5-2009

# Screening Preschoolers for Autism with Behavior Rating Scales

Amber Desiré Gross
*Western Kentucky University*, amber.brown791@wku.edu

# SCREENING PRESCHOOLERS FOR AUTISM WITH BEHAVIOR RATING SCALES

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, KY

In Partial Fulfillment
Of the Requirements of the Degree
Specialist in Education

By
Amber D. Gross

May 2009

SCREENING PRESCHOOLERS FOR AUTISM WITH
BEHAVIOR RATING SCALES

Date Recommended ____3/30/09_____

___Carl Myers_____
Director of Thesis

___Sandy A. Bowersox_____

___Lakeisha Meyer_____

_____
Dean, Graduate Studies and Research Date

Acknowledgements

There are many individuals whom I would like to thank for contributing to the success of this project as well as to my success in graduate school. First, I would like to thank Dr. Carl Myers, my thesis director, for all of the countless hours he has put in reading, editing, and conferencing with me to make this project great. His constructive criticism, positive feedback, and constant support have helped to make graduate school a positive experience for me. I also want to thank my other two amazing committee members, Dr. Lakeisha Meyer and Dr. Sandy Bowersox, for their helpful suggestions and excitement over this project. Most importantly, I would like to thank my husband Johnathan for his unwavering belief in me and my abilities as well as for the sacrifices he has been willing to make so I could pursue my career. I know I could not have made it through the last three years without him and his support! Lastly, I would like to thank my parents for always telling me "you can be anything you want as long as you put your mind to it." Hearing those words and believing them is how I made it here. Thank you all!

Table of Contents

List of Tables

Pages

SCREENING PRESCHOOLERS FOR AUTISM WITH
BEHAVIOR RATING SCALES

Amber Desiré Gross May 2009 58 pages

Directed by: Dr. Carl Myers, Dr. Sandy Bowersox, and Dr. Lakeisha Meyer

Department of Psychology Western Kentucky University

Early identification of autism is important in order to maximize the potential of intervention efforts. However, the identification of autism can require extensive training and experience. Psychologists routinely use behavior rating scales to assess children for various social, emotional, or behavioral problems when completing evaluations in the school or clinical setting. The focus of the current study is to determine whether a behavior rating scale can accurately distinguish between referred preschoolers with autism and referred preschoolers without autism. Parents of 82 preschoolers referred to a nonprofit child development clinic because of behavioral or developmental concerns completed two behavior rating scales as part of the initial evaluation. The findings revealed that while statistically significantly differences were found between the scores for the two groups on a number of scales, the practical implications were unclear. The use of behavior rating scales as screeners for autism in preschoolers would result in a very high false positive rate.

## Introduction

The description of autism and the criteria needed to make an accurate diagnosis have changed significantly over the past 66 years since the term was first used to describe the disorder. Those criteria have become more detailed and specific for the classification of Autistic Disorder but additional autism spectrum disorders (e.g., Aspergers, Pervasive Developmental Disorder – Not Otherwise Specified) have been included as well. With the broadening of the conceptualization of autism spectrum disorders, there has been an increase in the prevalence rate of autism. It is unclear how much of the increased prevalence rate is due to the expansion of the diagnostic criteria and how much is due to an actual increase of individuals with the disorder. Regardless of the reasons behind the increase, it has spurred awareness of the importance of early detection because the outcomes for individuals who are diagnosed early are more promising.

A full comprehensive evaluation is needed to provide an accurate diagnosis of autism. One part of that evaluation is using one or more autism diagnostic instruments. The two most commonly used and discussed in research are the Autism Diagnostic Observation Schedule-Generic (ADOS-G; Lord, Rutter, DiLavore, & Risi, 1999) and the Autism Diagnostic Interview-Revised (ADI-R, Rutter, LeCouteur, & Lord, 2003). Both instruments provide valuable information; however, the instruments take significant time and training to master as well as to administer. It would be impractical to administer these instruments to all children to assess for the possibility of autism. A possible solution to this problem has been to give autism screening instruments, which take less time and training to administer. The problem with using an autism screener is that the

possibility of autism must already be suspected. It would not be practical to administer an autism screener to all children.

Lord and Corsello (2005) suggested the possibility of using a behavior rating scale to screen for autism. Behavior rating scales often provide information on a wide range of problem behaviors, including autism. Furthermore, psychologists routinely use behavior ratings as part of their evaluations. However, there is very little research on the utility of behavior ratings scales as screeners for autism. It is the purpose of this research project to add to the current literature on this topic by examining the value of using the *Behavior Assessment System for Children, 2nd Edition* (BASC-2, Reynolds & Kamphaus, 2004) and the *Child Behavior Checklist* (CBCL, Achenbach & Rescorla, 2000) as screeners for autism. Parents of referred preschool children with and without autism completed both behavior rating scales. Scores for both groups of children were analyzed to determine the practical value of using the instruments as screeners for autism.

## Literature Review

### *The Diagnosis and Prevalence of Autism*

The term autism was first coined by Eugen Bleuler in 1911 (Volkmar & Klin, 2005). Bleuler used the term to describe idiosyncratic, self-centered thinking, which is very different from how autism is viewed today. According to Lovaas (1987), the individual to first recognize and describe autism as a syndrome, Leo Kanner, borrowed Bleuler's term in 1943. Kanner's description of autism included four characteristics: (a) lack of desire for social interactions, (b) acute communication disturbances, (c) resistance to environmental changes, and (d) normal intelligence. Despite Kanner's recognition and description of autism as a syndrome in the early 1940's, it was not formally defined by the American Psychiatric Association (APA, 1980) until 1980 in the publication of the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III). The diagnostic label used in the DSM-III was Infantile Autism, which fell under the broader category of Pervasive Developmental Disorders. The term "pervasive" was used to refer to an individual's problems with, or inability to develop basic skills involving, communication and interaction with others (APA, 1980). The diagnostic criteria for Infantile Autism provided in the DSM-III were limited. In fact, there were only six:

> A. Onset before 30 months of age. B. Pervasive lack of responsiveness to other people (autism). C. Gross deficits in language development. D. If speech is present, peculiar speech patterns such as immediate and delayed echolalia, metaphorical language, pronominal reversal. E. Bizarre responses to various aspects of the environment, e.g., resistance to change, peculiar interest in or attachments to animate or inanimate objects. F. Absence of delusions,

hallucinations, loosening of associations, and incoherence as in Schizophrenia. (APA, 1980, pp. 89-90)

In the span of seven years, which is the time from when the DSM-III was published to the time the DSM-III-R (APA, 1987) was published, the diagnostic criteria for Autism changed drastically. Instead of the phrase, "Infantile Autism," the DSM-III-R used the term, "Autistic Disorder." The diagnostic criteria were expanded significantly, providing more examples for those diagnosing the disorder. The diagnostic characteristics were divided into four categories: (a) qualitative impairment in social interactions, (b) qualitative impairment in communication (verbal/nonverbal), (c) restricted repertoire of activities and interests, and (d) onset during infancy or childhood (APA, 1987). Under the first three diagnostic areas, there were a total of 16 specific examples and criteria provided to assist in the diagnosis of an Autistic Disorder. In order for an individual to meet the criteria for an Autistic Disorder, eight of the 16 criteria must be present.

The DSM-IV (APA, 1994) made changes to the diagnostic criteria of Autistic Disorder, but they are not extensive changes. The primary diagnostic characteristics were divided into three areas: (a) qualitative impairment in social interaction, (b) qualitative impairments in communication, (c) and restricted repetitive and stereotyped patterns of behavior, interests, and activities. Within each area, there are four specific criteria and examples that assist in the diagnosis of an autistic disorder. An individual must meet six of the 12 criteria to be identified with an autistic disorder. Also, prior to age three an individual must show delays in one of the following areas: (a) social interaction, (b) language as used in social communication, or (c) symbolic or imaginative play. A final

diagnostic criterion added to the DSM-IV was the symptoms could "not be accounted for by Rett's Disorder or Childhood Disintegrative Disorder" (APA, 1994, p. 71). There were no changes made to the diagnostic criteria of an Autistic Disorder in the DSM-IV-TR (APA, 2000).

Autism prevalence rates have increased over the years. The reason for the increase is currently unknown; however, there is much debate about whether the increase is due to the change in diagnostic criteria or other environmental factors. The DSM-III (APA, 1980) described the disorder as "very rare" with a prevalence of two to four cases per 10,000. The DSM-IV-TR (APA, 2000) reported the median prevalence rate for autism in epidemiological studies was 1 in every 2,000 children. However, recent government reports show that the average rate of autism in the United States has risen to 1 in 150 children (Centers for Disease Control & Prevention [CDC] 2007). The rise in prevalence rates has increased awareness of the disorder among parents and professionals in the field as well as the need for services (Charman & Baird, 2002)

*The Importance of Early Diagnosis*

Because the prevalence and awareness of autism has increased, so has the need for research in the area of early diagnosis and intervention. The sooner a child and family are provided with support and assistance, the more promising the outcome for that child. The early intervention principle is especially important to young children who are diagnosed with autism. With children who have autism, waiting to begin interventions after the age of five can result in fewer positive outcomes than if the interventions were started prior to age three (Woods & Wetherby, 2003). The earlier a child with autism is identified and accurately diagnosed, the sooner effective interventions can be provided

(Smith & Dillenbeck, 2006; Stone et al., 1999; Woods & Wetherby, 2003). Unfortunately, most children are not diagnosed with autism until they are significantly older than the age of two (Coonrod & Stone, 2005). However, current research indicates the possibility of children to be accurately diagnosed with autism as young as 18 months to 2-years-old (Coonrod & Stone, 2005; Matson, Wilkins, & González, 2008; Smith & Dillenbeck, 2006). These findings are encouraging because most specialized and intensive intervention programs require a formal diagnosis to participate (Stone et al., 1999). Participation in such intensive programs can lead to dramatic improvements in cognitive, educational, and behavioral outcomes (Lovaas, 1987; Schreibman, 2000).

