5-2009

Behavior Rating Scales as Screeners for Autism? A Closer Look at the CAB-P and CBCL/1.5-5

Brandy Mickele McReynolds
Western Kentucky University, brandy.mcreynolds734@wku.edu

BEHAVIOR RATING SCALES AS SCREENERS FOR AUTISM? A CLOSER LOOK AT THE CAB-P AND CBCL/1.5-5

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
Brandy Mickele McReynolds

May 2009

BEHAVIOR RATING SCALES AS SCREENERS FOR AUTISM?
A CLOSER LOOK AT THE CAB-P AND CBCL/1.5-5

Date Recommended _May 14, 2009___

___Carl Myers_____
Director of Thesis

___Lakeisha Meyer_____

___Elizabeth Jones_____

_____
Dean, Graduate Studies and Research Date

Acknowledgements

First and foremost I would like to thank the members of my thesis committee, Dr. Carl Myers, Dr. Libby Jones, and Dr. Lakeisha Meyer for their support, which has made the completion of this project possible. I am particularly thankful for the flexibility my thesis committee members allowed me for the proposal of my thesis from the state of Hawaii. I would especially like to thank Dr. Carl Myers, my thesis director, for his support, encouragement, flexibility, humor, and most of all, never-ending patience, for the countless hours he has contributed to the success and completion of this project! These factors have greatly and positively impacted my experience and success in completing my thesis. In addition, I would like to thank my mother who has never faltered in her love and support during my academic career. She has always encouraged me to follow my dreams, regardless of their direction, and her confidence in me has always been my source of hope, strength and determination. To all that have contributed, I sincerely thank you for all you have done! Your efforts have been greatly appreciated!

Table of Contents

Pages

List of Tables

Pages

BEHAVIOR RATING SCALES AS SCREENERS FOR AUTISM?
A CLOSER LOOK AT THE CAB-P AND CBCL/1.5-5

Brandy Mickele McReynolds May 2009 63 pages

Directed by: Dr. Carl Myers, Dr. Libby Jones, and Dr. Lakeisha Meyer

Department of Psychology Western Kentucky University

In recent years, autism diagnoses have steadily increased, putting a substantial emphasis on early identification as a crucial component for intervention. Autism diagnoses, however, often require a thorough and comprehensive assessment from a highly trained practitioner. Although ideal, such assessments are often time consuming and expensive, creating a need for a quicker, more simplistic method of screening for autism. Clinicians customarily used behavior rating scales to identify a number of various problem behaviors and/or disorders. The purpose of the present study is to examine the utility of two common behavior rating scales in accurately discriminating between a group of preschoolers with autism and a group of referred preschoolers with autism. Parents/guardians of 74 preschoolers with and without autism, who had been referred to a child development clinic due to behavioral or developmental concerns, completed both behavior rating scales as part of a comprehensive assessment. Although analyses revealed significant differences between the two groups of participants on two of the scales from one of the instruments, these findings demonstrate little clinical value for screening purposes.

Introduction

Autism is a developmental disorder found within the broader category of Pervasive Developmental Disorders (PDD), as stated within the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2000). According to the Individuals with Disabilities Education Act of 2004, this disorder is broadly defined by skill deficits in social interaction, verbal and non-verbal communication, and academic performance in school. Stuart, Flis, and Rinaldi (2006) state that the general characteristics of autism also include unusual sensory seeking, repetitive behavior, stereotyped movements, and resistance to changes in the environment. Key behaviors, however, are difficulties with social reciprocity and communication with others.

The term Autism Spectrum Disorder (ASD) is used to describe a broad range of behavioral characteristics and developmental levels ranging from severe to mild. For example, those at the lower functioning end of the spectrum may have little or no language or communication ability, struggle with initiating social interaction, and display bizarre stereotypic and repetitive behavior, such as rocking back and forth, flapping hands, or hoarding insignificant objects. At the higher functioning end of the spectrum, those with autism may lead more typical lives, but lack age appropriate social skills or be highly disrupted by changes in routine or certain disturbing sensory stimuli (Stuart et al., 2006). It has also been found that autism in males is four times more common than autism in females (Marker, Weeks, & Kraegel, 2007).

Within the past few decades, there has been a striking rise in the number of children diagnosed with Autism Spectrum Disorder, which has caused concern among

practitioners and families affected by the disorder. Appropriately diagnosing children with autism has remained a crucial undertaking for practitioners. Outcomes for children with autism depend on specialized interventions to address their specific needs, and the earlier children with autism are diagnosed, the better the prognosis for those with this disorder. A diagnosis of autism requires a thorough and comprehensive assessment from a highly trained practitioner with much prior experience with autism. Such specialized assessments, however, cannot be given to every referred child on a routine basis. Thus, there is a need for a more simplistic method to screen for the possibility of autism. Behavior rating scales provide an inexpensive means of assessing a child or adolescent's behavior, social skills, and emotional problems and competencies while requiring much less training and time to administer. Although behavior rating scales are commonly used by psychologists to screen or assess a broad range of problem behaviors and/or disorders, very little research has been conducted to examine the use of behavior rating scales to screen for autism.

The present study will examine archival ratings from two behavior rating scales, the Clinical Assessment of Behavior – Parent Form (CAB-P, Bracken & Keith, 2004) and the Child Behavior Checklist (CBCL/1.5-5, Achenbach & Rescorla, 2000) with two groups of children: (a) children with autism, and (b) children who were referred, but not classified as having autism. Participants were the parents/guardians of preschool-aged children who were referred for an evaluation because of developmental or behavioral concerns. Mean scores for the various scales on the two behavior rating scales were compared for the two groups of referred children. In addition, for any scale that showed a significant difference between the groups, the sensitivity and specificity of those scales

regarding “correct” classification of the children as autistic or not autistic were determined. Through these analyses, the usefulness of the behavior rating scales for screening preschool children for the possibility of autism will be evaluated.

Literature Review

The *Diagnostic and Statistical Manual of Mental Disorders, Text Revision* (DSM-IV-TR, American Psychiatric Association, 2000) defines autism as the presence of significant impairments in socialization and communication, as well as profound manifestations of repetitive behavior and restricted interests and activities. The presentation of autistic characteristics may occur at different rates and severity levels depending on the age and developmental level of an individual. However, there must be delays or impaired functioning of at least one of the following before age three: (a) imaginative play, (b) social interaction, or (c) language used to communicate with others (APA, 2000). Each of the three primary diagnostic categories of behaviors will be described in more detail.

Socialization

The DSM-IV-TR (APA, 2000) describes children with autism as lacking interest in relationships with others and having little willingness to participate in social activities or initiate social interactions, when compared to children without autism. Bellini and Hopf (2007) describe specific social skill deficits of children within the autism spectrum. These include deficits in participating in or initiating social endeavors, using expressive language to communicate feelings, and the ability to understand the emotions and opinions of others. They state that it is not uncommon for those with autism to make socially inappropriate remarks, or to dominate conversations with topics solely of their own interest.

Conroy, Boyd, Asmus, and Madera (2007) also describe problem social behaviors that those with autism may display. These problem social behaviors may include

aggression, self-injury, tantrums, damaging the physical environment, or disruptive acts. However, because of their communication deficits, these problem social behaviors often serve a specific purpose or function intended to obtain or avoid things such as sensory stimulation, attention from others, tangible objects, or simply to escape undesired demands.

Children with autism, as compared to other children with other developmental disabilities, were found to have specific social skill deficits (Ventola et al., 2007). Deficits in children with autism include eye contact, overall quality of social interactions, shared enjoyment with other children, joint attention, relationships with peers, willingness to share with peers, willingness to offer comfort to peers, and responding to others with empathy. In general, younger children with autism use fewer conventional gestures, specifically nodding and shaking their head in agreement or disagreement, when compared to those with other developmental delays. One social skill deficit that both very young children with autism and very young children with other developmental disabilities do share is their deficit in pretend play skills. However, when compared to children with other developmental disabilities, between the ages of three and a half and five years old, children with autism display significantly fewer pretend play skills than those with other developmental disabilities (Ventola et al., 2007).

Communication

The DSM-IV-TR (APA, 2000) describes that numerous aspects of both verbal and nonverbal communication skills are affected and that those with autism may have a delay in, or a complete lack of, the development of language. When those with autism do have language, they may display marked inability to initiate or sustain a conversation with

others. These individuals may also demonstrate the use of idiosyncratic language or a stereotyped and repetitive use of language. Those with autism who do have language may display abnormal use of pitch, intonation, rhythm, rate, and stress of words. In addition, individuals with autism may be unable to comprehend simple directions or questions. Research by Ventola et al. (2007), comparing the defining communication characteristics of those with autism spectrum disorder to those with other developmental delays, have found that young children with ASD have more difficulty with using pointing gestures to show interest, displaying a range of facial expressions to communicate emotions, and acquiring imitation skills of both verbal and nonverbal behaviors.

Repetitive Behaviors

The DSM-IV-TR (APA, 2000) describes that those with autism tend to display behaviors, and take interest in activities, characterized by restricted, repetitive, and stereotyped patterns. The motor mannerisms of those with autism may be stereotyped and repetitive as well, and may include hand movements such as clapping and finger flicking, or body movements such as rocking, dipping, or swaying. Those with autism may also display odd body posture or walking on tiptoes. There may also be an abnormal intensity or focus on one particular object or topic, or a persistent focus on parts of objects (e.g., spinning the wheels on a car). Those with autism may display specific routines or rituals that have no apparent function, and may be extremely inflexible in changing their routines or rituals.

