5-2008

Comparing Parent Ratings of Referred Preschoolers on the Child Behavior Checklist and Behavior Assessment System for Children - Second Edition

Jennifer L. Bour
Western Kentucky University, bourjl@wku.edu

COMPARING PARENT RATINGS OF REFERRED PRESCHOOLERS ON THE CHILD BEHAVIOR CHECKLIST AND BEHAVIOR ASSESSMENT SYSTEM FOR CHILDREN – SECOND EDITION

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
Jennifer L. Bour

May 2008

COMPARING PARENT RATINGS OF REFERRED PRESCHOOLERS ON THE CHILD BEHAVIOR CHECKLIST AND BEHAVIOR ASSESSMENT SYSTEM FOR CHILDREN – SECOND EDITION

Date Recommended _____

_____
Director of Thesis

_____

_____

_____
Dean, Graduate Studies and Research Date

Acknowledgements

I would first like to thank my thesis director, Dr. Carl Myers. I greatly appreciate all your hard work and patience while working on this project. You have provided me with so much constructive feedback and guidance during my undergraduate and graduate years, and for that I am truly thankful! I would also like to thank my thesis committee members, Dr. Elizabeth Jones and Dr. Melissa Hakman, for taking the time to provide me with ideas and helpful suggestions to help better my thesis. I also thank the Regional Child Development Clinics for allowing me to collect my research data. Without their efforts, this research project would have been very difficult. Most importantly, I would like to thank my friends and parents who have been a wonderful support system to help keep me going through graduate school and with this project. I would especially like to thank my sister, Julie, who also went to graduate school ten years ago for School Psychology. She has given me such wonderful guidance and support since she has been through similar experiences. She has been my rock throughout this entire experience.

Table of Contents

Pages

List of Tables iii

Abstract iv

Introduction 3

Literature Review 5

Method 21

Results 28

Discussion 42

References 49

Appendix: Human Subjects Review Board Approval 53

List of Tables

Pages

Table 1: General Features of the *BASC-2 PRS-P* and *CBCL/1.5-5*……. 15

Table 2: Pairs of *BASC-2* and *CBCL* Scales Used for Comparisons…… 27

Table 3: Correlations Between Corresponding Scales on the *BASC-2 PRS-P* *CBCL/1.5-5* (n=95)…… 29

Table 4: Correlations Between Corresponding Scales for the Current Study and for a Non-referred Sample from the BASC-2 Manual…… 30

Table 5: Mean T Scores and *t*-test Results for Comparable *BASC-2* and *CBCL* Scales (n=95)…… 32

Table 6: Classification Consistency of Ratings Between the *BASC-2 PRS-P* and *CBCL/1.5-5*……. 36

Table 7: Percentage of Ratings Less Than or Greater Than One Standard Deviation Between Standard Scores on the Corresponding BASC-2 and CBCL Scales …….. 41

COMPARING PARENT RATINGS OF REFERRED PRESCHOOLERS ON THE CHILD BEHAVIOR CHECKLIST AND BEHAVIOR ASSESSMENT SYSTEM FOR CHILDREN – SECOND EDITION

Jennifer Lynn Bour May 2008 54 pages

Directed by: Dr. Carl Myers, Dr. Elizabeth Jones, and Dr. Melissa Hakman

Department of Psychology Western Kentucky University

It is essential for school psychologists assessing children to use instruments that are reliable and valid. The focus of the current study is to determine whether or not the parent preschool versions of two popular behavior rating instruments, the *Behavior Assessment System for Children – Second Edition* (*BASC-2*; Reynolds & Kamphaus, 2004) and *Child Behavior Checklist* (*CBCL*; Achenbach & Rescorla, 2000), are consistent measures of similarly-named behavioral constructs in preschool-aged children. Parents of 95 preschoolers referred to a nonprofit child evaluation clinic because of behavioral or developmental concerns completed both the *BASC-2* and *CBCL* during an initial evaluation session. The findings revealed that while significantly high correlations occurred, the mean standard scores were significantly different for nine of the 15 construct comparisons. For six of the 15 comparisons, the mean difference was greater than one standard deviation. Additionally, the *CBCL* usually resulted in higher ratings than the *BASC-2*. Such findings imply that the two instruments are not equivalent, and results from these two behavior rating scales need to be interpreted with caution. The determination of whether or not a construct is considered clinically significant may depend solely on the instrument completed by the parent.

Introduction

Behavior rating scales may be one of the most widely used assessment procedures in evaluating social-emotional problems in children. Because evaluations of preschoolers have become a relatively new area of assessment for psychologists (Skovgaard, Houmann, Landorph, & Christiansen, 2004), it is important to evaluate the validity of these types of instruments for use with young children.

Assessing children from the preschool population may present a number of challenges for professionals. Due to their young age and limited independence, it is difficult to obtain accurate information from the child directly. Consequently, parents must play a prominent role in the assessment process. A common way to obtain input from parents, as well as from other informants (e.g., teachers), is through the use of behavior rating scales. These instruments are often used by professionals because they are considered to be time efficient and easily administered (Elliot & Busse, 1993).

Broad-band behavior rating scales measure a variety of constructs, such as internalizing and externalizing behaviors in children. Externalizing behaviors are most often and easily observed by anyone because they are overt behaviors (e.g., hitting, kicking, etc.). In contrast, internalizing behaviors (e.g., anxious or depressive behaviors) are more difficult to observe by professionals and parents. Parents or teachers may not be aware of children's internalizing behaviors until they are asked to rate the specific behaviors on a behavior rating scale. Behavior rating scales are advantageous in this area because they help identify internalizing types of problems that children may be experiencing based on the reports of parents and teachers.

Although there are a variety of methods available for professionals to assess social-emotional behaviors in children, behavior rating scales are commonly used and are efficient for assessing problematic behaviors (Konold, Walthall, & Pianta, 2004). Two common behavior rating scales that are often used in evaluating social-emotional behaviors in school-aged children are the *Behavior Assessment System for Children – Second Edition* (*BASC-2*, Reynolds & Kamphaus, 2004) and the *Child Behavior Checklist* (*CBCL*, Achenbach & Rescorla, 2001). Both scales appear to be known for their sound psychometric properties, but most research examining the two instruments have studied the forms developed for school-aged children. Little research has been conducted on the preschool versions of these instruments.

The preschool versions of the *BASC-2* and *CBCL* assess a number of similarly named constructs. It is important for professionals to know if the two instruments are consistent measures of those constructs. This thesis will discuss the assessment of preschoolers, advantages and disadvantages of behavior rating scales, and research conducted on the *BASC-2* and *CBCL*. The preschool versions of the *CBCL* and *BASC-2* behavior rating scales were evaluated for the current research project by having parents of referred preschoolers complete both instruments. Scores from the *BASC-2* were compared to equivalent scales on the *CBCL* in a variety of ways.

Literature Review

Preschool Assessment

Psychologists often work with children of varying ages; however, most school psychologists are more familiar with various assessment instruments for the school-age population than with the preschool population. The assessment of preschool-age children is considered to be a newer area of assessment in the field of psychology (Skovgaard et al., 2004). With this in mind, it is important for school psychologists to expand their knowledge of assessment practices and measures for a younger population of children. However, there appears to be a limited amount of research and information available pertaining to the psychological assessment of preschool-age children.

There are a number of issues that professionals need to be aware of and take into consideration when evaluating preschool children. For instance, using instruments that are considered to be psychometrically sound for preschool children is essential. Both early identification of problems and beneficial intervention practices for preschool-age children are dependent upon having reliable assessment procedures for screening and identification purposes (Lidz, 2003; Merrell, Blade, Lund, & Kempf, 2003). However, given that preschool children's development is highly variable, many preschool instruments are often less accurate and reliable than instruments utilized with school-age children.

Additionally, it is imperative that no single assessment instrument be used in isolation when making evaluation decisions. Rather, it is considered best practice to use a multimethod assessment approach to gain a full understanding of the child and his or her needs (Lidz, 2003). In conducting a multimethod assessment, school psychologists

obtain information from a variety of sources. Valuable information about the child may be discovered by completing observations in several different environments (e.g., school and home), as well as in conducting interviews with parents and other professionals, such as teachers or daycare workers (Sattler, 2001). However, the most popular forms of obtaining assessment information on young children are through behavioral rating scales and parental interviews. When possible, research supports the importance of obtaining information from both parents using questionnaires and interviews, rather than from only one parent (Bingham, Loukas, Fitzgerald, & Zucker, 2003; Skovgaard et al., 2004). Parents are one of the main sources for referrals, and they provide valuable perceptions about their child's behaviors in the home environment.

