over time, authorities question whether there is enough evidence to support a true increase in incidence of ASDs (the number of new cases of ASD occurring over a specified period of time). A recent review of published epidemiological surveys of ASDs (Fombonne, 2007) indicated that, while the reported prevalence has increased, the evidence does not strongly support the conclusion that the actual incidence of ASDs, or the number of new cases, has increased.
Many phenomena have been theorized as contributing to the prevalence increase. Changes in the conceptualization of ASDs to include individuals from all levels of functioning and individuals with other neuropsychiatric and medical disorders have arguably loosened the criteria necessary for an ASD diagnosis, contributing to the prevalence increase. In addition, social and educational policy changes, such as the introduction of the 1990 Individuals with Disabilities Education Act (IDEA), have also led to increased diagnostic evaluation referral rates and significant diagnostic practice changes. Most notably, children who likely would have been diagnosed with ID prior to the policy changes are now being diagnosed with ASDs. In fact, several recent publications indicate evidence of simultaneous decreases in the population prevalence of ID as prevalence of ASDs has been increasing (Lecavalier, Snow, & Norris, in press).

This uncertainty surrounding ASD prevalence trends necessitates the use of comprehensive screening methods and a thorough assessment process when evaluating individuals suspected of having an ASD. Filipek and colleagues (1999) highlighted the need for a two-step process that involves screening all children for developmental disorders as part of their routine early childhood evaluations by primary care providers (known as Level 1 screening), followed by a more in-depth diagnostic evaluation of children identified from those screenings as being at-risk (known as Level 2 evaluations).

According to Filipek and colleagues (1999), Level 2 diagnostic evaluations should be performed by an interdisciplinary team of experienced clinicians with expertise in the
differential diagnosis and treatment of developmental disorders. The assessment should include a comprehensive medical and neurological evaluation, a speech-language-communication evaluation, and a psychological evaluation. The psychological evaluation should be based on clinical and DSM-IV criteria and should include the use of an empirically validated diagnostic instrument with at least moderate sensitivity and good specificity for autism. The evaluation should also include a standardized parental interview; a direct, structured observation of the child’s social and communicative behavior and play; as well as cognitive, adaptive behavior, and academic assessments (Filipek et al., 1999).

In terms of specific instruments for use in a diagnostic assessment, it is conventionally accepted that the current gold standard for assessment in research, particularly research funded by the National Institutes of Health, is the Autism Diagnostic Observation Schedule-Generic (ADOS-G, Lord, et al., 2000) in combination with the Autism Diagnostic Interview – Revised (ADI-R, Lord, Rutter, & Le Couteur, 1994).

The ADOS-G is a semi-structured, standardized observational assessment tool designed for diagnosing children and adolescents with autism. It consists of 3 separate modules designed for use with children functioning at a variety of developmental levels and chronological ages and a fourth module for adults. Module selection is based upon the individual’s language functioning. Each module contains structured activities and materials designed to observe and assess social interactions, communication, and other
behaviors relevant to autism spectrum disorders. The instrument produces an algorithm with cutoff scores for AD and for ASD (with ASD cutoff scores being lower than AD cutoff scores). The assessment typically takes about 30 minutes to complete with a child and trained administrator (Lord, Rutter, DiLavore, & Risi, 1999).

The ADI-R is a semi-structured clinical interview, given to a parent or another caregiver by a trained administrator, that collects information on developmental history and milestones, age of onset, communication and language, social interaction, and repetitive, restricted, and stereotyped interests and behaviors. It collects information on current functioning and functioning between the ages of 4 and 5 years, conceptualized to be the age range at which ASD symptoms are at their peak. Given the length of the ADI-R (typical assessment times range from 2 – 3 hours), it is not practical for use in most clinical assessment settings due to insurance and personnel constraints. This makes application of research results to clinical populations arguably questionable. The ADI-R also requires a significant amount of time from parents when used as part of research studies.

This time-consuming, burdensome nature of the ADI-R and similar semi-structured diagnostic interviews makes parent-completed rating scales a very attractive alternative. They typically can be completed by a parent or caregiver at home, at the parent’s leisure, and require a very negligible amount of time from the clinician to score. When attempting to select the most appropriate rating scale to use, clinicians must closely
consider the psychometric properties of the available instruments. Cicchetti (1994) proposed several guidelines and criteria for use by clinicians when deciding among the many instruments currently available. These guidelines and rules of thumb cover appropriate standardization and norming procedures, test reliability (including internal consistency, test-retest, interexaminer, and temporal reliability), and test validity (including content-related, face, discriminant, clinical, concurrent, factorial, and criterion validity).

Using criteria similar to the above-mentioned Cicchetti guidelines, several of the currently available parent-rated scales were recently reviewed for screening accuracy. Norris and Lecavalier (in press) reviewed the literature in order to examine the current state of Level 2, caregiver-completed rating scales designed for the screening of ASDs. The instruments that had enough published supporting evidence to be included in their review were the Social Communication Questionnaire (SCQ, Berument et al., 1999), the Gilliam Autism Rating Scale/Gilliam Autism Rating Scale – Second Edition (GARS/GARS-2, Gilliam, 2006), the Social Responsiveness Scale (SRS, Constantino & Gruber, 2005), the Autism Spectrum Screening Questionnaire (ASSQ, Ehlers, Gillberg, & Wing, 1999), and the Asperger Syndrome Diagnostic Scale (ASDS, Myles, Bock, & Simpson, 2001). Results indicated that the SCQ was the only instrument that had been subjected to comparison with other instruments, comparison with a wide range of diagnostic groups, and independent validation and replication. Overall, the studies investigating the SCQ indicated that it was a very good instrument for use in screening
ASDs in individuals functioning above the moderate range of ID and over the age of 3. The SRS and ASSQ had promise, but need additional examination, and the GARS/GARS-2 and the ASDS had inadequate psychometric properties and were not recommended for use in screening for ASD (Norris & Lecavalier, in press).

Given the limited number of empirically validated rating scales currently available, the field is in need of additional options in terms of clinically useful, efficient, and psychometrically sound parent-report measures for making diagnostic decisions about ASD. Ideally, these instruments would be used in combination with the ADOS-G as is the current conventionally accepted gold standard of the ADI-R in combination with the ADOS-G. A relatively new instrument worthy of investigation for this purpose is the PDD Behavior Inventory (PDDBI, Cohen & Sudhalter, 2005). My aim in this study was to investigate the diagnostic validity of that instrument.