*Autism Diagnostic Assessment Instruments*

A multidisciplinary approach is preferred when evaluating a child for autism, which requires significant time, training, and expertise in the area of autism (Charman & Baird, 2002). Information regarding the child's cognitive, social, emotional, language, motor, and sensory functioning must be obtained and assessed by completing a comprehensive evaluation (CDC, 2007). A comprehensive evaluation for autism consists of several components such as direct observations of the child, interviews with parents and caretakers, psychological testing, and often the use of autism diagnostic instruments. There have been a growing number of diagnostic instruments developed to assist in the diagnosis of autism. The Autism Diagnostic Observation Schedule-Generic (ADOS-G, Lord, Rutter, DiLavore, & Risi, 1999) and the Autism Diagnostic Interview – Revised (ADI-R, Rutter, LeCouteur, & Lord, 2003) are regarded as well developed and the most respected autism diagnostic instruments (Fililpek et al., 1999; Tanguay, 2000).

According to Lord et al. (2000), the current version of the ADOS-G is a modification and combination of two earlier versions: the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1989) and the Pre-Linguistic Autism Diagnostic Observation Scale (PL-ADOS; DiLavore, Lord, & Rutter, 1995). The ADOS-G is a standardized instrument that uses a direct observation approach to the diagnosis of autism (Lord et al., 1999). Specifically, the evaluator observes a child's social and communicative behaviors during structured situations. This method of assessment is designed to elicit specific diagnostic characteristics of a child suspected of having an autism spectrum disorder. The instrument can be used with individuals who vary on their levels of development, language abilities, and age. The ADOS-G consists of four different modules and the child is given the module that fits his or developmental level and language skills. Each module takes about 30 to 40 minutes each to complete. Scores obtained from the ADOS-G can be used to determine autism or a broader diagnosis of pervasive developmental disorder (Lord et al., 1999).

The Autism Diagnostic Interview – Revised, developed by Rutter et al. (2003), is another standardized instrument used to assist in the diagnosis of individuals with autism. This assessment has a semi-structured interview format and is administered to the child's parent or caregiver. The interview questions are designed to obtain a thorough developmental history and description of the individual covering three domains: communication, shared social interactions, and repetitive stereotypic behaviors. The ADI-R can be used to evaluate individuals with a developmental age of two years up to adulthood. This assessment typically takes an experienced administrator two hours to complete. The standardization sample for the ADI-R consisted of 25 children with

autism and 25 non-autistic children who ranged in mental age from 21 to 74 months. It was determined that the diagnostic algorithms developed for the ADI-R were successful at discriminating between children with and without autism spectrum disorders (Rutter et al., 2003).

Each instrument used for the identification of autism provides valuable information that is crucial for an accurate diagnosis. Some authors suggest using more than one diagnostic instrument (e.g., ADOS-G and ADI-R) when conducting a thorough evaluation (Lord & Corsello, 2005; Tanguay, 2000; Tomanik, Pearson, Loveland, Lane, & Shaw, 2007). However, because some of the instruments are more specialized, individuals administering the instruments are required to have considerable training and familiarity with autism spectrum disorders (Filipek et al., 1999; Lord & Corsello, 2005). Also, the instruments take a significant amount of time to administer, making it unfeasible to administer those instruments to each and every child. Thus, the possibility of autism would have to be recognized prior to using a specialized diagnostic instrument. Preferably, some type of screening instrument could be utilized to prompt a more comprehensive, specialized evaluation.

*Autism Screeners*

Screenings are brief assessments that can provide valuable information that is useful in determining if there are concerns that warrant completing a more comprehensive diagnostic assessment (Coonrod & Stone, 2005). They do not require a lot of training or experience to administer, and they take significantly less time than most diagnostic instruments. When screening for autism, there are two different levels of screenings (Filipek et al., 1999). Level one screenings would be used on all children to

determine if any type of developmental delay exists. Level two screenings would be used to discriminate between children with autism and those with other developmental disorders (Coonrod & Stone, 2005; Filipek et al., 1999). Some of the autism specific level two and level one screeners that will be discussed for this project are the Social Responsiveness Scale (SRS), Childhood Autism Rating Scale (CARS), Gilliam Autism Rating Scale (GARS), Social Communication Questionnaire (SCQ), Screening Tool for Autism in Two-Year Olds (STAT), Autism Behavior Checklist (ABC), Checklist for Autism in Toddlers (CHAT), and the Modified Checklist for Autism in Toddlers (M-CHAT). These autism screening instruments were selected based on their ability to fit the description of level one and level two screening measures for autism.

The Social Responsiveness Scale (SRS, Constantino & Gruber, 2005) was designed to be useful in screening and diagnosing individuals ages 4 to 18 with autism spectrum disorders. The instrument assesses the individual's social interactions, communication skills, and repetitive behaviors that are common characteristics of autism. The authors claim the SRS is sensitive enough to estimate the severity of autism. The normative sample for the SRS consisted of 1636 individuals ages 4 to 18. The instrument consists of 65 items that are rated by a parent or teacher on a 4-point Likert scale (i.e., not true, sometimes true, often true, almost always true). The completion and scoring time for the SRS is between 20 and 30 minutes. Because the SRS is relatively new, little research has been published about it. Conway (2007) noted that although the test developers provided sufficient support for the psychometric integrity of the SRS, they did not provide adequate information on the item selection process.

The Childhood Autism Rating Scale (CARS, Schopler, Reichler, & Renner, 1988) was developed to help screen children over the age of two for the possibility of autism. The resulting scores indicate whether the child falls in the normal, mild to moderate or severe autism range. The CARS was developed over a 15-year period and used a normative sample of 1,500 children. It consists of 15 items that are rated by an examiner on a seven-point scale (i.e., 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0). The examiner can complete these ratings by directly observing the child, talking to the parent, or reviewing any of the child's records. It is important that the examiner know the scoring criteria on all items before making observations. Children who obtain a score below 30 fall in the nonautistic range, while those with a score above 30 are considered to be in the autistic range. Children with score ranging from 30-36.5 would be considered mild to moderate. Any child with a score of 37 or higher would be considered severe. Although extensive reliability and validity scores are reported in the CARS manual, they are slightly dated and should be updated in future revisions (Welsh, 1992). The CARS is also based on out-of-date classification information from the DSM-III, meaning children are not being assessed on the current conceptualization of autism spectrum disorder (Magyar & Pandolfi, 2007)

The Gilliam Autism Rating Scale (GARS, Gilliam, 1995) was developed to assist in the screening and diagnosis of autism in individual ages 3 to 22. The GARS was normed on a sample of 1,092 individuals with autism. The GARS has 42 items that are rated 0 to 3 based on the frequency of the behavior. There are 14 additional items about the child's development prior to age three that are rated either yes or no. The GARS is rated by either the parent or teacher. The entire assessment takes an estimated 10 minutes

to complete. The GARS provides both standard scores and percentile ranks. A child obtaining an Autism Quotient score of 69 or below has a very low probability of autism. However, as the score increases, so does the probability of autism. For example, a child with a score of 123 has a high probability of autism. According to Sikora, Hall, Hartley, Gerrard-Morris, and Cagle (2008), the GARS was unable to differentiate between those children who had received Autism classification and children who received Non-spectrum classification, which could result in obtaining more false negative results.

The Social Communication Questionnaire (SCQ, Rutter, Bailey, & Lord, 2003) was designed to estimate the social and communication functioning of children suspected of having an autism spectrum disorder. The SCQ consists of two different forms (Lifetime & Current) comprised of 40 items each. The items used on the SCQ were designed to be closely aligned with the Autism Diagnostic Interview – Revised. The Lifetime form focuses on the child's developmental history as a whole, while the Current form concentrates on the previous three months. The forms take about 10 minutes to complete and can be used to assess individuals four years and older with a mental age greater than two years. Although the original standardization sample included individuals as young as four years of age, some studies have indicated that lowering cutoff scores can actually improve sensitivity toward screening younger individuals with the SCQ (Allen, Silove, Williams, & Hutchins, 2007; Corsello et al., 2007; Wiggins, Bakeman, Adamson, & Robins, 2007). However, additional studies with lower cutoff scores and younger participants should be conducted to verify these findings.

The Screening Tool for Autism in Two-Year Olds (STAT, Stone & Ousley, 1997) was developed to assist in the early identification of autism in children 24 to 36 months

of age. The STAT uses 12 interactive activities in the areas of play, imitation, and communication to elicit and observe autism characteristics. The STAT's interactive nature also allows the service provider to see behaviors firsthand instead of relying on parental reports (Stone, Coonrod, Turner, & Pozdol, 2004). Each item is scored pass/fail depending on the provided criteria and the entire screener takes about 20 minutes to administer. The total STAT score is compiled by adding the scores from each of the four domains together. Higher scores indicate a greater likelihood of autism. One study indicated the STAT has strong psychometric properties and is capable of being a good level two screening for autism (Stone et al., 2004). However, the psychometric properties of the STAT have only been tested in clinic-based settings instead of the community-based settings for which it was designed. Further research was recommended in this area (Coonrod & Stone, 2005; Stone et al., 2004).

