DIAGNOSTIC ACCURACY OF A UNIVERSAL BEHAVIOR SCREENER FOR IDENTIFICATION OF ATTENTION PROBLEMS IN FIRST GRADE

by

Erin Ezell

July, 2018

Director of Thesis: Christy M. Walcott, PhD

Major Department: Psychology

Teacher-rated inattention predicts reading outcomes, and children with both reading and elevated inattention may require special intervention. Therefore, screening of inattention seems warranted. It is examined whether teacher ratings on a universal behavior screener, the *Social Emotional Academic Behavior Rating Scale* (SAEBRS), has a similar diagnostic accuracy to the *ADHD-IV-Rating Scale, Inattention Scale* for first-grade students (n = 273). The SAEBRS-Total Behavior (SAEBRS-TB) and SAEBRS-Academic Behavior (SAEBRS-AB) scores were utilized for this study. Based on a receiver operator characteristic analysis, the SAEBRS-TB and SAEBRS-AB yielded high diagnostic accuracy. A SAEBRS-TB cut score of 40 yielded acceptable sensitivity and specificity, and a SAEBRS-AB cut score of 11 yielded optimal sensitivity and acceptable specificity. Therefore, the SAEBRS appears to be an accurate method to screen for children who are at-risk for elevated inattention.

DIAGNOSTIC ACCURACY OF A UNIVERSAL BEHAVIOR SCREENER FOR IDENTIFICATION OF ATTENTION PROBLEMS IN FIRST GRADE

A Thesis

Presented to the Faculty of the Department of Psychology

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

MA in School Psychology

by

Erin Ezell

July, 2018

DIAGNOSTIC ACCURACY OF A UNIVERSAL BEHAVIOR SCREENER FOR IDENTIFICATION OF ATTENTION PROBLEMS IN FIRST GRADE

by

Erin Ezell

APPROVED BY:

DIRECTOR OF THESIS: _____
Christy M. Walcott, PhD

COMMITTEE MEMBER: _____
Lori Curtindale, PhD

COMMITTEE MEMBER: _____
Karl L. Wuensch, PhD

INTERIM PSYCHOLOGY DEPARTMENT CHAIR: _____
D. Erik Everhart, PhD

DEAN OF THE
GRADUATE SCHOOL: _____
Paul J. Gemperline, PhD

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER I: INTRODUCTION & LITERATURE REVIEW 1

Student Behavior in the United States Education System 2

Universal Screening 5

Elevated Inattention 7

Reading and Inattention 7

Math and Inattention 9

Screening for Elevated Inattention 10

Conners 10-Item Scale 10

Strengths and Weaknesses of Attention-Deficit Hyperactivity 11

Swanson, Nolan, and Pelhman IV Scale 12

ADHD-IV- Rating Scale 13

Options for Universal Behavior Screening 14

Behavior and Emotional Screening System 15

Direct Behavior Ratings 16

Office Disciplinary Referrals 17

Strengths and Difficulties Questionnaire 18

Student Risk Screening Scale 19

Systematic Screening for Behavior Disorders 19

Social, Emotional, and Academic Behavior Rating Scale 20

Statement of the Problem 21

Hypotheses 21
CHAPTER II: METHODS 23
Participants 23
Measures 23
SAEBRS 23
Reliability 24
Validity 24
ADHD-RS-IV 25
Reliability 25
Validity 26
Procedures 27
Data Analysis Plan 27
CHAPTER III: RESULTS 31
Data Screening 31
Accuracy of the Data Entry 31
Descriptive Statistics 31
Hypothesis Testing 32
Hypothesis 1 33
Hypothesis 2 34
Hypothesis 3 36
CHAPTER IV: DISCUSSION 37
Implications and Directions for Future Research 39
Limitations 40

Conclusions 40

REFERENCES 42

APPENDIX A: IRB APPROVAL 56

LIST OF TABLES

1. 2x2 Contingency Table 29
2. Descriptive Statistics 31
3. Areas under the receiver operating characteristic curves (AUC) 33
4. Predictive Accuracy of SAEBRS-TB Cut Scores with ADHD-IV-RS 34
5. 2x2 Contingency Table (ADHD-IV-RS vs. SAEBRS-TB) 34
6. Predictive Accuracy of SAEBRS-AB Cut Scores with ADHD-IV-RS..... 35
7. 2x2 Contingency Table (ADHD-IV-RS vs. SAEBRS-AB) 35

LIST OF FIGURES

1. Comparison of ROC Curves 32

CHAPTER I: INTRODUCTION & LITERATURE REVIEW

Recent legislation requires schools to implement a Response to Intervention (RTI) model based on multi-tiered systems of support (MTSS; President's Commission on Excellence in Special Education, 2002). The goal of RTI is to use early detection to identify students who are struggling academically or behaviorally and provide appropriate interventions to prevent these challenges from becoming more serious and detrimental to students' success (Eagle, Dowd-Eagle, Snyder, & Holtzman, 2015). Under the traditional system, students may not receive extra services until a problem becomes severe, and they meet criteria for a special education qualification. RTI helps schools identify children earlier than previous methods using systematic and scientific universal screeners of all children. Therefore, RTI gives students who are at-risk the opportunity to receive less intensive intervention services, which may prevent more serious academic and behavior problems (e.g., substance abuse, academic failure, social rejection) in the future (Eklund et al., 2009).

Under MTSS, RTI approaches are delivered according to increasing levels (tiers) of student support required. Tier 1 involves students who are receiving general education services and do not need extra support in the classroom; Tier 1 typically encompasses about 80% of the student body. Furthermore, Tier 1 involves quarterly, universal screening approaches to monitor all students' progress as well as quality core instruction and basic interventions (Cook, Volpe, & Livanis, 2010). Interventions at the Tier 1 level can involve assisting a teacher with classroom management skills, modifying a schoolwide reading curriculum, or implementing a schoolwide behavior system, such as Positive Behavior Interventions and Supports (PBIS; Burke, et al., 2012).

Tier 2, which usually entails about 15% of the student body, is designated for students who are flagged during universal screenings and need extra support but do not qualify for special education (Sugai & Horner, 2006). The aim of Tier 2 is to catch students' problems early, provide basic targeted interventions within the general education environment, and prevent long-term negative impact. The goal of a Tier 2 intervention is to support students in meeting grade level benchmarks and catching up to their same-aged peers through evidence-based interventions. Examples of Tier 2 interventions are extra reading supports, behavior charts, and home-school notes. After continuous progress monitoring, a problem-solving team may decide to place a student back into Tier 1, continue Tier 2 interventions, or consider evaluating the student for more intensive Tier 3 services. Special education services, such as speech therapy, resource classes, and self-contained classes, are considered Tier 3 services. Ideally, if early intervention is working well at Tiers 1 and 2, only about 5% of children should need services at the intensity of Tier 3. Tier 3 services typically require students to have an Individualized Education Plan (IEP) because they offer specially designed instruction. RTI is intended to address both academic and emotional-behavioral difficulties, and this study will address a Tier 1 universal behavior screener to assist in identifying students for Tier 2 interventions to prevent future academic challenges (Sugai & Horner, 2009).

Student Behavior in the United States Education System

Twenty percent of children exhibit symptoms consistent with a diagnosable mental disorder (Merikangas et al., 2010), and researchers estimate that between 1% and 5% of children have a severe emotional and/or behavioral disorder (Costello, Mustillo, Erkanli, Keller, & Angold, 2003); however, children often do not receive mental health services until their symptoms have become severe. In fact, behavioral health services are only offered to about 15 to

20% of children per year with emotional and behavioral concerns (Ringel & Sturm, 2001), and only 1% of school-aged children are identified with an emotional and/or behavior disorder in the school setting and receive special education for these disorders. Nevertheless, this group of children utilizes over 50% of teachers and administrations' time and resources (United States Public Health Service, 2000). For example, in a school of 500 children, five to 25 children will have a severe emotional and/or behavior disorder. Yet, although it is estimated this group of children would utilize over 50% of staff time and resources, only one to five of these children would receive mental health services.

Children who receive mental health services in the schools are most often referred by their teachers; however, teacher referral is not necessarily the most efficient means of determining students who need mental health services (Tilly, 2008). Sometimes Office Disciplinary Referrals (ODRs) are also utilized to identify at-risk children. Like teacher referrals, ODRs are also not a systemic and reliable way to measure behavior, as teachers may not be consistent in their reasoning for writing referrals. Consequently, ODRs and teacher referrals do not serve as a preventative practice, as they may not occur until the behavior has become severe (Eklund et al., 2009). Furthermore, children who have internalizing behavior problems often do not receive ODRs (McIntosh, Campbell, Russell, & Zumbo, 2009); however, these children are still at-risk for academic concerns (Kalberg, Lane, Driscoll, & Wehby, 2011). Based on the prevalence of behavioral problems in schools, universal screening is warranted to identify students who need supports.

Using a preventive model, such as the RTI model described above, to catch problems early may prevent children from developing more serious or longer-term problems. Universal screeners identify children not adequately responding to Tier 1 instruction and basic behavioral

supports and detect children who would otherwise not receive services until the problem behavior becomes severe (Kratochwill, 2007). For example, previous research demonstrates that children who were identified by the *Behavioral and Emotional Screening System* (BESS) as at-risk earned lower report card grades than their peers, whether or not they were deemed at-risk according to traditional teacher referral methods. Consequently, universal behavior screening not only identifies children who are at-risk for emotional and behavior problems, but it may also help identify children who are at-risk for academic problems (Eklund et al., 2009).

Academic universal screeners generally encompass quick assessments that are given to every child at least three times per year. Behavior universal screeners typically include rating scales completed by teachers for each student. Each assessment and rating scale has an identified cut-off point to determine which students are deemed at-risk. Once a student is identified as at-risk, he or she is ideally provided an intervention to address his or her needs. Academic universal screeners are widely used in the elementary setting. Conversely, behavior screeners are not as prevalent, and many behavior scales tend to focus on externalizing behaviors (Cook, Volpe, & Livanis, 2010). Because disruptive behaviors are easier to observe in the classroom than internalizing problems (Lane, Parks, Kalberg, & Carter, 2007), traditional behavior screening methods utilized in schools often are able to identify children with externalizing problems, such as aggression and hyperactivity (McIntosh, Ty, & Miller, 2014). However, since children with internalizing behavior problems are less likely to cause classroom disruptions, it is less probable that this group of children is identified by their teachers as needing extra behavioral supports. Thus, children with other behavior problems that impact academics, such as inattention, may not be detected by some universal screeners (Kalberg et al., 2011). Therefore, they may not receive the necessary interventions to prevent future academic problems.

