Spring 2017

The Consistency of Teacher Ratings on the Behavior Assessment System for Children-3 and the Child Behavior Checklist 1.5-5

Carly A. Rentsch
Western Kentucky University, carly.rentsch456@topper.wku.edu

THE CONSISTENCY OF TEACHER RATINGS ON THE BEHAVIOR ASSESSMENT SYSTEM FOR CHILDREN-3 AND THE CHILD BEHAVIOR CHECKLIST 1.5-5

A Specialist Project
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

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

By
Carly A. Rentsch

May 2017

THE CONSISTENCY OF TEACHER RATINGS ON THE BEHAVIOR ASSESSMENT SYSTEM FOR CHILDREN-3 AND THE CHILD BEHAVIOR CHECKLIST 1.5-5

Date Recommended Feb. 17, 2017

Dr. Carl Myers, Director of Specialist Project

Dr. Samuel Kim

Dr. Daniel McBride

Dean, The Graduate School

4/4/17
Date

ACKNOWLEDGMENTS

I would first like to thank my thesis director, Dr. Carl Myers, who continued to impress me with his helpfulness and motivation when it comes to my project. During both my undergraduate and graduate years, you have provided me with such valuable information not only through assisting me with my project, but also by offering your time and generosity. I would also like to recognize two other members of my thesis committee, Dr. Samuel Kim and Dr. Daniel McBride. Your constructive feedback and recommendations have helped me to develop a project with adequate detail and understanding. A special thank you to the Head Start teachers who contributed their time and effort into completing each rating scale and making this project possible. I would also like to thank my internship supervisor who provided me with materials and allowed me to use a vast amount of time at my placement to work on this project. Also, thank you to second year School Psychology graduate students, Katie Marcum and Laura Duncan, at Western Kentucky University for assisting with my project and helping administer and score rating scales. Other individuals who have made an impact on my project include my family members, graduate peers, and close friends. I appreciate your willingness to read over my project and provide your feedback as well as compliments.

CONTENTS

LIST OF TABLES

THE CONSISTENCY OF TEACHER RATINGS ON THE BEHAVIOR ASSESSMENT SYSTEM FOR CHILDREN-3 AND THE CHILD BEHAVIOR CHECKLIST 1.5-5

Carly A. Rentsch May 2017 57 Pages

Directed by: Carl Myers, Samuel Kim, and Daniel McBride

Department of Psychology Western Kentucky University

The assessment of children's social-emotional skills, especially in the preschool years, is essential, as it yields early identification of problems and allows for appropriate interventions to be tried. School psychologists and other professionals use a variety of assessment methods (e.g., observations, interviews, behavior rating scales) to determine a child's social-emotional abilities. Two popular behavior rating scales used frequently by professionals are the *Behavior Assessment System for Children-Third Edition* (*BASC-3*) and the *Child Behavior Checklist 1.5-5* (*CBCL 1.5-5*). The current study examines the consistency of results from the two instruments. Fifty-six Head Start teachers from two regions of the country completed both the *BASC-3* and the *CBCL 1.5-5* at the same point of time while thinking of a specific student who displays behavioral concerns. The findings revealed that most of similarly named scales from the two instruments correlated significantly. However, 40% of those comparisons resulted in significantly different mean scores. Approximately half of the comparisons resulted in adequate classification consistency (i.e., either average or clinically significant). Overall, the findings imply that the two instruments do not always measure similarly named behavioral constructs in a consistent manner.

Introduction

Children's ability to regulate emotions and interact with others is related to adjustment in the classroom and academic achievement, making it an important topic for school psychologists. Emotional development includes the knowledge and skills necessary to regulate emotions, whereas social development is how effectively one can interact with others (Epstein, 2009). Although these are two separate terms, they are typically combined into the joint term, social-emotional competence, to represent both social and emotional skills. Social-emotional competence is the ability to understand and adapt to the social and emotional aspects of one's life, allowing one to successfully manage important life tasks such as learning and solving everyday problems (Elias et al., 1997). Social-emotional competence includes self-awareness, impulsivity control, working cooperatively to complete tasks, and caring for the self and others (Elias et al., 1997).

Social-emotional and behavioral based difficulties in preschool children represent a rising concern for early childhood professionals. Many children who show signs of emotional and behavioral problems in early childhood years will continue to have such problems over time (Gimpel & Holland, 2003). Young children who demonstrate social-emotional competence are better able to connect with others and emotionally understand diverse situations. In contrast, children who exhibit negative emotions often are less eager to learn in the classroom and respect those around them. As specific examples, Denham, Bassett, Zinsser, and Wyatt (2014) found that classroom adjustment in preschool predicted emotional regulation in kindergarten, as well as early school success. Another study completed by Nix, Bierman, Domitrovich, and Gill (2013) observed the differences

between two groups of kindergarteners, one who had received a preschool intervention, which promoted literacy and social-emotional skills, and another that had received the regular Head Start instruction. Findings indicated significantly higher scores on vocabulary, emergent literacy skills, and social-emotional abilities for those who received the intervention.

Given the importance of social-emotional skills for positive long-term outcomes, it is essential that social-emotional problems be accurately assessed to increase the likelihood of interventions being applied. It is also important to assess social-emotional skills in children early, preferably in the preschool years, so that any school problems can be addressed before they become entrenched in behaviors that disrupt students' school experiences. When these skills are assessed and addressed, children have a chance to improve in a number of ways, including the understanding of their emotions and interacting more appropriately with others.

This thesis project will discuss the different methods of assessment used to identify characteristics of social-emotional behavior with an emphasis at the preschool level. The project will describe two popular rating scales, the *Behavior Assessment System for Children-3* (Reynolds & Kamphaus, 2015) and the *Child Behavior Checklist 1.5-5* (Achenbach & Rescorla, 2000), which were used in this project. Head Start teachers completed both of the rating scales on children in their classrooms, and the consistency of the similarly named scales from the two rating instruments was evaluated.

The *Behavior Assessment System for Children-3* and the *Child Behavior Checklist 1.5-3* are behavior rating scales that are used routinely by school psychologists and other professionals to assist them regarding special education placement decisions, diagnoses,

or assessment of the severity of behavioral concerns. Therefore, it is essential for professionals to understand the validity and reliability of these instruments. According to Merrell (2008), there is a need for more research studies examining the reliability and validity of behavior rating scales. That way, professionals can be knowledgeable of the quality of the rating scales being used to assist in important decisions. Comparing two rating scales will help determine if they are consistently measuring what they purport to measure and help evaluate the construct validity of the instruments. Also, determining the consistency of similarly named scales from two popular instruments will provide professionals with an understanding of the characteristics of each scale, including any differences between the instruments when completed by the same individuals (Myers, 2013).

This thesis project is designed to address three research questions. The first of these questions is, "How well do similarly named scales on the *Behavior Assessment System for Children-3* and the *Child Behavior Checklist 1.5-5* correlate?" The second question is, "How consistent are mean scores on the similarly named scales from the two instruments?" Finally, this project will address the question, "What is the overall classification consistency (i.e., average vs. clinically significant) of scores between the two instruments on similarly named scales?" Evaluating the results from these three research questions together will provide important information to school psychologists and other professionals regarding the consistency and comparability of the two instruments.

Literature Review

This literature review will provide findings and relevant information from a number of different studies related to social-emotional competence. In addition, this review will provide individuals with an understanding of the different characteristics of social-emotional behavior, including two broad categories of these behaviors, externalizing and internalizing. The benefits and importance of early identification of social-emotional problems will also be discussed. More specifically, the current review will describe the social-emotional characteristics of children who attend Head Start preschools. This literature review will also provide readers with an overview of three different assessment methods to use when assessing preschoolers: interviews, observations, and rating scales. Finally, the review will conclude with a discussion of the advantages and disadvantages of behavior rating scales and similar studies that have compared different behavior rating scales.

Externalizing and Internalizing Behaviors

Social-emotional behaviors can be classified as externalizing or internalizing. Externalizing behaviors are overt behaviors that are noticeable by others, including acting out and aggression (Gimpel & Holland, 2003). Some externalizing behaviors are expected in early childhood and are commonly referred to as the "terrible twos" or "terrible threes." It is not unusual to see noncompliant, active, and aggressive behaviors during the early developmental period (Campbell, Shaw, & Gilliom, 2000). In contrast, internalizing behaviors are inwardly directed, often going unnoticed by others. Internalizing behaviors include socially withdrawn behaviors, such as shyness, nervousness, or fearfulness. When externalizing or internalizing behaviors are excessive,

they might represent the early stages of developmental disorders (American Psychiatric Association, 2013). Externalizing problems are seen in children who have disorders such as Attention-Deficit/Hyperactivity Disorder or Conduct Disorder, while internalizing problems include disorders such as Depression, Anxiety, and Social Phobias (Gimpel & Holland, 2003). An important distinction that can be made between externalizing and internalizing symptoms is that externalizing symptoms are more evident and difficult to overlook, whereas internalizing symptoms are often covert and difficult to detect (Merrell, 2008).

In a school setting, children who exhibit externalizing problems may be getting out of their seat during class time, fighting with classmates, or ignoring classroom rules (Zionts, Zionts, & Simpson, 2002). These behaviors can create problems not only for the child, but also for other individuals in the same environment because such behaviors can be annoying and disruptive (Merrell, 2008). Externalizing behaviors were found to be more evident in young children than internalizing behaviors, perhaps because externalizing behaviors are easier to observe and assess (Burlaka, Bermann, & Graham-Bermann, 2014). It has also been noted that externalizing behaviors in early childhood have received more research attention than internalizing behaviors (Mesman, Bongers, & Koot, 2001). Poor parenting skills and children's temperament have been suggested as factors in the development of externalizing disorders (Campbell, Shaw, & Gilliom, 2000).

Examples of behaviors of a child with internalizing problems include crying, physical complaints, withdrawing from friends, and disinterest in school (Zionts et al., 2002). Although there may be many factors accounting for a child's internalizing

symptoms, a study by Burlaka et al. (2014) was able to identify a few major contributors. Findings indicated that internalizing problems in young children are more likely when there is less social support at home and the parents are less educated. Although it may be rare for a preschooler or kindergartner to demonstrate internalizing problems to the extent that a psychological diagnosis is warranted, clinical interventions may be necessary for high levels of internalizing problems (Gimpel & Holland, 2003).