Challenges in preschool assessment. A number of challenges may arise when working with preschool-age children. Depending on the age and characteristics of the child, several factors may affect the testing process. For instance, when school psychologists administer standardized tests, preschool children may have a difficult time with standardized directions and procedures due to variability in acculturation. Young children have short attention spans, difficulty staying seated for long periods of time, and difficulty understanding and following verbal instructions, which inhibit optimal performance on standardized tests. With all of these pitfalls in mind, school psychologists need to make an effort to alter the testing environment, as well as their own mannerisms, to make it more suitable for younger children being tested (Sattler, 2001). In addition, the preschool-age child may have limited independence, which leads to a more prominent, supportive role for the parents in the assessment process. Behavior rating scales can help ease the challenges in the assessment process with such a young

population of children because they do not rely on the children's cooperation and comprehension, but more on how raters perceive their behaviors. Behavior rating scales will be discussed in further detail later in this literature review.

Preschool assessment methods. There are a number of assessment methods available for children that measure a variety of behaviors; however, there are some limitations for what may be used with younger children. For instance, self-report inventories and client interviews are common forms of assessment, but these procedures may only be beneficial when working with older children because preschool children are severely limited in their reading ability and in the amount of information they are able to give about their own development (Lidz, 2003).

Standardized testing is a popular form of assessment for young children. There are many standardized tests available for preschool children that allow for a number of comparisons to be made. School psychologists need to be aware of exactly what they are intending to measure before selecting a standardized test because certain instruments measure different constructs. For instance, the *Bayley Scales of Infant Development-Second Edition* is not considered to be an intelligence test but more for measuring developmental milestones in very young children, whereas the *Cognitive Abilities Scale-Second Edition* may be more appropriate for measuring a young child's intellectual functioning (Lidz, 2003). Additionally, standardized assessments also have their pitfalls. As mentioned previously, preschool-age children may exhibit many behaviors that are not conductive to optimal test performance on standardized measures. Young children may also struggle in being able to solely focus their attention to the task at hand.

Psychologists need to take these disadvantages into consideration when evaluating preschool children since these issues may affect their results.

Third party behavior rating scales are another popular form of assessment. A person familiar with the child rates the child's behaviors. These ratings yield scores that allow the comparison of that child's behavior ratings to many other children in the normative sample. Because behavior rating scales are the focus of this thesis, they will be discussed more in-depth in the next section.

Behavior Rating Scales

Behavior rating scales are common assessment tools used to measure the behaviors of children. Rating scales are beneficial for the preschool population due to their ease of administration and their reliance on reports from individuals closely involved with the child. As with any assessment instrument, there are limitations to using behavior rating scales, which are detailed later in this review. Behavior rating scales involve one or more individuals (e.g., teacher, parent) who rate the child's behavior in a particular setting (e.g., school or home environment). The raters should be individuals who have regular contact and interaction with the identified child.

Behavior rating scales are most often used to measure children's social-emotional functioning, which is crucial during the preschool years. Unresolved social-emotional difficulties often continue throughout childhood (Lavigne et al., 2001). Behavior rating scales assist in the early identification of these problems, leading to early interventions that help improve children's level of functioning (Mereydith, 2001). Typically, behavior rating scales use a Likert scale format. By using a Likert scale, raters are required to indicate the extent of their agreement to a given statement based on their perceptions of

the child's behavior. As an example, for a statement like, "Is shy around other children," the rater is able to choose from a selection of ratings based on their level of agreement (e.g., Never, Sometimes, Often, or Almost Always). Compiled ratings on groups of statements that are thought to comprise certain constructs can then be compared to a normative sample to determine how typical or atypical a child's social-emotional functioning is at that point in time.

Constructs of behavior rating scales. Scales can be narrow- or broad-band focused. There are a variety of narrow-band behavior rating scales available that measure specific types of psychological and behavioral constructs. For instance, the *Childhood Autism Rating Scale* (Schopler, Reichler, & Ro, 1988) is specifically designed to assess characteristics of autism, while the *Social Skills Rating System* (Gresham & Elliott, 1990) is used to measure a child's social skills. In addition, there are behavior rating scales that assess a broad range of constructs, called broad-band scales. This type of behavior rating scale typically evaluates a range of internalizing and externalizing types of behaviors in children. As described by Merrell et al. (2003), externalizing behaviors are considered to be overt or excessive behaviors. Aggression, hyperactivity, or delinquent behaviors would be considered externalizing behaviors. In contrast, children with internalizing types of behaviors, such as depression or anxiety problems, can be easily overlooked by teachers and other professionals. Internalizing behavioral symptoms are not easily observed. Because these children do not overtly demonstrate behavioral problems like children with externalizing problems do, they are not as commonly referred for problems within classrooms. Behavior rating scales are considered to be especially beneficial for assessing internalizing types of problems, in the sense that they help identify behaviors

that are not easily observed by professionals (Merrell et al., 2003). Often, professionals and parents may overlook internalizing behaviors when observing children. Using a behavior rating scale helps raters focus on types of behaviors that may not normally be considered by an individual observing a child.

Purposes of behavior rating scales. By assessing a range of psychological and behavioral constructs, broadband behavior rating scales may be used for a number of purposes and have many benefits. A common use for a behavior rating scale is as a screening instrument (Wrobel & Lachar, 1998). By using a behavior rating scale as a screener, professionals can get a quick "estimate" or gain some insight about a child's behaviors. These scales may aid in the determination if a child is at-risk for developing behavioral problems, which may then lead to preventative types of early interventions (Merrell, 2003). As a screener, they may also aid in decision making. After administering a behavior rating scale, a professional may decide whether a more extensive evaluation is needed.

Behavior rating scales are often used as part of a comprehensive evaluation. Behavior rating scales provide standardized ratings of a child's behaviors that contribute to a broader picture of a child's level of functioning. Knowing whether or not ratings on certain behavioral constructs (e.g., hyperactivity, anxiety, depression) are considered clinically significant is especially important when specific behavioral diagnoses are being considered. Merrell (2003) suggests using the "aggregation principle" when utilizing behavior rating scales. This principle recommends obtaining ratings from a variety of individuals from different settings to provide multiple perspectives on a child's problem behaviors. The compiled data is analyzed for consistencies across raters and situations.

Additionally, rating scales may be used by those who implement interventions for children (Elliot & Busse, 1993). Teachers, or other professionals, may be able to use the information found from a behavior rating scale in tailoring interventions. For instance, if the results yield that a student is distant, avoids others in class, and appears to be highly anxious, the teacher may be able to implement an intervention to work on those difficulties. Another purpose for using behavior rating scales is to monitor the effects of interventions (Merrell, 2003). Monitoring a student's progress during a behavioral intervention is important to ensure that the intervention is benefiting the student or if modifications need to be made. Also, using a behavior rating scale to re-evaluate the child's behavior after an intervention has been completed is useful because the information obtained demonstrates if progress has been accomplished and maintained.

Advantages of behavior rating scales. There are a number of advantages for using behavior rating scales, as reported by Elliot and Busse (1993) and Merrell (2003). Behavior rating scales are considered to be efficient, in the sense that they do not require a significant amount of time or effort to administer. Conducting observations or interviews to assess a broad range of internalizing and externalizing behaviors would take an extensive amount of time. In addition, the rating scales are simple to administer and score. Another advantage is that the use of rating scales allows for more involvement from educators and parents with assessment and interventions, because they allow input from multiple sources. They are considered reliable and valid when used to assess the behavior of school-age children (Merrell, 2003).

Behavior rating scales are considered advantageous when assessing children because young children are not able to provide much information about their own

development and behaviors. Additionally, the rating scales may aid in distinguishing between "normal" and "abnormal" levels of behavior in young children. For instance, assessing for Attention-Deficit/Hyperactivity Disorder (ADHD) in young children is considered to be a difficult task, due to the fact that it is hard to distinguish between what may be considered normal and abnormal levels of hyperactivity, impulsivity, and inattentiveness because such behaviors are often seen in preschoolers. Instead of relying solely on personal opinions, psychologists rely on behavior rating scales to provide normative data on the frequency of problem behavior (Sciutto & Terjesen, 2000).

Limitations of behavior rating scales. Behavior rating scales are instruments that are often used to assess behavioral symptoms in children of varying ages. Drotar, Stein, and Perrin (1995) noted several shortcomings. One shortcoming is the generalization and use with cultural groups on which the instruments are not normed. For instance, the *CBCL* has been translated into many different languages and used in a number of cultures, which can invalidate the test. Not every item on the scale may be interpreted in a similar matter from culture to culture. Additionally, the way that children's parents may interpret their child's symptoms may vary among cultures. For example, parents of one culture may have a different idea of when to be concerned for certain behavioral problems than parents from another culture. Differences in levels of concern for certain behaviors may be seen within any culture as well.