The Autism Behavior Checklist (ABC, Krug, Arick, & Almond, 1980) is an autism behavior rating scale that is designed to assist in identifying individuals 18 months to 35 years with autism. The ABC consists of 57 items that assess autism characteristics in the following areas: sensory, relating, body and object use, language, and social and self-help. The items on the ABC were formulated by researching existing autism screenings instruments. The authors selected the final items with the assistance of additional autism professionals (Coonrod & Stone, 2005). The ABC takes about 10 to 20 minutes for the parent or caregiver to complete. Once completed, a trained professional scores and interprets the results. All items on the ABC are scored (1 - 4) and only those endorsed items are calculated to obtain overall scores. Individuals obtaining a score of 67 or higher have a high possibility of autism, while individuals obtaining scores of 53

and lower have a low possibility of autism. One study found that the ABC could accurately identify 81.6% of children within their sample of children with autism (Marteleto & Pedromônico, 2005). However, those authors lowered the cutoff score to 49 to obtain those results.

The Checklist for Autism in Toddlers (CHAT, Coonrod & Stone, 2005) was designed to be a level one screening tool used by healthcare providers during the 18-month health checkup to determine if a child was at risk for autism. The CHAT consists of nine yes/no questions for the parent and five interactive items where the healthcare provider observes the child directly. Out of all 14 items, there are five key items on the CHAT that determine the child's risk for autism (high risk, medium risk, suspected). Studies indicate that the CHAT is an adequate screening tool for distinguishing children with autism from the normal population, but is not sufficient in making finer distinctions between those children who fall within the autism spectrum (Filipek et al., 1999; Gillberg, Nordin, & Ehlers, 1996). The Modified Checklist for Autism in Toddlers (M-CHAT, Coonrod & Stone, 2005) eliminated the interactive section for the healthcare provider to make observations and relies completely on parental report of behaviors. The M-CHAT consists of 23 items total, nine of which are the same from the original CHAT. The other 14 items cover additional autism characteristics. A child is considered at risk for autism on the M-CHAT if the child fails any three items on the entire checklist or fails two of the six critical items (Coonrod & Stone, 2005; Ventola et al., 2007). The M-CHAT shows potential as a screening tool for autism, but requires further research on its psychometric properties (Coonrod & Stone, 2005; Dumont-Mathieu & Fein, 2005).

*Screening for Autism with Behavior Rating Scales*

While numerous autism screeners already exist, they vary greatly in terms of quality and usefulness. Furthermore, someone still needs to recognize the child as potentially having autism in order to select an autism screener. One possible solution for practitioners could be the use of a behavior rating scale as a screener for autism. Behavior rating scales assess a wide range of problem behaviors and often include items assessing autistic-like behaviors. Behavior rating scales are similar to the autism screening instruments described earlier in that they do not require a significant amount of training and they are quick to administer. In addition, practitioners already use behavior rating scales routinely as part of their evaluations on referred children.

Behavior ratings scales often contain dozens of brief statements describing various problem behaviors and the person completing the scale indicates the applicability of that behavior to the child. The person completing the scale, usually a parent or teacher, must know the child well. Generally, ratings of applicability of each behavioral descriptor are based on the frequency of the behavior (e.g., rarely, sometimes, often, frequently). Scores for various domains of behavior are provided based on normative samples. In this manner, "behavior rating scales provide a standardized format for the development of summary judgments" about the child's level of problem behaviors in the home and/or school setting (Merrell, 2008, p. 97). Rating scales capitalize on obtaining information from people (e.g., parents or teachers) who are very familiar with the child over a long period of time in the child's natural environments of home and school (Merrell, 2008).

Although no empirical support was provided, Lord and Corsello (2005) suggested behavior rating scales have potential value as screeners for autism. It appears only two

studies have examined the use of behavior rating scales as screeners for children with autism. Duarte, Bordin, Oliveira, and Bird (2003) conducted a study with the purpose of assessing the validity of the school-age version of the CBCL in identifying children with autism. The study consisted of 101 participants divided into three different groups of children: (a) 36 children with autism, (b) 31 children with other psychiatric disorders (OPD), and (c) 34 non-referred schoolchildren used as a control group. Participants for the autism and OPD groups were selected from mental health clinics and matched based on age and gender. The schoolchildren were chosen through random selections from two public schools close to the mental health clinics. The children's ages ranged from 4 to 11. The majority of the participants were male. Duarte et al. (2003) used the Brazilian adaptation of the previous edition of the CBCL/4-18 (Achenbach, 1991) for their study. Experienced psychologists and psychiatrists completed evaluations to determined autism and OPD diagnoses using the ICD-10 criteria. The data collected were analyzed using factor analysis and logistic regression.

When the autism and schoolchildren groups were compared, the Thought Problems and Autistic/Bizarre scales were the scales that best distinguished between the two groups. "Comparing autistic children with OPD children, the Thought Problems, Autistic/Bizarre, and Aggressive Behavior scales, taken individually, distinguished autistic from OPD children" (Duarte et al., 2003, p. 705). However, the combination of the Autistic/Bizarre and Aggressive Behavior scales were found to distinguish between the autism and OPD groups the best of all. Sensitivity and specificity were also calculated for the scales that were determined to best differentiate between the groups. For the autism versus schoolchildren groups, the Thought Problems scale resulted in a

sensitivity of 94.3% and a specificity of 100%. On the Autistic/Bizarre scale, the sensitivity was 94.3% and the specificity was 94.1%. For the autism versus OPD groups, the Aggressive Behavior plus Autistic/Bizarre scales resulted in a sensitivity of 91.4% and a specificity of 96.7%. Duarte et al. (2003) concluded that their findings did provide some beginning support for the validity of the CBCL/4-18 to identify autism among Brazilian children.

Sikora et al. (2008) conducted a study to determine if a behavior rating scale, the CBCL, could be as clinically useful as an autism specific screening measure like the Gilliam Autism Rating Scale. Sikora et al. hypothesized that the CBCL would be as useful in screening for autism as the GARS, if not better. The study consisted of 147 children with an age range of 36-71 months ($M$ = 53.54). The sample contained more boys (109) than girls (38) and was primarily Caucasian (77.6%), with 6% African American, 9% Asian American/Pacific Islander, 6% Hispanic, and 1.5% being other ethnicities. The children were evaluated by the Autism Program at the Child Development and Rehabilitation Center (CDRC) at a university in Oregon. Those evaluations occurred between August 2003 and June 2005. The entire evaluation process consisted of the following measures: Autism Diagnostic Observation Scale-Generic (ADOG-G), Autism Screening Instrument for Educational Planning-Second Edition (ASIEP-2), Vineland Adaptive Behavior Scales (VABS), Gilliam Autism Rating Scale (GARS) and Child Behavior Checklist preschool version (CBCL/1½-5). Once evaluated, the children were divided up into three groups based on ADOS-G classification: Autism (79 children), Autism Spectrum Disorder (ASD, 18 children), and Non-spectrum (50 children).

All of the children in the study were referred by their primary care physician to the CDRC for an evaluation (Sikora et al., 2008). They were typically brought to the clinic by their parents, although sometimes by a foster parent or caseworker. The caregivers were given a comprehensive interview and several forms to complete (e.g., VABS, GARS, CBCL). Once the forms were completed, they were collected and scored. During the same appointment, the ADOS-G was given by two licensed clinicians and scored immediately. Results of the behavior rating scales were not reviewed until after the ADOS-G was administered. Once all the data had been collected and entered, statistical analysis of the GARS AQ and the DSM-oriented and syndrome scales of the CBCL were completed (Sikora et al., 2008). As a way to reduce bias Sikora et al. decided to use the ADOS-G classifications for grouping variables instead of exiting diagnoses.

According to Sikora et al. (2008), results of the Chi-square analysis yielded no significant sex differences among the groups, while the one-way ANOVA analysis did yield a significant difference with age. The Non-spectrum group was found to be older than the Autism and ASD groups. Pearson correlations determined that all scales on the CBCL and the GARS AQ were positively correlated with the strongest correlations being with the Withdrawn and Pervasive Developmental Problems scales. As previously mentioned, the GARS AQ was not found to be useful in distinguishing between the groups of children in this study. When comparing children with autism to referred children without autism, the mean scores on the CBCL/1½-5 Withdrawn and Pervasive Developmental Problems scales were statistically significantly different between the two groups. For the CBCL Withdrawn scale, the sensitivity was 64.56% and specificity was

62.0%. For the Pervasive Developmental Problems scale, the sensitivity was 79.75% and the specificity was 42.0%. Based on the results, Sikora et al. concluded that a behavior rating scale (CBCL) does appear to be as good and better than an autism specific (GARS) screening instrument.

*Purpose*

Behavior rating scales are commonly used by psychologists as a way to assess a wide variety of problem behaviors. Behavior rating scales are able to distinguish between referred and nonreferred children on a variety of clinical domains (Achenbach & Rescorla, 2000; Reynolds & Kamphaus, 2004). However, it is not clear how well a behavior rating scale can distinguish between a child with autism and other referred children. Two studies have examined the ability of the CBCL to distinguish between children with autism and referred, but non-autistic children. Duarte et al. (2003) found a combination of the Autistic/Bizarre and Aggressive Behavior scales was useful in distinguishing between the two groups. However, they used a Brazilian adaptation of the outdated school-age version of the CBCL. Thus, their results cannot be generalized to the current version of the CBCL with children from the United States.