The PDDBI is a parent-completed questionnaire consisting of 188 items divided into 10 domains. Each domain results in a T-score, and domain scores are used to generate several composite T-scores, including an overall Autism Composite T-score. These T-scores are used to compare patients to a standardization sample of children with autism (where a T-score of 50, SD of 10, is the mean score of a child with autism of a given age). It typically takes about 45 minutes for a parent or caregiver to complete the questionnaire, and it can be completed outside of the clinical setting (Cohen & Sudhalter, 2005).
To date, the three studies investigating the PDDBI that are published in peer-reviewed journals have all been performed by the developers of the instrument. Cohen, Schmidt-Lackner, Romanczyk, and Sudhalter (2003) developed the PDDBI primarily as a means of assessing changes in maladaptive behaviors and core adaptive social and language skills relevant to PDDs in the context of treatment studies. Their scale-development sample consisted of 311 parents of children ages 1 to 17 years. They found internal consistency to be adequate to good for all subscales (alpha coefficients ranged from 0.79 – 0.97, median of 0.89, for the parent version). They examined construct validity using principal components analyses, largely confirming a priori-defined category-item groupings and the overall structure of the instrument.

In a second publication, Cohen (2003) examined the criterion-related validity of the PDDBI using a subset of the Cohen et al. (2003) study sample (N=84, ages 3-6). The PDDBI Autism Composite was found to correlate significantly with the three main ADI-R current behavior scores (Qualitative impairments in reciprocal social interaction: $r = 0.58$; Communication: $r = 0.40$; and Repetitive behaviors and stereotyped patterns: $r = 0.35$; all $p<.01$ or better) as well as the Childhood Autism Rating Scale ($r = 0.53$, $p < .0001$; CARS; Schopler, Reichler, DeVellis, & Daly, 1980). The author also compared the maladaptive subscales of the PDDBI with comparable subscales on the Nisonger Child Behavior Rating Form (NCBRF; Aman, Tasse, Rohajn, & Hammer, 1996) and found that the relevant subscales all had significant correlations ($p<.001$). Lastly, he
compared the adaptive subscales on the PDDBI to relevant subscales of the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984). Results indicated that all of the relevant PDDBI subscales correlated with the comparable Vineland subscales (p<.001).

In a third publication, Cohen, Gomez, Gonzalez, Lennon, Karmel, and Gardner (2010) examined the PDDBI’s ability to assist in the differential diagnosis of AD, PDD-NOS, and non-ASD as defined by ADOS-G and ADI-R criteria. Their sample consisted of 73 children between the ages of 2-5 years, 46 with ASD and 27 at risk for developmental problems. The age range of the sample was limited to five years and under in order to equate the two diagnostic groups for mean age, and the IQs of participants, as assessed by the Griffiths Mental Development Scales (GMDS, Griffiths, 1984), were limited to a GMDS performance quotient (PQ) of 75 or greater in order to eliminate cases with severe developmental delay. The Autism and PDD-NOS groups both had significantly lower mean PQ scores than the non-spectrum group, so the authors used PQ as a covariate in their profile analyses. They used repeated measures multivariate analyses of covariance (MANCOVA), followed by univariate analyses of covariance (ANCOVAs) on measures showing significant effects, in order to compare across diagnostic groups and informants. The Autism and non-spectrum groups significantly differed on several of the subscales and composite scores, including the Autism Composite (F=57.1, p<.001, effect size = .51). They also completed Receiver Operating Characteristic (ROC) analyses on the PDDBI composite measures in order to devise cut-off scores and document sensitivity,
specificity, and positive and negative predictive validity. For the Autism Composite, they compared the diagnostic groups of ASD (autism or PDD-NOS) versus non-spectrum. Sensitivity, Specificity, AUC, PPV, and NPV for their sample were 100, 79, 94, 89, and 100, respectively. The optimal Autism Composite cutoff score in their sample was 32. Because of the above-mentioned differences in PQ scores between the non-spectrum group and the autism and PDD-NOS groups, they repeated the ROC analyses on a subset of subjects (N=20 per group) individually matched on PQ. They found little change in their results, except that the optimal Autism Composite cutoff score moved from 32 to 42 (other statistics in the matched subset analysis were: Sensitivity=90, Specificity=95, AUC=97, PPV=95, NPV=90). They concluded that their overall results could not be explained by group differences in PQ, that the instrument has excellent sensitivity and specificity, and that the PDDBI is useful in the differential diagnosis of ASDs (Cohen et al., 2010).

While Cohen and colleagues have reported in three separate publications that their instrument has good psychometric properties, no independent research group has published data on the performance of the PDDBI. This means that the effectiveness of the PDDBI as a diagnostic tool has not been independently validated through replication. Independent validation of research results is a key component of scientific inquiry, and without it, one cannot say with certainty that a given research group’s findings were not biased in some way. By investigating the PDDBI with a different sample, I attempted to determine whether Cohen and colleagues’ results were replicable and that the instrument,
in fact, is effective in the differential diagnosis of ASDs. Given the good psychometric performance of the instrument as indicated by Cohen and colleagues in the three publications discussed above, I anticipated that I would find similarly promising results in my sample.
Chapter 2: Method

Participants

The participants were 84 children and adolescents between the ages of 3 years, 1 month and 11 years, 9 months (mean age = 6 years, 6 months; SD = 2 years, 3 months) who had been referred from 2006 to 2008 to a specialty clinic in a large Midwestern children’s hospital for diagnostic evaluation to rule out autism spectrum disorders. These participants were matched from a larger chart review of consecutive patients evaluated at the children’s hospital during that time. Participants were individually matched within 12 months of age and 10 nonverbal IQ points into an ASD group (17 with AD, 19 with PDD-NOS, 6 with Asperger’s Disorder; total n = 42) and a non-ASD group (primarily consisting of diagnoses of Developmental Delay, Language Impairment or Intellectual Disability; total n = 42). Participants in both groups were also limited to nonverbal IQs less than 115. The groups were matched on age and nonverbal IQ because ASDs affect people across the life span and across the full range of cognitive abilities. This variability in the ASD population makes it crucial that samples from that population are comparable in terms of age and IQ with non-ASD, clinical controls (Lord & Corsello, 2005). Groups tightly matched on these characteristics were included in the current study in an effort to help reduce any potential effects of differences in cognitive functioning and development on the comparison of PDDBI scores across groups. The final matched sample had
nonverbal IQs ranging from 46 to 111 (mean nonverbal IQ = 80.1, SD = 16.2) and was
78.6% male (n = 66 males). The ethnic composition was the following: 76.2% White (n = 64), 14.3% Black (n = 12), 2.4% Asian (n = 2), 1.2% Hispanic (n = 1), and 6% Biracial (n = 5).

The ASD group consisted of 36 males and 6 females. The mean age was 6 years, 5
months (SD = 2 years, 3 months). The mean nonverbal IQ was 80.5 (range= 46 - 111, SD
= 16.6). ADOS-G data were available for 39 of the participants in this group. Of those
with ADOS-G data, 32 of the 39 participants (76.2%) met the ASD cutoff, while 19 of
the 39 participants (45.2%) met the full AD cutoff.