Another limitation is that behavior rating scales do not provide any information on a problem's etiology, although, they do provide insight on a student's current level of functioning (Elliot & Busse, 1993). Despite standard scores developed from normative samples, rating scales do not necessarily provide an objective measure of a student's

problem behaviors; rather, the behavior is measured on how the rater perceives the behavior (Konold et al., 2004). As a result, the ratings may be at-risk for being influenced by a number of things, such as the rater having a poor memory or lack of motivation to complete the instrument accurately. Another limitation is a risk of intentional or unintentional bias in the ratings. For instance, a rater (e.g., teacher) may have a preconceived notion of the child, thereby making the ratings biased in a positive or negative sense (Elliot & Busse, 1993). A parent or teacher could intentionally try to make a child more or less likely to qualify for a classification label or services by exaggerating or minimizing their ratings.

In addition to rater bias, there are other notable issues of reliability and validity with behavior rating scales. For instance, interrater reliability is more of an issue with rating scales than it is with cognitive assessments (Elliot & Busse, 1993). This especially becomes an issue when school psychologists obtain ratings from several individuals (e.g., teacher and parents). One rater may perceive a child's behaviors differently than another person rating the same type of behavior. Some behaviors, however, may be situationally specific. A child may actually be behaving differently in different situations (e.g., school and home), which would lead to the discrepancy in ratings on behavior rating scales. It is difficult to know whether differences in ratings are due to differences in the rater's perceptions or the child's actual behaviors. Another issue is that many of the behavior rating scales have similarly named constructs (e.g., aggression, hyperactivity). As a result, it would be easy to conclude that different instruments are measuring the same type of behavior. However, the instruments may actually be measuring different aspects of the similarly-named constructs and, therefore, one should not conclude that similarly

labeled scales on different behavior rating instruments are measuring the "same" behavior (Elliot & Busse, 1993; Frick & Kamphaus, 2001).

Summary. It is evident that behavior rating scales have a number of advantages and disadvantages. By using behavior rating scales, as well as other assessments, school psychologists help aid in the early identification of social-emotional or behavioral problems in young children. Early identification of problematic behaviors is an essential step before being able to implement early intervention strategies. By intervening at a young age, school psychologists may help prevent future negative outcomes for children (Merrell, Caldarella, Streeter, Boelter, & Gentry, 2001).

The BASC-2 and CBCL

There are a variety of instruments available for assessing children's social and emotional behaviors. Two popular instruments are the *Behavior Assessment System for Children: Second Edition* (*BASC-2*, Reynolds & Kamphaus, 2004) and the *Child Behavior Checklist* (*CBCL*, Achenbach & Rescorla, 2001). School-age versions of both assessment instruments are widely used by school psychologists because they exhibit good psychometric properties with the school-age population (Vaughn, Riccio, Hynd, & Hall, 1997). Both instruments were designed for aiding professionals in the evaluation and diagnosis of certain emotional and behavioral disorders in children. Both also contain internalizing and externalizing composite scales that evaluate a variety of internalizing and externalizing types of behaviors in young children. Table 1 provides an overview of general aspects of the two instruments. Because a comparison of the *BASC-2* and *CBCL* is the focus of this thesis, research comparing the two instruments will now be reviewed.

Table 1

General Features of the BASC-2 PRS-P and CBCL/1.5-5

Dimension	BASC-2	CBCL
Age Range	2-5 years	1½ - 5 years
Norms	1,200 children	700 non-referred children
Number of Items	134	99
Rating System	4-point scale (Never, Sometimes, Often, Almost Always)	3-point scale (Not True, Somewhat True, Very True)
Scales	8 "clinical" scales, 4 "adaptive" scales, and 4 composite scores	7 "syndrome" scales, 5 "DSM-Oriented" scales, and 3 composite scores

Few studies compare the current version of the *BASC* (i.e., *BASC-2*) with the *CBCL* and even fewer studies have compared the preschool versions. Vaughn et al. (1997) compared the original versions of the *BASC* (Reynolds & Kamphaus, 1998) and *CBCL* (Achenbach, 1991) to determine their effectiveness in discriminating between the different ADHD subtypes in school-age children. The sample consisted of 73 children with ages ranging from 6.7 to 11.9 years. These children were referred to a neuropsychology clinic that was university-based. The *Parent Rating Scale (PRS)* and *Teacher Rating Scale (TRS)* were used for the *BASC*, and the *Parent Report Form (PRF)* and *Teacher Report Form (TRF)* were used for the *CBCL*. Results indicated that both scales demonstrated the ability to correctly identify children who have ADHD; however, the *BASC PRS* and *TRS* were best for identifying the Inattentive subtype of ADHD.

Another study also evaluated the original *BASC* and *CBCL*'s utility in being able to correctly identify children with or without ADHD, and the subtypes inattentive and combined (Ostrander, Weinfurt, Yarnold, & August, 1998). The sample utilized for this research consisted of 301 children between the ages of 6 and 11 years. In this study, though, only the parent versions of both instruments were utilized. Overall, the results differed from the findings of Vaughn et al. (1997). It was found that the *BASC* was better able to discriminate between children with or without ADHD than the *CBCL* scale. However, when distinguishing between the inattentive and combined subtypes, the *CBCL* was a better predictor for students who may have the inattentive subtype, whereas the *BASC* was better at predicting the combined subtype of ADHD.

Doyle, Ostrander, Skare, Crosby, and August (1997) compared the original school-aged versions of *BASC* and *CBCL*. The main focus of this study was to determine if the *BASC-PRS* was considered to be a comparable instrument to the *CBCL*. The sample consisted of 156 children between 6 and 11 years who were considered to be at-risk for Conduct Disorder. It was determined that the following scales evidenced the strongest relationships: the Aggression scale on both instruments, the *BASC* Conduct Problems and *CBCL* Delinquency scales, and the *BASC* Depression and Anxiety scales with the *CBCL* Anxiety/Depression scale. Overall, however, it was supported that the school-age version of the original *BASC-PRS* was comparable to the previous versions of the *CBCL*.

An unpublished comparison of the preschool versions of the original *BASC* with the current *CBCL* was conducted by Sidebottom (2005), who completed her study with referred preschool-age children. The participants included 50 parents of preschoolers

between the ages of 30 and 71 months, with a mean age of 43.9 months. The participants were referred to a nonprofit child development clinic. This research was interested in a comparison of 13 similarly-named scales on both assessment instruments. The results revealed that 9 of the 13 comparisons did have significantly different means. The Anxiety scale of both instruments was the only scale that yielded a significantly different mean score. Additionally, 11 out of the 13 comparisons resulted in the *CBCL* having higher mean scores than the *BASC*. It was found that all of the scales were significantly and positively correlated and 9 of the 13 comparisons had correlations that were considered to be at least at a moderately strong level (> .50). The comparisons between the somatization, aggression, atypicality, and withdrawn scales had correlations less than .50. Sidebottom hypothesized that all corresponding scales would have differences between the standard scores less than one standard deviation (< 10 points); however, this hypothesis was refuted because 18 to 60 percent of the participants' standard scores did not fall within one standard deviation of each other.

Apparently, only one study has compared the preschool versions of the latest editions of the *BASC-2* and *CBCL* instruments. Research comparing the parent version of the *CBCL* for ages 1.5-5 and the *BASC-2 Preschool Rating Scale-Parent (PRS-P)* was reported in the *BASC-2* manual (Reynolds & Kamphaus, 2004). The comparison used parent ratings of 53 children that were 2 to 5 years old. The results indicated that the similarly-named scales on both instruments tended to have high correlations. For example, the Behavioral Symptoms Index (BSI) from the *BASC-2* and the *CBCL/1.5-5* Total Problems score had a correlation coefficient of .78. The ratings were of typical

children without behavioral problems, which is not representative of the population of preschool children referred for evaluations.

Purpose

The assessment and identification of social and emotional problems in preschool children is becoming more prevalent and is considered a newer area of assessment for psychologists (Skovgaard et al., 2004). There are a variety of methods that professionals use to evaluate preschool-aged children; however, the method given the most consideration in this thesis is the use of behavior rating scales. Behavior rating scales are considered to be beneficial when assessing behavioral problems because they assist in the early identification of social and emotional problems in young children. Additionally, behavior rating scales are advantageous in comparison to other assessment methods for social and emotional behaviors because they can help identify behaviors that are most often difficult to observe (e.g., internalizing behaviors). However, researchers have noted that little information is known about the identification of behavioral problems and the use of rating scales with the early childhood population (Sciutto & Terjesen, 2000).