Universal Screening

As previously mentioned, universal screening plays a vital role in the RTI process to identify children who are at-risk of future academic and/or behavior problems (Cook et al., 2010; Jenkins, Hudson, & Johnson, 2007). The goal of universal screeners is to be a fast and efficient method to gain knowledge about students. Academic universal screeners are widely available for reading, writing, and math in elementary school, and they may be more prevalent because academic screeners are easy to access. Academic screeners are typically administered three times during the school year: fall, winter, and spring. Universal academic screeners allow teachers to carefully monitor student progress and act accordingly (Kilgus, Riley-Tillman, Chafouleas, Christ, & Welsh, 2014). For example, if only 40% of third-grade students are scoring at the grade level benchmark for oral reading fluency, school personnel can make an informed decision to implement instruction geared towards improving oral reading fluency. However, if 80% of the students meet the benchmark, teachers can use more individualized instruction to assist the 20% of students who did not meet the benchmark.

Curriculum-based measurement (CBM) is an example of a well-established and widely used academic screening tool (Kilgus et al., 2014). CBMs are a type of general outcome measure (GOM) that provides collective information about a child's performance, as opposed to providing specific information linked to one academic standard (Deno, Fuchs, Maston, & Shin, 2001). GOMs are efficient, effective, and useful through the RTI framework. At the Tier 1 level, the teacher can quickly administer probes to easily gain data on every student. The students' scores can be individually compared to local and/or national norms. Students who are at-risk can be placed in Tier 2 interventions, and GMOs serve as a practical way to progress monitor during the intervention (Hintze, Owen, Shapiro, & Daly, 2000). Although GOMs are relatively

established for academics and evidence-based universal behavior screeners exist, behavioral GOMs have not been established but are necessary for progress monitoring (Kilgus et al., 2014).

Universal screening for behavior is not utilized as widely as academic screening, but it is becoming increasingly prevalent (Burke et al., 2012). Glover & Albers (2007) identified three vital criteria universal screeners must contain: appropriateness for intended use, technical validity, and usability/practicality. Appropriateness entails the screener having research-based evidence to support the screeners' purpose. The screener must also be developmentally relevant to the school's population and needs. Additionally, it is essential that the screener is feasible for the school and provides the desired information. Technical validity involves ensuring that the screener has recent norming data based on a large population of students and has been normed on a population that is similar to the school's students. The social validity of the scale is important, as acceptance of the procedure by teachers may produce more accurate screening results (Glover & Albers, 2007).

Under the usability/practicality standard, Glover and Albers (2007) emphasize that the scale must be cost-effective; therefore, the cost of the scale per student is taken into consideration. Some scales may involve specialized training to administer and interpret the results, whereas other scales do not involve intensive training. Furthermore, universal screeners are not intended to disrupt classroom instructional time or to be an additional burden to teachers. The usability/practicability standard also calls for the screener to produce results that assist with providing appropriate interventions. The goal of the screener is to identify children who are at-risk and provide them with supports that will improve student outcomes and prevent future behavior problems (Glover & Albers, 2007).

In addition to universal screeners being appropriate for intended use, containing technical validity, and being usable and practical (Glover & Albers, 2007), Severson, Walker, Hope-Doolittle, Kratochwill, & Gresham (2007) identified three methods of behavioral screening: multiple gating procedures, teacher evaluation and ratings of every student on common behavioral criteria, and teacher nomination of problem students, which is followed by teachers completing an evidence-based rating scale to provide more comprehensive information. The three previously mentioned methods are deemed to be more effective than teachers randomly referring students (Severson et al., 2007). The current study will explore teacher ratings of every student on common behavioral criteria.

Elevated Inattention

Despite not always being an obvious and observable trait in the classroom, attention is important for academic success. In fact, elevated levels of early inattention predict future academic problems (Rabiner, Coie, & Conduct Problems Prevention Research Group, 2000). Demonstrated using both clinic-based and population-based studies, robust literature suggests a negative link between teacher-rated attention problems and academic functioning. Even children who are not diagnosed with ADHD, but display subclinical levels of attention problems, have academic challenges (Rodriguez et al., 2007). Although attention problems are not typically a specific focus of universal screening, research indicates that is may be necessary to screen for attention problems.

Reading and inattention. Early reading problems may predict negative long-term academic outcomes, and studies have demonstrated that elevated inattention in children is associated with reading difficulties (McGee, Prior, Williams, Smart, & Sanson, 2002; Rabiner et al., 2000). Teacher-rated elevated inattention in early childhood is shown to be a useful predictor

in later reading outcomes (McGee et al., 2002). A study of 13,087 children from Sweden, Denmark, and Finland indicated that elevated inattention and hyperactivity may forecast reduced academic achievement. In this group of children, elevated inattention appeared to have a greater influence on academic impairment compared to elevated hyperactivity (Rodriguez et al., 2007). Another study revealed that children with elevated inattentive symptoms from ages six to 11 also displayed declining academic achievement at ages 11 and 17 (Breslau et al., 2010). Additionally, a study of 204 preschool children suggested that elevated inattention at the preschool age can impair emergent reading skills (Sims & Lonigan, 2012). Furthermore, it is estimated that kindergarteners' levels of inattention, as rated by the teacher, is significantly correlated with lower scores on phonemic awareness and rapid naming assessments during first grade (Walcott, Sheemaker, & Bielski, 2010).

Moreover, in a study of 4,148 children in England, children's math and reading progress were tracked from the beginning of kindergarten to the end of first grade (Merrell & Tymms, 2001). At the end of kindergarten, each child was rated by his/her teacher using a behavior scale based on the diagnostic criteria of ADHD in the DSM-IV. When reading and math abilities at the start of kindergarten were controlled, teacher-rated elevated inattention predicted impaired academic achievement during kindergarten and first grade. Consistent with previous research, children with elevated symptoms of predominately inattentive and combined ADHD experienced more significant impairment than children only exhibiting elevated symptoms of hyperactivity (Merrell & Tymms, 2001). Unfortunately, research has shown that typical tutoring interventions may not be as effective for children who have both elevated inattention and reading problems (Rabiner, Malone, & Conduct Problems Prevention Research Group, 2004). Thus, different types of Tier 2 interventions may be needed for this dually-affected population.

Math and inattention. Not only do children with elevated inattention have difficulties with reading, but research also indicates that elevated inattention can impact math performance. The research conducted by Fuchs et al. (2005) indicated that first-grade teacher ratings of inattention predicted the development of first-grade mathematical skills, and the ratings were also related to mathematical outcomes in the third grade (Fuchs et al., 2006). Raghubar et al. (2009) studied the effect of inattention on mathematical outcomes in 291 third and fourth graders. To assess for elevated inattention, teachers completed the *Strengths and Weakness of ADHD-Symptoms and Normal Behavioral Rating Scale* (SWAN; Swanson et al., 2001). This scale consists of 18 items at are based on the ADHD criteria in the *Diagnostic and Statistical Manual of Mental Disorders- Fourth Edition, Text Revision* (DSM-IV-TR; American Psychiatric Association, 2000). The scale compares children's attentive, hyperactive, and impulsive behaviors to their same-aged peers (Raghubar et al., 2009). Based on standardized subtests of reading and arithmetic, children were grouped into one of four groups: math and reading learning disabilities, math learning disability, reading learning disability, and no learning disability. Consistent with previous math research (Fuchs et al. 2005; Fuchs et al., 2006), and as indicated by research conducted on reading problems and inattention, the study found that a relationship did not exist between hyperactivity and math computation; however, a relationship did exist between elevated inattention and math computation abilities. There was not a distinct difference in teacher-rated inattention among the groups of children of children with learning disabilities. However, children in the math learning disability group were rated as having higher levels of inattention compared to children in the no learning disability group. Additionally, children who had increased ratings of inattention performed multi-digit computation tasks with less accuracy compared to their more attentive peers. Although elevated inattention was not associated with

operation switch errors, children with elevated inattention were more likely to make math fact and procedural errors in multi-digit arithmetic (Raghubar et al., 2009). The large-scale European cohort study mentioned above (Rodriguez et al., 2007) also found that students who had elevated inattention had significantly greater odds of having teacher reported deficits in math.

Screening for Elevated Inattention

Given that teacher-rated inattention predicts later academic outcomes, universal screening of teacher-rated inattention seems warranted. The current method for detecting inattention is via ADHD rating scales. Several ADHD screeners exist, but it is not feasible for teachers to rate every student on these diagnostic scales in addition to completing other universal behavior rating scales. The reasons for this are twofold; first, it is too time consuming for teachers to complete specific behavior screeners in addition to broadband behavior screeners, and second, ADHD rating scales are meant for specific diagnostic assessment that may require parental permission. Consequently, schools may find it advantageous to use one universal behavior rating scale that assesses multiple areas but still captures inattentive symptoms. Most current universal screening options consider various elements of social behavior (e.g., behaviors required for academic success, emotional behavior concerns, social/disruptive behaviors). However, typically, universal behavior screeners do not directly address academic behaviors including inattention. Although ADHD rating scales are not feasible for universal screenings, below is a review of common ADHD rating scales from which we chose a gold standard.

Conners 10-Item Scale. The *Conners 10-Item Scale* (Conners, 1973; Conners, 1990) involves a parent and teacher version. The parent or teacher rates each of the listed behaviors on a four-point rating scale, with a rating of zero being ‘Not at all True’ and a rating of three being

'Very much True'. The 10 items on this scale are from the Hyperactivity Index of the full Conners scale. (Westerlund, Ek, Holmberg, Naswall, & Fernell, 2009).