Prevalence Rates

Over the past few decades, autism rates have been dramatically increasing in the United States. According to Blaxill, Baskin, and Spitzer (2003), autism prevalence rates ranged from 0.61 to 0.68 per 1,000 individuals in 1987, and jumped to 2.13 to 2.98 per 1,000 individuals by 1994. The National Institute of Mental Health (NIMH, 2003) indicated precise autism prevalence rates are unknown but estimated that autism rates ranged from two to four of every 1,000 individuals in the United States at that time. Prevalence rates for disorders on the autism spectrum were recorded at being between two and six per 1,000 individuals by the Centers for Disease Control and Prevention (CDC, 2006). This prevalence rate included Asperger's Disorder, Autistic Disorder, and Pervasive Developmental Disorder Not Otherwise Specified. According to the CDC (2006), the most current estimates of prevalence rates for autism spectrum disorders in school-aged children are based on two surveys of parent-reported data that covered the years 2003 and 2004. These surveys include the National Health Interview Survey, revealing 5.7 per 1,000 children diagnosed with autism spectrum disorders, and the National Survey of Children's Health, revealing 5.5 per 1,000 children diagnosed with autism spectrum disorders. Based on these two multi-state surveys, the CDC (2006) supports the estimate of over 300,000 school-aged children having autism in the 2003-2004 school year.

There has been much disagreement about the origin of the increase in autistic symptoms and diagnoses in recent years within the United States. Reasons for such skyrocketing prevalence rates may be due to a variety of causes, including actual increases in the disorder itself, use of better diagnostic tests, possible misdiagnoses,

classification guidelines that are more inclusive, or a combination of these reasons (Steuernagel, 2005). Gernsbacher, Dawson, and Goldsmith (2005) and Wing and Potter (2002) claim that such dramatic increases in autism diagnoses are primarily due to the broadened diagnostic criteria, greater knowledge of autism as a spectrum, greater general awareness of the disorder among parents and professionals, improved specialist services, and improved case findings related to autism. Wing and Potter (2002) also acknowledge there may be a true increase in autism. Bertrand et al. (2001) noted that the diagnosis of autism using DSM-IV-TR criteria is based on subjective opinions about the quality of behaviors displayed by a child. This subjective decision-making has been the cause of inconsistencies in the past use of diagnostic labels, and have consequently affected efforts enabling researchers to accurately pinpoint baseline prevalence rates (Bertrand et al., 2001). Regardless of the exact reasons for the dramatic increase in autism prevalence rates, the many children diagnosed with autism spectrum disorders presents numerous educational challenges to parents and schools.

Importance of Early Identification

Early identification of autism is crucial to the prognosis of this disorder. Early identification is advantageous because the earlier a child is diagnosed with autism, the earlier the child is able to receive an appropriate targeted intervention, and thus the outcome for this child will be more positive (Lovaas, 2003; Maurice, Green, & Luce, 1996). The needs of children with autism differ from children with general developmental delays. Tomanik, Pearson, Loveland, Lane, and Shaw (2007) explain that children with autism who are not diagnosed early miss out on valuable opportunities from beneficial treatment, which results in a poorer long-term outcome. Without early

intervention, a poorer long-term outcome occurs because of a "cascade of effects that result from early deficits and interfere with later functioning" (Coonrod & Stone, 2005, pp. 708-709). As further stated by Coonrod and Stone (2005), "Mounting evidence from early intervention research suggests that participation in specialized intervention programs at young ages is important for optimizing the long-term outcomes of children with autism" (p. 707). These targeted interventions provide children with autism structure, and emphasize the development of communication skills.

Early intervention is thought to be advantageous because very young children have heightened brain plasticity. This plasticity in a child's brain may result in greater opportunities for overcoming neurodevelopmental deficits within the early years of life with early intervention (Coonrod & Stone, 2005). Early identification of autism is also imperative for increasing parent awareness and understanding of an autistic disorder because it helps parents become aware of their child's special needs at an earlier point in time. Raising children with autism is challenging and parents that are equipped with knowledge of their child's disorder, as well as professional advice and adequate support, are better able to adjust to the parenting difficulties that may arise in their child's life (Charman, 2003).

Assessment and Diagnosis of Autism

In order to diagnose autism, a comprehensive evaluation must be carried out. According to Gillberg, Nordin, and Ehlers (1996), autism may be reliably diagnosed around the age of 30 months, whereas screening for autism may take place as early as 18 months of age. Recently, however, there has been progress in identifying autism spectrum disorders in even younger children (Charman, 2003; Matson, Wilkins, &

González, 2008; Zwaigenbaum et al., 2007). Ideally, a comprehensive evaluation of autism should include observations of the child's behavior across several settings, parent interviews, and the use of assessments specialized for the diagnosis of autism. Marker et al. (2007) state that a comprehensive evaluation of autism should be individualized and should focus on the development and skill deficits of the child. Tomanik et al. (2007) explain that the most important skill deficits to examine are impairments in the child's communication, daily living skills, and socialization, as compared to other children whose behavior does not fall within the autism spectrum. They describe that most deficits in children with autism are reported in the areas of communication and socialization, and many children who have autism do not always display severe or salient deficits in all three areas.

Specialized Autism Diagnostic Instruments

A broad range in the possible characteristics and severity of characteristics makes autism difficult to diagnose. Accurately diagnosing autism can also be a challenging endeavor for clinicians because there are other developmental disorders that share similar characteristics. For example, Lord and Risi (1998) explain that children with autism who have hyperactive and impulsive behaviors or developmental delays may be misdiagnosed with Attention Deficit Hyperactivity Disorder or Mental Retardation, when characteristic socialization, communication, and stereotyped behaviors are missed. Because symptoms of autism vary greatly across ages and developmental level, diagnoses of autism are further complicated. When children do not clearly meet the criteria for autism, the potential for misdiagnoses is high. Although autism is difficult to distinguish from other disorders, several specialized assessment instruments exist that can be used to assist in

the diagnosis of autism. Two prominent assessment instruments, that have strong psychometric characteristics, include the Autism Diagnostic Observation System (ADOS, Lord, Rutter, DiLavore, & Risi, 1999), and the Autism Diagnostic Interview-Revised (ADI-R, Rutter, LeCouteur, & Lord, 2003).

Autism Diagnostic Observation System. The ADOS (Lord et al., 1999) is a structured observation system that is specifically designed to focus on the quality of a child's social and communication behaviors. This instrument is an interactive tool that can be used with children who have an estimated mental age of 3 years old or greater and is based on a standardized set of social contexts. The ADOS emphasizes the use of natural interaction and the use of prompts when necessary. The ADOS is considered a reliable and valid instrument, but its reliability and validity are greatly dependent on the examiner's training, clinical skills, and experience. According to Tomanik et al. (2007), behaviors that occur during testing may cause classification errors due to the ADOS only providing a single sample of a child's current behavior.

Autism Diagnostic Interview-Revised. The ADI-R (Rutter et al., 2003) is a standardized semi-structured interview that typically takes 90 minutes to two hours to complete. A trained interviewer administers the ADI-R by interviewing the child's parent or primary caregiver. This instrument is designed to distinguish between behaviors of children with autism and behaviors of typically developing children and can be used with children with mental ages of 18 months or older. For older children, the ADI-R often requires the parents to focus on their children when they were four to five years old because certain diagnostic features are most evident during this age period. The main objective of this instrument is for the interviewer to elicit behavioral descriptions

from the child's parents and then to code these descriptions according to pre-defined scoring criteria. The ADI-R was standardized on a sample of 25 children with autism and 25 children without autism. Children from this standardization sample ranged between 21 and 74 months of age. Differentiation between children with and without autism was successfully established through the diagnostic algorithms created for the ADI-R (Rutter et al., 2003). However, the ADI-R is more useful in detecting lower functioning autism than detecting Asperger's Disorder and higher functioning autism. Others have described the ADI-R as "widely recognized as a gold standard parent-report interview for establishing a clinical diagnosis of autism" (Constantino et al., 2003, p. 430).

Accurately diagnosing autism can be a difficult task for even a qualified and experienced clinician. Individuals that administer these specialized assessment instruments and interpret their scores to help diagnose autism must undergo highly specialized training and have extensive background experiences with children with autism in order to correctly recognize and diagnosis this disorder. Although specialized assessment instruments now exist to aid in diagnosing autism, someone has to recognize initially the possibility of autism in order to initiate the administration of such an instrument. Such instruments would not be given on a routine basis. There are a number of autism screeners that exist that do not require the extensive experience and training to administer. An overview of selected screeners will be presented in the next section.

Autism Screeners

According to Coonrod and Stone (2005), screeners for autism typically require much less time, experience to administer, and training than do specialized diagnostic assessments. Screeners do not provide diagnoses. Instead, screeners are intended to

indicate levels of risk for a disability. Screeners also allow an individual, who may have less prior experience with the child, to complete a checklist of observable behaviors related to a specific disorder. Screeners are intended to evaluate large numbers of children, most of whom do not have developmental disorders; therefore, screeners are meant to be brief and easy to administer (Coonrod & Stone, 2005).