Early Identification of Social-Emotional Problems

It is important to assess young children for social-emotional problems, preferably before elementary school. Preschoolers' emotional and behavioral problems do not always decline naturally, resulting in more emphasis on prevention and intervention efforts geared toward preschool children (Gimpel & Holland, 2003). Assessing the social-emotional needs of preschool children is advantageous because it allows for early interventions, helping the child to be better prepared for elementary school and making it less likely that they are placed in restrictive settings. According to Squires, Bricker, Heo, and Twombly (2001), identifying any social-emotional problems in young children is critical when it comes to intervening because it also reduces the likelihood of "placement in special education programs, residential treatment, and incarceration later in life" (p. 406). Denham et al. (2003) looked at aspects of emotion and their relation to overall social competence and concluded that assessing emotional competence at the preschool age contributes to both concurrent and future social skills. Preschoolers are at risk for encountering difficult life struggles, such as psychopathology and academic failure, if their social-emotional competence is not assessed and treated successfully (Denham et al., 2012).

A study completed by Rhoades, Warren, Domitrovich, and Greenberg (2011) examined the social-emotional competencies and academic achievement of 341 economically disadvantaged children in a public school across three years: preschool, kindergarten, and first grade. Students who received an emotion curriculum in preschool demonstrated greater social-emotional competencies in first grade, suggesting that preschooler's emotional knowledge is a significant predictor of later academic success. A similar study completed by Zhai, Raver, and Jones (2015) examined the social-emotional development of 414 third graders after being exposed to social-emotional learning services in preschool. Findings indicated that frequent exposure to social-emotional learning opportunities yielded improved social-emotional and academic development in later years including increased social skills, better student-teacher relationships, and stronger academic skills. As suggested by the reviewed studies, exposing preschoolers to social-emotional interventions allows for more positive outcomes in the school age years.

Social-Emotional Skills in Head Start Children

Children from economically disadvantaged communities are at risk for a variety of mental health problems because they are exposed to a large amount of risks (Domitrovich, Cortes, & Greenberg, 2007). A number of studies have been completed suggesting that low-income children are more likely to have developmental delays and externalizing behavior problems (Denham et al., 2012), as well as internalizing problems (Burlaka et al., 2014). A study completed by Webster-Stratton and Hammond (1998) found that base rates of aggressive and disruptive behaviors are notably higher for low-income populations compared to general populations of preschoolers. Research has also been conducted regarding positive and negative emotional expressiveness and regulation

and how it relates to preschoolers' adjustment to school (Herndon, Bailey, Shewark, Denham, & Bassett, 2013). Their findings highlighted the importance of promoting emotional competence in preschoolers, especially when it comes to their use of emotion language to assist in emotion regulation. Assessing the emotional and social behaviors of these at-risk children allows for a better understanding of how they may be influenced by their community and also helps in the preparation of the development of school readiness.

A program known for its devotion to serving low-income preschoolers, who may be vulnerable when it comes to emotional and behavioral problems, is Head Start. Head Start is a federally sponsored early childhood program that is required to serve all children eligible, including those with specific disabilities (Domitrovich et al., 2007). The program works to assist children who might not have the chance to adequately prepare for kindergarten or future education programs at home. The Head Start curriculum emphasizes cognitive and school readiness skills, abiding by the No Child Left Behind legislation (Fantuzzo, Bulotsky-Shearer, Fusco, & McWayne, 2005).

Because early intervention is touted as advantageous, there have been several large-scale intervention projects that have targeted internalizing and externalizing problems of preschoolers enrolled in Head Start programs. These interventions are designed to teach young children how to appropriately interact with others, recognize emotions, and develop social-emotional skills. One such intervention project is Research-based, Developmentally Informed (REDI), a randomized controlled preschool prevention intervention where children are exposed to social-emotional curriculum (Nix et al., 2016). Results of this intervention revealed it is an effective intervention when it comes to later social-emotional functioning into elementary schools because the children demonstrate

more developmentally appropriate social-emotional behavior. Another large-scale intervention is Incredible Years Teacher Classroom Management and Child Social and Emotion Curriculum (Webster-Stratton, Reid, & Stoolmiller, 2008). This intervention includes teachers using the Dinosaur School curriculum, which offers positive classroom management skills and lessons focused on building social-emotional competence (Webster-Stratton et al., 2008). Findings indicated that students displayed more social competence, better emotion regulation, and fewer conduct problems when exposed to this curriculum.

Another large-scale intervention project designed to prevent or reduce behavior and emotional problems as well as build social-emotional competence is the Promoting Alternative Thinking Strategies (PATHS) curriculum (Domitrovich et al., 2007). In a study that evaluated the effectiveness of the PATHS curriculum, findings revealed that preschoolers exposed to PATHS had greater emotion knowledge skills compared to those who were not exposed (Domitrovich et al., 2007). Providing children from disadvantaged lifestyles with the appropriate tools and strategies to promote social-emotional skills can result in favorable and positive outcomes when it comes to their social-emotional development and adjustment.

Preschool Assessment Methods

In order for individual child interventions to begin, assessment methods must be used to identify children with social-emotional concerns (Gimpel & Holland, 2003). There are several ways of assessing children's social-emotional skills, including interviews, direct observations, behavior rating scales, sociometric techniques, self-reports, and projective techniques (Merrell, 2008). However, sociometric techniques,

self-reports, and projective techniques rarely provide useful information at the preschool level (Merrell, 2008) and will not be reviewed in this paper. Instead, interviews, observations, and behavior rating scales will be described. As this specialist project focuses on behavior rating scales, advantages and disadvantages of that assessment method will also be reviewed.

Interviews. Interviews with parents and teachers can provide school psychologists with an accurate picture of a child's emotional and behavioral well-being. The primary complaint or referral for services almost always comes from the adults in the child's life including parents or teachers; therefore, it is important to clarify who has concerns about the child and what those concerns are (Gimpel & Holland, 2003). It is common to interview one or both parents, other family members, early intervention personnel, early childhood educators, and any other individuals who see the child on a regular basis (Noonan & McCormick, 2006).

Parents are desirable sources to interview because they have unique relationships with their child that cannot be duplicated by other informants (McConaughy & Ritter, 2014). During interviews, parents should be asked about family factors and stressors that may exacerbate the child's problems (McConaughy & Ritter, 2014). Examples of topics to include in an interview with a parent are changes in family structure or relationships (e.g., divorce), upsetting events or changes at home (e.g., moving), psychological problems of family members including mental health services for such problems, significant losses experienced by the child (e.g., loss of a loved one or loss of a pet), and medical traumas or serious illnesses of the child or family members (McConaughy & Ritter, 2014). These interviews should also address components such as cognitive and

social functioning, peer relationships, emotional development, and any interests and strengths of the child (Gimpel & Holland, 2003).

Teachers are also helpful informants because they are key sources to school-based problem-solving consultation (McConaughy & Ritter, 2014). Examples of interview questions for a teacher include questions regarding the teacher's primary concerns about the child, the child's academic performance, instructional strategies typically used, interventions tried in the classroom, and any special services that have already been provided to the child (McConaughy & Ritter, 2014). Teachers' perceptions of the child will most likely be different compared to responses from the parents. Teachers are not likely to have background information about the child or be completely aware of what is going on at home. However, interviewing teachers is advantageous because it allows for an understanding of school-based concerns and recognition of academic strengths and weaknesses of the child. The primary disadvantage of the interviewing assessment method is that it is very time consuming to thoroughly interview all relevant individuals and that the quality of information obtained is dependent upon the interviewer's skills at asking appropriate questions.

Observations. Behavioral observations are helpful when evaluating any problems the child may be undergoing at home or at school. Because young children often communicate more through behavior than words, observation serves as a cornerstone of psychological assessment when it comes to social-emotional and behavioral functioning of a child (Gimpel & Holland, 2003). One type of observation used to assess behavioral problems is a systematic observation. Systematic observations involve collecting data under standardized procedures where times and places for observations are carefully

selected (Hintze, Volpe, & Shapiro, 2002). These observations allow for quantitative descriptive information when it comes to behavior and its social and non-social context (Yoder & Symons, 2010). Systematic observations have also been used to measure children's behavioral regulation over time in research studies (e.g., Ponitz, McClelland, Matthews, & Morrison, 2009).

The second type of observation used to assess behavioral problems is termed naturalistic observation. In contrast to systematic observations, naturalistic observations are observations where the observer enters specific settings and observes without any predetermined behaviors in mind (Hintze et al., 2002). An advantage of this type of observation is that the relationship of the child's behavior with important environmental antecedents and consequences can be examined (Hintze et al., 2002). Using naturalistic observations has become common as part of an overall assessment strategy, and provides data that suggest how systematic observations could be conducted (Hintze et al., 2002). When it comes to preschoolers, observations in natural settings is often preferred because it directly measures the behavior of interest and provides data that are less likely to be distorted by the expectations and biases of parents and teachers (Merrell, 2008). However, caution should be exercised when using observations on preschoolers. The social-emotional behavior of preschoolers is often inconsistent and may change quickly in response to situational variables (Merrell, 2008). Thus, it is important to conduct several observations on young children in order to obtain a more accurate interpretation of his or her behavior. Like interviews, the quality of the observation data is dependent on the skills of the observer.

Behavior rating scales. Another way to assess social-emotional behavior is by the use of behavior rating scales. Behavior rating scales are instruments completed by appropriate individuals, usually parents or teachers, who provide ratings of specific behaviors based upon their previous observations and interactions with the child (Campbell & Hammond, 2014). Behavior rating scales are useful when assessing preschoolers' social-emotional skills because they allow for identification of specific problems that might not be revealed during observations or interviews. There are two general types of instruments when it comes to rating scales: narrowband measures and broadband measures. Narrowband rating scales are instruments that assess a child's level of functioning in a specific area of behavior (Shapiro & Kratochwill, 2000). For example, the *Child's Depression Inventory* is a scale designed to measure behaviors associated with depression. In contrast, broadband rating scales are used to assess multiple domains of behavior (Achenbach, McConaughy, & Howell, 1987). The two behavior rating scales reviewed in this project, the *Behavior Assessment System for Children, 3rd Edition (BASC-3,* Reynolds & Kamphaus, 2015*)* and the *Child Behavior Checklist 1.5-5* (*CBCL 1.5-5,* Achenbach & Rescorla, 2000*),* are both broadband instruments because they provide an assessment of many dimensions of behavior.