In one study, the *Conners 10-Item Scale* was administered to the parents and teachers of first graders (N = 422) in Sweden. Parents completed the questionnaire at the beginning of the school year, and teachers completed the scale after seven months of the school year elapsed. The same group of students was screened in fourth grade; their parents participated in a structured interview that included the *Conners 10-Item Scale*, the teachers completed the *Conners 10-Item Scale*, and the students were assessed for ADHD (Holmberg, Sundelin, & Hjern, 2012). In this study, pervasive ADHD entailed children who exhibit clinically significant ADHD symptoms, based on the DSM-IV criteria, at both home and school. Children who were identified as having ADHD based on the *Conners 10-Item Scale* in the first grade were approximately seven times more likely to be diagnosed with pervasive ADHD in the fourth grade. Out of the children identified as meeting the DSM-IV criteria for ADHD, compared to the parents' ratings, the fourth-grade teachers were twice as likely to correctly identify them based on their *Conners 10-Item Scale* (Holmberg et al., 2012). It is important to note that the researchers in the study did not differentiate between ADHD the three subtypes of ADHD: predominately inattentive, predominately hyperactive, and combined.

Strengths and Weaknesses of Attention-Deficit Hyperactivity-Symptoms and Normal-Behaviors. The *Strengths and Weaknesses of Attention-Deficit Hyperacidity-Symptoms and Normal-Behaviors* (SWAN) is normed on diverse populations (Smalley et al., 2007). The scale consists of 30 items. To eliminate socio-cultural and statistical bias, a dimensional discrimination from average rates to extreme rates is used. The rater compares the child or adolescent to other same-aged peers on a score from – 3 (below average) to +3 (above average),

where 0 is normal (Brites, Salgado-Azoni, Ferreira, Lima, & Ciasca, 2015). In large-scale population studies, SWAN identifies a prevalence of ADHD in 4% of the sample, which is relatively consistent with an estimated international prevalence rate of 5% (Lai et al., 2013; Robaey, Amre, Schachar, & Simard, 2007).

Previous research comparing the teacher ratings of the SWAN and the *Conners' Continuous Performance Test- Second Edition* demonstrated a 0.90 correlation (Cornish et al., 2005). SWAN has been used in diverse cultures and languages. Research indicates that the French version of the scale has excellent stability (0.86), excellent specificity (0.88), good internal consistency (Cronbach's alpha > 0.80), and strong validity (Robaey et al., 2007). Strong reliability and validity have also been found in other countries, such as China (Lai et al., 2013) and Spain (Kudo, Altamirano, & Mearns, 2012; Lakes, Swanson, & Riggs, 2012).

Swanson, Nolan, and Pelhman IV Scale. The *Swanson, Nolan, and Pelhman IV Scale* (SNAP-IV) has a long form and a short form. The long form primarily measures ADHD and oppositional defiant disorder (ODD) symptoms, but it also provides data about symptoms that overlap with other DMS-IV childhood psychiatric disorders (Bussing et al., 2008). The shorter version of the SNAP-IV, known as the *Multimodal Treatment* (MTA; Swanson et al., 2001), consists of 26 items, with nine items related to ADHD symptoms of inattention, nine related to ADHD symptoms of hyperactivity/impulsivity, and eight items related to the symptoms of ODD. Each item is rated on a scale from 0 (not at all) to 3 (very much). The average score is calculated for each subscale: inattentive, hyperactive/impulsive, combined ADHD, and ODD. Average subscale scores that are greater than the 95th percentile are considered clinically significant (Brites et al., 2015). The MTA version has acceptable reliability with a coefficient alpha of 0.97

for overall ratings and subdomain alpha ranging from 0.92 to 0.96. The scale also has acceptable internal consistency (Bussing et al., 2008).

A limitation of the scale is that the sample used to determine the cut-off points primarily involved Hispanic low-come elementary students (Gaub & Carlson, 1997); thus, research needs to be conducted to determine the generalizability of the SNAP-IV to other populations (Brites, Salgado-Azoni, Ferreira, Lima, & Ciasca, 2015). Furthermore, there are no differences in the cut-off points between age and gender, which decreases the feasibility of the scale in clinical settings (Collett, Ohan, & Myers, 2003). However, research conducted by Bussing et al. (2008) suggests that there are no significant differences in cut-off points between gender, age, and race on the MTA scale for elementary school students. Conversely, based on teacher rating of 1,205 elementary school students, a medium effect size indicated that teacher rated African American students significantly higher than White students, despite the minimal difference in parent ratings (Brites et al., 2015). Additionally, boys were rated higher than girls for hyperactivity/impulsivity, inattention, and ODD, which is consistent with previous research on ADHD (Bussing et al., 2008).

ADHD-IV- Rating Scale. One popular screener is the *ADHD-IV-Rating Scale* (ADHD-IV-RS; DuPaul, Power, Anastopoulos, & Reid, 1998). The ADHD-IV-RS is a norm-referenced and reliable scale with evidence of validity for the purpose of identifying symptoms of ADHD (DuPaul et al., 1998). The ADHD-IV-RS consists of two subscales, hyperactivity-impulsivity and inattention, with parent and teacher forms. Each subscale consists of nine questions that are related to the DSM-IV criteria for ADHD. Although the other scales ask questions related to inattention, hyperactivity-impulsivity and inattention are often grouped into one score. Conversely, the ADHD-IV-RS Inattention Scale is an empirical method to screen for DSM-IV

symptoms of elevated inattention (Goodman, 1997). This is deemed to be the gold standard because it is brief, mapped on to the DSM-IV criteria for inattention symptoms, has a separate inattention score, and has strong psychometrics.

Options for Universal Behavior Screening

To screen for students at-risk for social-emotional and behavior concerns, schools implement various screening methods. Some procedures include multiple gating procedures, evaluation of extant data (e.g., ODRs), and evaluation of teacher ratings on various behavior rating scales (Severson et al., 2007). There are limitations to the current universal behavior screening methods. Some screeners require a large amount of teacher time to complete, lack technical adequacy, and are not based on behaviors that have contextual relevant to academic success (Kilgus, Chafouleas, & Riley-Tillman, 2013).

In regards to family rights, Chafouleas, Kilgus, and Wsaeallach (2010) identified four ethical dilemmas associated with implementing school-based behavior screeners: community acceptance (e.g., social stigma associated with mental health issues; belief that schools are not the appropriate place for mental health interventions), family rights (e.g., screening may be invasive to family privacy), identification (e.g., elevated level of false positive from screeners may lead to misidentification; possible disproportionate identification of specific cultures and races), and service delivery provision (e.g., schools may not have appropriate mental health resources; careful selection of screening tools). Parental consent is not sought for academic screening; however, there are nebulous criteria of whether parental consent ought to be sought for behavior screening. Chafouleas et al. (2010) propose that universal screening is under the umbrella of typical classroom instruction. It is recommended that parents are informed of the screening procedures and purpose of the screeners in advance, staff members are trained on

confidentially, and parents have the option to opt out of the screening process (National Research Council and Institute of Medicine, 2009). To determine a practical method to screen for attention problems, several different universal screeners were considered.

Behavioral and Emotional Screening System. The *Behavioral and Emotional Screening Scale* (BESS) is used to measure emotional and behavioral strengths and weaknesses for children between the ages of three and 18. The BESS is based on normative data from the *Behavioral Assessment System for Children, Second Edition* (BASC-2; Reynolds & Kamphaus, 2004). Items on the BASC-2 *Teacher Rating Scale* with the strongest psychometrics were compiled into the shorter version of the scale, the BESS. The BESS Teacher Child/Adolescent form is intended for kindergarteners through 12th graders and contains 27 items. The teacher rates how often a student has displayed certain behaviors (0 = *Never*, 1 = *Sometimes*, 2 = *Often*, 3 = *Almost Always*). Scores are summed into four dimensions: Adaptive Skills, Externalizing Problems, Internalizing Problems, and School Problems (Reynolds & Kamphaus, 2004). It is estimated the BESS requires less than five minutes per child to complete, and thus, requires about one hour to complete for an entire class (Dowdy, Ritchey, & Kamphaus, 2010).

Based on confirmatory and exploratory factor analyses, an elevated total score can indicate that a student is at-risk for maladaptive behaviors, which could be useful for universal screening (DiStefano, Greer, & Kamphaus, 2013). Research indicates that the BESS has acceptable test-retest reliability of 0.80 to 0.91 (Dowdy et al., 2010; Kamphaus & Reynolds, 2007), and the overall score and dimension scores remain consistent over time (DiStefano & Morgan, 2010). Internal consistency ranges from a Cronbach's alpha of 0.83 to 0.96. As a measure of validity, previous research indicates that BESS scores predict various behavioral and academic outcomes, such as attendance, ODRs, special education status, and grades (Kilgus et

al., 2014). Additionally, the BESS correlates with other rating scales (Kamphaus & Reynolds, 2007), such as the *Conner's Teacher Rating Scale-Revised* (Conners, 1997).

Direct Behavior Ratings. The *Direct Behavior Ratings Single Item Scales* (DBR-SIS) involve a teacher providing behavior ratings for a student based on a pre-determined period that may last for a few minutes (Chafouleas et al., 2010) to hours (Kilgus et al., 2012). The teacher rates the child on a scale from 0 (*never*) to 10 (*always*; Kilgus et al., 2014). The ratings only take a few seconds per child (Chafouleas et al., 2010). Due to the flexibility of the DBR-SIS, the teacher can efficiently rate multiple children each day and/or rate one child several times throughout the day (Kilgus et al., 2014). When DBR-SIS is used as a universal screener, it is recommended that five to 10 data points are collected to determine which students are at-risk and to have reliable data (Chafouleas et al., 2013).

Research suggests that the DBR-SIS shows concurrent validity for kindergarteners based on moderate to high correlations with the *Social Skills Rating System* (SRSS; Gresham & Elliott, 1990) Social Skills and Problem Behavior subscales (Chafouleas, Kilgus, & Hernandez, 2009). Additionally, teachers rated their second-grade students (n = 118) on disruptive behavior, academic engagement, and compliance five to 10 times using DBR-SIS during the spring, and concurrent validity for the DBR-SIS was exhibited through high correlations of the teacher ratings compared to the BESS (Kilgus et al., 2012).