The CDC (2007) states that during well-child doctor visits, developmental screenings should occur to identify children with potential developmental disabilities. Specifically for autism, the CDC states that at ages 18, 24, and 30 months, all children should be screened for an autism spectrum disorder in well-child doctor visits. Pediatricians are to screen for the possibility of autism using the CDC's (2007) list of autism characteristics that young children might present:

- Not play "pretend" games (pretend to "feed" a doll)
- Not point at objects to show interest (point at an airplane flying over)
- Not look at objects when another person points to them
- Have trouble relating to others or not have an interest in other people at all
- Avoid eye contact and want to be alone
- Have trouble understanding other people's feelings or talking about their own feelings
- Prefer not to be held or cuddled or might cuddle only when they want to
- Appear to be unaware when other people talk to them, but respond to other sounds
- Be very interested in people, but not know how to talk to, play with, or relate to them

- Repeat or echo words or phrases said to them, or repeat words or phrases in place of normal language (echolalia)
- Have trouble adapting to changes in routine
- Have unusual reactions to the way things smell, taste, look, feel, or sound
- Loses skills they once had (for instance, stop saying words they were once using)

There are a few published screeners that can be used to aid in the identification of autism. Four such screeners include the Social Responsiveness Scale (SRS, Constantino & Gruber, 2005), the Childhood Autism Rating Scale (CARS, Schopler, Reichler, & Renner, 1988), the Gilliam Autism Rating Scale (GARS, Gilliam, 1995), and the Checklist for Autism in Toddlers (CHAT, Baron-Cohen, Allen, & Gillberg, 1992). These screeners are relatively easy to administer, and require much less time and training than do diagnostic assessment instruments for autism.

Social Responsiveness Scale. The SRS (Constantino & Gruber, 2005) is a quantitative rating scale that can be completed within 15-20 minutes. There are two versions of the SRS: one for teachers and one for parents. The SRS can be used with individuals between the ages of 4 and 18 to assess the individual's social interactions, communication skills, and repetitive behaviors that are often associated with autism. This instrument purports to examine autism's entire range of severity in a 65-item questionnaire based on a standardization sample of 1,636 individuals. The SRS focuses on a child's ability to engage in emotionally appropriate reciprocal social interactions. The SRS examines behaviors such as a child's awareness of the emotional and interpersonal cues of others, how they interpret those cues, if they respond appropriately to what he or she interprets, and if they are motivated to engage in social interactions

with others. Early research on the SRS indicates that it produces comparable results to the ADI-R, has strong inter-rater agreement, and is minimally correlated with a child's IQ (Constantino et al., 2003).

Childhood Autism Rating Scale. The CARS (Schopler et al., 1988) is based on behavioral observation and/or interview and can be used with children older than two years of age, including adolescents and adults. This instrument covers common behavioral characteristics of autism within 14 different domains. In addition, there is a 15th overall category that requires a rating of one's impression of the existence and severity of autism. All 15 of the problem areas are rated on a scale of 1-4, 1 representing age appropriate behavior, and 4 representing severe abnormality in age appropriate behavior. Total scores of 37-60 represent "severe autism," scores of 30-36 represent "mild or moderate autism," and scores of 29 or under represent "non-autism." At the time of its publication and for several years afterwards, the CARS was considered one of the best instruments available for the screening of autism (Gillberg et al., 1996).

Gilliam Autism Rating Scale. The GARS (Gilliam, 1995) may used with individuals between the ages of 3 and 22. Gilliam states that the GARS is a parent questionnaire designed to estimate the presence and severity of symptoms of autism. The GARS consists of four scales and 56 total questions. The four scales include Social Interaction, Communication, Stereotyped Behaviors, and Developmental Disturbances. The first three scales rate a child's current, typical behaviors, and the fourth scale rates past severe maladaptive behaviors. A rater does not complete the Communication subscale if the child does not talk, use sign language, or any other form of communication. Individuals completing the questionnaire are asked to rate the frequency

of each behavior on a 4-point scale, from "Never Observed" to "Frequently Observed." The totals of all four scales are summed for an "Autism Quotient" that is the primary summary score for the GARS. The Autism Quotient, explains Gilliam, is intended to measure the likelihood that a child has autism, with higher scores indicating a higher probability of autism.

Research evaluating the GARS has been highly critical. According to Lecavalier (2005) and South et al. (2002), the GARS does not significantly differentiate between children diagnosed with and without Autism Spectrum Disorder due to the instrument's established cutoff point. South et al. (2002) indicated a sensitivity of only .48 and a "lack of convergent validity between the GARS and either expert clinicians or gold-standard research diagnostic measures" (p. 598). Levavalier's (2005) factor analysis of the items on the GARS indicated about half of the items were associated with the factor measuring repetitive and stereotyped behaviors, which the author stated was an overemphasis on such behaviors, given that DSM-IV diagnostic criteria emphasizes social behaviors as most critical to the diagnosis. Sikora, Hall, Hartley, Gerrard-Morris, and Cagle (2008) reported similar findings indicating that the GARS is a poor screener of autism. In their study, the GARS was unable to distinguish between children with autism and other referred children without autism.

Checklist for Autism in Toddlers. The CHAT (Baron-Cohen et al., 1992) is a screener for both parents and practitioners used to help identify the early signs of autism in children 18 months of age. The CHAT assesses the child's attainment of developmental milestones and can be completed in five to ten minutes. There are two parts to the CHAT. Part A is a questionnaire for parents that includes nine yes/no

questions addressing the child's development in the areas of rough and tumble play, social interest, motor development, social play, pretend play, pointing to ask for something, functional play and showing objects. Part B consists of five questions that are recorded by a general practitioner through observations of the child. These five questions address the child's eye contact, gaze monitoring, pretend play, protodeclarative pointing and making a tower of blocks.

In a study involving siblings of children with autism by Baron-Cohen et al. (1992), the CHAT was found to be a reliable instrument as a screener for autism. Results revealed that test-retest and inter-rater reliability were adequate to good. At 18 months, clinically significant scores on the CHAT were only found among children who later developed the full syndrome of autism. In another study by Baron-Cohen (1995), the CHAT was used to detect autism among 1600 children who were not severely mentally retarded and the results revealed a low incidence of false positives.

Aside from the authors of the CHAT, limited research evaluating the instrument is available. Coonrod and Stone (2005) noted that some researchers were adapting the use of the instrument in idiosyncratic ways, making generalization of results difficult. For example, the CHAT has been modified for use with Chinese children (Wong et al., 2004). Additional research and replication of results from the early studies of the CHAT has been requested (Coonrod & Stone, 2005; Scambler, Rogers, & Wehner, 2001).

Behavior Rating Scales as Autism Screeners?

Behavior rating scales are popular tools used by psychologists and are routinely administered as part of a comprehensive evaluation. Some authors have suggested that behavior rating scales may also be important in the screening and diagnosis of autism.

For example, Lord and Corsello (2005) noted that while behavior rating scales should not be used for diagnostic purposes, they have potential value as screening tools for autism. They did not cite, however, any support for their assertion regarding the value of behavior rating scales as screeners.

Merrell (2008) describes behavior rating scales as a way of assessing a child or adolescent's behavior, social skills, and emotional problems and competencies. Behavior rating scales are comprised of numerous questions pertaining to an individual's behavioral characteristics and are presented in a standardized format. An informant that is familiar with the individual, often a parent or teacher, completes the behavior rating scale. Behavior rating scales are an indirect measure of behavior because they are based on an informant's perceptions of a specified behavior rather than a first hand observation of the behavior.

Merrell (2008) reported a number of advantages of behavior rating scales that include: (a) they are inexpensive, (b) they provide data related to low incidence behaviors that might not be seen during a direct observation, (c) they provide more reliable data than unstructured interviews or projective-expressive techniques, (d) they can be used to gather information about individuals who cannot supply information about themselves, (e) they consider behaviors over a period of time in an individual's school or home settings, and (f) they focus on informants' perceptions of behavior who are very familiar with the individuals.

While the use of behavior rating scales offer many advantages, Merrell (2008) also reported disadvantages when used for assessment purposes. Such disadvantages include: (a) not providing actual observational data, (b) rating an individual either

positively or negatively based on a positive or negative characteristic the individual may possess unrelated to the questions within the rating scale, (c) the tendency for raters to respond to questions either too generously or too critically, (d) different rating scales measure slightly different constructs, and (e) different rating scales are based upon different normative populations, and if standardization procedures differ, similar judgments on different rating scales may be interpreted as different results.

There is little research available in the current literature on the use of behavior rating scales as screeners for autism. A thorough review of the literature resulted in only two articles related to this topic. Both articles evaluated versions of the Child Behavior Checklist (CBCL, Achenbach, 1991; Achenbach & Rescorla, 2000) as screeners for autism. The authors of both articles confirmed that little research is available on the topic of behavior rating scales as screeners for autism.

Duarte, Bordin, Oliveira, and Bird (2003) conducted a study comparing three groups of children. These three groups consisted of: (a) 36 children with autism and related conditions, (b) 31 children with either attention deficit hyperactivity disorder, depressive disorder, conduct/oppositional defiant disorder, or separation anxiety/obsessive compulsive disorder, and (c) 34 non-referred schoolchildren used as a control group. The groups of children were matched by age and gender. The children's ages ranged from 4 to 11 years, with an average age range per group of 7 to 7.8 years. Most children in the study were males. The Brazilian adaptation of an earlier version of the CBCL (CBCL/4-18, Achenbach, 1991) was used in their study. Evaluations of the two referred groups of children were conducted in two mental health clinics in Sao Paulo City, Brazil, which included several contacts and interviews with patients and parents.

Parents of the schoolchildren control group were only administered the CBCL/4-18. The study identified CBCL/4-18 scales, and combinations of scales, that might differentiate among the three groups of children in order to compare children with autism to school children and to compare children with autism to children who have Other Psychiatric Disorders (OPD).