Sikora et al. (2008) found the Withdrawn and Pervasive Developmental Problems scales on the CBCL/1½-5 to be useful in distinguishing between children with autism and referred, but non-autistic children. However, the difference between the mean scores on those two scales was not that large between the two groups. The difference was 7.37 T score points for the Withdrawn scale and 4.96 T score points for the Pervasive Developmental Problems scale. Furthermore, the mean scores were rated as at least borderline clinically significant (a T score $\geq$ 65) for both groups (i.e., children with

autism and referred, but non-autistic children). For example, the mean T score on the Pervasive Developmental Problems scale for children with autism was 75.04, while the mean T score for the referred, but non-autistic group of children was 70.09. A T score of 70 or more is considered clinically significant on the CBCL. Thus, such a difference may be statistically different, but the difference is not clinically meaningful on a practical level.

The purpose of this study is to determine if a behavior rating scale can be useful as a screener for autism in referred preschool aged children. Although several behavior rating scales have been developed for use with preschoolers, two particular instruments of interest to this research are the preschool parent versions of the *Behavior Assessment System for Children, 2nd Edition* (BASC-2, Reynolds & Kamphaus, 2004) and the *Child Behavior Checklist* (CBCL, Achenbach & Rescorla, 2000). These instruments were selected because versions of both instruments are very popular for school-age children. Feil, Severson, and Walker (2002) indicated that the CBCL has become the model rating scale in measuring child and adolescent social and emotional behavior. Additionally, Merrell (2008) noted the BASC-2 represents "the best of the newer generation of behavior rating scales" (p. 114). Although these two instruments are highly regarded at the school-age level, less is known about the preschool versions (Merrell, 2008).

There will be two research questions to address the purpose of this study.

1. Are there specific scales on the preschool parent versions of the BASC-2 or CBCL that distinguish between preschoolers with autism and other clinically referred children without autism? Based on the characteristics of children with autism, it is hypothesized that the BASC-2 scales of Atypicality and Withdrawal would result in

higher scores for the ASD group than the Non-Spectrum group. It is also hypothesized that the BASC-2 scales of Adaptability, Social Skills, Functional Communication, and Adaptive Skills would result in lower scores for the ASD group than the Non-Spectrum group. Sikora et al. (2008), found significant differences between ASD and Non-Spectrum groups on the CBCL scales of Withdrawn and Pervasive Developmental Problems. Based on these results, it is hypothesized that the CBCL scales of Withdrawn and Pervasive Developmental Problems would results in higher scores for the ASD group than the Non-Spectrum group.

2. If there are specific scales on either instrument that distinguish between preschoolers with and without autism, what are the sensitivity, specificity, positive predictive value, and negative predictive value rates of each scale?

## Method

*Participants*

Previously, Bour (2008) examined the consistency of parent ratings on the BASC-2 PRS-P and the CBCL/1.5-5 with a sample of 95 children who had been referred for a behavioral or developmental evaluation at a nonprofit child development clinic. The intent of this research was to reevaluate Bour's data and compare scores of children with autism to referred children without autism. All but two of the referred children in Bour's sample were determined eligible for early intervention services. At the preschool level, children do not have to be classified as having a specific developmental disability in order to be eligible for services. Eligibility is simply defined as being delayed at least two standard deviations below the mean in at least one developmental area or at least 1.5 standard deviations below the mean in at least two developmental areas. There are five developmental areas for eligibility purposes: cognitive, motor, communication, social-emotional, and adaptive behavior. Thus, the majority of children in Bour's sample did not receive any specific diagnosis. However, there were 36 children identified as having an Autism Spectrum Disorder (ASD). Another 13 children were not formally diagnosed as having an ASD, but showed some autistic-like characteristics that made such a diagnosis possible. Because that group of children were not formally diagnosed, but yet showed some characteristics of autism, it was decided to exclude those 13 children from this study. The exclusion of those 13 children left 46 children in the referred, but Non-Spectrum group.

All evaluations at the nonprofit child development clinic were conducted by a Ph.D. level psychologist with over 20 years of experience in the early childhood field and

with children with autism. The diagnoses of ASD were based on professional clinical judgment (based on parent interviews, observations, and interactions with the children) and the results from the Autism Diagnostic Interview – Revised (ADI-R, Rutter et al., 2003). Sikora et al. (2008) stated that parent interview, clinical judgment, and the use of a specialized autism assessment instrument are the "gold standard" for diagnosing autism. Constantino et al. (2003) noted that the ADI-R "is widely recognized as a gold standard parent-report interview for establishing a clinical diagnosis of autism" (p. 430).

Basic demographic information regarding the ASD and Non-Spectrum groups was collected and has been displayed in Table 1. Both the ASD and Non-Spectrum groups were comparable in regards to gender and ethnicity. The majority of participants in both groups were boys and were predominately rated by their mothers. The average age of the participants in the ASD group was 4.2 months higher than those participants in the Non-Spectrum group. However, an independent samples *t*-test was used to evaluate the age difference between the groups and it was found not to be significantly different ($t(80) = 1.94$, $p > .05$). Although the Non-Spectrum group had more parents with a high school education or less, it was most likely due to the higher number of Non-Spectrum participants. According to the United States Census Bureau (U.S. Census Bureau, 2009), the demographic data collected in this study are similar to the general population in Kentucky in regards to ethnicity (Caucasian = 90.2%, African American = 7.5%, Hispanic = 2.0%, Asian = 1.0%). The United States Census Bureau data indicated that 74.1% of individuals living in Kentucky have a high school diploma or less, while 17.1% have college degrees. Thus, our sample of parent/guardian raters was, on average, better educated than typical Kentuckians.

Table 1

*Participant Characteristics of the Autism Spectrum Disorder (ASD) and Non-Spectrum Groups*

| | ASD | Non-Spectrum |
|---|---|---|
| Gender | | |
| Males | 31 (86.1%) | 38 (82.6%) |
| Females | 5 (13.9%) | 8 (17.4%) |
| Age | | |
| Mean | 36.8 | 32.6 |
| SD | 11.9 | 7.9 |
| Ethnicity | | |
| Caucasian | 30 (83.3%) | 37 (80.4%) |
| African American | 5 (13.9%) | 6 (13.0%) |
| Hispanic | 1 (2.8%) | 2 (4.3%) |
| Asian | 0 (0.0%) | 1 (2.2%) |
| Rater of Child | | |
| Mother | 32 (88.9%) | 34 (73.9%) |
| Father | 3 (8.3%) | 4 (8.7%) |
| Female Guardian | 1 (2.8%) | 8 (17.4%) |
| Parent Education | | |
| ≤ High School | 18 (50.0%) | 28 (60.9%) |
| ≥ Some College | 18 (50.0%) | 18 (39.1%) |

*Instruments*

*Behavior Assessment System for Children-Second Edition.* The Behavior Assessment System for Children, second edition (BASC-2, Reynolds & Kamphaus, 2004) is a revision of the 1992 version of the BASC (Reynolds & Kamphaus, 1992). The BASC-2 consists of five different components that can be used separately or in combination with one another. Those components consist of a teacher rating scale (Teacher Rating Scale), a parent rating scale (Parent Rating Scale, PRS), a self-report scale (Self-Report of Personality), a developmental history form (Structured Developmental History), and a classroom observation form (Student Observation System). Both the teacher and parent rating scales have different forms for ages 2 to 5, 6 to 11, and 12 to 21. The focus of this investigation is on the BASC-2 Parent Rating Scale-Preschool (BASC-2 PRS-P); therefore, only that form will be reviewed.

The BASC-2 PRS-P is a standardized rating assessment of a child's behavioral and adaptive functioning within both the home and community setting (Reynolds & Kamphaus, 2004). The form consists of 134 items that are rated by the parent on a four-point scale: Never, Sometimes, Often, and Always. The BASC-2 PRS-P provides the examiner with standard scores for "clinical" scales, "adaptive" scales, and composites. The BASC-2 PRS-P contains eight different clinical scales (i.e., Aggression, Anxiety, Attention Problems, Atypicality, Depression, Hyperactivity, Somatization, and Withdrawal) and four different adaptive scales (i.e., Activities of Daily Living, Adaptability, Functional Communication, and Social Skills). The instrument provides T scores ($M$ = 50, $SD$ = 10) for the clinical scales, adaptive scales, and composites, which are based on totals of raw scores for multiple items comprising specific scales. The

manual states it only takes about 10 to 15 minutes to complete (Reynolds & Kamphaus, 2004).

According to Reynolds and Kamphaus (2004), the BASC-2 demonstrates criterion-related validity by distinguishing between referred and non-referred groups of children. For the PRS-P, construct validity was established through comparisons to the original BASC and to the CBCL/1½-5. Correlations between corresponding scales on the BASC and the BASC-2 were extremely high, with most above .90. Comparisons between the BASC-2 PRS-P and the CBCL/1½-5 with a non-referred sample of 53 children resulted in correlations for corresponding scales ranging from .65 to .84.

BASC-2 PRS-P provides internal consistency, test-retest and interrater reliability coefficients (Reynolds & Kamphaus, 2004). The internal consistency reliability coefficient determines how consistent the results are across items within the same test. Coefficient alpha reliabilities for the general norm sample on the clinical scales ranged from .77 to .86, the adaptive scales ranged from .77 to .85, and the composites range from .85 to .93. The test-retest reliability is determined by giving the same person the instrument twice and correlating their scores. The range of test-retest reliability coefficients across all scales of the BASC-2 PRS-P was .66 to .88 (median of 40 days between ratings). The only scale that did not have a test-retest reliability coefficient in the .70s or .80s was Depression (.66). The test-retest reliability coefficients for the composites ranged from .79 to .86. The interrater reliability looks at how much different raters (i.e., mother and fathers) agree. The interrater reliability for all of the scales of the BASC-2 PRS-P was .56 to .90. Most of the interrater reliability coefficients for both clinical and adaptive scales fell in the .70s and .80s except for Aggression (.59), Anxiety

(.56), Social Skills (.64), and Functional Communication (.90). The interrater reliability coefficients for the composites ranged from .70 to .87.