Kilgus et al. (2014) conducted a large-scale study utilizing the DBR-SIS as a universal screener to measure Disruptive Behavior, Academic Engagement, and Respectful Behavior of 1,108 students. The sample included first-grade teachers (n = 31), fourth-grade teachers (n = 25), and seventh-grade teachers (n = 23). Concurrent validity for the three tested domains of the DBR-SIS predicted BESS and SRSS ratings. The Academic Engagement ratings, followed by

Disruptive Behavior, correlated the strongest with the BESS for all grade levels. Respectful Behavior cut scores were consistent across the three grade levels. However, cut scores for Academic Engagement and Disruptive Behavior differ across the grade levels. Based on the findings, elevated Disruptive Behavior ratings in the first grade are most indicative of at-risk status. Conversely, Academic Engagement had the strongest correlation for fourth and seventh grades (Kilgus et al., 2014).

Office Disciplinary Referrals. Office Disciplinary Referrals (ODRs) are commonly used to identify children who are at-risk. Schools may analyze this data on a systems level to determine trends, such as time of day, grade level, and violation. This information can provide information about the school's needs and focus areas for intervention (Sugai, Sprague, Horner, & Walker, 2000). Students who receive a high number of referrals or referrals related to extreme behaviors may be noted as at-risk by the school. Sugai et al. (2000) found that if a student earns zero or one ODR within a school year, he/she is considered low risk. However, earning two to five ODRs is considered moderate risk, and if a student receives greater than six ODRs, high risk is indicated (Sugai, Sprague, Horner, & Walker, 2000).

Although ODRs are commonly used to screen for some behavior risks, ODRs have not been thoroughly researched as a universal behavior screener, and several variables of ODR have not been considered throughout research studies. For example, researchers have not investigated how the level of infraction alters the level of behavioral risk (Burke et al., 2012). Kern and Manz (2004) also found that numerous school variables, including teacher perception, teacher tolerance, and perceived level of administrative support, influence the validity of ODRs (Kern & Manz, 2004). Furthermore, by the time students receive an ODR, they have displayed frequent problem behaviors or displayed at least one severe problem behavior (Burke et al., 2012), and

children who have internalizing behavior problems often do not receive ODRs (McIntosh, Campbell, Russell, & Zumbo, 2009). Therefore, it would be ideal for a universal screener to detect children who are at-risk before an ODR is necessary and to identify children with internalizing behavior concerns (Burke et al., 2012; McIntosh et al., 2009).

Strength and Difficulties Questionnaire. The *Strength and Difficulties Questionnaire* (SDQ) is a free behavior screener for children ages three to 16 years old (Goodman, 1997). Parent and teacher SDQ forms for children who are three years old consist of 22 items. The scale intended for children ages four to 16 consists of 25 items. Raters indicate their responses based on a three-point Likert-scale (0 = *Not True*, 1 = *Somewhat True*, and 2 = *Certainly True*). The screener can be completed by the student's parent and/or teacher, and students who are 11 years or older can also complete a self-report form (Goodman, 1997).

The SDQ assesses four behavior domains: conduct problems, hyperactivity, peer problems, and emotional symptoms. An overall total difficulties score is calculated from these four domains. Moreover, the screener consists of a prosocial behavior domain (Goodman, 1997). When using the SDQ in a large sample with low risk, the four domains can be combined into internalizing problems (emotional and peer symptoms), externalizing (conduct and hyperactivity), and prosocial scales. Though it is acceptable to combine these scales for screening purposes, it is not recommended to combine the scales for diagnostic purposes (Goodman, Lamping, & Ploubidis, 2010). The SDQ identified children with previously diagnosed psychiatric disorders with a specificity of 0.95 and a sensitivity of 0.66 (Goodman, 1997). The SDQ shows good validity for child psychiatric conditions in over 60 languages (Woerner et al., 2004). The teacher scale shows high internal reliability, with a Cronbach's alpha coefficient of 0.88 for hyperactivity (Goodman, 2001).

Student Risk Screening Scale. The *Student Risk Sisk Screening Scale* (SRSS) is used as a universal screener aimed at identifying children with antisocial behaviors (Lane, Parks, Kalberg, & Carter, 2007). The SRSS is free and efficient. Completion of the screener for an entire class takes an average of 10 minutes (Lane et al., 2010). Using a rating scale (0 = *Never*, 1= *Occasionally*, 2 = *Sometimes*, and 3 = *Frequently*), teachers rate the students on seven items: (a) steal; (b) lie, cheat, sneak; (c) behavior problem; (d) peer rejection; (e) low academic achievement; (f) negative attitude; and (g) aggressive behavior. A total score is calculated based on the seven items, and students are placed into one of three categories: low (0-3); moderate (4-8), and high (9-12; Drummon, 1994; Lane et al., 2015). Using the SRSS, students scores can be analyzed over time (Lane et al., 2008).

SRSS scores have established internal consistency and test-rest reliability in K-12 settings (Lane et al., 2012). SRSS scores for elementary school students predicted end of year behavioral performance, based on ODRs and reading performance (Menzies & Lane, 2012; Oakes et al., 2010). Elementary school students' SRSS scores also predicted self-control skills measured by the *Social Skills Rating System* subscale scores (Menzies & Lane, 2012). Compared to the *Systematic Screening for Behavior Disorders* (SSBD), SRSS scores demonstrated the same sensitivity and specificity in detecting students who were at-risk for internalizing and externalizing behavior challenges (Lane et al., 2010; Lane et al., 2009). However, based on ROC analysis, the SRSS has stronger accuracy in predicting externalizing problems compared to internalizing problems (Lane et al., 2010).

Systematic Screening for Behavior Disorders. Another method of universal behavior screening is the SSBD three step multiple gate procedure (Walker et al., 1990). SSBD analyses whether or not a student is at-risk for internalizing or externalizing behavior problems in the

elementary and middle school settings (Caldarella, Young, Richardson, Young, & Young, 2008; Walker et al., 1990;). During Stage One, teachers rank order each of their students on a list of externalizing and internalizing behaviors from 'Most Like" to 'Least Like'. Each student with the highest score on Internalizing and Externalizing Behaviors moves to the next gate. In some cases, teachers may request that two additional students also move to the second gate (Lane et al., 2009).

In the Second Stage, teachers rate each of their qualifying students on the *Critical Events Index* (CEI) and the *Combined Frequency Index* (CFI). On the CEI (33 items), teachers check whether the child exhibits certain high intensity, low frequency behaviors (e.g., sets fires). Teachers indicate on the CFI if a student exhibits specific low-intensity, high frequency adaptive and maladaptive behaviors. Both the CEI and CFI are nationally normed rating scales; therefore, any student that scores above the cut-off score moves to Stage Three (Lane et al., 2010). During Stage Three, a school professional that is not the student's teacher (e.g., school psychologist) systematically obverses the student's social behavior and academic engagement in a variety of setting to compare their observational to the teacher's report (Lane et al., 2010).

Social, Emotional, and Academic Behavior Rating Scale. The *Social, Emotional, and Academic Behavior Rating Scale* (SAEBRS) has been identified as a universal behavior screener for kindergarten through 12th grade students. The SAEBRS is a reliable and valid measure that consists of three categories: Social Behavior, Academic Behavior (AB), and Emotional Behavior. The teachers' ratings are based on a 4-point scale (0 = N*ever*, 1 = *Sometimes*, 2 = *Often*, and 3 = *Almost Always*; Kilgus et al., 2013). The SAEBRS is user-friendly as it typically takes a teacher a few minutes to complete per child (Kilgus, Sims, von der Embse, & Taylor,

2016). Furthermore, the behaviors assessed by the SAEBRS are contextually appropriate as they are linked to academic success (Glover & Albers, 2007).

Statement of the Problem

With the implementation of MTSS, the use of universal behavior screeners is becoming increasingly prevalent. The purpose of these screeners is to identify at-risk children early and provide the necessary intervention supports to prevent future academic, behavior, social, and emotional behavior problems. Children with elevated inattention may require specialized intervention because elevated inattention is significantly linked to academic outcomes. Although the ADHD-IV-RS Inattention Scale is a norm-referenced, reliable, and valid scale (DuPaul et al., 1998), it is not feasible for teachers to rank every student on this scale in addition to other universal behavior rating scales. Consequently, schools may find it advantageous to use one universal behavior rating scale that assesses multiple areas but still captures inattentive symptoms. Therefore, teacher ratings on a universal screener, SAEBRS-TB, and a universal screener subscale, the SAEBRS-AB, will be compared to teacher ratings on the ADHD-IV-RS Inattention Scale to determine if the SAEBRS accurately identifies children with elevated inattention. Additionally, using the ADHD-IV-RS Inattention Scale as the gold standard, a cut-off score for screening children with elevated inattention on the SAEBRS-TB and on the SAEBRS-AB will be determined.

Hypotheses. It is hypothesized that the SAEBRS-TB and SAEBRS-AB will yield a moderate classification accuracy as evidenced by an AUC value greater than or equal to 0.80. It is also hypothesized that the identified cut-scores on the SAEBRS-TB and SAEBRS-AB will provide high negative predictive power and sensitivity and adequate positive predictive power

and specificity. Furthermore, it is hypothesized that teacher ratings for students on the SAEBRS-TB and SAEBRS-AB will not differ significantly.

CHAPTER II: METHODS

Participants

The current study was a subset of a larger study that investigated the effectiveness of a Computer-Assisted Intervention (CAI) for children with comorbid elevated inattention and below benchmark reading scores. The Institutional Review Board granted permission for the research study (See Appendix A). The study sample included 16 first-grade general education teachers and their 273 first-grade students from one elementary school in a rural city in eastern North Carolina. All the teachers were women. The average class size was 18 students. Each of the student participants primarily participated in Tier 1, general education services. The student sample included 47% girls (*n* = 129) and 53% boys (*n* = 144). Out of the total students, at the time of Winter Benchmarking, 45% of the participants were six years old (*n* = 122), 51% were seven years old (*n* = 138), and 5% were eight years old (*n* = 13). Out of the entire school population, approximately 67% of the students received free lunch and less than 3% received reduced lunch. The race of the first-grade population at this school included less than 1% Asian, 22% Hispanic, 32% Caucasian or White, 43% African American or Black, and 2% who identify with two or more races (State Board of Education, 2016).