Behavior rating scales are less direct than observations of behavior or interviewing because they measure perceptions of specific behaviors, rather than a direct measure of behaviors (Merrell, 2008). Behavior rating scales typically use a Likert scale format, where raters record their perceptions of the frequency of specific behaviors. These ratings are then compared to a normative sample, allowing for the child's behavior to be compared to the behavior of other same aged children. The results obtained from

the rating scales allow school psychologists to review assessment data on a wide range of behavioral constructs, helping them to understand a child's emotional state. It has been reported that school psychologists typically use behavior rating scales as part of their evaluations (Shapiro & Heick, 2004).

Disadvantages of Behavior Rating Scales

Behavior rating scales have been described as efficient, effective, and simple, making them a popular choice when it comes to assessing behavior problems (Merrell, 2008). However, one must be cautious and familiar with the disadvantages of using behavior rating scales. One disadvantage is that behavior rating scales do not identify the etiology of the problems or the functions of inappropriate behaviors (McConaughy & Ritter, 2014). For example, a rating scale might allow a psychologist to identify that a child is easily frustrated and is quick to become emotional after events. However, the scale will not assist in finding the reason for any emotional sensitivity. Also, behavior-rating scales do not dictate choices for designing interventions, and it is vital to collect other data to make such decisions (McConaughy & Ritter, 2014).

Another disadvantage of behavior rating scales is that they can produce variation in the reliability and validity of a measure depending on the time element involved with making the rating (Merrell, 2008). Ratings made at different points in time tend to decrease reliability of results. Furthermore, there is a tendency for recent events and recent behaviors to be rated more drastically by raters (Worthen, Borg, & White, 1993). For example, temper tantrums often occur with many toddler- and preschool-aged children. If a parent completed a rating scale after a recent tantrum, scores on externalizing scales would likely be more elevated than if the rating scale was completed

after a period of calm behaviors. When interpreting the results of behavior rating scales, a professional should be aware of unusual events going on within the household and at school. Assessing the child at a vulnerable time can produce inaccurate or misleading results.

Merrell (2008) notes that another disadvantage of behavior rating scales is related to the construction of the rating format. He goes on to describe that most rating scales use a Likert format where the rater indicates the frequency of specific behaviors (e.g., never, sometimes, often, almost always). This Likert format often makes it difficult to obtain consistent responses. For example, it is unlikely that everyone will interpret the word, sometimes, the same. One individual might believe that sometimes means that a behavior rarely ever occurs and another individual might interpret sometimes as the behavior occurring a few times a month. Parents need a certain level of reading ability to complete the instruments. Furthermore, it may also be difficult for raters to understand what behaviors the descriptors on the rating scales entail. Raters can interpret descriptors such as oppositional, defiant, and withdrawn differently, especially across cultures. More accurate ratings are obtained when tangible and understandable definitions are provided (Merrell, 2008).

Although rating scales are generally considered reliable and valid, they are not truly objective measures of problems because they are just based on the informants' perceptions (McConaughy & Ritter, 2014). Informants, especially parents and teachers, can perceive behaviors differently. For example, a special education teacher who routinely encounters problem behaviors might rate a child's behavioral characteristic as fairly mild while a parent who is not used to that same behavioral characteristic may rate

it as fairly severe. Indeed, research has demonstrated that parents and teachers often provide significantly different ratings on the same children, with parents' ratings higher for both adaptive scales and problem behavior scales (Major, Seabra-Santos, & Martin, 2015). Furthermore, a parent or teacher may intentionally or unintentionally record biased responses. For example, a teacher may be inclined to provide negative responses about a child because of an incident that upset her involving the child. Also, a parent may provide inaccurate responses on the rating scale, in an effort to obtain or avoid special education services for his or her child. Another disadvantage is that a set of parents or teachers may provide different responses on a rating scale depending on the amount of time they spend with the child. For example, one parent may be home with the child each week whereas the other may be traveling for business, or one teacher might be the homeroom teacher and another may only see the student for one period per day.

Advantages of Behavior Rating Scales

Although many disadvantages have been described, behavior rating scales have shown to be advantageous in several ways when assessing children for social-emotional problems. One advantage is that behavior rating scale data are valuable for school psychologists when determining a child's eligibility for specific school-based services (Campbell & Hammond, 2014). Other advantages noted by McConaughy and Ritter (2014) include: (a) information on the scales is quantifiable, reliable, and valid, (b) scales are composed of multiple items that yield a broad range of potential problems, and (c) information is organized in a systematic way by groupings of syndromes. Perhaps most importantly, norms provide a standard for judging the severity of problems by providing a comparison to a large national sample of children the same age and gender

(McConaughy & Ritter, 2014). As such, results from a behavior rating scale demonstrate the severity level of a child's behaviors on behavioral constructs (e.g., aggression) by providing a normative comparison to other children, allowing for ease of understanding for parents and teachers.

Behavior rating scales also provide data on low frequency, but important, symptoms that might not be noticed through direct observation (Merrell, 2008). One example is aggressive behaviors. Such behaviors might not be present during an observation, but are considered on rating scales. There is also the advantage of time efficiency and low cost. Each rating scale may only cost a couple of dollars and informants can complete ratings in less than twenty minutes (Campbell & Hammond, 2014). The use of behavior rating scales has been recognized as valuable and cost effective, especially when it comes to identifying social-emotional functioning in large groups of children (Lidz, 2003).

Another advantage of using behavior-rating scales is that individuals of close contact to the child complete the scales. Respondents are generally persons that are familiar with the child's development and behavior and have also seen him or her interact in a naturalistic environment (Campbell & Hammond, 2014). Because young children are not able to provide a great amount of information about their own development and behavior, it is usually more appropriate to obtain this information from a close source. Rating scales provide judgments and observations from persons highly familiar with the child's behavior (Merrell, 2008).

It is important for a school psychologist to be familiar with the advantages and usefulness of behavior rating scales, while also being cautious of any potential problems.

It is also vital that several different methods be used during assessments in order to make the most accurate, fair interpretations. Rating scales should only be part of a multi-method, multi-source, multi-setting assessment designed to obtain relevant information about a child (Gimpel & Holland, 2003).

Comparability of Scales

One way to determine the validity of a behavior rating scale is to compare its ratings to other established rating scales that measure the same behaviors. When comparing two scales that measure the same construct, a strong correlation is anticipated. For example, in a study completed by Myers, Bour, Sidebottom, Murphy, and Hakman (2010), scores on the parent versions of the *Behavior Assessment System for Children-2* and the *Child Behavior Checklist 1.5-5* behavior rating scales were compared on clinically-referred preschoolers. The study resulted in positive strong correlations (r = .60's – .80's) for some of the scales on the two instruments, specifically those related to attention and hyperactivity. These findings were similar to those reported by Vaughn, Riccio, Hynd, and Hall (1997), who found ratings on the *Behavior Assessment System for Children* and the *Child Behavior Checklist* to be correlated significantly (.60's - .90's) when assessing a number of children for Attention-Deficit/Hyperactivity Disorder. Studies like these yield convergent validity, meaning that similarly named scales on each instrument correlate significantly with each other, which allows professionals to feel more confident using either instrument for assessment purposes.

Although strong correlations on similar scales between behavior rating scale instruments are desired, that is not always the case. The Myers et al. (2010) study also demonstrated that correlations between similarly named scales on the parent versions of

the *Behavior Assessment System for Children-2* and the *Child Behavior Checklist 1.5-5* were sometimes only at a moderate level. Furthermore, mean scores were frequently significantly different between the two instruments on similarly named scales, sometimes resulting in very different interpretations of results (i.e., average range vs. clinically significant). A similar study completed by Myers (2013) yielded extensive differences when comparing parent ratings on the *Child Behavior Checklist 1.5-5* with parent ratings on the *Clinical Assessment of Behavior.* The scores on similarly named scales from the two instruments were drastically different and suggested that the *Child Behavior Checklist 1.5-5* either overestimates the severity of behaviors or the *Clinical Assessment of Behavior* underestimates the severity of the same behaviors.

Summary

A young child's social-emotional well-being is an indicator of future behavior and adjustment. Several studies have documented the effects of early emotional behavior on self-regulation and academic success with the consistent finding that children who demonstrate positive social-emotional behaviors at a young age continue to engage in desired or prosocial behaviors (e.g., Nix et al., 2016; Rhoades et al., 2011; Zhai et al., 2015). When considering a child's social-emotional behavior, it is helpful to intervene early so that the child has a better chance of success in future settings. When assessing a child's social-emotional behavior, using a comprehensive approach is vital. Using observations, interviews, and behavior rating scales, evaluators can determine if a young child shows signs of behavioral and social-emotional problems. Behavior rating scales are a popular choice when it comes to assessing a child's emotional behavior. Broadband rating scales assess a wide range of behaviors representing externalizing and internalizing

behaviors. Although disadvantages are present, behavior rating scales are commonly used instruments that have been shown to be useful in identifying social-emotional problems in children.

Purpose

The reviewed literature provides evidence to support the importance of the identification of at risk children for social-emotional problems. While interviews and observations can provide useful information on a child's social-emotional status, both methods require an extensive amount of time from a trained professional. Broadband behavior rating scales are widely used to assess social-emotional skills and have several advantages including time efficiency, low cost, and norm-referenced ratings of students' behaviors. Unfortunately, similarly named constructs across behavior rating scales do not always give consistent results (Campbell & Hammond, 2014; Myers, 2013; Myers et al., 2010).

Findings of consistent results on similarly named scales suggest that the behavior rating scales are measuring the same types of behaviors (Myers et al., 2010). Alternatively, findings of inconsistent results suggest that the rating scales could be assessing different types of behaviors within that construct or that raters are interpreting the items differently across instruments. Knowing the consistency of scores on similarly named constructs across behavior rating scales helps provide school psychologists and other professionals with information to consider when making a decision between instruments to use to assess children. It is important that professionals understand how their choice of an instrument might influence their interpretation of results (Myers et al., 2010).