The most recent editions of the *BASC-2 PRS-P* and *CBCL/1.5-5* have extended their age ranges to include even younger preschoolers; however, there appears to be little evidence to support the use of either scale for preschoolers as young as 18 months of age. Research studies reported in the *BASC-2* and *CBCL* manuals, as well as from reviewed studies, indicated that school-age versions of both instruments demonstrate reliable and valid psychometric properties. However, due to the lack of research for the more recent preschool versions, it is important both behavior rating scales are evaluated to ensure that they are reliable and valid measures for assessing social and emotional problem behaviors

in preschool children. The purpose of the current study was to determine if the *BASC-2 PRS-P* and the *CBCL/1.5-5* were consistent social-emotional measures of preschoolers' problematic behaviors. There appears to be no published research available about the *BASC-2 PRS-P* and the *CBCL/1.5-5* with a referred population of preschool children. For the current study, a comparison between parent ratings of referred preschool children was made between the similarly-named scales from each instrument, which helped determine if the *BASC-2 PRS-P* and *CBCL/1.5-5* reliably and consistently measure corresponding behavioral constructs.

Hypotheses

1. Correlations between the standard scores on the corresponding constructs on the *BASC-2 PRS-P* and the *CBCL/1.5.5* would be positive, significant, and at a moderately strong level, defined as a correlation greater than .50 (Cohen, 1988).
2. Mean scores on each of the corresponding scales on the *BASC-2 PRS-P* and the *CBCL/1.5-5* would not be significantly different from one another. Because both instruments have corresponding scales that measure similar constructs, one would assume that the mean standard scores would all be consistent with both instruments.
3. Corresponding scales from the *BASC-2 PRS-P* and *CBCL/1.5-5* would provide consistent classification outcomes. Scores ≥ 1.5 standard deviations above the mean are considered clinically significant. Thus, scores from corresponding scales from both instruments were evaluated to determine the overall percentage of classification agreement (i.e., both scores < 1.5 standard

deviations above the mean plus both scores ≥ 1.5 standard deviations above the mean). Percentages of 80 or above were considered as acceptable classification consistency (Lidz, 2003).

4. Differences between the standard scores on the corresponding constructs would be less than one standard deviation. Both instruments use T scores, which have a mean of 50 and a standard deviation of 10 points. As previously stated, it was assumed that both instruments would have similar scores on corresponding constructs; therefore, it was expected that the standard scores would fall within one standard deviation of each other.

Method

Participants

The participants for this study were 95 parents or guardians of preschoolers who were referred to a nonprofit child development clinic for a behavioral developmental evaluation. The children ranged in age from 24 to 70 months, with a mean age of 35.1 months (*SD* = 10.9 months). Only one parent or guardian of each child was asked to participate in this study. Typically, the instruments were completed by the children's mothers (81%), while 9.5% were completed by female guardians and another 9.5% completed by the children's fathers. The majority of the preschool children were Caucasian (84%) while 12% were African American, 3% were Hispanic, and 1% was Asian. Additionally, 85.3% of the children were boys, whereas 14.7% were girls. The participants' education level was also ascertained, and 61% had a high school education or less, while 39% had at least some college education. The United States Census Bureau (U.S. Census Bureau, 2008) indicates the obtained demographic data were roughly comparable to the general population in Kentucky where 90.2% are Caucasian, 7.5% African American, 2.0% Hispanic, and 1.0% Asian. The Census Bureau also noted 74.1% of Kentuckians to have a high school degree or less and 17.1% to have a college degree.

Instruments

Behavior Assessment System for Children, 2nd Edition (BASC-2). The *BASC-2* (Reynolds & Kamphaus, 2004) has five different forms or assessment modalities. There are two behavior rating scales (teacher and parent) which are intended to be used for children ages 2 to 21 years. A Self-Report of Personality is also part of the *BASC-2* that is intended to be utilized with children between the ages of 8 and 25 years. A fourth

component of the *BASC-2* is a structured developmental history form. The developmental history is most often completed by a professional, in which detailed background information is obtained from a child's parent or guardian; however, it may also be given to the parent or guardian and completed in the form of a questionnaire. Finally, the *BASC-2* includes a form for observing student behaviors in the classroom setting. This direct observation form is used to observe both positive and negative types of behaviors in a classroom (Reynolds & Kamphaus, 2004).

For the purposes of this study, only the behavior rating scale designed to be completed by parents of preschool children, called the Parent Rating Scale – Preschool (PRS-P) was described and examined. The original *BASC* was developed in 1998 with an age range of 4 to 18 years; however, the most recent version extended its age range. The *BASC-2 PRS-P* was normed with a sample of 1,200 preschool children and is intended for use with children 2 years of age up through 5 years, 11 months. The *BASC-2* uses T scores, which have a mean of 50 and a standard deviation of 10. The *BASC-2 PRS-P* contains 134 behaviors and the parent or guardian rates the occurrence of each item using a 4-point rating system: Never, Sometimes, Often or Almost Always. These behavioral items comprise eight "clinical" scales, which evaluate maladaptive behaviors (i.e., Hyperactivity, Aggression, Anxiety, Depression, Somatization, Atypicality, Withdrawal, and Attention Problems). Additionally, the results from the *PRS-P* provide four "adaptive" scales (i.e., Adaptability, Social Skills, Activities of Daily Living, and Functional Communication), as well as four composite scores (i.e., Externalizing Problems, Internalizing Problems, Adaptive Skills, and the Behavioral Symptoms Index).

Reynolds and Kamphaus (2004) reported that the *BASC-2* has satisfactory reliability estimates for the preschool population. More specifically, for test-retest reliability, a sample of 87 preschool children with a median age of 4 years, 6 months was used to obtain test-retest data. Each child in the sample was rated twice by the same parent/guardian with 9 to 70 days between each rating. The children included in the sample were from the general population or had a previous clinical diagnosis. As noted in the *BASC-2* manual, test-retest reliabilities were adjusted for restriction of range (Reynolds & Kamphaus, 2004). The reliability coefficients for test-retest reliability ranged from the low .80s to the lower .90s. Moreover, when evaluating interrater reliability, a sample of 40 preschoolers with a median age of 3 years, 11 months was assessed by two different parents/guardians. The time that lapsed between ratings was between 0 to 70 days. Interrater reliability coefficients were also adjusted for restriction of range. The estimates for interrater reliability are somewhat lower than the test-retest coefficients. For interrater reliability, the median reliability coefficient yielded for the preschool children was .74.

The *BASC-2* manual (Reynolds & Kamphaus, 2004) contains evidence of adequate internal consistency coefficients for the *PRS-P* for the general norm sample of preschoolers. More specifically, combined internal consistency coefficients for the clinical and adaptive scales for ages 2 to 3 years ranged from .77 to .88, and for ages 4 to 5 years, coefficients ranged from .70 to .87. Combined internal-consistency coefficients for the composite scales for ages 2 to 3 years ranged from .85 to .93, and for ages 4 to 5, coefficicents ranged from .87 to .93.

Child Behavior Checklist (CBCL/1.5-5). The *Achenbach System of Empirically Based Assessment (ASEBA)* is an assessment approach often used by professionals to evaluate a variety of behaviors (i.e., adaptive, social/emotional, maladaptive behaviors, etc.) in people ranging in age from 1½ to over 90 years (Rescorla, 2005). A widely used behavior rating scale, which is a component of *ASEBA*, is the *CBCL*. Like the *BASC-2*, the *CBCL* also contains a variety of forms for parents/guardians and teachers to complete; however, the primary focus for this research is the parent form used for children 1½ through 5 years of age.

The *CBCL* was revised and updated in the year 2000. The original preschool *CBCL* version was for children between the ages of 2 and 4. There was also a version of the *CBCL* that was intended for children between the ages of 4 and 18 years. Currently, the most recent revision of the *CBCL* has extended the preschool version's age range to 1 ½ to 5 years of age (CBCL/1.5-5, Achenbach & Rescorla, 2000). Additionally, empirically-based scales, as well as DSM-oriented scales, were developed for the new version of the *CBCL/1.5-5* (Rescorla, 2005). The *CBCL* yields T-scores, which have a mean of 50, with a standard deviation of 10. Also, only the composite scores on the *CBCL* go down to a T score of 50.

The parent version of the *CBCL/1.5.-5* contains 99 different problems or behaviors, with item 100 allowing the respondent to write about any specific problems that were not previously mentioned in the other items. The parent is required to rate the occurrence of each item using a 3-point rating scale: 0 (Not True), 1 (Somewhat or Sometimes True), and 2 (Very True or Often True). The specific behavioral items comprise seven different "syndrome" scales (i.e., Emotionally Reactive,

Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, Aggressive Behavior), as well as five "DSM-Oriented" scales (i.e., Affective Problems, Anxiety Problems, Pervasive Developmental Problems, Attention Deficit/Hyperactivity Problems, Oppositional Defiant Problems). The *CBCL/1.5-5* also provides scores for the composite areas of Internalizing Problems, Externalizing Problems, and Total Problems.