The non-ASD group consisted of 30 males and 12 females. The mean age was 6 years, 6
months (SD = 2 years, 4 months). The mean nonverbal IQ was 79.6 (range= 50 - 108, SD
= 15.9). Primary psychological and medical diagnostic categories of the participants in
this group were the following: Disruptive Behavior Disorders and Attention Deficit
Disorders (n = 11), Language Disorders (n = 16), Other Developmental Disorders (n = 4),
Intellectual and Cognitive function disorders (n = 9), Stereotypic movement disorder (n =
1), and Disorder of Childhood – NOS (n = 1). ADOS-G data were available for 31 of the
participants in this group. Of those with ADOS-G data, 3 participants (7.1%) met the AD
cutoff and 8 participants (19%) met the ASD cutoff.
The subject characteristics of the ASD and non-ASD groups were compared. There were no significant differences between the groups on age, gender, ethnicity, IQ, or adaptive behavior. There were highly significant differences between the groups on the ADOS-G (see Table 1).

Informants
Parents of the participants completed the PDDBI at home prior to their diagnostic assessment. For the ASD group, the mean parental age was 36.9 years. For the non-ASD group, the mean parental age was 34.8 years. The two groups of parents did not significantly differ in age ($t=1.57, p=.121$). In terms of education, 33.3% of the parents in the ASD group graduated from college, while 31.9% of the parents in the non-ASD group graduated from college.

Instruments
PDD Behavior Inventory
The PDDBI is a parent-completed questionnaire consisting of 188 items. Fourteen separate subscale scores, called Domain or Composite scores, are generated from the PDDBI. Each domain and composite results in an age-normed T-score. An overall Autism Composite T-score is derived from these domain scores by subtracting the sum of T-scores of the domains that assess receptive/expressive social communication abilities from the sum of T-scores of domains that assess approach and withdrawal problems. The resulting Autism Composite T-score and the other domain and composite T-scores are
used to compare patients to a standardization sample of children with autism, where a T-score of 50, SD of 10, is the mean score of a child with autism of a given age (Cohen & Sudhalter, 2005).

Autism Diagnostic Observation Schedule-Generic

The ADOS-G consists of 4 separate modules designed for use by trained administrators with children and adults functioning at a variety of developmental levels, language levels, and chronological ages. Each module contains structured activities and materials designed to observe and assess the patient’s social interactions, communication, and other behaviors relevant to autism spectrum disorders. The instrument produces an algorithm with cutoff scores for AD and for ASD. Higher scores signify more pathology; hence, ASD cutoff scores are lower than AD cutoff scores (Lord, Rutter, DiLavore, & Risi, 1999).

Procedure

Each participant was referred to a major children’s hospital’s developmental disorders diagnostic clinic for evaluation for possible ASDs. Participants were evaluated by a multidisciplinary team including a psychologist specializing in developmental disorders, a developmental pediatrician, and a speech and language pathologist. Following each of their individual evaluations of the participant, the team came to a consensus diagnosis, using DSM-IV-TR criteria, based on interviews with the caregiver(s); extensive intelligence, speech, and behavioral testing of the participant (usually including the
ADOS-G); and diagnostic and adaptive behavior measures completed by the parent or caregiver prior to the evaluation (including the PDDBI).

In terms of intelligence testing, the majority of participants were assessed using the Stanford-Binet – Fifth Edition (Roid, 2003), the Leiter International Performance Scale – Revised (Roid & Miller, 1997), or both. Other intelligence tests used, depending on the age and language skills of the participant, included the Wechsler Intelligence Scale for Children – Fourth Edition (Wechsler, 2003), Wechsler Preschool and Primary Scale of Intelligence – Third Edition (Wechsler, 2002), and Mullen Scales of Early Learning (Mullen, 1997).

Adaptive functioning was assessed primarily using the Scales of Independent Behavior-Revised (Bruininks, Woodcock, Weatherman, & Hill, 1996). A small portion of the sample was assessed using the Vineland Adaptive Behavior Scales – Second Edition, Parent/Caregiver Rating Form (Sparrow, Cicchetti, & Balla, 2005). Both instruments were completed independently by the parent or caregiver prior to the in-person assessment appointment.

As mentioned above, the two diagnostic groups were individually matched on age and nonverbal IQ. IQ has been shown to moderate the presentation of core ASD symptoms including social and communication impairment and restrictive, repetitive behaviors and interests (Lecavalier et al., in press). Additionally, children with lower nonverbal IQs
have been found to have the most severe ASD symptoms, especially in social communication (Lecavalier et al., in press). This relationship between IQ and ASD symptomatology necessitates the inclusion of ASD and non-ASD groups that are evenly matched on IQ levels so that IQ differences do not confound the analyses.

Statistical Analyses

Independent samples t-tests were performed on the Autism Composite raw score, T-score and all of the PDDBI subscale and composite scores, with ASD diagnosis as the grouping variable. Receiver Operating Characteristic (ROC) analyses, procedures used to evaluate the performance of a classification tool by gauging the relative tradeoff between true and false positives, were also performed in order to determine an optimal ASD cutoff value for the Autism Composite T-score as well as to determine the Area Under the Curve (AUC). AUC represents the probability that the Autism Composite T-score will assign the correct diagnostic grouping for a randomly selected participant. AUC values can range from .50 (indicating a poor instrument generating an equal rate of true and false positives) to 1.0 (indicating a perfect instrument generating all true positives and no false positives). Statistics obtained for each potential cutoff score included sensitivity (the proportion of participants diagnosed with ASD who met or exceeded that Autism Composite cutoff score), specificity (the proportion of participants diagnosed as not having ASD whose Autism Composite score fell below that cutoff score), Positive Predictive Value (PPV; the probability of having an ASD diagnosis, given a score that meets or exceeds that cutoff score), Negative Predictive Value (NPV, the probability of
not having an ASD diagnosis, given a score that falls below that cutoff score), and Efficiency (the overall correct classification rate at that cutoff score) of the Autism Composite raw and T-scores at several potential cutoff scores.

In order to investigate the Autism Composite T-score performance in lower- and higher-IQ participants separately, analyses were repeated on a subgroup of the entire sample with nonverbal IQs of 70 or below ($n = 24$) and on a second subgroup of the entire sample with nonverbal IQs above 70 ($n = 50$). Participants with and without ASD in the two IQ-based subgroups were individually matched using the same criteria as in the main sample. Within these two subgroups, the ROC analysis was repeated and $t$-tests were repeated on the Autism Composite raw and T-score, subscale scores, and composite scores. All 84 participants from the entire sample could not be included in the IQ-based, matched subgroups because, in the entire sample, several pairs of participants had been matched across the nonverbal IQ of 70 threshold and a match for those participants within their own side of the threshold could not be made for the subgroup analyses.