The results from Duarte et al.'s (2003) analyses indicated the Thought Problems and Autistic/Bizarre scales were able to distinguish between the group of children with autism and the control group of non-referred schoolchildren. Surprisingly, the Externalizing scores of those two groups (i.e., children with autism and schoolchildren) were reported to be similar. The Thought Problems, Autistic/Bizarre, and Aggressive Behavior scales were all found to distinguish between children with autism from OPD children. The authors noted that the CBCL/4-18's Total scale and the Internalizing scale did not distinguish between children with autism and OPD children. When sensitivity and specificity were examined, the Thought Problems scale had the highest sensitivity (94.3%) and specificity (100%) when the group of children with autism was compared to the control group. When the group of children with autism was compared to the OPD group, it was a combination of the Aggressive Behavior plus the Autistic/Bizarre scales that had the highest sensitivity (91.4%) and specificity (96.7%). Mean scores obtained for each of the three groups on the various CBCL scales were not provided.

Duarte et al. (2003) concluded that their study provides evidence that the CBCL/4-18 can be useful in identifying autism and related conditions among Brazilian children. The study found that the Autistic/Bizarre plus Aggression combination was better at identifying children with autism from children with OPD and that the Thought

Problems scale was better at distinguishing children with autism from typical schoolchildren. However, because this study examined a Brazilian adaptation of an outdated version of the CBCL/4-18, the generalization of these results to the current version of the CBCL with other populations (e.g., children in the U.S.) is severely limited.

Sikora et al. (2008) conducted a study examining ratings from the CBCL/1.5-5 (Achenbach & Rescorla, 2000) for three groups of children: (a) Autistic (n = 79), (b) Autism Spectrum Disorder (ASD; n = 18), and (c) referred, but non-autism spectrum (n = 50). It was not clarified how they distinguished between their Autistic and ASD groups. The 147 children (38 girls, 109 boys) in Sikora et al.'s study ranged in age from 36-71 months (M = 53.54, SD = 10.59). All children participated in an evaluation that included the Autism Diagnostic Observation Schedule - Generic (ADOS-G, Lord et al., 2000).

Sikora et al. (2008) then examined the T scores from the DSM-oriented and Syndrome scales of the CBCL/1.5-5 for the three groups of participants. There were significant differences between the mean scores of the Autistic and Non-Spectrum groups on the CBCL scales of Pervasive Developmental Problems (4.96 T score points), Anxious/Depressed (4.62), Withdrawn (7.37), and Aggressive Behavior (5.68). Sensitivity and specificity were also examined by comparing results from the CBCL to diagnostic classification categories (Sikora et al., 2008). Only the CBCL Withdrawn and Pervasive Developmental Problems scales were used in the sensitivity and specificity analysis. For the Withdrawn scale, the authors found a sensitivity of 64.6% and a specificity of 62.0%. A sensitivity of 79.8% and a specificity of 42.0% were found for the Pervasive Developmental Problem scale. Sikora et al. described the two CBCL scales

as having “fairly poor specificity” (p. 447). Regardless, the authors concluded that, “results from the present study suggest that the CBCL, and the Withdrawn and Pervasive Developmental Problems scales in particular, can be used to screen for ASDs” (p. 448).

As a critique of their interpretation of results, the practical implications of their findings should also be examined. While the mean score differences between the three groups were statistically significantly different, the differences may not be that helpful to practitioners. As previously mentioned, the mean T score differences ranged from 4.62 to 7.37 points, which is not a very large difference (about ½ standard deviation). Furthermore, mean scores for both groups tended to fall in the same classification category. For example, on the Pervasive Developmental Problems scale, the group with autism had a mean T score of 75.04 and the referred, but non-spectrum group had a mean score of 70.08. A T score in the range of 65 to 69 is considered borderline significant and scores 70 or greater are considered clinically significant (Achenbach & Rescorla, 2000). Thus, because both groups had mean scores that were considered clinically significant, it would be difficult to explain to practitioners how to interpret such results on an individual basis.

Purpose of Present Research

In recent decades, there has been a steady increase in the number of children diagnosed with Autism Spectrum Disorder. Autism Spectrum Disorder negatively impacts those affected in many areas, with primary concerns in the areas of socialization and communication. Early identification of this disorder is imperative to the prognosis and outcomes for those with autism. Yet, diagnoses of autism involve thorough and comprehensive assessments conducted by highly trained individuals who have much

experience with autism and the instruments used to diagnosis this disorder. Such assessments can be expensive and highly time consuming. Thus arises the need for a quicker, less involved method to screen for this disorder.

Behavior rating scales have been suggested as potentially viable screeners for autism (Lord & Corsello, 2005). However, only two studies to date have supplied research on behavior rating scales as a method to screen for autism. The study by Duarte et al. (2003) examined an outdated version of the CBCL with a Brazilian population, leaving results difficult to generalize to the United States population. The study by Sikora et al. (2008) examined the CBCL/1.5-5 as a screener for autism, and the results revealed significant differences on a couple of scales between referred groups of children with and without autism. However, the statistically significant differences seem to have little practical value, as mean scores for both groups were in the clinically significant range.

The purpose of the present study is to replicate Sikora et al.'s (2008) findings for the CBCL/1.5-5 as a screener for autism and to expand on those findings by also examining the Clinical Assessment of Behavior – Parent Form (CAB-P, Bracken & Keith, 2004) as a screener for autism. The CAB-P was chosen because it is a more recently published behavior rating scale and it specifically contains an "Autism Spectrum Behaviors" scale. Obtaining evidence indicating that the CBCL/1.5-5 and/or the CAB-P are effective screeners for autism would support future use of these instruments in early childhood settings.

The goals of the present study are to address the following research questions:

1. Do any of the scales on either the CBCL/1.5-5 or the CAB-P show significant differences between groups of referred preschool-aged children with and without autism?

Hypothesis One: The Attention – Deficit/Hyperactivity scale, Autism Spectrum Behaviors scale, and Social Skills scale on the CAB-P will be different between the ASD and Non-Spectrum groups. It is hypothesized that the ASD group will score higher than the Non-Spectrum group on the Attention-Deficit/Hyperactivity scale. This hypothesis is supported by the DSM-IV-TR (APA, 2000) because it describes children with autism as having high levels of ADHD characteristics. Due to the title of the scale, Autism Spectrum Behaviors, it is hypothesized that the ASD group will score higher than the Non-Spectrum group on this scale. It is hypothesized that the ASD group will score lower than the Non-Spectrum group on the Social Skills scale. This hypothesis is also supported by the DSM-IV-TR (APA, 2000) where it is described that children with autism lack interest in relationships with others and the willingness to participate in social activities and initiate social interactions.

Hypothesis Two: The Withdrawn scale, Pervasive Developmental Problems scale, and the Attention Deficit Hyperactivity Problems scale on the CBCL/1.5-5 will be different between the ASD and Non-Spectrum groups. It is hypothesized that the ASD group will score higher than the Non-Spectrum group on the Withdrawn scale and Pervasive

Developmental Problems scale. This hypothesis is supported by research from Sikora et al. (2008) where these scales were found to be significantly higher for their ASD sample. It is hypothesized that the ASD group will score higher than the Non-Spectrum group on the Attention Deficit Hyperactivity Problems scale. As previously mentioned, this hypothesis is supported by the DSM-IV-TR (APA, 2000) where it described children with autism having many ADHD characteristics.

2. How accurately do the CBCL/1.5-5 and CAB-P identify children as having, or not having, autism?

Hypothesis Three: The CBCL/1.5-5 and CAB-P will not accurately identify children as having, or not having, autism. This hypothesis is based on findings from Sikora et al. (2008) where the Withdrawn and Pervasive Developmental Problems scales were found to be statistically significantly different between ASD and Non-Spectrum groups, but the mean scores for both groups were still in the clinically significant range. If there are only a few points difference between the two groups, measures of sensitivity and specificity are not likely to be very helpful. Only scales where a significant difference is found between groups will be evaluated for sensitivity, specificity, positive predictive value, and negative predictive value. Cutoff points at 1.0, 1.5, and 2.0 standard deviations above the mean will be used to evaluate if different levels might result in better screening outcomes.

Method

Participants

Data for this study were obtained from the files at the Regional Child Development Clinics (RCDC) in Bowling Green, Kentucky. RCDC is a non-profit agency that conducts diagnostic evaluations on children from birth through age eight. All participants for this study were previous clients of RCDC who received evaluations of their children because of behavioral (e.g., severe tantrums) or developmental (e.g., delayed speech) concerns. Demographic information was collected for each participant in order to understand the population sampled better and is presented in Table 1. There were a total of 34 participants with Autism Spectrum Disorder (ASD), 28 (82.4%) of which were male, and 6 (17.6%) of which were female. These children were classified as having autism by a psychologist with over 20 years of experience evaluating preschoolers and preschoolers with ASDs. In addition to clinical judgment, the ADI-R was used as part of the assessment of children with an ASD. According to South et al. (2002), assessments by experienced clinicians are a key component in the reliable diagnosis of ASDs. Lord et al. (2000) also stated that the current "gold standard" for diagnosing ASDs involves the use of clinical judgment. Both Constantino et al. (2003) and Lord, Rutter, and Le Couteur (1994) claimed that the ADI-R is a gold standard tool for establishing a clinical diagnosis of autism.

There were also a total of 40 Non-Spectrum participants, 30 (75.0%) of which were male, and 10 (25.0%) of which were female. Specific diagnoses of the children in the Non-Spectrum group were unavailable. At the preschool level, a non-categorical diagnosis system is used for eligibility purposes. The director of RCDC and Western

Table 1

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

	ASD	Non-Spectrum
Gender		
Males	28 (82.4%)	30 (75.0%)
Females	6 (17.6%)	10 (25.0%)
Age (months)		
Mean	43.8	33.1
Range	25-71	25-52
SD	13.9	5.3
Ethnicity		
Caucasian	31 (91.2%)	37 (92.5%)
African American	0 (0.0%)	2 (5.0%)
Hispanic	2 (5.9%)	1 (2.5%)
Asian	1 (2.9%)	0 (0.0%)
Rater of Child		
Mother	34 (100.0%)	36 (90.0%)
Female Guardian	0 (0.0%)	4 (10.0%)
Parent Education		
≤ High School	22 (64.7%)	23 (57.5%)
≥ Some College	12 (35.3%)	17 (42.5%)

Kentucky University's Human Subject Review Board granted permission for this project (see Appendix).