Measures

SAEBRS. The SAEBRS is a reliable and valid universal mental health screening tool used to screen kindergarten through 12th grade students for emotional and behavioral problems (Kilgus et al., 2014). The SAEBRS consists of 19 items. A Total Behavior (TB) score is calculated, based on all the items, along with a score for the Social Behavior, Academic Behavior (AB), and Emotional Behavior Scales. Teachers rate the frequency of the student's behaviors throughout the last month on a 4-point rating scale (0 = *Never*, 1 = *Sometimes*, 2 =

O*ften*, and 3 = *Almost Always*). The TB scale (19 items) scores range from 0 to 57. A TB score that is less than or equal to 36 indicates that the child is at-risk. It is recommended that if a child is at-risk based on his/her TB score, it is necessary to determine the specific subscales in which the child is at-risk. For SB (six items), a student who scores between zero and 12 is at-risk. An at-risk score on Social Behavior many indicate that the child has difficulty maintaining age-appropriate relationships with children and adults. A student is at-risk on Emotional Behavior (seven items) if his/her score is less than or equal to 17. Students who are at-risk for Emotional Behavior display challenges with emotional regulation, coping skills, and adjusting to change. The primary scale of interest in this study, the SAEBRS-AB (six items), asks the teacher to rate each student on the following items: Interest in Academic Topics, Preparedness for Instruction, Production of Acceptable Work, Difficulty Working Independently, Distractedness, and Academic Engagement. Scores less than or equal to a nine indicate that the student is at-risk for AB (Kilgus et al., 2106).

Reliability. Using Cronbach's alpha, at the elementary school level, the SAEBRS demonstrates internal consistency reliability for Social Behavior (r = .89), for AB (r = .92), and for Emotional Behavior (.82), and for TB (r = .94). Compared with the BESS, the SAEBRS demonstrates strong significant correlations (-.72 to -.94; $p < .001$) with each of the scales at the elementary and middle school level. Specifically, the TB scale has the strongest correlation (elementary = -.94 and middle = -.94), and AB has the highest subscale correlation (elementary = -.88 and middle = -.88; Kilgus et al., 2016).

Validity. Based on the established cut scores for the TB (≤36) and AB (≤9), they both yield optimal diagnostic accuracy (AUC = .94 to .98) at the elementary and middle school levels when using the BESS as the gold standard (Kilgus et al., 2016). The SAEBRS-TB and all

subscales have demonstrated strong concurrent validity compared to the SSIS (r =.36-.90; Gresham & Elliott, 2008) and to the BESS (*r* = .74-.94; (Kamphaus & Reynolds, 2007; Kilgus et al., 2016). Academic outcomes, based on reading Curriculum-Based Measurements and the Missouri Assessment Program-Communication Arts (MAP-CA), were significantly correlated with each of the SAEBRS scales. SAEBRS-TB and AB demonstrated the strongest correlation with academic outcomes (Kilgus et al., 2016).

ADHD-RS-IV. The ADHD-IV-RS is a screener for ADHD symptoms. It is a normed-reference scale with a parent and teacher version. The school version of the ADHD-IV-RS inattention scale was normed on 2,000 randomly selected students from ages four to 19 who attended kindergarten through 12th grade. There were 1,001 teachers involved in the data collection. Ethnic group and region of representation were based on the 1990 US census data. Of the total participants, 90.8% of them attended general education classes. Both parents and teachers rated African-American children higher on the ADHD-IV-RS Inattention Scale compared to Caucasian and Latino students (DuPaul et al., 1998).

Reliability. Internal reliability (*n* = 71) for the school version of the ADHD-IV-RS Inattention Scale was 0.96. To determine test-retest reliability, Pearson product-moment correlation coefficients (*n* = 52) was calculated based on teachers' responses collected at the beginning and end of a four-week period. For the inattention scale, the coefficient was 0.90. Interrater agreement was calculated by comparing parent and teacher ratings on a sample of 62 students in kindergarten to 12th grade, and the interrater agreement was 0.45, which is in the moderate range. The moderate score suggests that inattentive symptoms may not appear the same in the home and in the school settings (DuPaul et al., 1998).

Interobserver agreement was evaluated in kindergarten through eighth-grade classrooms. A research assistant observed students on three separate occasions for 45 minutes. The research assistant used a modified version of the ADHD Behavior Code and a second observer attended 30% of the observation sessions. The interobserver agreement was 88%, averaged across the hyperactive-impulsivity and inattention subscales (Barkley, 1990).

Validity. To determine the criterion-referenced validity, the ADHD-IV-RS Inattention Scale was compared to the *Conners Teacher Rating Scale-39* (CTRS-39) through observation of off-task and fidgety behaviors, along with accuracy of academic tasks. The school version of the ADHD-IV-RS Inattention Scale was significantly correlated with both off-task and fidgety behaviors. A significant negative correlation occurred between scores on the ADHD-IV-RS Inattention Scale and accuracy on academic tasks (Pappas, 2016).

To determine discriminate validity, parents and teachers rated a group of children (n = 92) that meet diagnostic criteria for ADHD- predominately inattentive or ADHD-combined or that did not meet criteria for ADHD. Teacher ratings indicated a statistically significant score among the three groups for the hyperactivity-impulsivity and the inattentive scales. The same group of children was used to evaluate predictive validity in the clinical setting, and it was found that teacher ratings of inattention were able to predict the ADHD- predominately inattentive group and the ADHD-combined group from the control group with 74% and 80% accuracy, respectively. Predictive validity in the school setting was assessed by using a group of children (n = 128) from two school districts. The children were referred for academic and/or behavior challenges, and the students were divided into the same groups as mentioned above. Similar to the clinic setting, teacher ratings of inattention predicted the ADHD-predominately inattentive

group from the control group 76% of the time, and they were able to predict the ADHD-combined group from the control group 80% of the time (Pappas, 2016).

Procedures

During Winter, 2016, each first-grade teacher completed the SAEBRS and the ADHD-IV-RS Inattention Scale for all the students in their class. They were given a two-week period to complete the scales. After reminders from the administration, all 16 packets were completed. Of the 277 students enrolled, forms were completed for 273 students, due to four of the students moving to different schools. From this group of children, a subset of students was chosen to be a part of a study that investigated computer-assisted interventions for children with comorbid below benchmark reading scores and elevated inattention.

Data analysis plan. To determine if SAEBRS-Total and/or AB scores had similar classification accuracy to the ADHD-IV-RS Inattention Scale, a receiver operator characteristic (ROC) curve analysis was used. ROC curve analysis was originally developed during World War II to assist with radar set gain. When the gain of the radar was set at zero, signals were not detected. Increasing the gain led to increased signal detection, but it also led to external noises that may be mistaken as a true signal. Consequently, increasing the signal too high created false positives that outweighed the true positives (Streiner & Cairney, 2007). After World War II, ROC curves began being used for signal detection studies in experimental psychology and in psychophysics (Radloff, 1977). Although it was not formally called signal detection theory at the time, medical settings also implemented similar practices (Goodenough, Rossmann, & Lusted, 1974; Swets, 1979). For example, when reading a brain scan, setting extremely high criteria for detecting brain tumors would lead to zero false positives; however, no abnormalities would be detected, and there would be a high rate of false negatives. Setting extremely low criteria for

brain tumor detection would lead to patients' risk level either being labeled as "definitely abnormal" or "almost definitely abnormal." Consequently, this process would likely detect all brain tumors, but a high false positive rate would be present. As a result, patients would participate in unnecessary follow-up procedures. Consequently, balancing the false positive and false negative rates creates a more ideal screener (Streiner & Cairney, 2007).

In the current study, the ADHD-IV-RS Inattention Scale was utilized as the gold standard, and the SAEBRS was the tested screener. The ADHD-IV-RS Inattention Scale was used as the gold standard because it is a commonly used as a diagnostic tool for ADHD based on symptoms in the DSM-IV. It is a norm-referenced scale that provides an established clinical cut-off score based on age and gender. Therefore, participants were dichotomized into two groups, not at-risk and at-risk, based on their teacher ratings on the ADHD-IV-RS Inattention Scale (Streiner & Cairney, 2007). Students who had a score less than one standard deviation above the mean were placed into the not-at-risk group, and students with a rating greater than or equal to one standard deviation above the mean were placed into the at-risk group.

Before a cut-score was determined on the SAEBRS-AB, the Area Under the Curve (AUC) was calculated as a measure of the probability that an individual identified at-risk on one screener would be identified as more at-risk on another (Lasko, Bhagwat, Zou, & Ohno-Machado, 2005). AUC ranges from 0 to 1.00. An AUC value of less than 0.50 yields unacceptable accuracy, values between 0.50 to 0.70 suggest low accuracy, values between 0.70 to 0.90 suggest moderate accuracy, and values between 0.90 to 1.00 suggest high accuracy (Streiner & Cairney, 2007). According to Swets (1979), sensitivity and specificity scores of 0.75 or greater yield adequate discriminatory power. Precisely, sensitivity scores higher than or equal to 0.80 are acceptable (Kilgus et al., 2013), and scores greater than or equal to 0.90 are

considered optimal (Streiner, 2003). Specificity scores greater than or equal to 0.70 are acceptable, and scores greater than or equal to 0.80 are optimal (Kilgus, Sims, von der Embse, & Taylor, 2016).