The *CBCL/1.5-5* was normed using a sample of 700 non-referred preschool children. The manual (Achenbach & Rescorla, 2000) reports test-retest reliability over an 8 day period, which resulted in the majority of the scale correlations being in the .80s and .90s. Additionally, internal-consistency coefficients ranged from .66 to .92 for the syndrome scales, .63 to .86 for the DSM-oriented scales, and .89 to .95 for the composite scales (Achenbach & Rescorla, 2000). The validity of the *CBCL/1.5-5* was demonstrated in the manual by providing evidence of the instrument's ability to accurately distinguish between referred and non-referred children.

Scale Comparisons

For the purposes of this study, fifteen corresponding scales and composites from the *BASC-2* and the *CBCL/1.5-5* were chosen for comparison. The comparisons were made on the basis of the scales' similarity in construct names and/or similarity in the behavioral symptoms measured. For instance, the *CBCL* contains an Aggressive Behavior scale, whereas the *BASC-2* has an Aggression scale. It is important to note that four of the *BASC-2* scales (i.e., Hyperactivity, Attention Problems, Anxiety, and Depression) and two of the *CBCL* scales (i.e., Anxious/Depressed and Attention Deficit/Hyperactivity Problems) were used in multiple comparisons due to the overlap in

construct names from both instruments. The *BASC-2 PRS-P* and *CBCL/1.5-5* scales for the fifteen comparisons are listed in Table 2.

Procedure

Staff from the nonprofit child development clinic located client files where both the *BASC-2 PRS-P* and *CBCL/1.5-5* were completed by parents/guardians of referred preschool-aged children. The staff then copied the score sheets after deleting the child's name from the score sheet. The basic information (chronological age, gender, ethnicity, etc.) was written on a demographic form and stapled to the score sheets. In this manner, test data were given to the investigator without knowledge or access to the children's or parents' names. All of the rating scales were scored using the computer scoring software sold by the tests' publishers. Because the *CBCL* uses gender-specific norms, gender-specific norms were also used when scoring the *BASC-2* protocols to enhance comparability.

Table 2

Pairs of BASC-2 and CBCL Scales Used for Comparisons

BASC-2 PRS-P	CBCL/1.5-5
Hyperactivity	ADHD Problems
Hyperactivity	Attention Problems
Attention Problems	Attention Problems
Attention Problems	ADHD Problems
Aggression	Aggressive Behavior
Anxiety	Anxious/Depressed
Anxiety	Anxiety Problems
Depression	Anxious/Depressed
Depression	Affective Problems
Somatization	Somatic Complaints
Atypicality	Pervasive Developmental Problems
Withdrawal	Withdrawn
Externalizing	Externalizing
Internalizing	Internalizing
Behavioral Symptoms Index	Total Problems

Results

Strength of Correlations – Hypothesis 1

It was hypothesized that the correlations between the standard scores on the corresponding constructs of the *BASC-2 PRS-P* and the *CBCL/1.5-5* would be positive, significant, and at a moderately strong level (> .50). To evaluate this hypothesis, paired-sample correlations between T scores were conducted on the corresponding constructs of both the *BASC-2* and *CBCL* to determine the strength of the relationships between instruments. Correlations between the T scores of all scales on the *CBCL/1.5-5* and on the *BASC-2 PRS-P* are provided in Table 3. Correlations for the 15 comparisons are in bold type in the table and are also listed separately in Table 4. A more stringent *p* value of .001 was used for significance for this particular data analysis to control for Type I error. Results revealed that all 15 comparisons were positively and significantly correlated. Fourteen of the 15 comparisons were considered to be correlated at a moderately strong level of $r > .50$. The comparison that did not achieve a correlation of greater than .50 was the Somatization scale on the *BASC-2* and the Somatic Complaints scale on the *CBCL*. The strongest correlations were achieved amongst two of the three composite scales: Behavior Symptoms Index -Total Problems (.90) and Externalizing (.90). The third composite scale comparison, Internalizing, only received a correlation of .63. Other scales with very strong correlations included the Aggression/Aggressive Behavior (.86) and Attention Deficit Hyperactivity (.79). Additionally, the current study's correlations were compared with correlations between the *BASC-2* and *CBCL* found in the *BASC-2* manual, which used a nonreferred sample. This comparison is

Table 3

Correlations Between Corresponding Scales on the BASC-2 PRS-P and CBCL/1.5-5 (n = 95)

	BASC-2 Scales										
CBCL Scales	1	2	3	4	5	6	7	8	9	10	11
Emotionally Reactive	.69*	.53*	.64*	.47*	.74*	.28	.61*	.28	.67*	.70*	.80*
Anxious/Depressed	.50*	.31	.55*	**.61***	**.59***	.35*	.53*	.22	.67*	.56*	.64*
Somatic Complaints	.29	.16	.31	.41*	.33*	**.44***	.40*	.26	.49*	.32	.40*
Withdrawn	.28	.42*	.17	.05	.25	.05	.71*	**.61***	.17	.23	.52*
Sleep Problems	.41*	.07	.40*	.27	.45*	.50*	.29	.10	.53*	.44*	.42*
Attention Problems	**.75***	**.66***	.54*	.13	.52*	.29	.46*	.08	.44*	.68*	.67*
Aggressive Behavior	.81*	.58*	**.86***	.29	.75*	.31	.54*	.24	.63*	.89*	.87*
Affective Problems	.55*	.39*	.52*	.14	**.61***	.28	.61*	.17	.49*	.57*	.66*
Anxiety Problems	.39*	.22	.44*	**.59***	.40*	.40*	.50*	.37*	.58*	.44*	.54*
Pervasive Dev. Prob.	.38*	.44*	.30	.23	.32	.17	**.73***	.69*	.32	.36*	.63*
Attention Deficit/Hyp.	**.79***	**.69***	.59*	.15	.58*	.20	.42*	.10	.44*	.73*	.70*
Oppositional Defiant	.81*	.56*	.75*	.27	.73*	.28	.49*	.15	.60*	.83*	.80*
Internalizing	.59*	.49*	.54*	.45*	.62*	.35*	.74*	.48*	**.63***	.60*	.79*
Externalizing	.86*	.62*	.84*	.28	.75*	.34*	.56*	.20	.64*	**.90***	.88*
Total Problems	.78*	.59*	.73*	.39*	.74*	.40*	.73*	.35*	.69*	.80*	**.90***

Note. Numbered BASC-2 scales are: 1=Hyperactivity, 2=Attention, 3=Aggression, 4=Anxiety, 5=Depression, 6=Somatization, 7=Atypicality, 8=Withdrawal, 9=Internalizing, 10=Externalizing, 11=Behavior Symptoms Index.
*$p < .001$.

Table 4

Correlations Between Corresponding Scales for the Current Study and for a Non-referred Sample from the BASC-2 Manual

BASC-2 Scale—CBCL Scale	BASC-2 Manual	Current Results
Hyperactivity—ADHD Problems	.79	.79
Hyperactivity—Attention Problems	.78	.75
Attention Problems—Attention Problems	.65	.66
Attention Problems—ADHD Problems	.59	.69
Aggression—Aggressive Behavior	.67	.86
Anxiety—Anxious/Depressed	.32	.61
Anxiety—Anxiety Problems	.39	.59
Depression—Anxious/Depressed	.47	.59
Depression—Affective Problems	.54	.61
Somatization—Somatic Complaints	.56	.44
Atypicality—Pervasive Dev. Problems	.42	.73
Withdrawal—Withdrawn	.42	.61
Externalizing—Externalizing	.83	.90
Internalizing—Internalizing	.68	.63
Behavioral Symptoms Index—Total Problems	.78	.90

provided in Table 4. A visual inspection of the results indicate a few differences between correlations within the *BASC-2* manual and the current results. For several of the corresponding scales, the results found within the *BASC-2* manual had much lower correlations than the current results. For instance, the biggest numerical difference was found for the comparison of the Atypicality and Pervasive Developmental Problems scales, in which the *BASC-2* manual correlation was .42 and the current results achieved a correlation of .73. The Anxiety and Anxious/Depressed comparison also had a substantial discrepancy between correlations, in which the manual reported a correlation of .32 and the current results achieved a correlation of .61. Eleven of the correlations from the current study were higher than those found within the *BASC-2* manual. A possible reason for these discrepancies may be due to the smaller sample size (n = 53) used in the manual. However, the most likely reason is that the sample from the *BASC-2* manual consisted of non-referred children. Thus, there was a restriction in the range of scores they obtained, resulting in lower correlation coefficients.