Information was obtained from all files where the age of the child was between 24 months and 71 months and both the CBCL/1.5-5 (Achenbach & Rescorla, 2000) and the CAB-P (Bracken & Keith, 2004) were administered. The mean age of participants with ASD was 43.8 months and the mean age of Non-Spectrum participants was 33.1 months. The 10 month difference between the groups is statistically significant, according to an Independent Samples *t*-test, $t(72) = 4.50$, $p < .001$. The two groups of children were very similar with regard to ethnicity, with over 90% consisting of Caucasian participants. Mothers or female guardians completed the behavior rating scales for all children; no fathers were part of the sample. The level of parent education for both groups was similar, with the majority of parent participants having a high school diploma or less. Thus, the two groups of children and parents were very similar, except that the ASD group was, on average, almost one year older.

Instruments

Clinical Assessment of Behavior. The Clinical Assessment of Behavior (CAB, Bracken & Keith, 2004) is a comprehensive behavior rating scale for parents and teachers intended to assess children and adolescents between the ages of 2 and 18 years. There are three forms of the CAB: the Parent Extended Form (CAB-PX), the Parent Form (CAB-P), and the Teacher Form (CAB-T). The CAB rating forms are intended to be completed by either one or both parents, and/or one or more of the child's teachers. The CAB-PX rating form consists of 170 questions, while both the CAB-P and CAB-T rating forms consist of 70 questions. According to the manual, the CAP-PX rating form can be

completed in 35-40 minutes, while the CAB-P and the CAB-T rating forms require only 10-15 minutes to complete. The CAB has a five-point response format (i.e., Always or Very Frequently, Often, Occasionally, Rarely, and Never) and can be administered individually or in a group setting. To score all three CAB rating forms, the computerized CAB scoring program must be used. This computerized scoring program for the three CAB rating scales generates scale and cluster raw scores, standard (T) scores, percentile ranks, 90% confidence intervals, Clinical Risk and Adaptive Strength classifications, and graphed profiles of the scores. T scores have a mean of 50 and a standard deviation of 10.

The CAB-P was administered to the parents at RCDC and only that version will be described in more detail. The CAB-P provides standardized scores for a number of "scales" and "clusters." The 10 "clinical clusters" include Anxiety, Depression, Anger, Aggression, Bullying, Conduct Problems, Attention Deficit/Hyperactivity, Autistic Spectrum Behaviors, Learning Disability, and Mental Retardation. There are two "adaptive clusters," which are called Executive Function and Gifted and Talented. There are two "emotional disturbance and social maladjustment scales," which are called Emotional Disturbance and Social Maladjustment. The two "clinical scales" consist of the broadband categories of Internalizing and Externalizing behaviors. The two "adaptive scales" are Social Skills and Competence. Finally, an overall score is provided, called the CAB Behavioral Index (CBI). The authors claim the CBI represents the best estimation of the child's overall level of adjustment (Bracken & Keith, 2004).

The standardization of the CAB-P, as described by Bracken and Keith (2004), is based on a sample of 2,114 parents, with 600 parents of children ages two through six

years. Normative data collection for all versions of the CAB occurred in 17 states. Overall, "the proportions of parents with 13 to 15 years education (38.7%) and those with 17 years or more of education (18.8%) in the normative sample were higher than the corresponding proportions in the U.S. population overall (30.1% and 7.8%, respectively)" (Bracken & Keith, 2004, p. 35). The authors go on to note that "the percentage of parents with 11 years or less of education in the normative sample (3.9%) was lower than the corresponding percentage in the U.S. population overall (11.7%)" (p. 35). Thus, on average, parent raters in the CAB normative sample may be more educated when compared to adults of similar age in the U.S. population.

According to Bracken and Keith (2004), there is high internal consistency for the CAB. Overall, the CAB-P scales have coefficient alphas ranging from .89 to .95. Coefficient alphas for the CAB-P clusters ranged from .88 to .94. However, the coefficient alphas were slightly lower for children two to six years of age. For the younger children, the scales ranged from .85 to .94, and the clusters ranged from .82 to .91. To assess test stability, a test-retest rating interval of 11 to 33 days, with a 19-day average, was used with 110 parent raters of children ages 2 to 18. Bracken and Keith (2004) did not state how many were in the two to six year age range. Results revealed no appreciable change in test scores for parents from the first to the second administration over an average of almost three weeks. The corrected test-retest correlations for the various scales and clusters on the CAB-P ranged from .80 to .93 with about half of the correlations in the .80s and half in the .90s. Inter-rater reliability (i.e., mothers vs. fathers) was also assessed for 31 pairs of parents of children between the ages of 2 and

13. The correlations measuring inter-rater reliability on the various scales and clusters ranged from .47 to .74 (Bracken & Keith, 2004).

Concurrent validity studies, as described by Bracken and Keith (2004), were conducted with the CAB-P and the original version of the Behavior Assessment System for Children (BASC, Reynolds & Kamphaus, 1992). Specifically, the parent ratings from a clinical sample of 254 children were used to compare the CAB-P and the BASC Parent Rating Scales. However, only 46 of those children were in the two to six year age range and separate results are not provided for that younger age group. Results found that on corresponding clusters between the CAB-P and BASC, the two instruments were generally comparable on mean T scores. Mean T score differences ranged from 0.04 (aggression scales) to 7.82 (conduct problems scales) points. The results also revealed that correlations between corresponding scales from these two instruments ranged from .53 (anxiety scales) to .81 (total scales). Such results suggest strong to very strong correlations.

Bracken and Keith (2004) describe a second concurrent validity study, using the same sample of children used in the CAB-P and BASC comparisons, which compared parent ratings between the CAB-P and the Devereaux Scales of Mental Disorders (DSMD, Naglieri, LeBuffe, & Pfeiffer, 1994). Results found that on corresponding clusters between the CAB-P and DSMD, the mean T scores for the two instruments were generally comparable. Mean T score differences ranged from 1.08 (anxiety scales) to 5.71 (attention scales) points. The results revealed correlations between corresponding scales from these two instruments ranged from .57 (autism scales) to .79 (attention scales).

Child Behavior Checklist. The *Achenbach System of Empirically Based Assessment* (ASEBA) is an assessment approach used to evaluate adaptive, social/emotional, and maladaptive behaviors in people ranging in age from 18 months to over 90 years (Rescorla, 2005). A widely used component of the ASEBA is the behavior rating scale called the *Child Behavior Checklist* (CBCL). The CBCL contains a variety of forms for different age levels and for different raters (i.e., parents, teachers, caregivers). The focus of this research is the parent version of the CBCL/1.5-5, which is completed by parents of children 18 months through 5 years of age. The CBCL/1.5-5 is a revision of the earlier CBCL/2-3 form (Achenbach, 1992). The CBCL/1.5-5 can be completed within 10-15 minutes. The CBCL/1.5-5 is comprised of a set of 99 behaviors and the rater is to indicate the applicability of each behavior to the child (i.e., 0 = not true, 1 = somewhat true, 2 = very true or very often), based on the prior 2 months (Achenbach & Rescorla, 2000).

Achenbach and Rescorla (2000) describe the various scales and scores on the CBCL/1.5-5. The CBCL/1.5-5 contains seven "syndrome" scales: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, and Aggressive Behavior. The CBCL/1.5-5 also contains five "DSM-oriented" scales (i.e., Affective Problems, Anxiety Problems, Pervasive Developmental Problems, Attention Deficit/Hyperactivity Problems, Oppositional Defiant Problems) where high scores would suggest diagnoses that should be considered by the practitioner. Finally, the CBCL/1.5-5 contains three composite scales (i.e., Internalizing, Externalizing, Total Problems). Summing the ratings from the four internalizing syndromes (i.e., Emotionally Reactive, Anxious/Depressed, Somatic Complaints,

Withdrawn) results in the Internalizing scale. The Externalizing scale is scored by summing the ratings for the two externalizing syndromes (i.e., Attention Problems, Aggressive Behavior). The Total Problems score is generated by summing the ratings from all 99 items. The CBCL/1.5-5 can be scored via a computer software program, which uses gender-specific norms. Scores are presented in terms of T scores and percentile ranks.

The standardization of the CBCL/1.5-5, as described by Achenbach and Rescorla (2000), is based upon a sample of 700 non-referred children who had relatively high problem scores, but who were not referred for mental health or special education services. The sample of children was taken from 40 U.S. states, 2 Canadian provinces, 3 Australian states, and Jamaica. The ethnicity was based on 59% Caucasian, 17% African descent, 9% Latino, and 15% mixed or other. Mothers completed 88% of the forms, fathers 10%, and other adults 2%.

Achenbach and Rescorla (2000) described studies examining reliability data for the CBCL/1.5-5. Test-retest reliability was based upon mothers' ratings of 68 children, who were not referred, on two separate occasions with a mean interval of 8 days. The scales revealed test-retest reliability coefficients ranging from .68 to .92. Across all scales of the CBCL/1.5-5, the mean coefficient was .85. The inter-parent agreement on the CBCL/1.5-5 had a mean reliability coefficient of .61 and the test authors stated there was no bias for one parent over another to report more problems.