In order to investigate reliability of the PDDBI subscales, coefficient (Cronbach’s) alpha was calculated on all 14 domain/composite scores in the entire sample.

Correlations were calculated separately for ASD and non-ASD groups within the entire sample in order to determine whether PDDBI score is affected by subject characteristics in the two diagnostic groups. Pearson correlations were used to determine the association
between age and PDDBI score. Because different instruments were used for cognitive and adaptive behavior assessments among participants, nonverbal IQ scores and adaptive behavior composite scores were converted into the following categories: 0 = below 25, 1 = 25-40, 2 = 41-55, 3 = 56-70, 4 = 71-85, 5 = 86-100; 6 = 101-115. Spearman’s rho correlations were used to determine the associations between nonverbal IQ, adaptive behavior, and PDDBI score within the two diagnostic groups.
Chapter 3: Results

Diagnostic Validity

Results of the \( t \)-tests on the Autism Composite raw score and T-score from the entire sample indicated that the ASD and non-ASD groups differed significantly on both scores (raw score: \( t = 3.05, p = .003 \); T-score: \( t = 3.07, p = .003 \), effect size (Cohen's \( d \)) = .68). To provide some visual perspective on the separation of the two groups with the PDDBI, I computed histograms for both the ASD and the non-ASD groups. These are juxtaposed in Figure 1. As can be seen, there was some separation of the groups based on the PDDBI T-scores. However, the figure is more remarkable for the extent of overlap between the ASD and the non-ASD groups. As mentioned above, \( t \)-tests were also performed on the 14 PDDBI subscale/composite scores. Given the relatively large number of comparisons performed, alpha for significance was set at .01 in order to correct for chance associations. At this significance level, \( t \)-tests on the subscale/composite scores indicated that, within the entire sample, the two diagnostic groups differed significantly on only one of the 14 subscales/composites, Social Pragmatic Problems (SOCPP; \( t = 3.73, p < .001 \)). See Table 2 for \( t \)-test results on all 14 subscale/composite scores.

In terms of the IQ-based subgroups, the Autism Composite raw score and T-score for participants with and without ASD differed significantly within the lower-IQ subgroup
(raw score: $t = 2.25, p = .04$; T-score: $t = 2.25, p = .04$). Within the higher-IQ subgroup, however, the scores did not significantly differ between the ASD and non-ASD groups (raw score: $t = 1.10, p = .28$; T-score: $t = 1.12, p = .27$). When evaluating the 14 PDDBI subscale/composite scores for the IQ-based subgroups, the two diagnostic groups differed significantly on one of the subscales/composites in the lower-IQ group (FEARS: $t = 4.38, p < .001$), which is not used in calculating the Autism Composite score. The two diagnostic groups did not differ significantly on any of the subscale/composite scores in the higher-IQ group.

To summarize, the ASD and non-ASD groups differed significantly on the Autism Composite raw and T-scores in the entire sample as well as in the lower-IQ subgroup. The two diagnostic groups also differed significantly on one additional subscale/composite score in the entire sample (SOCPP) and one additional subscale/composite in the lower-IQ subgroup (FEARS). The two diagnostic groups did not significantly differ on the Autism Composite or any of the other subscale/composite scores in the higher-IQ subgroup.

Receiver Operating Characteristic Curves Analyses

The ROC analysis of Autism Composite T-scores in the entire sample resulted in an AUC of .695 ($p = .002$; 95% CI = .582 - .808). Table 3 shows the Sensitivity, Specificity, PPV, NPV, and Efficiency at several different Autism Composite T-score cutoff scores, calculated using DAG_STAT (Mackinnon, 2000). The optimal cutoff score in this
sample is 45. At that cutoff, Sensitivity was .74 (95% CI = .58-.86) and Specificity was .62 (95 % CI = .46-.76). PPV was .66 (95% CI=.51-.79), NPV was .70 (95% CI=.53-.84), and Efficiency was .68 (95% CI=.57-.78). The optimal Autism Composite T-score cutoff of 32 that was reported by Cohen et al (2010) is also listed in the table. At a cutoff of 32, Sensitivity was .98 (95% CI=.87-1.0), Specificity was .19 (95% CI=.09-.34), PPV was .55 (95% CI=.43-.66), NPV was .89 (95% CI=.52-1.0) and Efficiency was .58 (95% CI=.47-.69). In other words, use of this cutoff score in the current sample would result in 41 out of the 42 ASD participants being correctly classified (with 1 false negative).

However, it would also result in 34 out of the 42 non-ASD participants being incorrectly classified (34 false positives and 8 true negatives). This represents a total of 58.3% of the current sample that would be correctly classified with a cutoff score of 32. The ROC analysis was repeated on the Autism Composite raw scores in order to determine if performance of the PDDBI was better without using the transformed scores. Virtually identical results were obtained (AUC = .694; \( p = .002 \); 95% CI = .581 - .808; optimal raw score cutoff was 82 with Sensitivity = .74 and Specificity = .62).

Within the lower-IQ subgroup (n = 24), the AUC increased to .774 (\( p = .023 \); 95% CI = .575 - .974). Table 4 shows the Sensitivity, Specificity, PPV, NPV, and Efficiency at several different Autism Composite T-score cutoff scores for this subgroup. The optimal cutoff score in this subgroup was still 45, but Sensitivity increased to .92 (95% CI = .62-1.0) and Specificity increased slightly to .67 (95 % CI = .35-.90). PPV was .73 (95% CI=.45-.92), NPV was .89 (95% CI=.52-1.0), and Efficiency was .79 (95% CI=.58-.93).
The optimal Autism Composite T-score cutoff of 32 that was reported by Cohen et al (2010) resulted in Sensitivity of 1.0 (95% CI=.74-1.0), Specificity of .25 (95% CI=.06-.57), PPV of .57 (95% CI=.34-.78), and Efficiency of .63 (95% CI=.41-.81). In other words, use of this cutoff score in the lower-IQ subgroup would result in all of the 12 ASD participants being correctly classified, but it would also result in 9 out of the 12 non-ASD participants being incorrectly classified (9 false positives and 3 true negatives). This represents a total of 62.5% of the lower-IQ subgroup that would be correctly classified with a cutoff score of 32.