Consistency of Mean Scores – Hypothesis 2

It was hypothesized that the mean scores on each of the corresponding scales on the *BASC-2 PRS-P* and the *CBCL/1.5-5* would not be significantly different from one another. To evaluate this hypothesis, paired sample *t*-tests were used to compare the mean scores on each of the corresponding scales. Mean scores for the corresponding scales are listed in Table 5. A more stringent *p* value of .01 was used as the criterion for significance for this data analysis to control for Type I errors. Significantly different mean scores were found in 9 of the 15 comparisons. Six of the 9 significantly different

Table 5

Mean T-Scores and t-test Results for Comparable BASC-2 and CBCL Scales (n = 95)

BASC-2 Scale—CBCL Scale	BASC-2	CBCL	*t* values
Hyperactivity—ADHD Problems	64.6	64.0	0.7
Hyperactivity—Attention Problems	64.6	64.8	-0.2
Attention Problems—Attention Problems	64.4	64.8	-0.5
Attention Problems—ADHD Problems	64.4	64.0	0.5
Aggression—Aggressive Behavior	58.0	69.1	-13.4**
Anxiety—Anxious/Depressed	46.5	57.9	-13.3**
Anxiety—Anxiety Problems	46.5	59.7	-14.2**
Depression—Anxious/Depressed	59.7	57.9	1.4
Depression—Affective Problems	59.7	63.2	-2.7*
Somatization—Somatic Complaints	48.7	59.0	-10.1**
Atypicality—Pervasive Developmental Problems	68.9	71.7	-2.7*
Withdrawal—Withdrawn	57.8	68.8	-10.4**
Externalizing—Externalizing	62.4	67.8	-8.3**
Internalizing—Internalizing	52.2	63.5	-11.0**
Behavioral Symptoms Index—Total Problems	66.6	67.4	-1.3

*$p < .01$. **$p < .001$.

mean scores were greater than one standard deviation apart, indicating fairly substantial differences in results.

The corresponding scales that were found to be the most similar were the multiple combinations of the *BASC-2* Hyperactivity and Attention Problems scales with the *CBCL* ADHD and Attention Problems scales. Additionally, the Depression-Anxious/Depressed and the Behavioral Symptoms Index-Total Problems scales from both instruments did not result in significantly different means. The scales that resulted in the largest difference between means were the *BASC-2* Aggression and Anxiety scales with the *CBCL* Aggressive Behavior, Anxious/Depressed, and Anxiety Problems scales.

Another interesting finding is that the mean scores for the *CBCL* were higher than the *BASC-2* for 12 out of 15 comparisons. The *CBCL* truncates the "syndrome" and "DSM-oriented" scales at 50, which means that an individual may not receive a score lower than 50, except for the overall composite scales. The *BASC-2* does not truncate its scores; therefore, this may account for the consistently higher mean scores on the *CBCL*. To test this possibility, all T scores, with the exception of composite scores, were re-coded on the *BASC-2*. Any scores on the *BASC-2* that were below 50 were re-coded to equal 50. Upon re-analysis of the data, *CBCL* mean scores still remained higher than the *BASC-2* truncated mean scores for 11 out of 15 *CBCL* scales. Additionally, after conducting paired sample *t*-tests with the re-coded data, it was found that the scale comparisons that had significant differences at the $p < .001$ value listed in Table 5 still remained significantly different at the $p < .001$ level. Differences emerged where the *p* value was at < .01 in Table 5. Results now revealed that the Depression-Affective Problems and the Atypicality-Pervasive Developmental Problems comparisons were no

longer significantly different. Interestingly, before the truncation, the Depression-Anxious/Depressed comparison was not found to be significantly different; however, after re-coding the *BASC-2* data, this comparison now became significantly different at the $p < .001$ level because the mean for the *BASC-2* Depression scale was significantly higher than the mean for the *CBCL* Anxious/Depressed scale. This post-hoc analysis would suggest that neither of the BASC-2's Depression and Anxiety scales are measuring the same behaviors as the Anxious/Depressed Scale from the CBCL. Overall, it appears the higher scores achieved on the *CBCL/1.5-5* are not due to the truncation of scores, but that the *CBCL/1.5-5* actually results in consistently higher T scores than the *BASC-2 PRS-P*.

Consistency of Ratings – Hypothesis 3

It was hypothesized that the ratings on the corresponding scales from the *BASC-2 PRS-P* and *CBCL/1.5-5* would be consistent. In addition to comparing mean ratings on the two instruments, consistency of results was evaluated by examining classification outcomes on an individual level. Specifically, standard scores for each scale comparison were categorized as clinically significant or not clinically significant based on T scores greater than or less than a criterion score of 65 using cross-tabulation tables. A T score of 65 was chosen because it is 1.5 standard deviations above the mean and is the criterion for "Borderline Significant" on the *CBCL*. The use of a T score of 70 ("Clinically Significant") was not used because it was thought a score two standard deviations above the mean would be too stringent of a criterion score. Data analyses involved calculating the percentage of ratings less than 65 on both instruments, the percentage of ratings at or above 65 on both instruments, and the percentage of ratings where a scale on one

instrument had a score greater than or equal to 65 but there was a corresponding scale with a score less than 65. A total percentage of overall consistency between ratings on the *BASC-2 PRS-P* and *CBCL/1.5-5* was then calculated. Overall consistency consisted of adding the two percentages in which both instruments consistently measured the same construct as either above or below the criterion score of 65. These results are in Table 6. On average, it was found that the *BASC-2* ratings were above the criterion score while the *CBCL* ratings on comparable scales were below the criterion score only 5.8% of the time. In contrast, the *CBCL* ratings were above the criterion score while the *BASC-2* ratings were below the criterion score 19.2% of the time. Thus, these results indicate that the likelihood of obtaining a significantly high score on the *CBCL* but not on the *BASC-2* is 3.3 times more likely than obtaining a significantly high score on the *BASC-2* but not on the *CBCL*.

Using the criterion of acceptable classification consistency of 80 percent or above (Lidz, 2003), only 4 of the 15 comparisons met that criterion level. The Behavioral Symptoms Index-Total Problems (87%) comparison resulted in the highest consistency percentage. Similar results were also found with the Hyperactivity-ADHD Problems (85%), Externalizing-Externalizing (84%), and the Anxiety-Anxious/Depressed (81%) comparisons. These results suggest that there was a higher level of agreement in classification ratings on those scales between the *BASC-2* and *CBCL*. Most comparisons (n = 8) resulted in overall classification consistencies between 70 and 80%. There were three comparisons with overall classification consistencies below 70%: Withdrawal-Withdrawn (57%), Internalizing-Internalizing (60%), and Somatization-Somatic Complaints (69%).

Table 6

Classification Consistency of Ratings Between the BASC-2 PRS and CBCL/1.5-5

BASC-2 Scales	CBCL Scales		
	ADHD Problems		
Hyperactivity	T < 65	T ≥ 65	Consistency
T < 65	43%	4%	85%
T ≥ 65	11%	42%	
	Attention Problems		
Hyperactivity	T < 65	T ≥ 65	Consistency
T < 65	34%	14%	77%
T ≥ 65	9%	43%	
	Attention Problems		
Attention Problems	T < 65	T ≥ 65	Consistency
T < 65	33%	14%	76%
T ≥ 65	11%	43%	
	ADHD Problems		
Attention Problems	T < 65	T ≥ 65	Consistency
T < 65	38%	8%	76%
T ≥ 65	16%	38%	

Table 6 (continued).

BASC-2 Scales	CBCL Scales		
	Aggressive Behavior		
Aggression	T < 65	T ≥ 65	Consistency
T < 65	46%	29%	71%
T ≥ 65	0%	25%	
	Anxious/Depressed		
Anxiety	T < 65	T ≥ 65	Consistency
T < 65	76%	18%	81%
T ≥ 65	1%	5%	
	Anxiety Problems		
Anxiety	T < 65	T ≥ 65	Consistency
T < 65	73%	21%	78%
T ≥ 65	1%	5%	
	Anxious/Depressed		
Depression	T < 65	T ≥ 65	Consistency
T < 65	61%	6%	78%
T ≥ 65	16%	17%	
	Affective Problems		
Depression	T < 65	T ≥ 65	Consistency
T < 65	50%	18%	73%
T ≥ 65	9%	23%	

Table 6 (continued).