When conducting validity studies for the CBCL, Achenbach and Rescorla (2000) used 563 referred children who were matched with 563 non-referred children from their national normative sample on the characteristics of gender, SES, ethnicity, and age. The

referred children scored significantly higher on all the problem scales of the CBCL than did non-referred children. When the percentage of referred and non-referred children were compared according to a dichotomous classification cutoff score, Achenbach and Rescorla reported that 77% of the referred sample scored in the clinically significant range on at least one syndrome scale while 26% of the non-referred sample scored in that range.

As evidence of construct validity, Achenbach and Rescorla (2000) presented a study comparing the older version of the CBCL/2-3 to a British rating scale called the Behavior Checklist. Correlations between the two instruments ranged from .56 to .77. A few studies were conducted with DSM diagnoses and the DSM scales on the CBCL/2-3 and reported by Achenbach and Rescorla (2000). In one such study, the CBCL Externalizing scores were found to correlate .49 with the sum of DSM Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD) symptoms that were assessed by diagnostic interviews with mothers. Another study of DSM diagnoses of disruptive disorders and preschoolers was found to have a correlation of .47 with scores on the CBCL/2-3 Aggressive Behavior Scale.

Procedure

Staff from the Regional Child Development Clinics collected the demographic data and behavior rating scale information from their files where both the CAB-P and CBCL/1.5-5 were completed by parents/guardians of children between 24 and 71 months of age. For each participating child in the current study, the staff from the RCDC deleted the names of these children from their behavior rating scale score sheets and proceeded to make copies. Basic demographic information for these children (i.e., chronological age,

gender, ethnicity), their diagnosis (if autistic), and parents' education level were written on a demographic form and stapled to the score sheets. By doing this, the participants remained anonymous to the investigator. Computer scoring software specifically developed to score the CAB-P and CBCL/1.5-5 by the instruments' publishers was used to score all rating scales. The data were coded into two groups: children with autism and children who were referred, but who were not classified as having autism (Non-Spectrum).

Results

The first research question examined whether or not any of the scales on either the CAB-P or CBCL/1.5-5 revealed significant differences between a group of children with autism and a group of children who were referred, but did not have autism (Non-Spectrum). The mean scores on all scales for both instruments were compared between the two groups. A series of *t*-tests were used for the comparisons and to distinguish if the means from the scales on the instruments were significantly different. Results for the CAB-P scales can be seen in Table 2 and results for the CBCL/1.5-5 scales can be seen in Table 3. All comparisons found significant at the alpha level of $p < .05$ are shown in the tables.

Due to the number of comparisons between scales of the CAB-P and CBCL/1.5-5, however, a more stringent significance level was necessary to avoid finding a significant difference by chance. The Bonferroni correction procedure was used to determine a more restrictive significance level. The Bonferroni procedure divides the .05 significance level by the number of comparisons made. For the CAB-P, there were 19 comparisons made and when .05 was divided by 19, the criterion level was found to be .002. For the CBCL/1.5-5, there were 15 comparisons made and the more stringent criterion level was found to be .003. As a result of the use of the Bonferroni correction procedure, the .002 and .003 significance levels were established. However, none of the comparisons had significance levels between .001 and .003 so the standard alpha level of $p < .001$ was used as the "true" criteria for significance.

Table 2

Mean T Scores on CAB-P Scales for ASD and Non-Spectrum Groups

CAB-P Scale	ASD	Non-Spectrum	*t* values
Anxiety	55.1	53.6	0.7
Depression	53.8	53.5	0.1
Anger	55.3	59.6	-1.6
Aggression	59.1	63.9	-1.6
Bullying	54.2	57.9	-1.6
Conduct Problems	57.2	59.4	-1.1
Attention-Deficit/Hyperactivity	58.4	58.2	0.1
Autistic Spectrum Behaviors	67.3	64.3	1.0
Learning Disability	58.5	55.9	1.4
Mental Retardation	62.0	58.7	1.4
Executive Function	42.8	43.2	-0.2
Gifted and Talented	39.3	41.7	-1.3
Emotional Disturbance	58.3	57.9	0.2
Social Maladjustment	53.7	61.2	-2.3*
Social Skills	37.7	37.8	-0.0
Competence	36.8	40.2	-1.5
Internalizing	55.1	55.3	-0.1
Externalizing	54.9	60.4	-1.9
CAB Behavioral Index	60.2	61.9	-0.6

*$p < .05$.

Table 3

Mean T Scores on CBCL/1.5-5 Scales for ASD and Non-Spectrum Groups

CBCL1/½-5 Scale	ASD	Non-Spectrum	*t* values
Emotionally Reactive	62.5	62.4	0.1
Anxious/Depressed	58.3	57.4	0.6
Somatic Complaints	57.6	55.0	1.6
Withdrawn	73.1	65.7	3.7**
Sleep Problems	56.2	62.1	-2.5*
Attention Problems	63.9	63.4	0.2
Aggressive Behavior	65.8	68.5	-0.8
Affective Problems	60.9	61.6	-0.3
Anxiety Problems	59.5	59.1	0.2
Pervasive Developmental Problems	74.8	68.1	3.8**
Attention Deficit Hyperactivity Problems	62.4	63.1	-0.3
Oppositional Defiant Problems	62.6	67.3	-0.8
Internalizing	64.8	61.2	1.9
Externalizing	64.4	66.7	-0.7
Total Problems	65.9	65.5	0.2

*$p < .05$. **$p < .001$.

For the CAB-P scales, only one significant difference was found between the groups and that difference was only at the alpha level of $p < .05$. At the alpha level of $p < .001$, there were no significant differences in mean T scores between Non-Spectrum and ASD participants on any of the CAB-P scales. Non-Spectrum participants were found to have a significantly higher mean T score on the Social Maladjustment scale than that of ASD participants only if the alpha level of $p < .05$ is used (see Table 2). Although the Attention-Deficit/Hyperactivity and Autistic Spectrum Behaviors scales were hypothesized to be significantly higher for the ASD group, scores were very comparable for the two groups. The Attention-Deficit/Hyperactivity scores were essentially the same for both groups. Surprisingly, the mean Autistic Spectrum Behaviors score was only three points higher for the ASD group than for the Non-Spectrum group and that difference was not statistically significant. The Social Skills scale was hypothesized to be significantly lower for the ASD group but the mean scores for both groups were essentially identical.

Very few differences emerged on the CBCL/1.5-5 scales between the two groups. Non-Spectrum participants were found to have a significantly higher mean T score on the Sleep Problems scale than that of ASD participants, but only at the alpha level of $p < .05$ (see Table 3). The Withdrawn, Pervasive Developmental Problems, and Attention Deficit Hyperactivity Problems scales were hypothesized to be significantly higher for the ASD group. This hypothesis was partially supported. The ASD participants were found to have significantly higher mean T scores than that of the Non-Spectrum participants on the Withdrawn and Pervasive Developmental Problems scales at the alpha level of $p < .001$. However, the Attention Deficit Hyperactivity Problems scale did not

result in a significant difference between mean T scores for the ASD and Non-Spectrum participants. The difference between the mean T scores on the Attention Deficit Hyperactivity Problems scale was less than one point.

The second research question investigated how accurately the CAB-P and the CBCL/1.5-5 identified children as having, or not having, autism. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were examined for all scales where there was a statistically significant difference ($p < .001$) between the two groups of children. Only two scales met that criteria and they were from the CBCL/1.5-5: Withdrawn and Pervasive Developmental Problems.

Sensitivity refers to a test's ability to identify accurately an individual as having a particular disease or disorder (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008). That is, of those determined to have a particular disease or disorder by some other established method, sensitivity is the percent that is correctly identified as having that disorder through the use of the test under consideration. Thus, sensitivity is the percentage of correctly classified true positive cases (Eiraldi, Power, & Karustis, 2000). With regard to the current research, sensitivity refers to the percentage of children with autism that the two CBCL/1.5-5 scales identify as having autism based on a predetermined cutoff score.

In a sense, specificity is the opposite of sensitivity and refers to a test's ability to identify accurately an individual as not having a particular disease or disorder (Parikh et al., 2008). That is, of those determined not to have a particular disease or disorder by some other established method, specificity is the percent that is correctly identified as not having that disorder through the use of the test under consideration. Thus, sensitivity is the percentage of correctly classified true negative cases (Eiraldi, Power, & Karustis,

2000). With regard to the current research, specificity refers to the percentage of non-spectrum children who receive a score below a predetermined cutoff score on the two CBCL/1.5-5 scales.

Predictive value, according to Corsello et al. (2007), refers to the likelihood that a child actually has, or does not have, a particular diagnosis. Parikh et al. (2008) describe PPV as the proportion of individuals that test positive for a specific disease or disorder on the test under consideration that actually have the disease or disorder. NPV is described as the proportion of individuals that test negative for a specific disease or disorder on the test under consideration that actually do not have the disease or disorder. With regard to the current research, PPV takes all participants who scored higher than the predetermined cutoff score on the CBCL/1.5-5 scales and determines the percent that were identified as having autism. NPV takes all participants who scored below the predetermined cutoff score on the CBCL/1.5-5 scales and determines the percent that were not identified as having autism.

The authors of the CBCL/1.5-5 do not claim that a specific score on any scale indicates a child has autism. For the purposes of this research, however, cutoff points at 1.0, 1.5, and 2.0 standard deviations above the mean were selected as the criterion scores on the behavior rating scales to evaluate sensitivity, specificity, and PPV/NPV. These criterion levels correspond to T scores of 60, 65, and 70. One standard deviation above the mean was selected because it is the top of what is considered the average range. The authors of the CBCL/1.5-5 describe 1.5 standard deviations above the mean to be borderline clinically significant and two standard deviations above the mean to be

clinically significant (Achenbach & Rescorla, 2000). Thus, those two cutoff points were chosen as well.