Thus, the purpose of this thesis project was to determine the consistency of two popular rating scales, the *Behavior Assessment System for Children-3rd edition (BASC-3,* Reynolds & Kamphaus, 2015*)* and the *Child Behavior Checklist 1.5-5* (*CBCL 1.5-5*, Achenbach & Rescorla, 2000). The *CBCL 1.5-5* and the previous edition of the *BASC-3*, the *BASC-2,* were described as representing some of the best broadband behavior rating scales currently available (Merrell, 2008). However, previous research noted inconsistencies between the *BASC-2* and the *CBCL 1.5-5* when completed by parents of clinically referred children (Myers et al., 2010). Given that the *BASC-3* was recently revised, additional independent research on the instrument, and its relationship to the *CBCL 1.5-5,* was needed.

This thesis project addressed three important research questions that provide important information regarding the consistency and validity of behavioral constructs across two popular behavior rating scales. These research questions were:

1. How well do similarly named scales on the *BASC-3* and *CBCL 1.5-5* correlate when preschool teachers complete the two instruments at the same point in time?
2. How consistent are mean scores on the similarly named scales from the two instruments?
3. What is the overall classification consistency (i.e., average vs. clinically significant) of scores between the two instruments on similarly named scales?

Method

Participants

Participants in this study included 56 full time Head Start teachers from two different regions in the United States. Twenty-four (43%) of the participants were from North Carolina and 32 (57%) were from Kentucky. Information collected about the teachers included their gender, race, and years of experience as a preschool teacher. The vast majority of the Head Start teachers were female (96.4%). Slightly more than half of the teachers were Caucasian (57%), while 38% were African American and 5% were Hispanic. The teachers' years of experience ranged from 1 to 33 years, with a mean of 12.4 years (*SD* = 7.3 years).

The demographic variables of gender, age, ethnicity, and presence of an identified disability were collected on the children whom the teachers selected to rate. Of the 56 students, the majority were male (70%) while 30% were female. The ages of the children ranged from 2 years, 7 months to 5 years, 5 months, with a mean age of 4 years, 1 month (*SD* = 8.4 months). About half of the preschool students rated were Caucasian (48.2%) while 32.1% were African American, 1.8% were Latino/Hispanic, 3.6% were Asian, and 14.3% were reported as Biracial. Only 21.4% of the children that were rated had an identified disability.

Instruments

Behavior Assessment System for Children, 3rd Edition (BASC-3). The *BASC-3* is a "multimethod, multidimensional system used to evaluate the behavior and self-perceptions of individuals' ages 2 through 25 years" (Reynolds & Kamphaus, 2015, p. 1). Versions of the *BASC-3* include teacher rating scales, parent rating scales, and self-report

of personality scales. For the purposes of this study, only the teacher scale will be reviewed, as the teacher form was used to gather data in this study.

The teacher rating scale, a broadband measure of both adaptive and problem behaviors in the school setting, includes three versions for three age groups: preschool (ages 2 through 5), child (ages 6 through 11), and adolescent (ages 12 through 21). Because the participants in this study are preschool teachers, the preschool teacher version (ages 2 through 5) was administered in this study and is reviewed in this section. On the instrument, the teacher rates 105 descriptors of behaviors on a four-point scale of frequency, (i.e., *Never, Sometimes, Often,* and *Almost Always);* this process typically takes about 15 minutes to complete (Reynolds & Kamphaus, 2015). Once the rater completes the *BASC-3*, it is then scored with the *BASC-3* scoring software. The 105 behaviors on the *BASC-3* are combined to comprise different clinical scales, adaptive scales, content scales, and composites. Each of these is listed in Table 1, which also includes the scales from the teacher version of the *Child Behavior Checklist 1.5-5*.

Results on the *BASC-3* are provided with standard scores called T scores, which have a mean of 50 and a standard deviation of 10. Reynolds and Kamphaus (2015) provide the following descriptions of the T score ranges. For the adaptive scales, low scores are of concern and the descriptive ranges include: Clinically Significant (30 and below), At-Risk (31-40), Average (41-59), High (60-69), and Very High (70 and above). For the clinical scales, content scales, and composites, high scores are of concern and the descriptive ranges are: Very Low (30 and below), Low (31-40), Average (41-59), At-Risk (60-69), and Clinically Significant (70 and above).

Table 1

Scales on the Preschool Teacher Versions of the BASC-3 and CBCL 1.5-5.

BASC-3	*CBCL 1.5-5*
Composites	**Composites**
Externalizing Problems	Externalizing Problems
Internalizing Problems	Internalizing Problems
Behavioral Symptoms Index	Total Problems
Adaptive Skills	
Clinical Scales	**DSM-Oriented Scales**
Hyperactivity	Depressive Problems
Aggression	Anxiety Problems
Anxiety	Autism Spectrum Problems
Depression	Attention Deficit/Hyperactivity
Somatization	Oppositional Defiant Problems
Atypicality	
Withdrawal	**Syndrome Scales**
Attention Problems	Emotionally Reactive
	Anxious/Depressed
Adaptive Scales	Somatic Complaints
Adaptability	Withdrawn
Social Skills	Attention Problems
Functional Communication	Aggressive Behavior
	Other
Content Scales	Stress Problems
Anger Control	
Bullying	
Developmental Social Disorders	
Emotional Self-Control	
Executive Functioning	
Negative Emotionality	
Resiliency	

Note. *BASC-3* = Behavior Assessment System for Children, 3rd edition, Teacher Rating Scale-Preschool. *CBCL 1.5-5* = Child Behavior Checklist, Caregiver-Teacher Form.

There have been several favorable reviews on previous *BASC* editions in the *Mental Measurement Yearbook*. Stein (2007) wrote that the items and scales on the *BASC* and *BASC-2* were carefully and thoroughly developed. Furlong (2007) reported that the *BASC-2* had good reliability and validity, and he thought it was a promising instrument. Another review by Johnson (2007) also noted the *BASC-2* had good technical characteristics that would make it a useful tool for assisting with diagnoses. These reviews provide information that the past editions of the *BASC-3* were considered effective measures of adaptive and maladaptive behavioral functioning in children.

The reliability of the *BASC-3* preschool teacher rating scale was analyzed through internal consistency, test-retest reliability, and interrater reliability, and the results presented in this section are from the *BASC-3* manual (Reynolds & Kamphaus, 2015). Internal consistency is the degree to which the items on each scale measure the same dimensions. Table 2 lists the coefficient alpha ranges that represent the consistency of the items for each scale.

Table 2

Internal Consistency Ranges for the BASC-3 Teacher Rating Form

	Ages 2-3	Ages 4-5	Clinical Sample (n = 65, ages 4-5)
Composite Scale	.89-.96	.92-.97	.93-.97
Clinical Scales	.77-.89	.81-.93	.87-.94
Adaptive Scales	.87-.90	.87-91	.85-.89
Content Scales	.78-.90	.79-.90	.78-.94

Note. Information is from Reynolds and Kamphaus (2015).

An evaluation of the *BASC-3*'s test-retest reliability, which is when the same child is rated twice over a short period of time, generally indicates adequate stability across time for all scale types and levels. Stability coefficients ranged from .71 to .93 over a mean period of 18 days (range 7-56 days). Interrater reliability includes the agreement of scores obtained from different raters who completed the scale at the same point in time. "The range of interrater reliability coefficients among scales varies widely across all levels, which is consistent with past BASC editions" (Reynolds & Kamphaus, 2015, pp. 125-126). Teacher-to-teacher interrater coefficients ranged from .42 to .83 at the preschool level.

The validity of the *BASC-3* teacher rating scale (TRS) was measured by the scale intercorrelations, factor analyses, and by measuring the pattern of correlations between the TRS scores with scores obtained on other behavioral measures. The information on validity in this section is from the *BASC-3* manual (Reynolds & Kamphaus, 2015). The intercorrelations of scores within the clinical scales and within the adaptive scales are positive, and the intercorrelations between scores from the clinical and adaptive scales are negative, as would be expected. The *BASC-3* was correlated with other behavior rating scales. The *BASC-3* is strongly correlated with the *BASC-2*, with most correlations above .90. In addition, the *BASC-3* TRS form and the *Child Behavior Checklist 1.5-5* caregiver-teacher form were compared, which is the comparison being reviewed for this specialist project. The sample included 90 typically developing preschool children, ages 2 through 5 years, who were in regular preschool classrooms. The mean age was 4.0 years (*SD* = 1.1 years) and the mean number of days between the completion of the two scales was 12.9 days with a range of 0 to 60 days. The comparison yielded mostly moderate

scores for composite and clinical scales, and scales measuring externalizing behaviors typically demonstrated higher correlations than those measuring internalizing behaviors. Table 3 demonstrates the correlations of similarly named scales on the preschool teacher versions of the *BASC-3* and the *CBCL 1.5-5*.

Table 3

Correlations Between Selected Scales on the Preschool Teacher Versions of the BASC-3 and the CBCL 1.5-5 (n = 90)

Hyperactivity - Attention Deficit Hyperactivity	.67
Attention Problems – Attention Deficit Hyperactivity	.58
Attention Problems - Attention Problems	.61
Aggression - Aggressive Behavior	.78
Anxiety - Anxious/Depressed	.55
Anxiety - Anxiety Problems	.53
Depression – Depressive Problems	.66
Depression - Anxious/Depressed	.65
Somatization - Somatic Complaints	.15
Withdrawal – Withdrawn	.59
Atypicality – Autism Spectrum	.65
Developmental Social Disorders – Autism Spectrum	.66
Externalizing - Externalizing	.76
Internalizing - Internalizing	.57
Behavior Symptoms Index - Total Problems	.77

Note. Correlations are from Reynolds and Kamphaus (2015).

Child Behavior Checklist for Ages 1.5-5. The Achenbach System of Empirically Based Assessment (ASEBA) includes various rating scales that assess a broad range of maladaptive behaviors from early childhood to adulthood. As part of the system, there is a broadband rating scale for young children called the *Child Behavior Checklist for Ages 1.5-5* (*CBCL 1.5-5*, Achenbach & Rescorla, 2000). Only the Caregiver-Teacher Form is reviewed for this study. The Caregiver-Teacher Form has 99 items (i.e., behavioral descriptors) that are designed to reflect problem behaviors that toddlers and preschool-aged children may exhibit. The frequency of all behaviors is rated on a 3-point scale: Not True, Somewhat True, and Very True or Often True (Achenbach & Rescorla, 2000). These items comprise various scales that were listed in Table 1. Like the *BASC-3*, the results from the ratings are converted to T scores. The average range for the *CBCL 1.5-5* includes T scores from 50 to 64. Scores 65 to 69 are considered to be in the "borderline clinical" range and scores 70 and above are in the "clinical" range (Achenbach & Rescorla, 2000).