Within the higher-IQ subgroup \((n = 50)\), the AUC decreased to .609 \((p = .187; 95\% \text{ CI} = .450 - .767)\). Table 5 shows the Sensitivity, Specificity, PPV, NPV, and Efficiency at several different Autism Composite T-score cutoff scores for this subgroup. The optimal cutoff score in this subgroup increased to 47, with Sensitivity of .56 (95% CI = .35-.76) and Specificity of .72 (95 % CI = .51-.88). PPV was .67 (95% CI=.43-.85), NPV was .62 (95% CI=.42-.79), and Efficiency was .64 (95% CI=.49-.77). The optimal Autism Composite T-score cutoff of 32 that was reported by Cohen et al (2010) resulted in Sensitivity of .96 (95% CI=.80-1.0), Specificity of .16 (95% CI=.05-.36), PPV of .53 (95% CI=.38-.68), NPV of .80 (95% CI=.28-1.0) and Efficiency of .56 (95% CI=.41-.70). In other words, use of this cutoff score in the higher-IQ subgroup would result in 24 out of the 25 ASD participants being correctly classified (with 1 false negative). However, it would also result in 21 out of the 25 non-ASD participants being incorrectly classified.
(21 false positives and 4 true negatives). This represents a total of 56.0% of the higher-IQ subgroup that would be correctly classified with a cutoff score of 32.

Cronbach’s Alpha

Internal consistency of the 14 PDDBI subscales ranged from .81 to .98, with an average value of .91 (see Table 6).

Relationship Between Participant Characteristics and PDDBI Scores

In the entire sample, neither the PDDBI Autism Composite raw score nor the Autism Composite T-score was significantly correlated with age in either diagnostic group (raw score in ASD group: \( r = -.11, p = .49 \); raw score in non-ASD group: \( r = -.04, p = .79 \); T-score in ASD group: \( r = -.11, p = .50 \); T-score in non-ASD group: \( r = -.04, p = .80 \)). Both scores were significantly correlated with nonverbal IQ in the ASD group (raw score: Spearman’s rho (\( \rho \)) = -.41, \( p = .008 \); T-score: \( \rho = -.40, p = .009; p < .01 \) for both correlations). Scores were not correlated with nonverbal IQ in the non-ASD group (raw score: \( \rho = -.01, p = .95 \); T-score: \( \rho = -.004, p = .98 \)). Both scores were also significantly correlated with adaptive behavior in the ASD group (raw score: \( \rho = -.54, p = .001 \); T-score: \( \rho = -.54, p = .001 \)). Scores were not correlated with adaptive behavior in the non-ASD group (raw score: \( \rho = -.14, p = .42 \); T-score: \( \rho = -.14, p = .42 \)).
Chapter 4: Discussion

This was the first study of the PDDBI conducted by a researcher not affiliated with the developers of the instrument. In addition, this study used a comparison group consisting of age- and nonverbal IQ-matched participants diagnosed with non-ASD developmental disabilities. As children with non-ASD developmental disabilities often present with symptoms similar to those of children with ASDs, this study design created an especially stringent comparison group used to assess the performance of the PDDBI.

As mentioned above, Cohen and colleagues (2010) performed an ROC analysis and found that the optimal Autism Composite T-score cutoff for a comparison of an ASD group (autism or PDD-NOS) versus a non-spectrum group was 32 (with sensitivity of 1.0 and specificity of .79). In the current sample, however, a cutoff of 32 resulted in excellent sensitivity (.98), but extremely poor specificity (.20). Performance at that cutoff score in the lower-functioning subgroup was slightly better, while performance in the higher-functioning subgroup was even worse. In the current sample, an Autism Composite T-score cutoff of 45 resulted in the best sensitivity and specificity. However, this cutoff falls well within the range (a mean of 50 and a SD of 10) stated by the developers of the instrument as being “characteristic of approximately 68% of cases with autism” (Cohen & Sudhalter, 2005, p. 25). In addition, Galscoe (2005) stated that, in
general, sensitivity of an instrument should be at least .70-.80 and specificity should be closer to .80 or higher. Given that the ultimate goal of using a screening instrument like this is to identify those children and adolescents who likely have an ASD, it could be argued that sensitivity should be closer to .80 or higher in this case. The Sensitivity and Specificity values obtained at the cutoff of 45, while being the best in this sample, were still relatively poor, as both fell well below .80 (Sensitivity = .74, Specificity = .62). For comparison purposes, Papanikolaou and colleagues (2009), in an analysis of the ADOS-G in differentiating between an Autism and PDD-NOS group versus a non-spectrum group, found sensitivity and specificity values of .85 and .75, respectively. In an analysis of the ADI-R in differentiating between an Autism group versus a PDD-NOS and non-spectrum group, Papanikolaou and colleagues found sensitivity and specificity values of .88 and .69, respectively (Papanikolaou et al., 2009).

*T*-tests on the entire sample indicated that the ASD and non-ASD groups differed significantly on the Autism Composite T-score. However, the means for the two groups (ASD mean = 50.4, SD=11.4; non-ASD mean = 42.7, SD = 11.7) only differed by 7.7 points. Also, just as with the optimal cutoff score in this sample, both means fell within the Autism Composite T-score range (a mean of 50 and a SD of 10) stated by the developers of the instrument as being “characteristic of approximately 68% of cases with autism” (Cohen & Sudhalter, 2005, p. 25). In fact, 50% of the non-ASD sample fell within that T-score range of 50 +/- 10. Additionally, while the histograms presented in Figure 1 show some separation of the two diagnostic groups based on the PDDBI T-
scores, the figure is more remarkable for the extent of overlap between the ASD and the non-ASD groups. In other words, while there was a statistically significant difference between the two diagnostic groups on the Autism Composite T-score, this difference was likely not practically or clinically significant in terms of the instrument’s ability to differentiate between children with and without ASD when used in clinical practice following the interpretation guidelines stated in the manual.

When looking at the performance of the instrument’s 14 domain/composite scores in differentiating between the two diagnostic groups in the entire sample, only one 12-item subscale (Social Pragmatic Problems; SOCPP) resulted in significantly different scores between the two diagnostic groups. This subscale contains three 4-question sections assessing social approach problems, social awareness problems, and inappropriate reactions to the approaches of others. Given that impairments in socialization are one of the three core-defining features of ASDs and are given the most weight in making a diagnosis of autistic disorder according to DSM-IV-TR guidelines (at least two social criteria need to be met, compared to just one each for the other two areas; APA, 2000), one would expect to see significant differences between subjects with and without ASD on this subscale assessing social problems. However, it is surprising that the diagnostic groups in this sample did not significantly differ on the PDDBI subscales most relevant to the other two core-defining features of ASDs (impairments in verbal and nonverbal communication, and restricted and repetitive patterns of behavior).
In terms of internal consistency of the domain/composite scores, Cronbach’s alpha ranged from .81 to .98, with an average value of .91. According to Cicchetti (1994), when coefficient alpha is .80 or above, the level of clinical significance can be considered good to excellent. However, a visual inspection of the items within the domains and composites of the PDDBI revealed a high level of item redundancy (i.e., multiple items assessing the same construct). This redundancy of items and the sheer number of items within each domain and composite would inherently lead to an increase in coefficient alpha, regardless of the actual reliability of the instrument.