It can be inferred from Kempter and Ritter (1991), Parikh et al. (2008), and Strik, Honig, Lousberg and Denollet (2001) that "excellent" levels of sensitivity, specificity, PPV, and NPV would be at 90% and above, with "good" levels falling between 80% and 89%. Furthermore, "fair" levels of sensitivity, specificity, PPV, and NPV fall between a range of 70% and 79%, and "poor" levels fall between 50% and 69%. "Very poor" levels of sensitivity, specificity, PPV, and NPV fall below 50%. Strik et al. (2001) described that, for screening purposes, higher rates of sensitivity and negative predictive values are more important than high rates of specificity and positive predictive values. With regard to the current research, having high percentages of sensitivity and negative predictive values ensures that few children with autism would be missed, but it would also result in high rates of non-spectrum children being referred for extensive autism evaluations. The results for the sensitivity, specificity, PPV, and NPV for the Withdrawn and Pervasive Developmental Problems scales at all three cutoff points can be seen in Table 4.

It was hypothesized that the scales would not accurately identify children as having, or not having, autism. Support for this hypothesis was mixed. Using a cutoff score of 1.0 standard deviation above the mean, both the Withdrawn and Pervasive Developmental Problems scales had excellent sensitivity, very poor specificity, poor PPV, and excellent NPV. At 1.5 standard deviations above the mean, the Withdrawn scale had good sensitivity, very poor specificity, poor PPV, and good NPV. The Pervasive Developmental Problems scale, at 1.5 standard deviations above the mean, had excellent sensitivity, very poor specificity, poor PPV, and fair NPV. At 2.0 standard

Table 4

Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) for the Significant CBCL/1.5-5 Scales

	Sensitivity	Specificity	*PPV*	*NPV*
1.0 Standard Deviation				
Withdrawn	97.1	27.5	52.4	91.7
Pervasive Developmental Problems	100.0	22.5	52.3	100.0
1.5 Standard Deviations				
Withdrawn	88.2	45.0	57.7	81.8
Pervasive Developmental Problems	91.2	25.0	50.8	76.9
2.0 Standard Deviations				
Withdrawn	82.4	62.5	65.1	80.6
Pervasive Developmental Problems	76.5	50.0	56.5	71.4

Note. Percentages above 90 are considered excellent, 80-89 are considered good, 70-79 are considered fair, 50-69 are considered poor, and below 50 are considered very poor.

deviations above the mean, the Withdrawn scale had good sensitivity, poor specificity, poor PPV, and good NPV. The Pervasive Developmental Problems scale, at 2.0 standard deviations above the mean, had fair sensitivity, poor specificity, poor PPV, and fair NPV. If the intent of using the CBCL/1.5-5 as a screener is to make sure that no child with autism is missed, the cutoff score of 1.0 standard deviations above the mean (T score = 60) fulfills that mission. However, approximately three-fourths of the children without autism are also flagged as possibly having autism by such a cutoff score. Using a more

stringent criterion of 2.0 standard deviations above the mean results in almost one-fourth of the children with autism being missed while half of the non-spectrum children being identified as needing further assessment.

Discussion

The purpose of this study was to examine the use of the CAB-P and the CBCL/1.5-5 as a screener for identifying autism in young children. Supporting evidence indicating that the CBCL/1.5-5 and CAB-P are effective screeners for autism would justify future use of these instruments in early childhood settings. To address this goal, the CAB-P and CBCL/1.5-5 scores for 34 children diagnosed with ASD and 40 referred, but Non-Spectrum children were analyzed.

Data were analyzed in order to examine whether any of the scales on either the CAB-P or CBCL/1.5-5 revealed significant differences between a group of children with an ASD diagnosis and a group of children who were referred for a clinical evaluation, but did not have an ASD. It was hypothesized that the Attention Deficit Hyperactivity scale, Autism Spectrum Behaviors scale, and Social Skills scale on the CAB-P, in addition to the Withdrawn scale, Pervasive Developmental Problems scale, and Attention Deficit Hyperactivity Problems scale on the CBCL/1.5-5 would be different between the ASD and Non-Spectrum groups. When analyzing these data, a stricter criterion for significance was used to avoid Type I errors. When the CAB-P and the CBCL/1.5-5 scales were compared with a stricter value of $p < .001$, only two significant differences emerged and those scales were both on the CBCL/1.5-5: Withdrawn and Pervasive Developmental Problems.

Clinical Assessment of Behavior – Parent Form

There were no significant differences found between the ASD and Non-Spectrum groups on any of the CAB-P scales, indicating the CAB-P was not effective in differentiating between the groups of ASD and Non-Spectrum participants. This finding

is particularly surprising because the CAB-P contains an Autistic Spectrum Behaviors scale. While the mean score for the ASD sample on the Autistic Spectrum Behaviors scale was in the mild clinical risk range, the scale was not able to distinguish between children with autism and other referred children, at least at the preschool level. It is unknown whether similar results would occur at the school-age level. Future research is needed to examine why such a finding occurred.

A cursory review of the 13 items on the CAB-P that contribute to a score on the Autistic Spectrum Behaviors scale indicate six items are related to social-communication skills (e.g., ability to follow directions, turn-taking in conversations), two items are related to stereotypic/repetitive behaviors (e.g., blank stares), and five items address general behavioral problems (e.g., hard to manage, noncompliant). Although speculative, perhaps the reason the Autistic Spectrum Behaviors scale does not distinguish between groups of children with autism and referred but non-spectrum children is the heavy emphasis on general behavioral problems. Furthermore, some of the other items on the Autistic Spectrum Behaviors scale may not be age-appropriate for a very young child (e.g., turn taking in conversations, following multi-step directions, able to take another's perspective, actively participates in discussions). That is, because few two-year-olds would be given credit for such skills, the scale does distinguish between those preschoolers with autism and other referred children without autism.

It is unclear why the CAB-P Attention Deficit/Hyperactivity and Social Skills scales revealed remarkably similar scores between the ASD and Non-Spectrum participants. Granted, the Non-Spectrum group was a referred population, and many of those children's parents probably initiated referrals because of hyperactive behaviors and

poor social skills. However, one obvious explanation is that these scales simply do not discriminate between ASD and Non-Spectrum individuals. Furthermore, these findings do not support statements in the DSM-IV-TR (APA, 2000) stating children with autism show more hyperactive types of behaviors. The lack of a difference between groups may simply be a function of the age of the participants; additional research is needed to see if any differences, particularly hyperactive behaviors, exist at the school-age level.

This study's findings suggest that the CAB-P, as the sole screening instrument, would be of limited clinical value to practitioners evaluating preschoolers with and without autism because the CAB-P does not seem to differentiate between these groups. Thus, if the CAB-P was used as a screener for autism, children with autism may potentially go undetected and consequently not referred for further evaluation of autism. This would mean a delay in the process of a timely identification of autism and the crucial implementation of appropriate early interventions for this disorder. Similarly, children without autism might be referred for further evaluation because of concerns about having autism, when the disorder does not exist.

Child Behavior Checklist/1.5-5

Preschool children with autism were found to have statistically significantly higher scores on the Withdrawn and Pervasive Developmental Problems scales of the CBCL/1.5-5 than the Non-Spectrum group. These findings are consistent with previous research from Sikora et al. (2008) who also found that the Withdrawn and Pervasive Developmental Problems scales showed statistically significant higher mean scores for the group with autism than the non-spectrum group.

Interestingly, the mean T scores on the Pervasive Developmental Problems and Withdrawn scales for Sikora et al.'s Autism and Non-Spectrum samples were remarkably similar to the current samples of participants, even though the mean age of their sample of children with autism was almost 7 months older and their non-spectrum sample was 24 months older than the current samples. For the group with autism, Sikora et al.'s mean score on the Pervasive Developmental Problems scale was 75.0 while the mean for the current sample was 74.8 and, on the Withdrawn scale, Sikora et al.'s mean score was 73.3 while the mean for the current sample was 73.1. For the Non-Spectrum group, Sikora et al.'s mean score on the Pervasive Developmental Problems scale was 70.1 while the mean for the current sample was 68.1 and, on the Withdrawn scale, Sikora et al.'s mean score was 66.0 while the mean for the current sample was 65.7. These findings indicate these two scales on the CBCL/1.5-5 provide very consistent results for varying ages of referred preschoolers with and without autism.

Because the Withdrawn and Pervasive Developmental Problems scales of the CBCL/1.5-5 were found to distinguish between children with autism and children without autism, further analyses were conducted to determine those two scales' sensitivity, specificity, positive predictive power, and negative predictive power. As the standard deviations above the mean (cutoff scores) increased, the sensitivity and negative predictive value of the Withdrawn and Pervasive Developmental Problems scales decreased. However, the sensitivity and negative predictive value were excellent for both the Withdrawn and Pervasive Developmental Problems scales in distinguishing children who had been diagnosed with ASD from Non-Spectrum children, when the cutoff score was at 1.0 standard deviation above the mean.

According to Strik et al. (2001), higher levels of sensitivity and negative predictive value are more important than higher levels of specificity and positive predictive values for screening purposes. High levels of sensitivity and negative predictive value are considered important for screening to make sure anyone with a particular disorder is not missed. Thus, it would seem that the current findings indicate that the CBCL/1.5-5 Withdrawn and Pervasive Developmental Problems scales would be an effective screener for autism when the cut off score is 1.0 standard deviation above the mean.