To determine a cut-score on a screener, sensitivity and specificity are calculated as part of a ROC curve analysis. Sensitivity, also known as the true positive (TP) rate, is the proportion of truly at-risk students identified as at-risk by the screener (Streiner, 2003). In the current study, sensitivity refers to the proportion of individuals who are at-risk, per the ADHD-IV-RS Inattention Scale, who were also identified as at-risk by the SAEBRS-AB. Specificity, also known as the true negative (TN) rate, refers to the students who were truly not at-risk identified as not at-risk by the screener (Streiner, 2003). In the current study, the specificity is the proportion of students who are not at risk, according to the ADHD-IV-RS Inattention Scale, who were also identified as not at-risk by the SAEBRS. The false positive (FP) rate, the rate of people who were truly not at-risk but falsely identified as not at-risk according to the screener (i.e., SAEBRS), is equal to (1 – specificity). The sum of the TP rate and false negative (FN) rate is always one; consequently, as the TP rate increases, the FN rate decreases. Similarly, the sum of the TN rate and the FP rate is also equal to one. Thus, sensitivity and specificity have a reciprocal relationship. Negative predictive power (NPP) is the probability that a student identified as not at-risk by the SAEBRS that were actually not at-risk. Positive predictive power (PPP) is the probability that a student identified as at-risk by the SAEBRS is actually at-risk (Streiner & Cairney, 2007). See Table 1 for a representation of this decision matrix.

Using SPSS, a ROC curve analysis determined the classification accuracy of the SAEBRS-TB and SAEBRS-AB using the ADHD-IV-RS Inattention Scale as the gold standard.

Also, using the ROC curve analysis, a cut-off score for identifying students at-risk for elevated inattention was determined for the SAEBRS-AB.

Table 1

2x2 Contingency Table

	ADHD-IV + Dx	ADHD-IV - Dx	
SAEBRS + Dx	TP	FP	PPP = TP/ (TP+FP)
SAEBRS - Dx	FN	TN	NPP = TN/ (TN+FN)
	Sensitivity = TP/ (TP+FN)	Specificity = TN/ (TN+FP)	

In order to test the third hypothesis, McNemar's test (McNemar, 1947) was conducted to determine if students' performance on SAEBRS-AB differ significantly from that of the SAEBRS-TB. The cut-off scores determined by the ROC curve analysis were used to dichotomize scores on the SAEBRS-TB and SAEBRS-AB at -risk or not at-risk.

CHAPTER III: RESULTS

Data Screening

Accuracy of the data entry. Before conducting the ROC curve analyses, 25% of the data were checked for accuracy by undergraduate students and a graduate student, and descriptive statistics were analyzed. The descriptive statistics were also analyzed to ensure that the continuous scores were within the expected range for the ADHD-IV-RS Inattention Scale and the SAEBRS and that the standard deviations were probable. Any outliers were fixed by going back to the original data source, and missing data were fixed by replacing the missing score by entering the mean on other subscale items.

Descriptive statistics. Table 2 displays descriptive statistics for all 273 student participants. There was a significant difference in the SAEBRS-TB scores for girls (M=47.70, SD=9.51) and boys (M=40.90, SD=12.21), t(272) = 5.08, p<.001, two-tailed. Similarly, on the SAEBRS-AB, there was a significant difference in the scores for girls (M=14.29, SD=9.51) and boys (M=11.49, SD=4.94), t(272)=4.98, p<.001. On the ADHD-IV-RS, there was also a significant difference in the scores for girls (M=4.08, SD=6.25) and boys (M=8.97, SD=8.16), t(272) = -5.53, p<.001. Across all scales, boys were more likely to be identified at risk than girls. This is consistent with previous research that suggests teachers are more likely to rate boys at-risk than girls (Caci, Morin, & Tran, 2016). To assess the normality of the data, skewness and kurtosis were determined for the total score of each scale by gender. Both the SAEBRS-TB and SAEBRS-AB scores were negatively skewed for the female sample, male sample, and total sample. Scores for the total sample as well as the female and male samples were positively skewed for the ADHD-IV-RS Inattention Scale. Based on criteria established by Gravetter & Wallnau (2014), skewness and kurtosis for the scales fall within acceptable limits.

Table 2

Descriptive statistics

	Min	Max	Mean	SD	Skewness	Kurtosis
SAEBRS-TB (Total)	4	57	44.25	11.40	-0.88	0.14
Female	17	57	48.01	9.08	-1.07	0.31
Male	4	57	40.91	12.21	-0.61	-0.29
SAEBRS-AB (Total)	0	18	12.84	4.85	-0.67	-0.54
Female	1	18	14.36	4..27	-1.12	0.30
Male	0	18	11.49	4.94	-0.41	-0.78
ADHD-IV-RS	0	27	6.70	7.71	1.13	0.24
Female	0	27	4.14	6.28	1.96	3.59
Male	0	27	8.67	8.61	0.69	-0.69

Note. Total sample (N = 273); Female participants (n = 128); Male participants (n = 145).

Based on the ADHD-IV-RS Inattention Scale, 78% of the students were not at-risk, 13% of the students were in the at-risk range, and 9% of the student' ratings were in the clinically significant range. Therefore, out of the 273 total students, 61 (22.34%) were in the at-risk or clinically significant range for inattention, and 212 students (77.66%) were not at-risk. Additionally, 10.16% of girls and 33.10% of boys were in the at-risk or clinically significant range for inattention. Based on the current SAEBRS classification, 25.80% of the students were at-risk on the TB scale and 25.30% of the students were at-risk for academic behavior problems.

Hypothesis Testing

The following sections explore the results obtained from the three research hypotheses based on the ROC curve analysis and McNemar's test.

Hypothesis 1. It was hypothesized that the SAEBRS-TB and SAEBRS-AB will yield a moderate classification accuracy as evidenced by an AUC value greater than or equal to 0.80. ROC curve analyses were used to determine if the SAEBRS-TB and SAEBRS-AB are accurate predictors of teacher-rated elevated inattention based on the ADHD-IV Inattention Rating Scale. Figure 1 displays a graph of the results. The AUC for both the SAEBRS-TB and SAEBRS-AB fell within the high range of diagnostic accuracy, equaling 0.91 (SE = 0.02, CI-95 = 0.88 – 0.95) and 0.95 (SE = 0.01, CI-95 = 0.92 – 0.98), respectively. Overall, findings suggest that the SAEBRS-AB offered the best diagnostic accuracy for predicting elevated inattention.

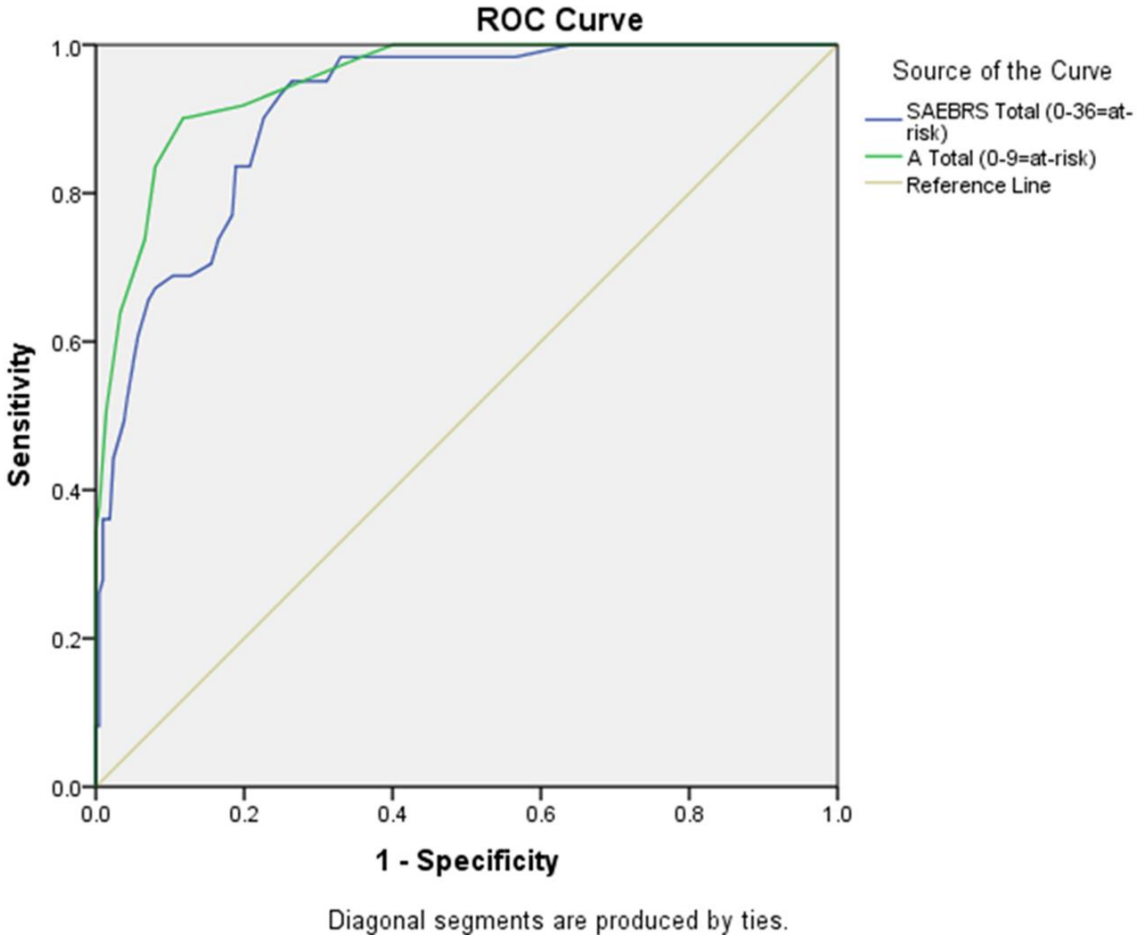

Figure 1. Comparison of ROC Curves.

Table 3

Areas under the receiver operating characteristic curves (AUC)

SAEBRS	AUC	Std. Error	p[1]	95% CI
TB	0.91	0.02	.000	0.88 to 0.95
AB	0.95	0.01	.000	0.92 to 0.98

[1]Comparison of observed AUC and the null hypothesis (AUC=.50)

Hypothesis 2. It was hypothesized that the identified cut-scores on the SAEBRS-TB and SAEBRS-AB will provide high negative predictive power and sensitivity and adequate PPP and specificity. To answer the second research question, two ROC curve analyses were conducted to determine the appropriate cut scores and to model the classification accuracy. To determine the optimal cut-off scores for elevated inattention on the SAEBRS-TB and SAEBRS-AB, conditional probability statistics were calculated. Sensitivity, the true positive rate, is the primary concern since it is vital for a universal screener to select all at-risk students. High NPP is also vital to ensure that students who are identified as not at-risk by the SAEBRS-AB are truly not at-risk.