Taken together, the results of the $t$-tests and ROC analyses do not indicate that the Autism Composite T-score differentiated between children with and without an ASD in a clinically meaningful, useful way in this sample. Therefore, the utility of the PDDBI Autism Composite T-score as a diagnostic tool seems questionable. This is a very different result than the developers of the instrument obtained in their analyses, particularly the conclusions made in the most recent publication (Cohen et al., 2010).

This discrepancy between results of that study (Cohen et al., 2010) and the current study may be due to several factors. Cohen and colleagues (2010) identified a significant difference between their diagnostic groups on nonverbal/performance IQ. When they investigated that difference using a small subset of their sample that was matched on IQ, the optimal Autism Composite cutoff score increased dramatically, from 32 in the entire (unmatched) sample to 42 in the matched subset. That matched subset cutoff was
actually quite close to the cutoff of 45 obtained in the current matched sample. Additionally, when investigating the possible effects of participant characteristics on Autism Composite T-scores in this sample, it was found that the Autism Composite T-score was significantly negatively correlated with nonverbal IQ and with adaptive behavior in the ASD group. Also, the Autism Composite T-score was significantly different between diagnostic groups in the low-IQ subgroup (p = .04) but not in the higher-IQ subgroup (p = .27). Taken together, these findings indicate that the PDDBI’s Autism Composite T-score is heavily influenced by the level of functioning of the patients that it is used to assess, especially those with ASD. In particular, the score appears to be more effective at distinguishing between ASD and non-ASD participants who are lower-functioning. This could be a possible explanation for the differences between the results obtained by Cohen and colleagues (2010) with their entire (unmatched) sample and the results obtained in this IQ-matched entire sample and IQ-based subgroup analyses.

Another issue worthy of consideration is that the instrument was developed primarily as a means of tracking change in core ASD diagnostic features in the context of treatment studies. In fact, the initial development sample consisted only of children with ASD (Cohen et al, 2003). However, it can be argued that the instrument should also be established as reasonably accurate in its initial measurement of diagnostic symptoms if it is to be considered adequate in tracking change in those symptoms over time. Cohen and colleagues (2003), as an apparent afterthought, included a very small ROC analysis on
the Autism Composite score between children with autism \((n = 135)\), PDD-NOS \((n = 28)\) and mixed receptive-expressive language disorder (MRELD; \(n = 20\)). When comparing the autism and MRELD diagnostic groups, a cutoff score of greater than 40 resulted in sensitivity and specificity of .91 and .80, respectively. But a sample size of only 20 in the MRELD group, with no attention paid to differences in IQ across the groups, was not a strong test of diagnostic validity.

Given that this instrument was not developed with much consideration given to its ability to differentiate between individuals with and without ASD, its poor performance when used for that purpose is not entirely surprising. The instrument did perform better in a group of just lower-functioning subjects, so future research may indicate that the instrument may be more useful as a diagnostic tool in lower-functioning populations. However, the argument that the PDDBI performs well as a diagnostic tool overall (Cohen et al., 2010) is questionable, based on the data obtained from this study.

Limitations and Future Directions

One limitation was that the sample was not large enough to allow for additional subgroup analyses, including those based on ASD subtypes, age groups, or more narrow subgroups based on level of functioning. Another limitation was that participants were not assessed using one standardized battery of IQ and adaptive behavior instruments. However, the stringent matching methods likely compensated for most variability that may have been present do to the use of different instruments.
Future research should investigate the Autism Composite T-score performance in larger, more narrowly defined IQ-based and age-based subgroups. The current study found that the PDDBI performs best in lower-functioning patients and does not perform well in higher-functioning patients (those with nonverbal IQs over 70). However, Cohen et al. (2010), using a higher-functioning sample limited to patients with nonverbal IQs over 75, found that the PDDBI performed well in higher-functioning patients. Future research should investigate this issue further and attempt to determine for which functioning level(s) and age range(s) this instrument performs best. Also, independent validation of the PDDBI’s utility as a measure of change in symptoms due to treatment remains to be evaluated.
APPENDIX:

TABLES AND FIGURES
Table 1: Characteristics of Sample

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>ASD</th>
<th>Non-ASD</th>
<th>Test value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (in months): mean (SD)</td>
<td>77.1 (27.4)</td>
<td>78.2 (27.6)</td>
<td>-.176</td>
<td>.861</td>
</tr>
<tr>
<td>Gender: no. (%) male</td>
<td>36 (85.7)</td>
<td>30 (71.4)</td>
<td>2.55</td>
<td>.11</td>
</tr>
<tr>
<td>Ethnicity: no. (%) Caucasian</td>
<td>28 (66.7)</td>
<td>36 (85.7)</td>
<td>5.53</td>
<td>.237</td>
</tr>
<tr>
<td>NVIQ: mean (SD)</td>
<td>80.5 (16.6)</td>
<td>79.6 (15.9)</td>
<td>.248</td>
<td>.805</td>
</tr>
<tr>
<td>VIQ\textsuperscript{a}</td>
<td>75.9 (18.7)</td>
<td>78.1 (20.0)</td>
<td>-.482</td>
<td>.631</td>
</tr>
<tr>
<td>FSIQ\textsuperscript{b}</td>
<td>77.4 (18.3)</td>
<td>77.0 (17.1)</td>
<td>.10</td>
<td>.920</td>
</tr>
<tr>
<td>Adaptive Behavior Composite Score\textsuperscript{c}</td>
<td>68.2 (22.8)</td>
<td>67.9 (24.8)</td>
<td>.050</td>
<td>.960</td>
</tr>
<tr>
<td>ADOS: mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social\textsuperscript{d}</td>
<td>6.8 (3.6)</td>
<td>3.6 (3.0)</td>
<td>4.09</td>
<td>.001</td>
</tr>
<tr>
<td>Communication\textsuperscript{d}</td>
<td>3.9 (2.5)</td>
<td>1.7 (1.5)</td>
<td>4.41</td>
<td>.001</td>
</tr>
<tr>
<td>Social + Communication\textsuperscript{d}</td>
<td>10.7 (5.6)</td>
<td>5.2 (4.0)</td>
<td>4.65</td>
<td>.001</td>
</tr>
<tr>
<td>Stereotyped Behavior\textsuperscript{e}</td>
<td>1.6 (1.6)</td>
<td>0.5 (0.9)</td>
<td>2.99</td>
<td>.004</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Sample size of 70 (34 in ASD and 36 in non-ASD); \textsuperscript{b}Sample size of 81 (41 in ASD, 42 in non-ASD); \textsuperscript{c}Sample size of 70 (36 in ASD and 34 in non-ASD); 58 using the Scales of Independent Behavior-Revised (SIB-R) and 12 using the Vineland Adaptive Behavior Scales (VABS)-2; \textsuperscript{d}Sample size of 70 (39 in ASD and 31 in non-ASD); \textsuperscript{e}Sample size of 55 (29 in ASD and 26 in non-ASD)
Table 2: $T$-test Results on PDDBI Subscale and Composite Scores