*Child Behavior Checklist.* The Child Behavior Checklist for ages 1½ to 5 years-old (CBCL/1½-5, Achenbach & Rescorla, 2000) is a popular part of the broader Achenbach System of Empirically Based Assessment. The CBCL/1½-5 is a revision of the 1992 version, the CBCL/2-3 (Achenbach, 1992). The CBCL/2-3 was developed for use on children ages 2-3 and another version of the CBCL was for ages 4-18. The current CBCL/1½-5 was revised to encompass a larger age range of preschoolers while the school age form is now for ages 6-18. The CBCL/1½-5 is considered a broadband behavior rating scale that looks at a wide range of behaviors and syndromes found in young children. The CBCL for preschoolers has two versions, one for parents and one for teachers. The focus of this investigation is on the parent version of the CBCL/1½-5; therefore, only that form will be reviewed.

The CBCL/1½-5 is a standardized rating assessment of a child's behavioral, emotional, and social functioning (Achenbach & Rescorla, 2000). The form consists of 99 items that are rated 0 (Not True), 1 (Sometimes True), or 2 (Often True). A Language Development Survey (LDS) is an additional feature of the CBCL/1½-5 that provides information about the child's vocabulary and word combination abilities, but will not be reviewed in this project. The CBCL/1½-5 provides the examiner with standard scores for "syndrome" scales as well as "DSM-oriented" scales. The CBCL/1½-5 contains seven different syndrome scales (i.e., Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, and Aggressive Behavior) and five different DSM-Oriented scales (i.e., Affective Problems, Anxiety Problems,

Pervasive Developmental Problems, Attention Deficit Hyperactivity Problems, and Oppositional Defiant Problems). The instrument provides T scores ($M$ = 50, $SD$ = 10) for both the syndrome and DSM-oriented scales, which are based on totals of raw scores for multiple items comprising specific scales. The manual states it only takes about 10 to 15 minutes to complete (Achenbach & Rescorla, 2000).

The CBCL/1½-5 manual provides criterion-related validity and construct validity (Achenbach & Rescorla, 2000). The criterion-related validity of the CBCL/1½-5 is supported by the instrument's ability to discriminate between referred and non-referred children according to DSM diagnoses. Construct validity refers to how well scores on the CBCL/1½-5 correlate with scores from another test measuring the same construct (convergent) or how scores on the CBCL/1½-5 are weakly correlated with scales on an another instrument measuring different constructs (discriminant). According to the manual, construct validity of the CBCL/1½-5 is supported by strong correlations between the CBCL/1½-5 and similar constructs on instruments such as the Richman Behavior Checklist, the Toddler Behavior Screening Inventory, and the Infant-Toddler Social and Emotional Assessment. Such instruments, however, do not appear to be commonly used in clinical practice. In particular, comparisons to the Richman scale seem questionable as it was developed in England in 1977.

The CBCL/1½-5 provides test-retest, cross-informant agreement, and internal consistency reliability coefficients (Achenbach & Rescorla, 2000). The test-retest reliability is determined by giving the same person the instrument twice and correlating his/her scores. The range of test-retest reliability coefficients across all scales of the CBCL/1½-5 was .68 to .92 (mean interval of 8 days). The only four scales that did not

have test-retest reliability coefficients in the .80s and .90s were Anxious/Depressed (.68), Attention Problems (.78), Affective Problems (.79), and Attention-Deficit/Hyperactivity Problems (.74). Cross-informant agreement looks at how much different raters (i.e., mothers and fathers) agree. The CBCL/1½-5 found that mothers' and fathers' mean inter-parent agreement was .61. The internal consistency reliability coefficients (Cronbach's Alpha) determine how consistent the results are across items within the same test. The coefficients ranged from .66 to .92 on the syndrome scales, from .63 to .86 on the DSM-oriented scales, and from .89 to .95 on the composite scales.

*Procedure*

According to Bour (2008), data were collected by the staff at the nonprofit child development clinic. The staff members were able to pull client files where the CBCL/1½-5 and the BASC-2 PRS-P were completed by parents or guardians of referred preschool-age children. After deleting the child's name, staff members made copies of each score sheet. Basic information (i.e., age, gender, ethnicity, diagnosis, parents' education level) was recorded on a demographic form and stapled to the corresponding score sheet. This enabled the investigator to obtain necessary information for the study, while maintaining the confidentiality of the participants and their families (Bour, 2008). Both the CBCL/1½-5 and the BASC-2 PRS-P were computer scored using the test publisher's software. To enhance comparability, both rating scales were scored using gender-specific norms. Permission to complete additional data analyses on Bour's data set was granted by Western Kentucky University's Human Subjects Review Board (see Appendix).

## Results

The first research question examined whether there were specific scales on the preschool parent versions of the BASC-2 or CBCL that distinguished between preschoolers with autism and other clinically referred children without autism. To address the first question, mean scores were determined for each scale on the BASC-2 PRS-P and the CBCL/1½-5 for each group (autistic and referred but non-spectrum). The comparability of scores on the BASC-2 and CBCL scales between the two groups were evaluated through a series of *t*-tests.

Table 2 displays the means of the BASC-2 scales for the Autism Spectrum Disorder (ASD) and the Non-Spectrum groups. A series of *t*-tests were used to test for significance using an a priori significance level of $p < .01$. A more stringent method of minimizing the possibility of a Type I error is to use the Bonferroni *t* test. With 16 comparisons, the use of the Bonferroni *t* test would result in a significance level of $p = .003$. However, for informational purposes, all significant results at the $p < .01$ level are shown in Table 2. Several significant differences were found between the ASD and the Non-Spectrum groups for the following BASC-2 scales: Hyperactivity, Aggression, Anxiety, Depression, Social Skills, Functional Communication, Externalizing, and Internalizing. On all scales where significant differences were found, the Non-Spectrum group obtained higher mean scores than the ASD group. For most of the scales, a higher score indicates more problematic behaviors. However, the Social Skills and Functional Communication scales are considered positive or adaptive types of scales where lower scores indicate fewer adaptive behaviors. Thus, the BASC-2 results indicate children with ASD have fewer problematic behaviors related to Hyperactivity, Aggression,

Table 2

*Mean T Scores on BASC-2 PRS-P Scales for ASD and Non-Spectrum Groups*

| BASC-2 Scale | ASD | Non-Spectrum | *t* values |
|---|---|---|---|
| Hyperactivity | 59.8 | 68.5 | -3.0* |
| Aggression | 50.2 | 64.2 | -4.3** |
| Anxiety | 42.9 | 49.2 | -2.8* |
| Depression | 54.0 | 64.1 | -2.9* |
| Somatization | 47.8 | 50.3 | -1.1 |
| Atypicality | 71.4 | 65.2 | 1.9 |
| Withdrawal | 61.1 | 54.9 | 2.4 |
| Attention Problems | 64.6 | 64.4 | 0.1 |
| Adaptability | 41.1 | 38.8 | 1.1 |
| Social Skills | 33.2 | 39.4 | -3.8** |
| Activities of Daily Living | 36.6 | 41.9 | -2.2 |
| Functional Communication | 33.9 | 37.2 | -2.8* |
| Adaptive Skills | 32.4 | 36.5 | -2.4 |
| Externalizing | 55.5 | 67.9 | -3.9** |
| Internalizing | 47.6 | 56.0 | -3.0* |
| Behavioral Symptoms Index | 63.8 | 68.4 | -1.5 |

*$p < .01$.  **$p < .003$.

Anxiety, Depression, Externalizing and Internalizing than referred but non-spectrum children. It was hypothesized that children with ASD would have fewer Social Skills, Adaptability, Adaptive Skills and Functional Communication skills than other referred children. This hypothesis was partially supported. Children with ASD did have significantly lower Social Skills and Functional Communication skills but did not have lower Adaptability and Adaptive Skills. Surprisingly, there were no significant differences between the ASD and Non-Spectrum groups on the Atypicality or Withdrawal scales, as hypothesized.

Table 3 displays the means of the CBCL scales for the Autism Spectrum Disorder (ASD) group and the Non-Spectrum group. Again, a series of *t*-tests were used to test for significance using an a priori significance level of $p < .01$. Using the more stringent Bonferroni *t* test, the 15 comparisons would also result in a significance level of $p = .003$. However, all significant results at the $p < .01$ level are shown in Table 3. There were significant differences found among the ASD and the Non-Spectrum groups for the following CBCL scales: Withdrawn, Aggressive Behavior, and Pervasive Developmental Problems. As hypothesized, those participants in the ASD group had higher means scores on the Withdrawn and Pervasive Developmental Problems scales than those in the Non-Spectrum group. One additional significant difference between the two groups was that the mean score on the Aggressive Behavior scale was higher for the Non-Spectrum group than the ASD group.