On the SAEBRS-TB, a cut score of 40.50 yields acceptable sensitivity (.84) and specificity (.81). Therefore, because 40.50 is not a whole number, a cut score of 40 was selected. There is a disproportional difference between PPP (.56) and NPP (.95), indicating that a cut score of 40 will lead to over-identification of children with elevated inattention.

Table 4

Predictive Accuracy of SAEBRS-TB Cut Scores with ADHD-IV-RS

Positive if Less Than or Equal to	Sensitivity	1- Specificity
36.50	.69	.13
37.50	.71	.16
38.50	.74	.17
39.50	.77	.18
40.50	.84	.19
41.40	.84	.20
42.50	.90	.23

Table 5

2x2 Contingency Table (ADHD-IV-RS vs SAEBRS-TB)

	ADHD-IV + Dx	ADHD-IV - Dx	
SAEBRS-TB + Dx	51	40	PPP = .56
SAEBRS-TB - Dx	10	172	NPP = .95
	Sensitivity = .84	Specificity = .81	

On the SAEBRS-AB, a cut score of 10.90 yields optimal sensitivity (.90) and acceptable specificity (.88). Consequently, because 10.90 is not a whole number, the cut score of 11 was selected. With a cut score of 11, the sensitivity remains optimal (.92), and the specificity also remains acceptable (.80). Similar to the SAEBRS-TB, there is a disproportionate balance of PPP (0.57) and NPP (0.97). Therefore, 42.9% of those identified at-risk by the SAEBRS-AB were not actually at-risk, and 2.9% of those identified as not at-risk by the SAEBRS-AB were actually at-risk.

Table 6

Predictive Accuracy of SAEBRS-AB Cut Scores with ADHD-IV-RS

Positive if Less Than or Equal to	Sensitivity	1- Specificity
7.50	.64	.03
8.50	.74	.07
9.50	.84	.08
10.40	.90	.12
10.90	.90	.12
11.50	.92	.20
12.50	.95	.28

Table 7

2x2 Contingency Table (ADHD-IV-RS vs SAEBRS-AB)

	ADHD-IV + Dx	ADHD-IV - Dx	
SAEBRS-AB + Dx	56	42	PPP = .57
SAEBRS-AB - Dx	5	170	NPP = .97
	Sensitivity = .92	Specificity = .80	

Hypotheses 3. In order to determine if teachers' ratings for students on the SAEBRS-TB differ significantly from their ratings on the SAEBRS-AB, a McNemar's test was conducted. The cut scores, indicated above, for each scale are used to classify each case as at-risk or as not at-risk. The results of this analysis suggest that there was not a significant difference (p = .35) between students who were identified as at-risk and not at-risk on the two scales.

CHAPTER IV: DISCUSSION

Universal behavior screeners are becoming increasingly prevalent through the implementation of MTSS, which requires systematic, school-wide screenings of academic and behavioral functioning. The purpose of universal screening is to identify at-risk children early and provide interventions that prevent or reduce the likelihood of future academic and behavior problems. Based on current research indicating that children with elevated inattention and concurrent reading problems may not respond to traditional, face-to-face tutoring interventions, it is advantageous that these students be identified early as they may require specialized intervention. Although the ADHD-IV-RS Inattention Scale is a norm-referenced, reliable, and valid scale for identifying students with teacher-rated attention problems (DuPaul et al., 1998), it is not appropriate or feasible for teachers to rank every student on this diagnostic scale in addition to other universal behavior screenings. Consequently, schools may find it advantageous to use one universal behavior rating scale that assesses multiple areas but still captures inattentive symptoms. Therefore, teacher ratings on a universal screener, SAEBRS-TB, and a universal screener subscale, the SAEBRS-AB, were compared to teacher ratings on the ADHD-IV-RS Inattention Scale to determine if the SAEBRS accurately identifies children with elevated inattention. Additionally, using the ADHD-IV-RS Inattention Scale as the gold standard, a cut-off score for screening children with elevated inattention on the SAEBRS-AB was determined.

ROC curve analyzes were used to determine if the SAEBRS-TB and SAEBRS-AB were able to accurately predict teacher rated attention problems based on the ADHD-IV-RS Inattention Scale. The AUC indicated both the SAEBRS-TB and SAEBRS-AB reliably identified students with elevated, teacher-rated inattention. Students who are identified as at-risk on the ADHD-IV-RS Inattention Scale had a high probability of being at-risk on the SAEBRS-TB and SAEBRS-AB. The findings indicated that the SAEBRS may be a useful Tier 1 universal

screener for identifying children with elevated inattention, along with other behavior problems. Thus, the SAEBRS did well in identifying "true positives." However, the PPP of these screeners was relatively low, suggesting that the SAEBRS would identify as "at-risk" several students who were not really presenting with elevated, teacher-rated inattention (e.g., false positives). The implications of this are discussed later in this paper.

For the second hypothesis, a ROC curve analysis was conducted to determine possible cut-scores for the SAEBRS-TB and SAEBRS-AB. Typically, when screening for overall behavior problems, scores at or below 32 are considered at-risk on the SAEBRS-TB (Kilgus et al., 2106). In the current study, examination of sensitivity and specificity to determine an optimal cut-off score to screen for elevated inattention yielded a cut score of 40 on SAEBRS-TB, which indicates acceptable sensitivity (Kilgus et al., 2013) and optimal specificity (Kilgus et al., 2016). A cut-off score of 40 indicates a low PPP (0.56), which is likely to occur when optimal sensitivity is favored over specificity for the purposes of screening (Jenkins, Hudson, & Johnson, 2007). For general universal screening of academic behavior, a score at or below nine indicates that a child is at-risk for behavior problems based on the SAEBRS-AB (Kilgus et al., 2016). In order to screen for attention problems, a cut-off score of 11 on SAEBRS-AB revealed optimal sensitivity and specificity (Streiner, 2003; Kilgus et al., 2016). Similar to the SAEBRS-TB, a cut-off score of 11 on the SAEBRS-AB does indicate a low PPP (0.57). Therefore, children may be over identified when using SAEBRS-AB to screen for elevated inattention. However, since the purpose is to use this as a universal screener, over-identification is considered acceptable (Streiner & Cairney, 2007). Because the SAEBRS-AB yields higher sensitivity and specificity than the SAEBRS-TB, the SAEBRS-AB appears to be the more screener for elevated inattention.

For the third hypotheses, a McNemar's test revealed that teacher ratings did not significantly differ for students who were identified as at-risk or not at-risk on the SAEBRS-TB and SAEBRS-AB, based on established cut-scores for students' elevated inattention.

Implications and Directions for Future Research

Because this study included a large population of first-grade students, it may be beneficial to replicate with students in other elementary school grades. Research conducted on diverse age groups may determine if this scale can be used to assist school personnel in identifying effective interventions for children with attention problems. First-graders were used for this study because this is also a time when early reading problems are first detected via universal screenings, and we have an interest in identifying children with both teacher-rated attention and early reading problems. It may also be practical to conduct research evaluating the short-term and long-term effectiveness of utilizing specialized interventions with children who are at-risk on the SAEBRS-AB.

In the school setting, it may be advantageous to train school psychologists, administrators, and/or teachers on evaluating and using SAEBRS data to determine who needs targeted academic and behavioral interventions as well as to monitor progress. Once trained to analyze the data, school personnel may find it beneficial to use the SAEBRS-AB to screen for attention problems through a multiple gating procedure (Severson et al., 2007). At Tier 1, schools can use the SAEBRS to universally screen all children for teacher-rated social, emotional, and academic problems. Students who are at-risk for academic problems may be further screened for attention problems, and then, provided appropriate academic and behavior interventions.

The SAEBRS is a universal screener and not a diagnostic tool (Kilgus et al., 2013), and the ADHD-IV-RS Inattention Scale is a symptom screener for ADHD (DuPaul et al., 1998); as such, these should not be used for diagnostic purposes in and of themselves. For psychologists or other professionals to identify a child as displaying elevated inattention or other symptoms of ADHD, additional interviews, observations and diagnostic testing would be necessary.

Limitations

This study presented with some limitations. First, missing data occurred due to student attrition. Second, the study did not control for the order in which teacher completed the rating scales. Thus, ordering effects may have influenced teachers' responses. Third, although the population consisted of a large group of first graders, data were collected in one school in a rural setting, and therefore, may not generalize to other students and settings. Replications of this work are needed to ensure the generalizability of the SAEBRS-AB to accurately predict elevated inattention among students of diverse demographics (e.g., age, geographical location, ethnicity). Fourth, the analyses in this study were dependent upon teacher ratings; thus, the ratings may reflect some biases. However, research shows that teacher-rated inattention is highly correlated with negative academic outcomes (Rabiner et al., 2004; Fuchs et al., 2005), suggesting that this is a viable method for identifying students at-risk.

Conclusions

Overall, the SAEBRS appears to accurately identify children with elevated inattention. The determined cut score of 40 on SAEBRS-TB is in the acceptable range, and a cut score of 11 of the SAEBRS-AB falls within the acceptable to optimal range. However, ROC curve analyses indicate that the SAEBRS displays a disproportionate balance among PPP and NPP, which commonly occurs when the optimal percentage of correct decisions is prioritized. Initial findings

indicate that the SAEBRS, especially the SAEBRS-AB, may be utilized to screen for children with elevated inattention. Further research can verify these findings across different age samples and geographic regions.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision ed.). Washingon, DC.

Barkley, R. A. (1990). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment.* New York, NY: Guilford.