The following information on reliability and validity is from Achenbach and Rescorla (2000). The reliability for the Caregiver-Teacher Form of the *CBCL 1.5-5* was determined through internal consistency, test retest reliability, and inter-rater agreement. As noted earlier, internal consistency is the degree to which the items on the scale measure the same dimensions. The items on the different scales had consistency coefficients ranging from .78 to .93, with Affective Problems being the lowest (.78) and Oppositional Defiant Problems the highest (.93).

Test-retest reliability data included a sample of 59 children and was obtained from an average 8-day period. Reliability remained high for most scales, with most coefficients

within the mid to upper .80's. The coefficients ranged from .57 to .91. With inter-rater reliability, the agreement of most teacher ratings were in the low .50's to upper .70's, but the Somatic Complaints area had a low coefficient of .21.

According to Achenbach and Rescorla (2000), the validity of the *CBCL 1.5-5* Caregiver-Teacher Form was determined by closely matching non-referred and referred samples of children, which allowed for the testing of problem scale scores to distinguish between the two groups. A multiple regression analysis demonstrated no significant effects due to age or socio-economic status. The *CBCL 1.5-5* was also compared to other measures of problems to determine construct validity. When compared to another instrument, called the *Richman Behavior Checklist*, correlations of .56 to .77 were found. As mentioned and illustrated earlier (see Table 3), when the *CBCL 1.5-5* was compared to the *BASC-3*, correlations of similarly named scales ranged from .15 to .78.

Procedure

The Institutional Review Board at Western Kentucky University approved this project (see Appendix). Researchers contacted the directors of two Head Start organizations in two different states through e-mail and asked if they and their teachers were interested in participating in the study. A detailed description of the thesis project was included in the emails. The directors agreed via email, and the researchers scheduled times with each program to collect information. Each Head Start organization had regional meetings attended by all of their teachers. All data were collected during those regional meetings. One researcher collected data from the teachers in North Carolina and two researchers collected data from the teachers in Kentucky.

When meeting with the groups of teachers, each teacher was handed a large envelope that included the following information: a brief description of the procedures, two consent forms (one to turn in and one to keep), a demographics form, and two behavior rating scales (i.e., *BASC-3* and *CBCL 1.5-5*). The teachers were asked to sign and date the consent form if they wished to participate. If they consented to the study, they were then asked to complete the demographics form, providing information about themselves as well as the child they planned to rate.

The participating teachers were read scripted verbal instructions on how to complete each rating scale, in order to provide consistent instructions across groups. Teachers were verbally encouraged to "think of a student with any level of behavioral concerns." The demographics form also provided written examples of the types of behavioral concerns they should consider, in order to encourage the teachers to think of internalizing types of problems as well as the externalizing problems. Specifically, the directions noted that behavioral concerns "can include any one or more of the following behaviors: withdrawn behaviors (e.g., not engaging in tasks), aggression, tantrums, very active, anxiety (e.g., nervousness or worry), poor social skills (e.g., not sharing or playing well with others), or noncompliance (e.g., not listening to the teacher)." Finally, the teachers were asked to complete each behavior rating scale, one after the other. The order of completion was randomized so that half of the teachers would complete the *BASC-3* first and the other half would complete the *CBCL 1.5-5* first. The participating teachers were asked to place the first completed behavior rating scale back in the envelope as soon as it was completed, before starting the second rating scale.

Data Analysis

The results addressing the research questions were analyzed in the following ways:

1. How well do similarly named scales on the *BASC-3* and *CBCL 1.5-5* correlate when preschool teachers complete the two instruments at the same point in time?

Pearson correlations were determined for all scales on the *BASC-3* compared to all scales on the *CBCL 1.5-5* to provide a broad range of comparisons. There are 15 pairs of similarly named scales of interest for this thesis project. Those comparisons will be the same ones listed in Table 3.

2. How consistent are mean scores on the similarly named scales from the two instruments?

A series of *t*-tests were used to determine if the mean scores for similarly named pairs of scales from the *BASC-3* and the *CBCL 1.5-5* were statistically significantly different. Due to the high number of comparisons (i.e., 15) and the chance of a Type I error, a Bonferroni correction procedure was applied to determine statistical significance (i.e., .05 / 15 = .003). Because the corrected p value was very close to the standard $p < .001$ value, a difference between mean scores was considered statistically significance if it met the $p < .001$ criterion.

3. What is the overall classification consistency (i.e., average vs. clinically significant) of scores between the two instruments on similarly named scales?

The overall classification consistency was determined between the similarly named scales by calculating percentages for the number of T scores that were: (a) at or above 65 on both instruments, (b) less than 65 on both instruments, (c) at or above 65 on the *BASC-3* but less than 65 on the *CBCL 1.5-5*, and (d) at or above 65 on the *CBCL 1.5-*

5 but less than 65 on the *BASC-3*. An overall consistency percentage was determined from the number of times both instruments resulted in scores above and below the cutoff score of 65. A minimum criterion score of 80% was used to evaluate the overall consistency percentages. While no minimum criterion was found in the literature for classification consistency, the 80% figure is what is recommend as a minimum level of adequate inter-rater agreement when two people conduct observations (Alessi & Kaye, 1983).

Results

Strength of Correlations

The first research question asks how well similarly named scales on the two instruments correlate when teachers complete them at the same point in time. Correlations between the corresponding scales should be positive and at a moderate to strong level. Taylor's (1990) description of the strength of correlations was used for this study as follows: Generally Weak ($r = \leq .35$), Moderate ($r = .36$ - $.67$), Strong ($r = .68$ – $.89$) and Very Strong ($r = \geq 90$). To answer the first research question, correlations between all scales on the *CBCL 1.5-5* and the *BASC-3* were determined. Because of the large number of scales from the *BASC-3*, two tables were used to report all the correlations. Table 4 provides the correlations between the *CBCL 1.5-5* and the *BASC-3* clinical and composite scales, and Table 5 lists correlations with the *BASC-3* adaptive and content scales.

Correlations are useful in determining the magnitude of the relationship between two variables. Although correlations determine the association of the two variables, they do not ensure causation (Taylor, 1990). For example, if one variable (e.g., Bullying on the *BASC-3*) correlated strongly with another variable (Aggressive Behaviors on the *CBCL*), this does not necessarily mean that bullying causes aggressive behaviors. The strength of the relationship between two variables is determined by looking at the reliability coefficient (r). The closer the coefficient r is to ± 1, regardless of the direction, the stronger the association between the variables (Taylor, 1990). If a relationship is not present, the correlation will be close to zero (Taylor, 1990).

Table 4

Correlations Between Scales on the Teacher Versions of the CBCL/1.5-5 and BASC-3 Clinical and Composite Scales (n = 56)

	BASC-3 Clinical and Composite Scales										
CBCL/1.5-5 Scales	1	2	3	4	5	6	7	8	9	10	11
Emotionally Reactive	.21	.07	.66*	.69*	.47*	.11	.34	.44*	.76*	.14	.53*
Anxious/Depressed	.02	-.07	.61*	.43*	.44*	.08	.26	.44*	.61*	-.03	.35
Somatic Complaints	.31	.26	.46*	.37	.56*	-.15	.04	.26	.58*	.24	.30
Withdrawn	.03	-.13	.19	.09	-.09	.26	.49*	.56*	.07	-.07	.37
Attention Problems	.65*	.05	-.04	.02	-.02	.77*	.50*	.05	-.02	.34	.51*
Aggressive Behavior	.62*	.79*	-.02	.48*	.16	.23	-.11	.01	.27	.80*	.58*
Depressive Problems	.02	-.05	.25	.28	.12	.14	.25	.38	.27	-.03	.29
Anxiety Problems	.01	-.15	.66*	.37	.45*	.10	.28	.50*	.61*	-.09	.33
Autism Spectrum	.13	.02	.38	.23	.15	.22	.45*	.62*	.31	.08	.49*
ADHD Problems	.73*	.23	-.03	.16	.04	.75*	.34	.04	.07	.50*	.58*
Oppositional Defiant	.47*	.69*	.05	.55*	.17	.15	-.09	-.06	.34	.66*	.49*
Stress Problems	.38	.26	.20	.47*	.25	.31	.35	.39	.40	.34	.61*
Internalizing	.14	.02	.60*	.51*	.36*	.18	.47*	.63*	.61*	.09	.58*
Externalizing	.77*	.71*	-.05	.42*	.14	.46*	.07	.01	.23	.82*	.67*
Total Problems	.61*	.44*	.29	.51*	.31	.44*	.31	.32	.47*	.58*	.74*

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

Table 5

Correlations Between Scales on the Teacher Versions of the CBCL/1.5-5 and BASC-3 Adaptive and Content Scales (n = 56)

	BASC-3 Adaptive and Content Scales										
CBCL/1.5-5 Scales	1	2	3	4	5	6	7	8	9	10	11
Emotionally Reactive	-.46*	-.05	-.03	-.25	.25	.04	.35	.62*	.42*	.54*	-.52*
Anxious/Depressed	-.31	-.06	-.12	-.22	.08	-.04	.29	.32	.18	.28	-.32
Somatic Complaints	-.11	.12	.03	.01	.31	.21	.11	.34	.18	.42*	-.16
Withdrawn	-.14	-.33	-.44*	-.42*	-.14	-.04	.65*	-.06	.09	.04	-.28
Attention Problems	.05	.06	-.31	-.11	-.05	.10	.41	.07	.53*	.09	-.17
Aggressive Behavior	-.18	-.01	.11	-.03	.71*	.76*	-.08	.56*	.61*	.59*	-.26
Depressive Problems	-.13	-.15	-.10	-.16	.03	-.08	.30	.14	.12	.14	-.27
Anxiety Problems	-.41	-.04	-.23	-.34	.05	-.13	.36	.32	.20	.21	-.40
Autism Spectrum	-.32	-.21	-.48*	-.47*	.04	.09	.70*	.17	.25	.24	-.43*
ADHD Problems	.04	.03	-.27	-.10	.16	.29	.31	.21	.63*	.24	-.17
Oppositional Defiant	-.18	.09	.20	.05	.66*	.59*	-.11	.55*	.52*	.61*	-.23
Stress Problems	-.24	-.19	-.12	-.25	.33	.31	.39	.43*	.44*	.37	-.29
Internalizing	-.36	-.17	-.26	-.37	.14	.04	.57*	.38	.31	.41	-.48*
Externalizing	-.12	.02	.00	-.05	.60*	.70*	.05	.52*	.72*	.54*	-.25
Total Problems	-.21	-.03	-.19	-.20	.43*	.47*	.32	.50*	.64*	.53*	-.36

Note. 1=Adaptability, 2=Social Skills, 3=Functional Communication, 4=Adaptive Skills, 5=Anger Control, 6=Bullying, 7=Developmental Social Disorders, 8=Emotional Self Control, 9=Executive Functioning, 10=Negative Emotionality, 11=Resiliency.
*$p < .001$.