Levels of specificity and positive predictive value, however, cannot be ignored. The current study resulted in specificity rates at 27.5% for the Withdrawn scale and 22.5% for the Pervasive Developmental Problems scale. Such results indicate that at the 1.0 standard deviation cutoff, approximately three-fourths of the Non-Spectrum group would have also been referred for an autism evaluation. Thus, the use of these scales as a sole screening tool for autism would lead to a high rate of false positives. Put another way, of the 74 children (both groups combined) evaluated as part of this study, 65 of them would have been referred for an additional autism evaluation. With such a low rate of specificity for the CBCL/1.5-5 scales, using this instrument as a screener for autism without any additional information may be of little practical value to those evaluating preschoolers. A high number of children without autism would unnecessarily be referred for a further evaluation of autism, which would also mean an enormous waste of time, financial resources, and effort, not to mention a large number of parents unnecessarily frightened due to the screening results.

Summary

Although very similar results were found with the CBCL/1.5-5 in the current study as compared to Sikora et al. (2008), differences exist in the interpretation of such results. Sikora et al. concluded that the Withdrawn and Pervasive Developmental Problems scales could be used to screen preschool children for autism. However, while both Sikora et al. and the current results identified statistically significant differences, on a practical level this difference may be too small to be clinically useful. The T score point difference between the two groups of children is less than ½ standard deviation. Also, the mean scores for both groups on both scales are significantly high (i.e., at least borderline clinically significant). Furthermore, an examination of the sensitivity and specificity of the Withdrawn and Pervasive Developmental Problems scales suggest that the vast majority of all referred children would be sent on for an additional autism evaluation. It would be hard to imagine how these results in isolation would be useful to a clinician when screening an individual child.

Limitations

Several important factors should be considered when interpreting the results of this study. The sample size for this study included a total of 74 participants from one geographic region of the country. Also, the majority of participants were of Caucasian ethnicity. Such a sample may not be representative of the general population of the United States. With this knowledge, results from this study may not generalize to more diverse samples or other geographic areas.

Another limitation of this study is that all the parent and guardian raters of the CAB-P and CBCL/1.5-5 scales were female. Only mothers and female guardians rated

the children being assessed. It is unknown what outcomes might be a result of ratings by fathers and male guardians. In addition, while the age of Non-Spectrum participants ranged from 25 – 52 months, the age range of ASD participants extended to 19 months older, ranging from 25 – 71 months. The Non-Spectrum group was significantly older; age differences may have influenced the score comparisons of the ASD and Non-Spectrum samples on the two instruments.

Future Research

Research still needs to be conducted to investigate the usefulness of the CAB and CBCL scales as screeners for autism in older children. In the study by Duarte et al. (2003), it was reported that support was found for certain scales on the older version of the CBCL as a useful screening instrument for autism in children aged 4 to 18 years. Future studies could investigate the current version of the CBCL as a screener for autism in children between the ages of 6 and 18. In addition, because the CAB-P did not effectively differentiate between ASD and Non-Spectrum groups of preschoolers within the current study, it would be important to investigate whether the CAB scales would produce similar results for older children. It might also be beneficial for future research to evaluate rigorously the items on the CAB-P Autism Spectrum Behaviors scale. Perhaps there are individual items, or subgroups of items, that are better at distinguishing between children with autism and other referred, but non-spectrum preschoolers.

Furthermore, future research might examine the combination of a behavior rating scale in conjunction with some other type of assessment or screening method as a way to distinguish between children with autism and other referred children without autism. As examples, some type of measure of joint attention, play interests, social awareness, or

social interactions combined with the Withdrawn and/or Pervasive Developmental Problems scales from the CBCL/1.5-5 might reveal key components that could be used to effectively screen for autism in young children.

References

Achenbach, T. (1992). *Manual for the Child Behavior Checklist/2-3.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. 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.

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

Baron-Cohen, S. (1995). *Mind blindness. An essay on autism and theory of mind*. Cambridge: MIT Press.

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

Bellini, S., & Hopf, A. (2007). The development of the Autism Social Skills Profile: A preliminary analysis of psychometric properties. *Focus on Autism and Other Developmental Disabilities, 22*, 80-87.

Bertrand, J., Mars, A., Boyle, C., Bove, F., Yeargin-Allsopp, M., & Decoufle, P. (2001). Prevalence of autism in a United States population: The Brick Township, New Jersey, investigation. *Pediatrics, 108,* 1155-1161.

Blaxill, M. F., Baskin, D. S., & Spitzer, W. O. (2003). Commentary: Blaxill, Baskin, and Spitzer on Croen et al. (2002), The changing prevalence of autism in California. *Journal of Autism and Developmental Disorders, 33*, 223-226.

Bracken, B. A., & Keith, L. K. (2004). *Clinical Assessment of Behavior professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Centers for Disease Control and Prevention. (2006). *Press briefing: CDC releases parent-reported estimates of autism prevalence*. Retrieved October 23, 2007, from http://www.cdc.gov/od/oc/media/transcripts/t060504.htm.

Centers for Disease Control and Prevention. (2007). *Autism information center: Symptoms*. Retrieved June 22, 2008, from http://www.cdc.gov/ncbddd/autism/symptoms.htm.

Charman, T. (2003). Screening and surveillance for autism spectrum disorder in research and practice. *Early Child Development and Care, 173*, 363-374.

Conroy, M. A., Boyd, B. A., Asmus, J. M., & Madera, D. (2007). A functional approach for ameliorating social skills deficits in young children with autism spectrum disorders. *Infants and Young Children, 20*, 242-254.

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

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.

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 rock and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, 48*, 932-940.

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.

Eiraldi, R. B., Power, T. J., & Karustis, J. L. (2000). Assessing ADHD and comorbid disorders in children: The Child Behavior Checklist and the Devereux Scales of Mental Disorders. *Journal of Clinical Child Psychology, 29*, 3-16.

Gernsbacher, M. A., Dawson, M., & Goldsmith, H. H. (2005). Three reasons not to believe in an autism epidemic. *American Psychological Society, 14,* 55-58.

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*. TX: Pro-ed.

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

Lecavalier, L. (2005). An evaluation of the Gilliam Autism Rating Scale. *Journal of Autism and Developmental Disorders, 35*, 795-806.

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. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Lord, C., & Risi, S. (1998). Frameworks and methods in diagnosing autism spectrum disorders. *Mental Retardation and Developmental Disabilities, Research Reviews, 4*, 90-96.

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

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

Lovaas, O. I. (2003). *Teaching individuals with developmental delays: Basic intervention techniques*. Austin, TX: Pro-Ed.

Marker, C., Weeks, M., & Kraegel, I. (2007). Integrating faith and treatment for children with high functioning autism spectrum disorders. *Journal of Psychology and Christianity, 26*, 112-121.

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.

Maurice, C., Green, G., & Luce, S. C. (1996). *Behavior intervention for young children with autism: A manual for parents and professionals*. Austin, TX: Pro-Ed.

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

Naglieri, J. A., LeBuffe, P. A., & Pfeiffer, S. I. (1994). *Devereux Scales of Mental Disorder*. San Antonio, TX: The Psychological Corporation.

National Institute of Mental Health. (2003). *Unraveling autism*. Retrieved November 2, 2007, from http://www.nimh.nih.gov/publicat/unravel.cfm

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

Rescorla, L. A. (2005). Assessment of young children using the Achenbach System of Empirically Based Assessment (ASEBA). *Mental Retardation and Developmental Disabilities Research Reviews, 11*, 226-237.

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

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

Scambler, D., Rogers, S. J., & Wehner, E. A. (2001). Can the Checklist for Autism in Toddlers differentiate young children with autism from those with developmental

delays? *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1457-1463.

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

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.

South, M., Williams, B. J., McMahon, W. M., Owley, T., Filipek, P. A., Shernoff, E., et al. (2002). Utility of the Gilliam Autism Rating Scale in research and clinic populations. *Journal of Autism and Developmental Disorders, 32*, 593-599.

Steuernagel, T. (2005). Increases in identified cases of autism spectrum disorders. *Journal of Disability Policy Studies, 16*, 138-146.

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.

Stuart, S. K., Flis, L. D., & Rinaldi, C. (2006). Connecting with families. Parents speak up about preschool services for their children with autism spectrum disorders. *Teaching Exceptional Children*, *39*, 46-51.

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.

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

Wing, L., & Potter, D. (2002). The epidemiology of autistic spectrum disorders: Is the prevalence rising? *Mental Retardation and Developmental Disabilities Research Reviews, 8,* 151-161.

Wong, V., Hui, L. S., Lee, W., Leung, L. J., Ho, P. P., Lau, W. C., et al. (2004). A modified screening tool for autism (Checklist for Autism in Toddlers [CHAT-23]) for Chinese children. *Pediatrics, 114*, 166-176.

Zwaigenbaum, L., Thurm, A., Stone, W., Baranek, G., Bryson, S., Iverson, J., et al. (2007). Studying the emergence of autism spectrum disorders in high-risk infants: Methodological and practical issues. *Journal of Autism and Developmental Disorders, 37*, 466-480.

Appendix

Human Subjects Review Board Approval Letter

WESTERN KENTUCKY UNIVERSITY
Human Subjects Review Board

In future correspondence please refer to HS09-121, February 16, 2009

Brandy McReynolds
c/o Dr. Carl Myers, Psychology. WKU

Dear Brandy:

Your revision to the research project, Behavior Rating Scales as Screeners for Autism? A Closer Look at the CAB and CBCL/1.5-5, 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 Expedited Review Level until June 1, 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 from 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 McReynolds HS09-121