BASC-2 Scales	CBCL Scales		
	Somatic Complaints		
Somatization	T < 65	T ≥ 65	Consistency
T < 65	64%	29%	69%
T ≥ 65	2%	5%	
	Pervasive Developmental Problems		
Atypicality	T < 65	T ≥ 65	Consistency
T < 65	19%	23%	74%
T ≥ 65	3%	55%	
	Withdrawn		
Withdrawal	T < 65	T ≥ 65	Consistency
T < 65	34%	40%	57%
T ≥ 65	3%	23%	
	Externalizing		
Externalizing	T < 65	T ≥ 65	Consistency
T < 65	44%	16%	84%
T ≥ 65	0%	40%	
	Internalizing		
Internalizing	T < 65	T ≥ 65	Consistency
T < 65	47%	39%	60%
T ≥ 65	1%	13%	

Table 6 (continued).

BASC-2 Scales	CBCL Scales		
	Total Problems		
Behavioral Symptoms Index	T < 65	T ≥ 65	Consistency
T < 65	34%	9%	87%
T ≥ 65	4%	53%	

Note. Consistency refers to the percentage of agreement where scores from corresponding scales were both either above or below the cutoff score (T = 65).

Standard Score Differences – Hypothesis 4

As yet another way to measure the consistency of ratings on an individual basis, it was hypothesized that the differences between the standard scores on the corresponding constructs of the *BASC-2* and *CBCL* will be less than one standard deviation (< 10 points) apart. To evaluate this hypothesis, the percentage of participants who scored less than, and greater than, one standard deviation between standard scores on corresponding scales from both instruments was determined. Results are presented in Table 7. It was found that all scale comparisons resulted in having standard scores greater than one standard deviation apart. Six scale comparisons had more than half of the standard scores greater than one standard deviation apart: Withdrawal-Withdrawn (54%), Somatization-Somatic Complaints (58%), Aggression-Aggressive Behavior (61%), Internalizing-Internalizing (65%), Anxiety-Anxious/Depressed (66%), and Anxiety-Anxiety Problems (68%). Only three of the scale comparisons had less than 20% of their score comparisons greater than one standard deviation apart: Attention Problems-Attention Problems (16%), Attention Problems-ADHD Problems (14%), and the overall composite comparison of Behavior Symptoms Index-Total Problems (9%). Overall, these results provide additional evidence that the *CBCL/1.5-5* and the *BASC-2 PRS-P* do not consistently measure most similarly-named constructs.

Table 7

Percentage of Ratings Less Than or Greater Than One Standard Deviation Between Standard Scores on the Corresponding BASC-2 and CBCL Scales

	Percent	
BASC-2 Scale—CBCL Scale	< 1 SD	≥ 1 SD
Hyperactivity—ADHD Problems	75	25
Hyperactivity—Attention Problems	73	27
Attention Problems—Attention Problems	84	16
Attention Problems—ADHD Problems	86	14
Aggression—Aggressive Behavior	39	61
Anxiety—Anxious/Depressed	34	66
Anxiety—Anxiety Problems	32	68
Depression—Anxious/Depressed	54	46
Depression—Affective Problems	51	49
Somatization—Somatic Complaints	42	58
Atypicality—Pervasive Developmental Problems	65	35
Withdrawal—Withdrawn	46	54
Externalizing—Externalizing	76	24
Internalizing—Internalizing	35	65
Behavioral Symptoms Index—Total Problems	91	9

Note. ADHD = Attention Deficit Hyperactivity Disorder.

Discussion

The *BASC-2* and *CBCL* are common instruments utilized by school psychologists to evaluate social-emotional behaviors in school-age children. However, most available research examines the forms intended for school-age children. Both instruments have recent revisions, which expanded their age range to include young children. Yet the appropriateness of assessing children as young as 18 months with behavior rating scales has received little attention in the literature. The current study examines the consistency of measurement between corresponding scales on the parent preschool forms of the *BASC-2 PRS-P* and the *CBCL/1.5-5* with a group of clinically-referred preschool children. The results from the current study provided a number of cautions about the use of the two popular preschool behavior rating scales.

Parents or guardians of 95 referred preschool children from a non-profit child development clinic were the participants for this study. The parents completed the preschool forms of the *BASC-2* and *CBCL* at the same point in time. By using this procedure, temporal, source and setting variance were controlled, thus leaving instrument variance as the only explanation for differences in results (Merrell, 2003). A total of 15 corresponding scales from the *BASC-2* and *CBCL* were chosen for comparison. To evaluate these scales, a number of analyses were conducted, including correlations between standard scores and a comparison of mean scores on either scale. Furthermore, the percentage of classification consistency between ratings was also evaluated to determine the level of agreement between the two instruments. The final analysis involved evaluating the size of the differences between standard scores on the corresponding scales for each individual.

The current research examined the strength of correlations between 15 corresponding scales of the *BASC-2* and *CBCL* preschool instruments. It was found that 14 of the 15 comparisons were significantly correlated at a moderately strong level (i.e., $r > .50$). The Somatization-Somatic Complaints correlation was less than .50, which may suggest that the scales are measuring somewhat different constructs. The lack of a moderately strong correlation for the somatization scales was consistent with Sidebottom's (2005) findings comparing the first version of the *BASC PRS-P* with the current version of the *CBCL/1.5-5*.

Test manuals typically provide correlations between instruments as evidence of construct validity. High correlations, however, do not necessarily mean equivalent results and the results from this research provide a good example of that caution. For instance, the aggression scales for both instruments had a very strong correlation (.86), yet were also found to have significantly different mean scores. Additional data analyses suggested the high correlations mask the stark differences between rating scales. An interesting and important finding is that 9 of the 15 comparisons were found to have significantly different mean scores. The differences in mean scores suggest that the instruments provide different results for many of the similarly-named scales. This is an important finding for psychologists, in the sense that the *BASC-2* and *CBCL* scales may not provide the same information for many of the corresponding scales when evaluating preschool children.

On average, the *BASC-2* and *CBCL* are consistent measures of attention and hyperactive behaviors in preschool children. Both instruments achieved strong and significant correlations for the corresponding scales, in addition to having small

differences between mean scores and a high classification consistency. These findings indicate that both the *BASC-2* and *CBCL* are indeed measuring similar constructs for the attention and hyperactivity scales. This finding is important for psychologists since ADHD behaviors are a common referral concern. The *BASC-2* and *CBCL* are also consistent measures of two composite indices: Externalizing behaviors and overall problem behaviors. If the overall problem behavior score is used for special education eligibility purposes, this is an important finding. Both measures seem to provide equivalent overall ratings.

Another interesting finding is that the Internalizing composites of both instruments achieved a low level of classification consistency, which would suggest a low level of agreement between ratings. Thus, Merrell et al.'s (2003) contention that behavioral rating scales are especially beneficial for assessing internalizing types of problems is not supported by these results. Additionally, the research revealed that the *BASC-2 PRS-P* Anxiety and Depression scales and *CBCL/1.5-5* Anxious/Depressed scale did not achieve acceptable and consistent results across most comparisons in this research. As a result, it is difficult to make conclusions about what constructs the *CBCL/1.5-5* Anxious/Depressed scale is measuring. It may prove to be beneficial if the *CBCL*'s Anxious/Depressed scale was separated into two different scales, as in the *BASC-2 PRS-P*, for measuring anxious and depressed behaviors. Separation of the Anxious/Depressed scale into two scales would also be consistent with a recommendation made by Frick and Kamphaus (2001). Frick and Kamphaus criticized the Anxious/Depressed scale on the school-age version of the *CBCL* because they

thought the combination of anxious and depressed behaviors into one scale did "not match current conceptualizations of childhood psychopathology" (p. 194).

Another important finding from the current research is that the *CBCL* consistently resulted in higher scores than the *BASC-2*. More specifically, 12 of the 15 comparisons resulted in higher mean scores on the *CBCL*. Re-coding of *BASC-2* data indicated that the discrepancy between ratings was not due to the *CBCL*'s truncation of scores at 50. Unfortunately, it is unknown which instrument provides more "accurate" scores. The *CBCL* may provide inflated scores or the *BASC-2* may underestimate the severity of behavioral problems. For instance, the *BASC-2* Aggression scale had a mean score (58.0) in the average range of functioning, whereas the *CBCL* Aggressive Behavior scale had a mean score (69.1) at a significantly high level. A psychologist may have a very different interpretation of a child's aggressive behaviors depending on which instrument was administered. These differences in scores between the two instruments have important implications for psychologists. Psychologists must be cautious in their score interpretation. A high score on the *CBCL* may not be that significant or an average score on the *BASC-2* may not reveal significant concerns that truly exist in the home environment. This finding further supports the idea of using a multi-method assessment approach and to not solely rely on the results of behavior rating scales since they may under or over-estimate behavioral problems in children (Lidz, 2003).