The results from Tables 4 and 5 provide many more comparisons than were the focus of this study. However, correlations from all comparisons are included to provide additional information about the relationship between scales on these two instruments. As can be seen from the tables, there are several comparisons that provide reasonable results. As examples, the *CBCL*'s Aggressive Behavior scale was significantly correlated with the *BASC*'s Bullying (r = .76) and Anger Control (r = .71) scales. Also, the *CBCL*'s Oppositional Defiant scale was significantly correlated with the *BASC*'s Aggression scale (r = .69). Another example is the *CBCL*'s Attention Problems and the *BASC*'s Executive Functioning scale, which produced a significant correlation (r = .53). This was reasonable, as attentional control is an executive functioning skill.

However, other results were not what might be expected. For example, the *BASC*'s Social Skills scale did not result in significant negative correlations with *CBCL*'s Autism Spectrum (r = -.21) or Withdrawn (r = -.33) scales, even though behaviors of autism and withdrawn behaviors would be expected to interfere with a child's social skills. It might also be expected that the *CBCL*'s Depressive Behaviors would correlate strongly with the *BASC-3*'s Withdrawal Scale, as withdrawal is often associated with depression; however, it does not (r =.38).

The correlations of interest to this study were pulled from Tables 4 and 5 and listed in Table 6 to highlight those results. For comparison purposes, Table 6 also includes the correlations between the two instruments with a typical sample of children as listed in the *BASC-3* manual (Reynolds & Kamphaus, 2015) and previously included in Table 3. Most correlations are remarkably similar. Comparisons with substantially lower correlations in the current study consist of the Depression - Depressive Problems,

Table 6

Correlations Between Similarly Named Scales on the Teacher Preschool Versions of the BASC-3 and the CBCL 1.5-5

BASC-3 scales – *CBCL* scales	Current Study	BASC-3 Manual
Hyperactivity - Attention Deficit Hyperactivity	.73*	.67
Attention Problems - Attention Deficit Hyperactivity	.65*	.58
Attention Problems - Attention Problems	.77*	.61
Aggression - Aggressive Behavior	.79*	.78
Anxiety - Anxious/Depressed	.61*	.55
Anxiety - Anxiety Problems	.66*	.53
Depression – Depressive Problems	.28	.66
Depression - Anxious/Depressed	.43*	.65
Somatization - Somatic Complaints	.56*	.15
Withdrawal - Withdrawn	.56*	.59
Atypicality - Autism Spectrum	.45*	.65
Developmental Social Disorders - Autism Spectrum	.70*	.66
Externalizing - Externalizing	.82*	.76
Internalizing - Internalizing	.61*	.57
Behavior Symptoms Index - Total Problems	.74*	.77

*$p < .001$.

Depression – Anxious/Depressed, and the Atypicality – Autism Spectrum scales, with *BASC* scales listed first. One comparison, Somatization – Somatic Complaints, resulted in a substantially higher correlation in the current study than reported in the *BASC-3* manual (i.e., .56 vs. .15).

The only pair of corresponding scales from the two instruments that did not produce a significant correlation in the current study was Depression – Depressive Problems. The *CBCL* includes an Anxious/Depressed scale, an Anxiety Problems scale, and a Depressive Problems scale, while the *BASC-3* contains only the Anxiety and Depression scales. *CBCL*'s Anxious/Depressed scale correlates higher with *BASC-3*'s Anxiety scale (r = .61) than its Depression scale (r = .43), suggesting the items on the Anxious/Depressed scale may be assessing more anxiety concerns than depression concerns.

While the Atypicality scale on the first two versions of the *BASC* was meant to assess behaviors of autism, the *BASC-3* has included a Developmental Social Disorders scale as a more direct assessment of Autism Spectrum Disorder. As such, the *CBCL*'s Autism Spectrum scale does have a higher correlation with the Developmental Social Disorders scale (r = .70) than the Atypicality scale (r = .45).

Consistency of Mean Scores

The second research question examined the consistency of mean scores of similarly named scales from the *BASC-3* and *CBCL 1.5-5*. Paired sample *t*-tests were used to determine any significant differences between the mean scores in each corresponding pair. The results are provided in Table 7. Asterisks were used to denote significant differences in mean scores.

Effect sizes (*ES*), included in Table 7, are the estimates of the magnitude of effects determined between variables (Cohen, 1992). Effect sizes are resistant to sample size influences and provide a more accurate measure of the magnitude of the effect between two or more variables (Ferguson, 2009). Basically, significant effect sizes ensure

Table 7

Mean T Scores for Comparable BASC-3 and CBCL/1.5-5 Scales (n = 56)

BASC-3 Scale - *CBCL* Scale	BASC-3	CBCL	*t* values	*ES*[a]
Hyperactivity – ADHD	68.6	72.4	3.1	.15
Attention Problems - ADHD	64.8	72.4	6.3*	.32
Attention Problems - Attention Problems	64.8	71.4	5.6*	.28
Aggression - Aggressive Behavior	69.4	68.6	0.6	.03
Anxiety - Anxious/Depressed	52.1	58.3	4.9*	.30
Anxiety - Anxiety Problems	52.1	58.3	5.1*	.28
Depression – Depressive Problems	63.1	61.3	1.1	.09
Depression – Anxious/Depressed	63.1	58.3	3.0	.22
Somatization – Somatic Complaints	45.4	55.0	7.1*	.44
Withdrawal - Withdrawn	62.0	64.2	1.3	.09
Atypicality - Autism Spectrum	69.7	66.4	1.8	.13
Dev. Social Disorders - Autism Spectrum	67.3	66.4	0.9	.05
Externalizing - Externalizing	70.2	69.3	0.9	.04
Internalizing - Internalizing	54.4	62.8	6.3*	.36
Behavioral Symptoms Index - Total Problems	71.2	68.6	2.8	.14

[a]Pearson's *r*. *$p < .001$.

that the effect is not due to sampling error. Unfortunately, there is no exact agreement on what magnitude of effect is considered significant (Ferguson, 2009). However, Cohen (1992) suggests a value of .10 - .29 to be a small effect, .30 - .49 to be a medium effect, and ≥ .50 to be a large effect when comparing independent means. For example, Anxiety on the *BASC-3* and Anxious/Depressed on the *CBCL 1.5-5* resulted in an effect size of .30, which indicates the significant difference between the two scales is at a medium effect level.

The majority of the pairs of corresponding scales do not indicate significant differences in mean scores; however, statistically significant differences occurred for six of the 15 (40%) corresponding pairs. The *BASC-3* Attention Problems scale resulted in significantly different scores from both the Attention Deficit/Hyperactivity and Attention Problems scales on the *CBCL 1.5-5*. Similarly, the *BASC-3* Anxiety scale resulted in significantly different scores from both the Anxious/Depressed and Anxiety Problems scales on the *CBCL 1.5-5*. Significant differences were also found with the Somatization-Somatic Complaints and Internalizing-Internalizing comparisons.

Classification Consistency

The final research question examined the overall classification consistency (i.e., average vs. clinically significant) of scores between the two instruments on corresponding scales. Percentages were calculated for the number of T scores that were (a) at or above 65 on both instruments, (b) less than 65 on both instruments, (c) at or above 65 on the *BASC-3* but less than 65 on the *CBCL 1.5-5*, and (d) at or above 65 on the *CBCL 1.5-5* but less than 65 on the *BASC-3*. To determine overall consistency

between scales on the two instruments, a minimum criterion score of 80% was used as a cutoff. Table 8 indicates the consistency of ratings considered average versus clinically significant for each corresponding pair.

Eight, or approximately half (53.3%), of the corresponding pairs of scales from the two instruments had classification consistencies above the criterion of 80%. Four of the corresponding pairs (26.7%) had overall classification consistencies between 70% and 79% while the other three pairs (20%) had classification consistencies ranging from 53% to 59%. Two corresponding pairs, Aggression – Aggressive Behaviors and Somatization – Somatic Complaints resulted in the highest overall classification consistency percentage (both at 91%). This suggests that for these scales, both instruments are similar in classifying whether a construct is average or clinically significant, even though the mean scores from the Somatization – Somatic Complaints comparison were significantly different (see Table 7). Consistent with the finding of significantly different mean scores, the Internalizing Problems – Internalizing Problems comparison only yielded an overall classification consistency of 59%, which is very poor.

The results from Table 8 also indicate that the *CBCL 1.5-5* tends to result in scores at or above 65 more often than the *BASC-3*. In 12 of the 15 comparisons (80%), the *CBCL* had a higher percentage of scores where only the *CBCL* had a score above 65 than when only the *BASC* had a score above 65. For example, the Attention Problems-ADHD comparison demonstrates that only the *CBCL* rates behaviors to be clinically significant 16% of the time compared to zero times only the *BASC-3* had a clinically significant score. An exception to this finding is the Depression scale from the *BASC-3*,

which resulted in many more clinically significant ratings compared to the *CBCL*'s Depressive Problems (30% vs. 14%) and Anxious/Depressed (34% vs. 12%) scales.