Table 3

*Mean T Scores on CBCL/1½-5 Scales for ASD and Non-Spectrum Groups*

| CBCL1/½-5 Scale | ASD | Non-Spectrum | *t* values |
|---|---|---|---|
| Emotionally Reactive | 62.7 | 64.8 | -0.9 |
| Anxious/Depressed | 54.8 | 59.2 | -2.5 |
| Somatic Complaints | 59.1 | 58.7 | 0.2 |
| Withdrawn | 75.0 | 63.9 | 4.9** |
| Sleep Problems | 59.2 | 63.6 | -1.5 |
| Attention Problems | 63.9 | 65.8 | -0.9 |
| Aggressive Behavior | 63.8 | 72.8 | -2.8* |
| Affective Problems | 63.3 | 62.9 | 0.2 |
| Anxiety Problems | 57.1 | 60.5 | -1.7 |
| Pervasive Developmental Problems | 76.2 | 68.7 | 3.4** |
| Attention Deficit Hyperactivity Problems | 62.2 | 65.9 | -1.8 |
| Oppositional Defiant Problems | 62.6 | 67.3 | -2.0 |
| Internalizing | 64.4 | 62.0 | 1.1 |
| Externalizing | 63.5 | 71.0 | -2.6 |
| Total Problems | 65.9 | 68.0 | -0.9 |

*$p < .01$.  **$p < .003$.

The second research question evaluated those scales where significant differences between the two groups were found by establishing percentages of correct classification (i.e., sensitivity, specificity, positive predictive value and negative predictive value) using multiple T score cutoff scores (i.e., 1.0 *SD*, 1.5 *SD*, & 2.0 *SD*). Sensitivity is defined as the ability of a test to classify an individual correctly as having a specific disorder (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008). If 100 preschoolers with autism were evaluated with a testing instrument and that instrument indicated that 85 of them had autism, then the sensitivity would be 85%. For this research project, sensitivity refers to the proportion of preschoolers with ASD who would be correctly identified as having ASD based on certain T score cutoff points using the identified scales on the CBCL and BASC-2. For example, the percentage of children classified as having autism scoring at least 60 (1.0 SD), 65 (1.5 SD), and 70 (2.0 SD) on the CBCL Withdrawn scale will provide sensitivity rates at those three levels.

Specificity is defined as the ability of a test to classify an individual correctly as not having a specific disorder (Parikh et al., 2008). A specificity of 95% would indicate that the testing instrument used was able to identify 95 preschoolers correctly as not having ASD out of 100 who did not have the disorder. For this research project, specificity refers to the proportion of preschoolers without ASD (Non-Spectrum) who would be correctly identified as not having ASD based on certain T score cutoff points using the identified scales on the CBCL and BASC-2. In this research project, the percentage of children in the Non-Spectrum group scoring less than 60 (1.0 SD), 65 (1.5 SD), and 70 (2.0 SD) on the CBCL Withdrawn scale will provide specificity rates at those three levels. The key difference between the two statistical measures is that

sensitivity focuses on those individuals with the disorder, while specificity focuses on those without the disorder.

Positive predictive value (PPV) is defined as the percentage of individuals with a positive test result who actually have the disorder (Parikh et al., 2008). In this research project, PPV is the proportion of preschoolers with autism who have a behavior rating scale score above (on problem behavior scales) or below (on adaptive scales) certain T score cutoff points. Negative predictive value (NPV) can be defined as the percentage of individuals with a negative test result who do not have the disorder (Parikh et al., 2008). For this research, NPV is the proportion of Non-Spectrum preschoolers who have a behavior rating scale score below (on problem behavior scales) or above (on adaptive scales) certain T score cutoff points. Positive predictive value focuses on test results deemed "positive" while negative predictive value focuses on test results deemed "negative."

The sensitivity, specificity, positive predictive value and negative predictive value were calculated for the scales on the BASC-2 and CBCL that demonstrated significant differences between the ASD group and the Non-Spectrum groups. Multiple cutoff scores (i.e., 1.0 *SD*, 1.5 *SD*, & 2.0 *SD*) were used during data analysis to examine if a certain severity level is best for distinguishing between children with ASD and other referred, but non-spectrum children. "For screening purposes, a high sensitivity and NPV are more important than a high specificity and PPV" (Strik, Honig, Lousberg, & Denollet, 2001, p. 427). Thus, the sensitivity and NPV results will primarily be examined for the possibility of the behavior rating scales as screeners for autism. However, even though sensitivity and NPV are considered more important for screening

instruments, low percentages on specificity and PPV cannot be ignored. For example, if a screening instrument indicated everybody had autism, the sensitivity would be 100%, but obviously, the use of such a screener would be pointless. Based on descriptions provided by Strik et al. (2001) and Kempter and Ritter (1991), percentages above 90 are considered excellent, percentages between 70 and 90 are considered good, while percentages below 70 are poor.

In Table 4, the cutoff of one standard deviation was used to distinguish between those children in the ASD group and those in the Non-Spectrum group on the BASC-2 and the CBCL. At the one standard deviation level, the BASC-2 Functional Communication and Social Skills scales were the only ones with good to excellent sensitivity and NPV among the BASC-2 scales. Thus, the Functional Communication and Social Skills scales could be deemed adequate as screeners of autism if one only looked at those two aspects of the results. Those two scales, however, had poor specificity and PPV. In fact, none of the BASC-2 scales had high specificity and PPV. At one standard deviation, the CBCL Withdrawn and Pervasive Developmental Problems scales had good to excellent sensitivity and NPV. Again, however, the specificity and PPV were quite poor. These results suggest that those four scales are good at identifying most of the children with ASD. However, the majority of the Non-Spectrum group also scored greater than one *SD* from the mean, indicating a high level of false positives.

In Table 5, the cutoff of 1.5 standard deviations was used to distinguish between those children in the ASD group and those in the Non-Spectrum group for the BASC-2 and the CBCL. At the 1.5 standard deviation level, only the BASC-2 Social Skills scale had good sensitivity and NPV among the BASC-2 scales. The specificity and PPV

Table 4

*Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) for the BASC-2 PRS-P Scales and CBCL1/½-5 Scales (1.0 SD)*

| BASC-2 Scale | Sensitivity | Specificity | *PPV* | *NPV* |
|---|---|---|---|---|
| Social Skills | 83.3 | 39.1 | 51.7 | 75.0 |
| Functional Communication | 91.7 | 26.1 | 49.3 | 80.0 |
| Hyperactivity | 50.0 | 30.4 | 36.0 | 43.8 |
| Aggression | 13.9 | 39.1 | 15.2 | 36.7 |
| Anxiety | 2.8 | 78.3 | 9.1 | 50.7 |
| Depression | 33.3 | 39.1 | 30.0 | 42.9 |
| Externalizing | 36.1 | 28.3 | 28.3 | 36.1 |
| Internalizing | 16.7 | 65.2 | 27.3 | 50.0 |
| CBCL1/½-5 Scale | | | | |
| Withdrawn | 97.2 | 32.6 | 53.0 | 93.8 |
| Pervasive Developmental Problems | 94.4 | 26.1 | 50.0 | 85.7 |
| Aggressive Behavior | 58.3 | 28.3 | 38.9 | 46.4 |

Table 5

*Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) for the BASC-2 PRS-P Scales and CBCL1/½-5 Scales (1.5 SD)*

| BASC-2 Scale | Sensitivity | Specificity | *PPV* | *NPV* |
|---|---|---|---|---|
| Social Skills | 77.8 | 60.9 | 60.9 | 77.8 |
| Functional Communication | 69.4 | 56.5 | 55.6 | 70.3 |
| Hyperactivity | 38.9 | 32.6 | 31.1 | 40.5 |
| Aggression | 8.3 | 60.9 | 14.3 | 45.9 |
| Anxiety | 2.8 | 89.1 | 16.7 | 54.0 |
| Depression | 13.9 | 50.0 | 17.9 | 42.6 |
| Externalizing | 22.2 | 43.5 | 22.2 | 41.7 |
| Internalizing | 5.6 | 78.3 | 16.7 | 51.4 |
| CBCL1/½-5 Scale | | | | |
| Withdrawn | 88.9 | 58.7 | 62.8 | 87.1 |
| Pervasive Developmental Problems | 94.4 | 32.6 | 52.3 | 88.2 |
| Aggressive Behavior | 38.9 | 37.0 | 32.6 | 43.6 |

percentages for the Social Skills scale increased greatly (up to 60.9% each) at this level, but such a percentage is still considered poor. At 1.5 standard deviations, the CBCL Withdrawn and Pervasive Developmental Problems scales continued to have good to excellent sensitivity and NPV. The specificity and PPV continued to be at a poor level, although the percentages for each increased.

A cutoff of two standard deviations was used to distinguish between those children in the ASD group and those in the Non-Spectrum group and those results are presented in Table 6. At the two standard deviation level, none of the BASC-2 scales had good sensitivity and NPV. Interestingly, the Social Skills scale had good specificity and PPV at this level. Such results suggest that referred, but non-spectrum children rarely score more than two standard deviations below the mean on the Social Skills scale. At two standard deviations, only the CBCL Pervasive Developmental Problems scale continued to have good sensitivity and NPV. However, the specificity and PPV continued to be at a poor level.

In summary, several scales did have high Sensitivity and NPV, particularly at the 1.0 and 1.5 standard deviation cutoff levels; however, the Specificity and PPV were quite poor. Most children in the Non-Spectrum group would be referred for a further evaluation for the possibility of autism using these cutoff levels. Out of all the scales at all cutoff levels, the Social Skills scale on the BASC-2 and the Withdrawn scale on CBCL appear to have the highest Sensitivity, Specificity, PPV, and NPV percentages at the 1.5 standard deviation cutoff level.