<table>
<thead>
<tr>
<th>PDDBI Subscale</th>
<th>ASD mean (SD)</th>
<th>Non-ASD mean (SD)</th>
<th>Mean Difference</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SENSORY</td>
<td>50.21 (9.56)</td>
<td>46.83 (9.59)</td>
<td>3.39</td>
<td>1.61</td>
</tr>
<tr>
<td>RITUAL</td>
<td>53.71 (10.88)</td>
<td>51.22 (13.19)</td>
<td>2.50</td>
<td>.94</td>
</tr>
<tr>
<td>SOCPP</td>
<td>52.03 (11.27)</td>
<td>42.93 (10.54)</td>
<td>9.10</td>
<td>3.73*</td>
</tr>
<tr>
<td>SEMPP</td>
<td>53.28 (9.73)</td>
<td>49.38 (9.81)</td>
<td>3.90</td>
<td>1.79</td>
</tr>
<tr>
<td>AROUSE</td>
<td>50.90 (10.23)</td>
<td>47.23 (11.01)</td>
<td>3.68</td>
<td>1.55</td>
</tr>
<tr>
<td>FEARS</td>
<td>55.53 (10.10)</td>
<td>50.18 (13.50)</td>
<td>5.35</td>
<td>2.01</td>
</tr>
<tr>
<td>AGG</td>
<td>55.08 (12.19)</td>
<td>55.33 (16.16)</td>
<td>-.250</td>
<td>-.078</td>
</tr>
<tr>
<td>REPRIT/C</td>
<td>53.38 (11.66)</td>
<td>47.27 (12.88)</td>
<td>6.11</td>
<td>2.24</td>
</tr>
<tr>
<td>AWP/C</td>
<td>54.43 (11.45)</td>
<td>49.44 (14.06)</td>
<td>4.99</td>
<td>1.75</td>
</tr>
<tr>
<td>SOCAPP</td>
<td>54.55 (9.44)</td>
<td>59.20 (10.44)</td>
<td>-4.65</td>
<td>-2.09</td>
</tr>
<tr>
<td>EXPRESS</td>
<td>54.83 (8.75)</td>
<td>56.53 (10.59)</td>
<td>-1.70</td>
<td>-.783</td>
</tr>
<tr>
<td>LMRL</td>
<td>53.80 (8.28)</td>
<td>53.00 (8.15)</td>
<td>.800</td>
<td>.436</td>
</tr>
<tr>
<td>EXSCA/C</td>
<td>55.38 (8.59)</td>
<td>58.51 (9.91)</td>
<td>-3.14</td>
<td>-1.52</td>
</tr>
<tr>
<td>REXSCA/C</td>
<td>55.30 (8.21)</td>
<td>57.56 (9.34)</td>
<td>-2.26</td>
<td>-1.16</td>
</tr>
</tbody>
</table>

SENSORY = Sensory/Perceptual Approach Behaviors; RITUAL = Ritualisms/Resistance to Change; SOCPP = Social Pragmatic Problems; SEMPP = Semantic/Pragmatic Problems; AROUSE = Arousal Regulation Problems; FEARS = Specific Fears; AGG = Aggressiveness; REPRIT/C = Repetitive, Ritualistic, and Pragmatic Problem Behaviors Composite; AWP/C = Approach/Withdrawal Problems Composite; SOCAPP = Social Approach Behaviors; EXPRESS = Expressive Language; LMRL = Learning, Memory and Receptive Language; EXSCA/C = Expressive Social Communication Abilities Composite; REXSCA/C = Receptive/Expressive Social Communication Abilities Composite; *$p<.001$;
Table 3: PDDBI Autism Composite T-score Sensitivity, Specificity, PPV, NPV, and Efficiency

<table>
<thead>
<tr>
<th>Cutoff score</th>
<th>Sens</th>
<th>Spec</th>
<th>PPV</th>
<th>NPV</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>32\textsuperscript{a}</td>
<td>.976</td>
<td>.190</td>
<td>.547</td>
<td>.889</td>
<td>.583</td>
</tr>
<tr>
<td>33</td>
<td>.976</td>
<td>.238</td>
<td>.562</td>
<td>.909</td>
<td>.607</td>
</tr>
<tr>
<td>34</td>
<td>.952</td>
<td>.262</td>
<td>.563</td>
<td>.846</td>
<td>.607</td>
</tr>
<tr>
<td>35</td>
<td>.881</td>
<td>.262</td>
<td>.544</td>
<td>.688</td>
<td>.571</td>
</tr>
<tr>
<td>36</td>
<td>.857</td>
<td>.310</td>
<td>.554</td>
<td>.684</td>
<td>.583</td>
</tr>
<tr>
<td>37</td>
<td>.857</td>
<td>.357</td>
<td>.571</td>
<td>.714</td>
<td>.607</td>
</tr>
<tr>
<td>38</td>
<td>.786</td>
<td>.381</td>
<td>.559</td>
<td>.640</td>
<td>.583</td>
</tr>
<tr>
<td>39</td>
<td>.786</td>
<td>.381</td>
<td>.559</td>
<td>.640</td>
<td>.583</td>
</tr>
<tr>
<td>40</td>
<td>.762</td>
<td>.429</td>
<td>.571</td>
<td>.643</td>
<td>.595</td>
</tr>
<tr>
<td>41</td>
<td>.762</td>
<td>.476</td>
<td>.593</td>
<td>.667</td>
<td>.619</td>
</tr>
<tr>
<td>42</td>
<td>.738</td>
<td>.500</td>
<td>.596</td>
<td>.656</td>
<td>.619</td>
</tr>
<tr>
<td>43</td>
<td>.738</td>
<td>.500</td>
<td>.596</td>
<td>.656</td>
<td>.619</td>
</tr>
<tr>
<td>44</td>
<td>.738</td>
<td>.548</td>
<td>.620</td>
<td>.676</td>
<td>.643</td>
</tr>
<tr>
<td>45\textsuperscript{b}</td>
<td>.738</td>
<td>.619</td>
<td>.660</td>
<td>.703</td>
<td>.679</td>
</tr>
<tr>
<td>46</td>
<td>.690</td>
<td>.619</td>
<td>.644</td>
<td>.667</td>
<td>.655</td>
</tr>
</tbody>
</table>