Table 8

Classification Consistency of Scores Considered Average or Clinically Significant Between Comparable BASC-3 and CBCL/1.5-5 Scales (n = 56)

	Consistency of Scores				
BASC-3 – CBCL/1.5-5	Both scales ≥ 65	Both scales < 65	Only *BASC* ≥ 65	Only *CBCL* ≥ 65	Overall
Hyperactivity - ADHD	57%	23%	6%	14%	80%
Attention Problems - ADHD	55%	29%	0%	16%	84%
Attention Prob. - Attention Prob.	52%	32%	4%	12%	84%
Aggression - Aggressive Beh.	62%	29%	0%	9%	91%
Anxiety - Anxious/Depressed	7%	75%	4%	14%	82%
Anxiety - Anxiety Problems	5%	81%	5%	9%	86%
Depression - Depressive Prob.	13%	43%	30%	14%	56%
Depression – Anxious/Depressed	9%	45%	34%	12%	53%
Somatization - Somatic	4%	87%	4%	5%	91%
Withdrawal - Withdrawn	32%	39%	11%	18%	71%
Atypicality - Autism	41%	29%	12%	18%	70%
Dev. Social Disorder - Autism	45%	32%	9%	14%	77%
Externalizing - Externalizing	68%	21%	2%	9%	89%
Internalizing - Internalizing	13%	46%	5%	36%	59%
Beh. Symptoms Index - Total	63%	16%	14%	7%	79%

Note. "Overall" refers to the consistency of agreement where scores from corresponding scales were both ≥ a T score of 65, or both scales were below a T score of 65.

Discussion

Given the importance of the development of social-emotional skills in early childhood, it is essential that these skills be assessed accurately and with reliable instruments. Numerous behavior rating scales exist to assess social-emotional skills and many of the broadband instruments purport to measure similar constructs. It is important to know if different instruments are measuring similarly named constructs in a consistent manner. The current study examined the relationship of corresponding scales on the preschool versions of the *BASC-3* and *CBCL 1.5-5* when Head Start teachers provided ratings of students that exhibited behavioral concerns. Although there was consistency with some corresponding scales on the two instruments, inconsistency of results for similarly named scales was also determined.

Consistency was assessed in three ways through three research questions: (a) How well do similarly named scales on the preschool teacher versions of the *BASC-3* and *CBCL 1.5-5* correlate?, (b) How consistent are mean scores on the similarly named scales from the two instruments?, and (c) What is the overall classification consistency (i.e., average vs. clinically significant) of scores between the two instruments on similarly named scales?

When determining the correlations of corresponding scales, results indicated several pairs to be highly correlated, which would be expected as the instruments claim to measure the same behavioral constructs. Similarly named scales that produced strong correlations (r = .68 or higher) include Hyperactivity – Attention Deficit Hyperactivity, Attention Problems – Attention Problems, Aggression– Aggressive Behavior, Developmental Social Disorder – Autism Spectrum, and the composite areas of Externalizing – Externalizing and Behavioral Symptoms Index – Total Problems, with

the *BASC-3* scales listed first in each pair. Corresponding pairs of scales that addressed internalizing types of behaviors generally had correlations considered at a moderate level (e.g., Withdrawal-Withdrawn, $r = .56$) or, in one comparison, at a non-significant level (i.e., Depression-Depressive Problems, $r = .28$). Thus, it is evident that the strongest correlations were found in externalizing types of behaviors, which are the behaviors that are noticed more often. Internalizing behaviors are less noticeable, especially within a classroom setting. For example, aggression in the classroom is more likely to be noticed by the teacher compared to withdrawn or depressive behaviors. The finding of higher correlations for externalizing types of behaviors is consistent with Burlaka et al. (2014) and Merrell (2008), who stated that externalizing behaviors are more consistently measured than internalizing behaviors.

When examining correlations amongst pairs in the current study compared to the study provided in the *BASC-3* manual (Reynolds & Kamphaus, 2015), most of the pairs were consistent and similar, with some pairs producing correlations much higher or lower in the current study. In the current study, two-thirds of the correlations were higher than correlations for the same comparisons in the *BASC-3* manual. While the sample of participants included in the study from the *BASC-3* manual was larger ($n = 90$), their sample included all typically developing preschoolers. The current study examines results based on a sample of students for which the instruments are more likely to be used because typically developing children are rarely referred for an assessment of their social-emotional skills. As such, the current results may provide a better reflection of how the two instruments are related.

While test publishers routinely report correlations between behavior rating scales, they often stop there with their comparisons. Significant correlations, however, do not mean each scale produces equivalent results. A high correlation simply means the relative rank order of high and low scores was similar between the two instruments. Determining if the scores obtained on similarly named scales from the two instruments were similar was the next analysis of the current study. The comparison of scores revealed the fact that high correlations do not always equate to consistent scores. For example, the Attention Problems scales on the *CBCL 1.5-5* and *BASC-3* were strongly correlated (r = .77); however, their mean scores were significantly different. Similar findings were also the case for other comparisons, such as Attention Problems - ADHD. This suggests that high correlations alone might mask differences between instruments when measuring a similarly named construct. Another type of finding is that similarly named scales might not be strongly correlated, but still produce mean scores that are not significantly different. For example, the Depression - Depressive Problems comparison indicated a weak correlation (r = .28) but the mean scores were not significantly different.

Although most of the 15 corresponding pairs did not produce significant differences, there were statistically significant differences for six of the pairs: Attention Problems – ADHD, Attention Problems – Attention Problems, Anxiety – Anxious/Depressed, Anxiety – Anxiety Problems, Somatization – Somatic Complaints, and Internalizing – Internalizing. Four of those six comparisons included just two scales from the *BASC-3*: Attention Problems and Anxiety scales. These results suggest the *BASC-3* measures those two constructs unlike the related scales on the *CBCL*. Furthermore, similar to the finding of lower correlations, these behavioral constructs are

mostly internalizing behaviors, which are those behaviors that are inner-directed and less observable. Therefore, ratings from teachers could vary amongst the different items, as the teachers may have difficulty rating behaviors that are not always evident. Results from Myers et al. (2010) also indicated significant differences in mean scores between similarly named scales (i.e., 9 out of 15) on parent versions of the *BASC-2* and *CBCL 1.5-5*. Many of those significantly different pairs were also constructs measuring internalizing behaviors.

The third way the results from the two instruments were assessed was by examining the classification consistency to determine how often corresponding scales from each instrument resulted in average or clinically significant scores. It is important to consider the classification consistency of the instruments because this is how scores are often interpreted. For instance, a psychologist may be more likely to consider the descriptive ranges for behaviors rather than just mention the scores. Only about half (53%) of the corresponding pairs on the instruments produced similar classification results above the overall 80% criterion. For these pairs of similarly named constructs, this indicates that two instruments are classifying behaviors similarly as average or clinically significant. For example, the Somatization – Somatic Complaints comparison resulted in significantly different mean T scores (i.e., 45.4 vs. 55.0), but those mean scores are both in the average range, resulting in the same conclusion about the child's functioning in that area. However, given almost half of the corresponding pairs did not classify the constructs at a similar level should be a concern to practitioners. Some pairs, like those measuring Depression and Internalizing behaviors, only agreed a little more than half the time whether or not the construct should be considered clinically significant. Thus, for

some constructs, the results from a behavior rating scale are only a little better than flipping a coin (i.e., at a chance level) to determine if clinical significance exists.

When considering the results from all three analyses, only a few behavioral constructs (i.e., Hyperactivity, Aggression, and Externalizing), were all-around consistent. More specifically, the pairings that produced strong correlations, no significant mean differences, and high overall classification consistency ratings were: Hyperactivity – ADHD, Aggression – Aggressive Behavior, and Externalizing – Externalizing. Other pairings that were considered acceptable in that they barely missed one criterion cutoff score include Withdrawn – Withdrawal, Developmental Social Disorder – Autism Spectrum, and Behavior Symptoms Index – Total Problems. The comparisons with the worst consistency were Depression – Depressive Problems, Depression – Anxious/Depressed, and Internalizing – Internalizing.

Five of the six comparisons of similarly named scales resulted in significantly different mean scores but had moderate to strong correlations and high classification consistencies. A specific example is Attention Problems – Attention Problems, which produced significantly different mean scores (i.e., 64.8 vs. 71.4), but had a strong correlation (r = .77) and a high classification consistency (84%). For this example, such results could be the outcome of many of the children selected by the teachers having attention problems and both instruments resulted in many scores in the clinically significant range.

It is interesting to note that the *CBCL* tends to result in higher mean scores and indicate more clinically significant ratings when compared to the *BASC-3.* Such a finding was also evident in the Myers et al. (2010) study when they examined preschool parent

versions of the *BASC-2* and the *CBCL 1.5-5*. However, while a consistent finding across studies, it is unknown which instrument provides a more accurate representation of behavior. Although it seems that the *CBCL 1.5-5* provides more elevated scores, it could also be that the *BASC-3* is providing more deflated scores. A curious exception to this finding though, is that the scales measuring Depression on the *BASC-3* tended to score higher compared to the *CBCL 1.5-5* and yielded more scores that were clinically significant. It is possible that the *BASC-3* items that measure depression are worded more vaguely, or reflect common behaviors seen in preschoolers, than items on the *CBCL 1.5-5*, prompting higher scores.

Strengths and Limitations

A strength of the current research was the procedure that required the teachers to complete both rating scales at the same point in time. This ensures that temporal, setting, and rater error variance were not confounding factors in this study (Merrell, 2008). All Head Start teachers were provided the same set of instructions, with half of the teachers completing the *BASC-3* first and the other half completing the *CBCL 1.5-5* first in order to control for any order effects. For generalization purposes, important demographic information was collected concerning the teachers and the students that were being rated. A fairly diverse sample of teachers and students was obtained. Another strength is the geographic diversity in the raters, with about half of the raters from two different regions of the country. Another strength of the current study is that the children being rated were ones that would be likely evaluated for social-emotional concerns. Thus, the results provide data on how the instruments work with children likely to get referred, rather than a typically developing sample of children, such as the sample in the BASC-3 manual.

A possible limitation of the current study, which occurs with most studies, is the limited number of participants. With only 56 participants, it is difficult to generalize to the entire population. Also, another potential limitation is that only children in Head Starts were being rated upon, which excludes children from a wide range of preschool settings and focuses on those who are less economically privileged. An additional limitation may be the gender of the raters. With the majority of the raters being female, it is unknown if female teachers are likely to interpret behaviors differently than males. However, most preschool teachers are females so the likelihood of getting male preschool teachers to participate in such a study is probably not necessary.

Future Research

The research field may benefit from further research on this topic focusing on different age groups and, consequently, different age level forms of the *BASC-3* and *CBCL* besides just the preschool age. Specifically, because most referrals of children with behavioral problems to school psychologists often come from the elementary level, targeting the elementary age may be beneficial. Given preschooler's behaviors are often inconsistent, more consistent findings may be found with an older population of children.