Table 6

*Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) for the BASC-2 PRS-P Scales and CBCL1/½-5 Scales (2 SD)*

| BASC-2 Scale | Sensitivity | Specificity | *PPV* | *NPV* |
|---|---|---|---|---|
| Social Skills | 41.7 | 89.1 | 75.0 | 66.1 |
| Functional Communication | 13.9 | 91.3 | 55.5 | 57.5 |
| Hyperactivity | 33.3 | 47.8 | 33.3 | 47.8 |
| Aggression | 8.3 | 63.0 | 15.0 | 46.8 |
| Anxiety | 0.0 | 89.1 | 0.0 | 53.3 |
| Depression | 11.1 | 60.9 | 18.2 | 46.7 |
| Externalizing | 8.3 | 50.0 | 11.5 | 41.1 |
| Internalizing | 0.0 | 82.6 | 0.0 | 51.4 |
| CBCL1/½-5 Scale | | | | |
| Withdrawn | 63.9 | 69.6 | 62.2 | 71.1 |
| Pervasive Developmental Problems | 80.6 | 45.7 | 53.7 | 75.0 |
| Aggressive Behavior | 30.6 | 47.8 | 31.4 | 46.8 |

## Discussion

The first research question examined whether specific scales on the BASC-2 or CBCL could distinguish between preschoolers with autism and other clinically referred children without autism. Results of this study found a number of statistically significant differences on various scales. At the more restrictive $p < .001$ level, significant differences were found on the Aggression, Social Skills, and Externalizing scales on the BASC-2 and the Withdrawn and Pervasive Developmental Problems scales on the CBCL. Children on the autism spectrum scored in the average range on the BASC-2 Aggression and Externalizing scales while the non-spectrum group scored significantly higher, in the at-risk range. Thus, even though significant differences were found, scores within the average range would not be helpful in screening for autism because most preschoolers will score within this range. Children with ASD had significantly lower mean scores on the Social Skills scale, as was hypothesized. However, even the Non-Spectrum group had a Social Skills mean score below the average range.

As predicted, the children with ASD showed significantly more Withdrawn and Pervasive Developmental Problems than the Non-Spectrum group on the CBCL. The current findings are very consistent with results by Sikora et al. (2008) who also found mean scores for the Withdrawn and Pervasive Developmental Problems scales were significantly higher for their group of preschoolers with autism than a group of referred, but non-autistic children. Sikora et al. (2008) had a mean score of 73.3 on the Withdrawn scale for their group of children with autism while the current results found a mean score of 75.0 on that scale for that group. Similarly, Sikora et al. had a mean score of 75.0 on the Pervasive Developmental Problems scale while current results had a mean of 76.2. It

is interesting to note how remarkably similar the two different groups of children with autism were rated by their parents, even though the mean age for their sample of children with autism was 50.7 months while the current sample's mean age was 36.8 months.

Curiously, neither the Atypicality or Withdrawal scales on the BASC-2 that were hypothesized to come out higher for the ASD group were found significantly different between the two groups. In particular, it is interesting to note that the Withdrawn scale on the CBCL had a mean score of 75.0 and the Withdrawal scale on the BASC-2 had a mean score of 61.1, which is almost a one and a half standard deviation difference. Although the two scales have the same construct name, they clearly are not measuring equivalent constructs. A cursory qualitative examination of the items on each scale suggest that one possible explanation for the difference is that the BASC-2 Withdrawal scale appears to be measuring social types of behaviors (e.g., difficulty making friends, shy, fear of strangers) where as the CBCL Withdrawn scale looks at more avoidance and unresponsive types of behaviors characteristic of autism (e.g., avoids eye contact, doesn't answer, unresponsive to affection).

The mean score on the BASC-2 Atypicality scale was not significantly higher for the ASD group than the non-spectrum group. This may be due to, as described by Reynolds and Kamphaus (2004), the Atypicality scale measuring both psychotic and autistic-like behaviors. Thus, in addition to autistic-like behaviors (e.g., unaware of others), it also measures "odd" behaviors such as acting confused and saying nonsensical things. The CBCL Pervasive Developmental Problems scale seems to assess autistic-like behaviors more directly (e.g., disturbed by change, rocks body, speech problem).

Bour (2008) compared the CBCL and BASC-2 scales at the preschool level to see if similarly named scales on both instruments measured the constructs in a similar way and did find that the Withdrawn/Withdrawal scales and the Pervasive Developmental Problems/Atypicality scales on the CBCL and the BASC-2 were statistically significantly different. Thus, previous research seems to suggest that the similarly named scales on the two instruments are measuring slightly different constructs.

Only two of the four adaptive behavior scales on the BASC-2 that were hypothesized to come out lower for the ASD group actually came out significantly lower (i.e., Social Skills and Functional Communication). There were no significant differences found on the Adaptability and Adaptive Skills scales. Both groups had mean scores below the average range on Adaptive Skills. Interestingly, the Adaptability scale had a mean score that was in the average range for preschoolers with ASD and below the average range for the Non-Spectrum group (although the difference was not significant). Typically, children with autism spectrum disorder do not adjust well to changes in their environment. It appears that the Adaptability scale on the BASC-2 does not adequately capture those characteristics of young children with autism.

According to the DSM-IV-TR (APA, 2000), hyperactive behaviors are a common characteristic of children with Autism Spectrum Disorder. However, on both the Hyperactivity scale on the BASC-2 and the Attention Deficit Hyperactivity Problems scale on the CBCL, the Non-Spectrum group had higher mean scores than the ASD group. Such results could imply that hyperactive behaviors may not be as prevalent in preschoolers with Autism Spectrum Disorder. Another possibility is that hyperactive

behaviors are a primary concern among Non-Spectrum preschool children who are referred to a clinical setting.

Typically, it is important for a screener for a disorder like autism to identify all possible children with the disorder so that a further evaluation can be conducted and early intervention services provided as needed. Thus, a high Sensitivity and Negative Predictive Value are important for screeners. The current study examined different cutoff levels on the CBCL and BASC-2 to determine whether those behavior rating scales might be useful as screeners for autism. While several scales did have high Sensitivity and NPV, the Specificity and PPV were so low as to bring into question the value of the scales as screeners. That is, most children, including the Non-Spectrum group, would be referred for a further evaluation for the possibility of autism. It was noted that the Social Skills scale on the BASC-2 and the Withdrawn scale on CBCL appear to have the highest Sensitivity, Specificity, PPV, and NPV percentages at the 1.5 standard deviation cutoff level. For screening purposes, those two scales at that level seem to have the most potential for capturing children with autism and excluding children without autism. However, approximately 40% of referred but non-autistic children would be referred for further evaluation of autism if those scales were to be used as a screener for autism. Individual practitioners or agencies would need to decide if such a high number of false positives is worth the time and expense of additional evaluations. Perhaps future research can determine if some other screening tool, used in conjunction with either the Social Skills or Withdrawn scales, would be the most effective screener for autism.

The applied practical implications of the current study are somewhat questionable. Even though mean differences on some scales were statistically

significantly different between the ASD and Non-Spectrum groups, the point difference was relatively small. For example, the mean score for the Non-Spectrum group was 68.7 on the CBCL Pervasive Developmental Problems scale while it was 76.2 for the ASD group, just one half of a standard deviation. Furthermore, the mean score for the Non-Spectrum group is still considered clinically significant. Thus, it is hard to translate these results into practical information for a psychologist on an individual case basis. As an example, if a referred child received a T score of 70 on the Pervasive Developmental Problems scale, it is not clear how to interpret such a score. Many referred preschool children without autism score that high as well. Perhaps the more important implication of these results is that psychologists should be careful about interpreting high scores as they can be misleading.

*Strengths and Limitations*

Lord and Corsello (2005) suggested that behavior rating scales might be useful as tools to screen for autism. However, little research has been conducted to examine if behavior rating scales are indeed useful as screeners. Sikora et al. (2008) appears to be the only published study to examine specifically this possibility in the United States and they examined the CBCL. This study appears to be the second to evaluate the CBCL and the first to evaluate the BASC-2. The current results examining the CBCL are largely consistent with Sikora et al.'s, although this study adds to the literature by also examining the practical implications behind using the CBCL as a screener for autism. The current results are important because they challenge Lord and Corsello's (2005) assertion that behavior rating scales are useful as screeners for autism.

A limitation of the current study is that the sample of preschoolers used in the study was obtained from the same non-profit child development clinic, which may make generalizability of results limited. As typical of most studies, it would be beneficial if the sample were larger and more diverse. Perhaps obtaining participants from various clinics in multiple regions across the United States would enhance the generalization of results. The age of the participants used in the study made it difficult to determine specific diagnoses, if any, of the preschoolers in the Non-Spectrum group. Comparing preschoolers with autism to other identified groups might provide additional information for differential diagnosis related to behavior rating scale results. The main limitation of this study was that there was no secondary validation of diagnoses for this sample of preschoolers (i.e., autism or non-spectrum). Having a second professional verifying the autism diagnoses would have strengthened these results.

*Future Research*

In the current study, there were some preschoolers excluded from this study because they had some characteristics of autism but not enough to be classified on the spectrum. Perhaps future research could explore the results of behavior rating scales with a larger group of children with only a few autistic characteristics. Following this group of children in the future may provide additional information on the characteristics of behavior rating scales for identifying autism. Future research could also examine the usefulness of behavior rating scales as screeners for autism in a large sample representative of the general population. Perhaps behavior rating scales such as the BASC-2 and CBCL make better screeners for autism in the general population instead of a referred group of children. Additional future research could also look at other behavior

rating scales other than those analyzed in the current study to determine their usefulness as screeners for autism. Future research could also examine combinations of different individual scales on these behavior rating instruments, or a combination of an individual scale plus another screening tool, as more useful screeners of autism.