Strengths and Limitations

A possible limitation of the current research is the overall representativeness of the obtained sample. The sample only consisted of "referred" children and their parents or guardians, which may not represent the population as a whole. However, the

participants were a sample of "real" referred children and the participants completed the scales under the conditions of an actual evaluation. Such children and circumstances replicate exactly how behavior rating scales will be used by practitioners. Having parents of nonreferred children complete both instruments, as done by Reynolds and Kamphaus (2004), provides questionable validity evidence. Additionally, the three validity indicators on the *BASC-2 PRS-P* protocols indicated a high level of acceptability on the current results. More specifically, the *F* Index, which measures an overly negative response pattern, achieved "acceptable" ratings for 93% of the participants. The Response Pattern and Consistency indicators were rated "acceptable" for 98% of the ratings. A large number of parents who completed the behavior rating scales were mothers; therefore, the fathers may be considered underrepresented. There does not appear to be research examining fathers' consistency of ratings. Another possible limitation may be that it was unknown how well the parents/guardians of the children actually understood the questions on either behavior rating scale. If a parent/guardian had difficulty comprehending what an item was asking, this could alter the obtained results. Parents' and guardians' reading and comprehension abilities were not assessed.

Despite the weaknesses, strengths of the study should not be underscored. The relatively large sample utilized for the current research may be considered a strength of the study. Additionally, the fact that time, rater, and setting variables were controlled is another important strength of this research. It is clear that any differences found between the two instruments could only be attributed to the instruments themselves.

Future Research

A possible area for future research would be to further evaluate the reliability and validity of using the *BASC-2 PRS-P* and *CBCL/1.5-5* for children who have specific diagnoses, like ADHD. In the current study, both instruments were found to be comparable when measuring ADHD-related behaviors. Such research may prove to be beneficial in further evaluating these instruments with preschoolers with ADHD or other diagnoses. Additionally, it would also be useful to design studies that could evaluate the appropriateness (e.g., social validity) of using behavior rating scales with such a young population of children. The most recent revisions of both the *BASC-2* and *CBCL* extended their age range downward so that younger children could be assessed. Parents' and professionals' views on the appropriateness of measuring certain constructs (e.g., anxiety, depression, ADHD) on children as young as 18 months of age should be ascertained. It is unclear what research support is available for measuring such constructs in children so young.

Additionally, it may prove to be beneficial to conduct a more thorough analysis of the individual items on the instruments' scales to evaluate how each instrument defines similarly-named constructs. This analysis may help clarify what behavioral constructs are actually being evaluated. Such clarification would be especially beneficial for the Anxious/Depressed scale on the *CBCL/1.5-5*. Finally, comparisons of the *BASC-2* or *CBCL* with other instruments is another area for future research. For instance, the *Clinical Assessment of Behavior* (*CAB*; Bracken & Keith, 2004) measures similar constructs as the *BASC-2* and *CBCL* in children as young as two years of age. It would

be interesting to see how the *CAB* compares in its measurement of behavioral constructs in referred preschoolers.

Summary

The current research has provided psychologists with a variety of information that may be considered beneficial, in which both consistencies and inconsistencies were identified in each instrument. However, it remains unclear which instrument provides a more accurate measure of behavioral constructs in preschool-aged children. This research has identified evidence that suggests that either instrument is an acceptable measure when evaluating ADHD-related behaviors, externalizing behaviors, or an overall level of problem behaviors. Additionally, it was found that the *CBCL/1.5-5* typically provides more elevated scores than the *BASC-2 PRS-P*, which could influence how a psychologist interprets scores from either instrument when evaluating preschool-aged children.

References

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.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Bingham, C. R., Loukas, A., Fitzgerald, H. E., & Zucker, R. A. (2003). Parental ratings of son's behavior problems in high-risk families: Convergent validity, internal structure, and interparent agreement. *Journal of Personality Assessment, 80*(3), 237-251.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Doyle, A., Ostrander, R., Skare, S., Crosby, R. D., & August, G. J. (1997). Convergent and criterion-related validity of the Behavior Assessment System for Children – Parent Rating Scale. *Journal of Clinical Psychology, 26*(3), 276-284.

Drotar, D., Stein, R. E. K., & Perrin, E. C. (1995). Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology, 24*(2), 184-192.

Elliot, S. N., & Busse, R. T. (1993). Behavior rating scales: Issues of use and development. *School Psychology Review, 22*(2), 313-321.

Frick, P. J., & Kamphaus, R. W. (2001). Standardized rating scales in the assessment of children's behavior and emotional problems (pp. 190-204). In C. E. Walker & M. C. Roberts (Eds.), *Handbook of clinical child psychology* (3rd ed.). New York: Wiley.

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System*. Circle Pines, MN: AGS Publishing.

Konold, T. R., Walthall, J. C., & Pianta, R. C. (2004). The behavior of child behavior ratings: Measurement structure of the Child Behavior Checklist across time, informants, and child gender. *Behavioral Disorders, 29*(4), 372-383.

Lavigne, J. V., Cicchetti, C., Gibbons, R. D., Binns, H. J., Larsen, L., & DeVito, C. (2001). Oppositional Defiant Disorder with onset in preschool years: Longitudinal stability and pathways to other disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 1393-1400.

Lidz, C. S. (2003). *Early childhood assessment*. Hoboken, NJ: John Wiley & Sons, Inc.

Merrell, K. W. (2003). *Behavioral, social, and emotional assessment of children and adolescents* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Merrell, K. W., Blade, R. L., Lund, J., & Kempf, K. K. G. (2003). Convergent and discriminant construct validity of the internalizing symptoms scale for children with the BASC-SRP-C. *Psychology in the Schools, 40*(2), 139-144.

Merrell, K. W., Caldarella, P., Streeter, A. L., Boelter, E. W., & Gentry, A. (2001). Convergent validity of the Home and Community Social Behavior Scales:

Comparisons with five behavior rating scales. *Psychology in the Schools, 38,* 313-325.

Mereydith, S. P. (2001). Temporal stability and convergent validity of the Behavior Assessment System for Children. *Journal of School Psychology, 39*(3), 253-265.

Ostrander, R. Weinfurt, K. P., Yarnold, P. R., & August, G. J. (1998). Diagnosing Attention Deficit Disorders with the Behavioral Assessment System for Children and the Child Behavior Checklist: Test and construct validity analyses using optimal discriminant classification trees. *Journal of Consulting and Clinical Psychology, 66*(4), 660-672.

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. (1998). *Behavior Assessment System for Children*. Circle Pines, MN: AGS Publishing.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children* (2nd ed.). Circle Pines, MN: AGS Publishing.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications (4th ed.).* La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Schopler, E., Reichler, R. J., & Ro, B. (1988). *Childhood Autism Rating Scale.* Circle Pines, MN: AGS Publishing.

Sciutto, M. J., & Terjesen, M. D. (2000, August). *A psychometric review of measures of ADHD in early childhood.* Paper presented at the Annual Meeting of the American Psychological Association, Washington, D.C.

Sidebottom, K. (2005). *A comparison of the BASC and CBCL with referred preschoolers.* Unpublished specialist thesis, Western Kentucky University, Bowling Green.

Skovgaard, A. M., Houmann, T., Landorph, S. L., & Christiansen, E. (2004). Assessment and classification of psychopathology in epidemiological research of children 0-3 years of age: A review of literature. *European Child and Adolescent Psychiatry, 13*(6), 337-346.

U.S. Census Bureau. (2008). Kentucky quickfacts from the U.S. Census Bureau. *U.S. Census Bureau.* Retrieved April 3, 2008, from http://quickfacts.census.gov/qfd/states/21000.html

Vaughn, M. L., Riccio, C. A., Hynd, G. W., & Hall, J. (1997). Diagnosing ADHD (predominantly inattentive and combined type subtypes): Discriminant validity of the Behavior Assessment System for Children and the Achenbach Parent and Teacher Rating Scales. *Journal of Clinical Child Psychology, 26*(4), 349-357.

Wrobel, N. H., & Lachar, D. (1998). Validity of self and parent-report scales in screening students for behavioral and emotional problems in elementary school. *Psychology in the Schools, 35*(1), 17-27.

Appendix

Human Subjects Review Board Approval

WESTERN KENTUCKY UNIVERSITY
Human Subjects Review Board
Office of Sponsored Programs
301 Potter Hall

In future correspondence please refer to HS07-056, October 6, 2006

Jennifer Bour
c/o Dr. Carl Myers
Department of Psychology, WKU

Dear Jennifer:

Your revision to your research project, "Comparing Parent Ratings of Preschoolers on the CBCL and BASC-2," 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 as data is being retrieved from a secondary source; (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 July 31, 2007.

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.

Sincerely,

Sean Rubino, M.P.A.
Compliance Manager
Office of Sponsored Programs
Western Kentucky University