Another consideration for future research is to compare other behavior rating scales as well as parent versions of various rating scales. With the *BASC-3* and the *CBCL* being just two of the popular behavior rating scales available, it would be interesting to determine consistency of other behavior rating scales (e.g., *Conners 3*). Future researchers could also examine the consistency of other types of rating scales, such as adaptive behavior rating scales that are commonly used for evaluations. For example, the *Adaptive Behavior Assessment System, Third Edition* (*ABAS-3*) and the *Vineland*

Adaptive Behavior Scales, Third Edition are two popular instruments that measure adaptive behavior that may or may not provide similar ratings.

Summary

In the current study, the consistency of the teacher versions of two popular preschool behavior rating scales, the *BASC-3* and the *CBCL 1.5-5*, was examined. Results indicated that although the two instruments measure some behavioral constructs similarly, several similarly named scales did not provide consistent results. It is highly recommended that professionals use multi-faceted assessment measures and strategies for making decisions, as the current results indicate that two popular ratings scales cannot always be trusted to provide similar results. Practitioners using the instruments evaluated in this study need to be aware that very different results and, consequently, different interpretations of behavior can occur depending simply upon which instrument is chosen during the evaluation. Future research could consider examining different rating scales and with different age ranges and raters to further examine the consistency of results to provide additional information to practitioners regarding the use of such instruments.

References

Achenbach, T. M., McConaughy, S., & Howell, C. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213-232.

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.

Alessi, G. J., & Kaye, J. H. (1983). *Behavior assessment for school psychologists*. Kent, OH: National Association of School Psychologists.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: Fifth edition.* Arlington, VA: American Psychiatric Publishing.

Burlaka, V., Bermann, E. A., & Graham-Bermann, S. A. (2014). Internalizing problems in at-risk preschoolers: Associations with child and mother risk factors. *Journal of Child and Family Studies, 24,* 2653-2660.

Campbell, J. M., & Hammond, R. K. (2014). Best practices in rating scale assessment of children's behavior. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp. 287-304). Bethesda, MD: National Association of School Psychologists.

Campbell, S., Shaw, D., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for maladjustment. *Development and Psychopathology, 12,* 467-488.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 115-159.

Denham, S. A., Bassett, H. H., Thayer, S. K., Mincic, M. S., Sirotkin, Y. S., & Zinsser, K. (2012). Observing preschoolers' social-emotional behavior: Structure,

foundations, and prediction of early school success. *Journal of Genetic Psychology*, *173*, 246-278.

Denham, S. A., Bassett, H. H., Zinsser, K., & Wyatt, T. M. (2014). How preschoolers' social-emotional learning predicts their early school success: Developing theory-promoting, competency-based assessments. *Infant & Child Development*, *23*, 426-454.

Denham, S., Blair, K., DeMulder, E., Levitas, J., Sawyer, K., Auerbach–Major, S., & Queenan, P. (2003). Preschool emotional competence: Pathway to social competence? *Child Development, 74*, 238-256.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of Primary Prevention, 28,* 67-91.

Elias, M., Zins, J., Weissberg, R., Frey, K., Greenberg, M., Haynes, N., … Shriver, T. (1997). *Promoting social and emotional learning: Guidelines for educators.* Danvers, MA: Association for Supervision and Curriculum Development.

Epstein, A. (2009*). Me.you.us: Social-emotional learning in preschool.* Ypsilanti, MI: HighScope Press.

Fantuzzo, J. W., Bulotsky-Shearer, R., Fusco, R. A., & McWayne, C. (2005). An investigation of preschool classroom behavioral adjustment problems and social–emotional school readiness competencies. *Early Childhood Research Quarterly*, *20*, 259-275.

Ferguson, C. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*, 532-538.

Furlong, M. (2007). Review of the BASC second edition. In K. F. Geisinger, R. A. Spies, J. F. Carlson, & B. S. Plake (Eds.), *The seventeenth mental measurements yearbook*. Lincoln, NE: Buros Center for Testing.

Gimpel, G. A., & Holland, M. L. (2003). *Emotional and behavioral problems of young children: Effective interventions in the preschool and kindergarten years.* New York, NY: Guilford Press.

Herndon, K., Bailey, C., Shewark, E., Denham, S., & Bassett, H. (2013). Preschoolers' emotion expression and regulation: Relations with school adjustment. *The Journal of Genetic Psychology, 6,* 642-663.

Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in the systematic direct observation of student behavior. In A. Thomas & J. Grimes (Eds.) *Best Practices in School Psychology IV* (pp. 993-1005). Bethesda, MD: National Association of School Psychologists.

Johnson, K. (2007). Review of the BASC second edition. In K. F. Geisinger, R. A. Spies, J. F. Carlson, & B. S. Plake (Eds), *The seventeenth mental measurements yearbook*. Lincoln, NE: Buros Center for Testing.

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

Major, S. O., Seabra-Santos, M. J., & Martin, R. P. (2015). Are we talking about the same child? Parent-teacher ratings of preschoolers' social-emotional behaviors. *Psychology in the Schools, 52*, 789-799.

McConaughy, S. H., & Ritter, D. R. (2014). Best practices in multimethod assessment of emotional and behavioral disorders. In P. Harrison & A. Thomas (Eds.), *Best*

practices in school psychology: Data-based and collaborative decision-making. (pp. 367-390). Bethesda, MD: National Association of School Psychology.

Merrell, K. W. (2008). *Behavioral, social, and emotional assessment of children and adolescents.* (3rd ed.). New York, NY: Taylor & Frances Group, LLC.

Mesman, J., Bongers, I., & Koot, H. (2001). Preschool developmental pathways to preadolescent internalizing and externalizing problems. *Journal of Child Psychology and Psychiatry, 42,* 679-689.

Myers, C. (2013). Comparing results from the Clinical Assessment of Behavior and Child Behavior Checklist with referred preschoolers. *Psychology in the Schools. 50,* 1-12.

Myers, C., Bour, J., Sidebottom, J., Murphy, S., & Hakman, M. (2010). Same constructs, different results: Examining the consistency of two behavior rating scales with referred preschoolers. *Psychology in the Schools, 47,* 205-216.

Nix, R. L., Bierman, K. L., Domitrovich, C. E., & Gill, S. (2013). Promoting children's social-emotional skills in preschool can enhance academic and behavioral functioning in kindergarten: Findings from Head Start REDI. *Early Education and Development, 24,* 1000-1019.

Nix, R. L., Bierman, K. L., Heinrichs, B. S., Gest, S. D., Welsh, J. A., & Domitrovich, C. E. (2016). The randomized controlled trial of Head Start REDI: Sustained effects on developmental trajectories of social-emotional functioning. *Journal of Counseling and Clinical Psychology, 85,* 140-159.

Noonan, M., & McCormick, L. (2006). *Young children with disabilities in natural environments*: *Methods & procedures.* Baltimore, MD: Paul H. Brooks Publishing Company.

Ponitz, C. C., McClelland, M. M., Matthews, S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45,* 605-619.

Reynolds, C., & Kamphaus, R. (2015). *Behavior assessment system for children, third edition.* San Antonio, TX: Pearson Education.

Rhoades, B. L., Warren, H. K., Domitrovich, C. E., & Greenberg, M. T. (2011). Examining the link between preschool social–emotional competence and first grade academic achievement: The role of attention skills. *Early Childhood Research Quarterly, 26,* 182-191.

Shapiro, E., & Heick, P. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavior/emotional problems. *Psychology in the Schools, 41,* 551-561.

Shapiro, E., & Kratochwill, T. (2000). *Conducting school-based assessments of child and adolescent behavior.* New York, NY: Guilford Press.

Squires, J., Bricker, D., Heo, K., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-completed screening measure. *Early Childhood Research Quarterly,* 16, 405-419.

Stein, S. (2007). Review of the BASC second edition. In K. F. Geisinger, R. A. Spies, J. F. Carlson, & B. S. Plake (Eds), *The seventeenth mental measurements yearbook*. Lincoln, NE: Buros Center for Testing.

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography, 6,* 35-39.

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,* 349-357.

Webster-Stratton, C., & Hammond, M. (1998). Conduct problems and level of social competence in Head Start children: Prevalence, pervasiveness, and associated risks. *Clinical Child and Family Psychology Review, 1,* 101-124.

Webster-Stratton, C., Reid, J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years teacher and child training programs in high-risk schools. *Journal of Child Psychology and Psychiatry, 49,* 471-488.

Worthen, B. R., Borg, W. R., & White, K. R. (1993). *Measurement and evaluation in the schools: A practical guide.* White Plains, NY: Longman.

Yoder, P., & Symons, F. (2010). *Observational measurement of behavior.* New York, NY: Springer Publishing Company.

Zhai, F., Raver, C. C., & Jones, S. M. (2015). Social and emotional learning services and child outcomes in third grade: Evidence from a cohort of Head Start participants. *Children & Youth Services Review, 56,* 42-51.

Zionts, P., Zionts, L., & Simpson, R. (2002*). Emotional and behavioral problems: A handbook for understanding and handling students.* Thousand Oaks, CA: Corwin Press, Inc.

Appendix: Institutional Review Board Approval Letter

INSTITUTIONAL REVIEW BOARD
OFFICE OF RESEARCH INTEGRITY

DATE:	October 17, 2016
TO:	Carly Rentsch, BA
FROM:	Western Kentucky University (WKU) IRB
PROJECT TITLE:	[906034-1] Specialist Project: THE CONSISTENCY OF TEACHER RATINGS ON THE BEHAVIOR ASSESSMENT SYSTEM FOR CHILDREN-3 AND THE CHILD BEHAVIOR CHECKLIST
REFERENCE #:	IRB 17-110
SUBMISSION TYPE:	New Project
ACTION:	APPROVED
APPROVAL DATE:	October 17, 2016
EXPIRATION DATE:	February 28, 2017
REVIEW TYPE:	Expedited Review

Thank you for your submission of New Project materials for this project. The Western Kentucky University (WKU) IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a *signed* consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

This project has been determined to be a Minimal Risk project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of February 28, 2017.

Please note that all research records must be retained for a minimum of three years after the completion of the project.

If you have any questions, please contact Paul Mooney at (270) 745-2129 or irb@wku.edu. Please include your project title and reference number in all correspondence with this committee.