References

Achenbach, T. M. (1991). *Integrative guide for the 1991CBCL/4-18, YSR, and TRF profiles.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1992). *Manual for the child behavior checklist/2-3.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families

Allen, C. W., Silove, N., Williams, K., & Hutchins, P. (2007). Validity of the Social Communication Questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders, 37*, 1272-1278.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., Revised). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

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

Bour, J. L. (2008). *Comparing parent ratings of referred preschoolers on the Child Behavior Checklist and Behavior Assessment System for Children – Second*

*Edition*. Unpublished specialist thesis. Western Kentucky University, Bowling Green.

Centers for Disease Control and Prevention. (2007, February 9). Autism spectrum disorders overview. Retrieved August 23, 2008, from http://www.cdc.gov/ncbddd/autism/overview.htm

Charman, T., & Baird, G. (2002). Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. *Journal of Child Psychology and Psychiatry, 43,* 289-305.

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., et al. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders, 33,* 427-433.

Constantino, J. N., & Gruber, C. P. (2005). *Social Responsiveness Scale.* Los Angeles, CA: Western Psychological Services.

Conway, F. (2007). Review of the Social Responsiveness Scale. In K. F. Geisinger, R. A. Spies, J. F. Carlson, & B. S. Plake (Eds.), *Buros Mental Measurement Yearbook (*Vol. 17, pp. 743-748). Lincoln, NE: University of Nebraska Press.

Coonrod, E. E., & Stone, W. L. (2005). Screening for autism in young children. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders: Vol. 2. Assessment, interventions, and policy* (3rd ed., pp. 707-729). Hoboken, NJ: Wiley.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Leventhal, B. L., et al. (2007). Between a ROC and a hard place: Decision making and making decisions about using SCQ. *Journal of Child Psychology and Psychiatry, 48:9*, 932-940

DiLavore, P., Lord, C., & Rutter, M. (1995). Pre-Linguistic Autism Diagnostic Observation Schedule (PL-ADOS). *Journal of Autism and Developmental Disorders, 25*, 355-379.

Duarte, C. S., Bordin, I. A. S., Oliveira, A., & Bird, H. (2003). The CBCL and the identification of children with autism and related conditions in Brazil: Pilot findings. *Journal of Autism and Developmental Disorders 33*, 703-707.

Dumont-Mathieu, T., & Fein, D. (2005). Screening for autism in young children: The Modified Checklist for Autism in Toddlers (M-CHAT) and other measures. *Mental Retardation and Developmental Disabilities Research Reviews 11*, 253-262.

Feil, G. E., Severson, H. H., & Walker, H. M. (2002). Early screening and intervention to prevent the development of aggressive, destructive behavior patterns among at-risk children. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 143-166). Bethesda, MD: NASP Publications.

Filipek, P. A., Accardo, P. J., Baranek, G. T., Cook, E. H., Dawson, G., Gordon, B., et al. (1999). The screening and diagnosis of autistic spectrum disorders, *Journal of Autism and Developmental Disorders, 29*, 439-484.

Gillberg, C., Nordin, V., & Ehlers, S. (1996). Early detection of autism. Diagnostic instruments for clinicians. *European Child & Adolescent Psychiatry 5*, 67-74.

Gilliam, J. E. (1995). *Gilliam Autism Rating Scale*. Austin, TX: Pro-ed.

Kempter, B. & Ritter, M. M. (1991). Unexpected high calcitonin concentrations after pentagastrin stimulation. *Clinical Chemistry, 37,* 473-474.

Krug D. A., Arick J. R. & Almond P. J. (1980). Behavior checklist for identifying severely handicapped individuals with high levels of autistic behaviors. *Journal of Child Psychology and Psychiatry, 21,* 221-229.

Lord, C., & Corsello, C. (2005). Diagnostic instruments in autistic spectrum disorders. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders: Vol. 2. Assessment, interventions, and policy* (3rd ed., pp. 730-771). Hoboken, NJ: Wiley.

Lord, C., Risi, S., Lambrecht, L., Cook, E., Leventhal, B., DiLavore, P., 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*, 205-223.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule.* Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., Good, S., Heemsbergern, J., Jordan, H., Mawhood, L., et al. (1989). Autism Diagnostic Observation Schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders, 19*, 185-212.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3-9.

Magyar, C. I., & Pandolfi, V. (2007). Factor structure evaluation of the Childhood Autism Rating Scale. *Journal of Autism Developmental Disorder, 37*, 1787-1794.

Marteleto, M. R. F., & Pedromônico, M. R. M. (2005). Validity of Autism Behavior Checklist (ABC): Preliminary study. *Rev Bras Psiquiatr, 27,* 295-301.

Matson, J. L., Wilkins, J., & González, M. (2008). Early identification and diagnosis in autism spectrum disorders in young children and infants: How early is too early? *Research in Autism Spectrum Disorders, 2*, 75-84.

Merrell, K. W. (2008). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). New York: Lawrence Erlbaum Associates.

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity, and predictive values. *Indian Journal of Ophthalmology, 56,* 45-50.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children.* Circle Pines, MN: American Guidance Service.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children-Second Edition.* Circle Pines, MN: AGS Publishing.

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

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

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services.

Schreibman, L. (2000). Intensive behavioral/psychoeducational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disorders, 30*, 373-378.

Sikora, D. M., Hall, T. A., Hartley, S. L, Gerrard-Morris, A. E., & Cagle, S. (2008). Does parent report of behavior differ across ADOS-G classifications: Analysis of scores from the CBCL and GARS. *Journal of Autism and Developmental Disorders, 38*, 440-448.

Smith, V. K., & Dillenbeck, A. (2006). Developing and implementing early intervention plans for children with autism spectrum disorders. *Seminars in Speech and Language, 27*, 10-20.

Stone, W. L., Coonrod, E. E., Turner, L. M., & Pozdol, S. L. (2004). Psychometric properties of the STAT for early autism screening. *Journal of Autism and Developmental Disorders, 34*, 691-701.

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

Stone, W. L., & Ousley, O. Y. (1997). *Screening Tool for Autism in 2-year-olds*. Unpublished manuscript, University of Nashville, TN.

Strik, J., Honig, A., Lousberg, R., & Denollet, J. (2001). Sensitivity and specificity of observer and self-report questionnaires in major and minor depression following myocardial infarction. *Psychosomatics, 42*, 423-428.

Tanguay, P. E. (2000). Pervasive developmental disorders: A 10-year review. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 1079-1095.

Tomanik, S. S., Pearson, D. A., Loveland, K. A., Lane, D. M., & Shaw, J. B. (2007). Improving the reliability of autism diagnoses: Examining the utility of adaptive behavior. *Journal of Autism Developmental Disorders, 37*, 921-928.

U.S. Census Bureau. (2009). Kentucky quick facts from the U.S. Census Bureau. *U.S. Census Bureau.* Retrieved February 13, 2009, from http://quickfacts.census.gov/qfd/states/21000.html

Ventola, P., Kleinman, J., Pandey, J., Wilson, L., Esser, E., Boorstein, H., et al. (2007). Differentiating between autism spectrum disorder and other developmental disabilities in children who failed a screening instrument for ASD. *Journal of Autism and Developmental Disorders, 37*, 425-436.

Volkmar, F. R., & Klin, A. (2005). Issues in the classification of autism and related conditions. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders: Vol. 1 Diagnosis, developmental, neurobiology and behavior* (3rd ed., pp. 5-41). Hoboken, NJ: Wiley.

Welsh, J. S. (1992). Review of the Childhood Autism Rating Scale. In J. J. Kramer, J. C. Conoley, & J. C. Conoley (Eds.), *Buros Mental Measurement Yearbook (*Vol. 11, pp. 169-173). Lincoln, NE: University of Nebraska Press.

Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). The utility of the Social Communication Questionnaire in screening for autism in children referred for early intervention. *Focus on Autism and Other Developmental Disabilities 22*, 33-38.

Woods, J. J., & Wetherby, A. M. (2003). Early identification and intervention for infants and toddlers who are at risk for autism spectrum disorder. *Language, Speech, and Hearing Services in Schools, 34*, 180-193.

# Appendix

## Human Subjects Review Board Approval

Amber D. Gross
c/o Dr. Carl Myers
Psychology
WKU

Dear Amber:

Your revision to the research project, *Screening Preschoolers for Autism with Behavior Rating Scales*, was reviewed by the HSRB and it has been determined that risks to subjects are: (1) minimized and reasonable; and that (2) research procedures are consistent with a sound research design and do not expose the subjects to unnecessary risk. Reviewers determined that: (1) benefits to subjects are considered along with the importance of the topic and that outcomes are reasonable; (2) selection of subjects is equitable; and (3) the purposes of the research and the research setting is amenable to subjects' welfare and producing desired outcomes; that indications of coercion or prejudice are absent, and that participation is clearly voluntary.

1. In addition, the IRB found that you need to orient participants as follows: (1) signed informed consent is not required; (2) Provision is made for collecting, using and storing data in a manner that protects the safety and privacy of the subjects and the confidentiality of the data. (3) Appropriate safeguards are included to protect the rights and welfare of the subjects.

This project is therefore approved at the Exempt Review Level until May 31, 2009.

2. Please note that the institution is not responsible for any actions regarding this protocol before approval. If you expand the project at a later date to use other instruments please re-apply. Copies of your request for human subjects review, your application, and this approval, are maintained in the Office of Sponsored Programs at the above address. Please report any changes to this approved protocol to this office. A Continuing Review protocol will be sent to you in the future to determine the status of the project. Also, please use the stamped form that accompanies this letter.

Sincerely,

Paul J. Mooney, M.S.T.M.
Compliance Manager
Office of Sponsored Programs
Western Kentucky University

cc: HS file number Gross HS09-136