Sens=Sensitivity; Spec=Specificity; PPV=Positive Predictive Value; NPV=Negative Predictive Value;
\textsuperscript{a}Cutoff proposed by Cohen et al (2010), \textsuperscript{b}Optimal cutoff in current sample
Table 4: IQ $\leq$ 70 subgroup: Selected PDDBI Autism Composite T-score Sensitivity, Specificity, PPV, NPV, and Efficiency

<table>
<thead>
<tr>
<th>Cutoff score</th>
<th>Sens</th>
<th>Spec</th>
<th>PPV</th>
<th>NPV</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>32$^a$</td>
<td>1.0</td>
<td>.250</td>
<td>.571</td>
<td>--</td>
<td>.625</td>
</tr>
<tr>
<td>35</td>
<td>.917</td>
<td>.250</td>
<td>.550</td>
<td>.750</td>
<td>.583</td>
</tr>
<tr>
<td>38</td>
<td>.917</td>
<td>.333</td>
<td>.579</td>
<td>.800</td>
<td>.625</td>
</tr>
<tr>
<td>42</td>
<td>.917</td>
<td>.417</td>
<td>.611</td>
<td>.833</td>
<td>.667</td>
</tr>
<tr>
<td>45$^b$</td>
<td>.917</td>
<td>.667</td>
<td>.733</td>
<td>.889</td>
<td>.792</td>
</tr>
<tr>
<td>46</td>
<td>.833</td>
<td>.667</td>
<td>.714</td>
<td>.800</td>
<td>.750</td>
</tr>
<tr>
<td>48</td>
<td>.750</td>
<td>.667</td>
<td>.692</td>
<td>.727</td>
<td>.708</td>
</tr>
<tr>
<td>52</td>
<td>.750</td>
<td>.750</td>
<td>.750</td>
<td>.750</td>
<td>.750</td>
</tr>
<tr>
<td>55</td>
<td>.667</td>
<td>.750</td>
<td>.727</td>
<td>.692</td>
<td>.708</td>
</tr>
<tr>
<td>56</td>
<td>.583</td>
<td>.833</td>
<td>.778</td>
<td>.667</td>
<td>.708</td>
</tr>
</tbody>
</table>

Sens=Sensitivity; Spec=Specificity; PPV=Positive Predictive Value; NPV=Negative Predictive Value;
$^a$Cutoff proposed by Cohen et al (2010), $^b$Optimal cutoff in current sample
Table 5: IQ > 70 subgroup: Selected PDDBI Autism Composite T-score Sensitivity, Specificity, PPV, NPV, and Efficiency

<table>
<thead>
<tr>
<th>Cutoff score</th>
<th>Sens</th>
<th>Spec</th>
<th>PPV</th>
<th>NPV</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>32&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.960</td>
<td>.160</td>
<td>.533</td>
<td>.800</td>
<td>.560</td>
</tr>
<tr>
<td>33</td>
<td>.960</td>
<td>.240</td>
<td>.558</td>
<td>.857</td>
<td>.600</td>
</tr>
<tr>
<td>35</td>
<td>.840</td>
<td>.280</td>
<td>.538</td>
<td>.636</td>
<td>.560</td>
</tr>
<tr>
<td>38</td>
<td>.680</td>
<td>.400</td>
<td>.531</td>
<td>.556</td>
<td>.540</td>
</tr>
<tr>
<td>42</td>
<td>.600</td>
<td>.520</td>
<td>.556</td>
<td>.565</td>
<td>.560</td>
</tr>
<tr>
<td>45</td>
<td>.600</td>
<td>.600</td>
<td>.600</td>
<td>.600</td>
<td>.600</td>
</tr>
<tr>
<td>46</td>
<td>.560</td>
<td>.600</td>
<td>.583</td>
<td>.577</td>
<td>.580</td>
</tr>
<tr>
<td>47&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.560</td>
<td>.720</td>
<td>.667</td>
<td>.621</td>
<td>.640</td>
</tr>
<tr>
<td>48</td>
<td>.480</td>
<td>.720</td>
<td>.632</td>
<td>.581</td>
<td>.600</td>
</tr>
<tr>
<td>52</td>
<td>.280</td>
<td>.800</td>
<td>.583</td>
<td>.526</td>
<td>.540</td>
</tr>
</tbody>
</table>

Sens=Sensitivity; Spec=Specificity; PPV=Positive Predictive Value; NPV=Negative Predictive Value;
<sup>a</sup>Cutoff proposed by Cohen et al (2010),<sup>b</sup>Optimal cutoff in current sample
Table 6: Coefficient Alpha for PDDBI Subscales and Composites

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Coefficient</th>
<th>N of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>SENSORY</td>
<td>.864</td>
<td>20</td>
</tr>
<tr>
<td>RITUAL</td>
<td>.879</td>
<td>12</td>
</tr>
<tr>
<td>SOCPP</td>
<td>.891</td>
<td>12</td>
</tr>
<tr>
<td>SEMPP</td>
<td>.842</td>
<td>12</td>
</tr>
<tr>
<td>AROUSE</td>
<td>.808</td>
<td>12</td>
</tr>
<tr>
<td>FEARS</td>
<td>.923</td>
<td>20</td>
</tr>
<tr>
<td>AGG</td>
<td>.928</td>
<td>20</td>
</tr>
<tr>
<td>REPRIT/C</td>
<td>.942</td>
<td>56</td>
</tr>
<tr>
<td>AWP/C</td>
<td>.969</td>
<td>108</td>
</tr>
<tr>
<td>SOCAPP</td>
<td>.935</td>
<td>36</td>
</tr>
<tr>
<td>EXPRESS</td>
<td>.975</td>
<td>32</td>
</tr>
<tr>
<td>LMRL</td>
<td>.892</td>
<td>12</td>
</tr>
<tr>
<td>EXSCA/C</td>
<td>.970</td>
<td>68</td>
</tr>
<tr>
<td>REXSCA/C</td>
<td>.973</td>
<td>80</td>
</tr>
</tbody>
</table>

SENSORY = Sensory/Perceptual Approach Behaviors; RITUAL = Ritualisms/Resistance to Change; SOCPP = Social Pragmatic Problems; SEMPP = Semantic/Pragmatic Problems; AROUSE = Arousal Regulation Problems; FEARS = Specific Fears; AGG = Aggressiveness; REPRIT/C = Repetitive, Ritualistic, and Pragmatic Problem Behaviors Composite; AWP/C = Approach/Withdrawal Problems Composite; SOCAPP = Social Approach Behaviors; EXPRESS = Expressive Language; LMRL = Learning, Memory and Receptive Language; EXSCA/C = Expressive Social Communication Abilities Composite; REXSCA/C = Receptive/Expressive Social Communication Abilities Composite
Figure 1: Histograms of Autism Composite T-scores for ASD ($n = 42$) and non-ASD ($n = 42$) groups