Breslau, N., Breslau, J., Peterson, E., Miller, E., Lucia, V.C., Bohnert, K., & Nigg, J. (2010). Change in teachers' rating of attention problems and subsequent academic achievement: A prospective analysis. *Psychological Medicine, 40,* 159-166. doi: 10.1542/peds.2008 1406

Brites, C., Salgado-Azoni, C. A., Ferreira, T. L., Lima, R. F., & Ciasca, S. M. (2015). Development and applications of the SWAN rating scale for assessment of attention deficit hyperactivity disorder: a literature review. *Brazilian Journal of Medical and Biological Research, 48,* 965-972. doi:10.1590/1414-431X20154528

Burke, M. D., Davis, J. L., Lee, Y. H., Hagan-Burke, S., Kwok, O., & Sugai, G. (2012). Universal screening for behavioral risk in elementary schools using SWPBS expectations. *Journal of Emotional and Behavioral Disorders, 20,* 38-54. doi:10.1177/1063426610377328

Bussing, R., Fernandez, M., Harwood, M., Hou, W., Garvan, C. W., & Eyberg, S. M. (2008). Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms: Psychometric properties and normative ratings from school district sample. *Assessment, 15,* 317-328. doi:10.1177/1073191107313888

Caci, H. M., Morin, A. J., & Tran, A. (2016;2012;). Teacher ratings of the ADHD-RS IV in a

community sample: Results from the ChiPARD study. *Journal of Attention Disorders, 20*, 434-444. doi:10.1177/1087054712473834

Caldarella, P., Young, E. L., Richardson, M. J., Young, B. J., & Young, K. R. (2008). Validation of the Systematic Screening for Behavior Disorders in middle and junior high school. *Journal of Emotional and Behavioral Disorders, 16*, 105-117.

Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., Christ, T. J., Black, A. C., & Kilgus, S. P. (2010). An investigation of the generalizability and dependability of Direct Behavior Ratings (DBR–SIS) to measure engagement and disruptive behavior of middle school students. *Journal of School Psychology, 48*, 219–246.

Chafouleas, S. M., Christ, T., Riley-Tillman, T. C., Briesch, A. M., & Chanese, J. (2007). Generalizability and dependability of Direct Behavior Ratings to measure social behavior of preschoolers. *School Psychology Review, 36*, 63-79.

Chafouleas, S. M., Kilgus, S. P., & Hernandez, P. (2009). Using Direct Behavior Rating (DBR) to screen for school social risk: A preliminary comparison of methods in a kindergarten sample. *Assessment for Effective Intervention, 34*, 214-223.

Chafouleas, S. M., Kilgus, S. P., & Wallach, N. (2010). Ethical dilemmas in school-based behavioral screening. *Assessment for Effective Intervention, 35*, 245–252.

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., Welsh, M. E., & Christ, T. J. (2013). Direct Behavior Rating as a school-based behavior screener for elementary and middle grades. *Journal of School Psychology, 51*, 367-385.

Collett, B. R., Ohan, J. L., & Myers, K. M. (2003). Ten-year review of rating scales. V: Scales assessing attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 1015-1037.

Conners, C. K. (1973) Rating scales for use in drug studies with children. *Psychopharmacology Bulletin* (Special issue, Pharmacotherapy of Children), *9*, 24–29.

Conners, C. K. (1990) *Manual for Conners Rating Scales*. Multi-Health Systems, Toronto, Canada.

Conners, C. K. (1997). *Conners' rating scales—revised.* North Tonawanda, NY: Multi-Health Systems.

Cook, C. R., Volpe, R. J., & Livanis, A. (2010). Constructing a roadmap for future universal screening research beyond academics. *Assessment for Effective Intervention, 35*, 197-205.

Cornish, K. M., Manly, T., Savage, R., Swanson, J., Morisano, D., Butler, N., . . . Hollis, C. P. (2005). Association of the dopamine transporter (DAT1) 10/10-repeat genotype with ADHD symptoms and response inhibition in a general population sample. *Molecular Psychiatry, 10*, 686-698. doi:10.1038/sj.mp.4001641

Costello, E. J., Mustillo, S., Erkanli, A., Keller, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60*, 837-844.

Deno, S. L., Fuchs, L. S., Maston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standard for students with learning disabilities. *School Psychology, 30*, 507-524.

DiStefano, C., & Morgan, G. B. (2010). Evaluation of the BESS TRS-CA using the Rasch rating scale model. *School Psychology Quarterly, 24*, 202-212. doi:10.1177/0734282912456969

DiStefano, C., Greer, F. W., & Kamphaus, R. W. (2013). Multifactor modeling of emotional and behavioral risk of preschool-age children. *Psychological Assessment, 25*, 467-476. doi:10.1037/a0031393

Dowdy, E., Ritchey, K., & Kamphaus, R. W. (2010). School-based screening: A population-based approach to inform and monitor children's mental health needs. *School Mental Health, 2*, 166-176. doi:10.1007/s12310-010-9036-3

Drummon, T. (1994). *The Student Risk Screening Scale (SRSS).* Grants Pass, OR: Josephine County Mental Health Program.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD Rating Scale-IV: Checklists, norms, and clinical interpretation.* New York, NY: Guilford.

Eagle, J. W., Dowd-Eagle, S. E., Snyder, A., & Holtzman, E. G. (2015). Implementing a multi-tiered system of support (MTSS): Collaboration between school psychologists and administrators to promote systems-level change. *Journal of Educational and Psychological Consultation, 25*, 160–177. doi:10.1080/10474412.2014.929960

Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early identification of behavioral and emotional problems in youth: Universals screening versus teacher-referral identification. *The California School Psychologist, 14*, 89-95.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493-513.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., & Capizzi, A. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29-43.

Gaub, M., & Carlson, C. L. (1997). Behavioral characteristics of DSM-IV ADHD subtypes in a school-based population. *Journal of Abnormal Child Psychology, 25,* 103-111.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45*(2), 117-135. doi:10.1016/j.jsp.2006.05.005

Goodenough, D. J., Rossmann, K., & Lusted, L. B. (1974). Radiographic application of receiver operating characteristic (ROC) curves. *Radiology, 110*(1), 89-95.

Goodman, A. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38,* 581-587.

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalizing and externalizing subscales instead of the hypothesized five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *Journal of Abnormal Child Psychology, 38,* 1179-1191.

Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 1337-1345.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 1337-1345.

Gravetter, F., & Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Belmont, CA: Wadsworth.

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System*. Bloomington, MN: Pearson Assessments.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills improvement system: Rating Scales.* Bloomington, MN: Pearson.

Hintze, J. M., Owen, S. V., Shapiro, E. S., & Daly, E. J. (2000). Generalizability of oral reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly*, *15*, 52-68.

Holmberg, K., Sundelin, C., & Hjern, A. (2012). Screening for attention-deficit/hyperactivity disorder (ADHD): Can high-risk children be identified in the first grade? *Child: Care, Health and Development, 39*, 268-276. doi:10.1111/j.1365-2214.2012.01382.x

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*, 582-600.

Kalberg, J. R., Lane, K. L., Driscoll, S., & Wehby, J. (2011). Systematic screening for emotional and behavioral disorders at the high school level: A formidable and necessary task. *Remedial and Special Education, 32*, 506–520. doi:10.1177/0741932510362508

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 behavioral and emotional screening system.* Minneapolis, MN: Pearson.

Kern, L., & Manz, P. (2004). A look at current validity issues of school-wide behavior support. *Behavioral Disorders, 30*, 47-59.

Kilgus, S. P., Chafouleas, S. M., & Riley-Tillman, T. C. (2013). Development and initial validation of the Social and Academic Behavior Risk Screener for elementary grades. *School Psychology Quarterly, 28*, 210-226.

Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & Welsh, M. E. (2012). Direct Behavior Rating scales as screeners: A preliminary investigation of diagnostic accuracy in elementary school. *School Psychology Quarterly, 27*, 41-50.

Kilgus, S. P., Eklund, K., von der Embse, N. P., Taylor, C. N., & Sims, W. A. (2016). Psychometric defensibility of the social, academic, and emotional behavior risk screener (SAEBRS) teacher rating scale and multiple gating procedure within elementary and middle school samples. *Journal of School Psychology, 58*, 21-39. doi:10.1016/j.jsp.2016.07.001

Kilgus, S. P., Riley-Tillman, T. C., Chafouleas, S. M., Christ, T. J., & Welsh, M. E. (2014). Direct behavior rating as a school-based behavior universal screener: Replication across sites. *Journal of School Psychology, 52*, 63-82. doi:10.1016/j.jsp.2013.11.002

Kilgus, S. P., Sims, W. A., von der Embse, N. P., & Taylor, C. N. (2016). Technical adequacy of the Social, Academic, and Emotional Behavior Risk Screener in an elementary sample. *Assessment for Effective Intervention, 42*, 46-59. doi:10.1177/1534508415623269

Kratochwill, T. R. (2007). Preparing psychologists for evidence-based school practice: Lessons learned and challenges ahead. *American Psychologist, 62,* 826-843.

Kudo, M., Altamirano, W., & Mearns, J. (2012). SWAN preschool rating scale (SWAN-P): Validity evidence for english and spanish versions. *International Journal of Educational and Psychological Assessment, 10,* 139-157.

Lai, K. Y., Leung, P. W., Luk, E. S., Wong, A. S., Law, L. S., & Ho, K. K. (2013). Validation of the Chinese strengths and weaknesses of ADHD-symptoms and normal-behaviors questionnaire in Hong Kong. *Journal of Attention Disorders, 17,* 194-202. doi:10.1177/1087054711430711

Lakes, K. D., Swanson, J. M., & Riggs, M. (2012). The reliability and validity of the English and Spanish strengths and weaknesses of ADHD and normal behavior rating scales in a preschool sample: continuum measures of hyperactivity and inattention. *Journal of Attention Disorders, 16,* 510-516. doi:10.1177/108754711413550

Lane, K. L., Harris, K., Graham, S., Weisenbach, J., Brindle, M., & Morphy, P. (2008). The effects of self-regulated strategy development on the writing performance of second grade students with behavioral and writing difficulties. *Journal of Special Education, 41,* 234-253.

Lane, K. L., Kalberg, J. R., Lambert, W., Crnobori, M., & Bruhn, A. (2010). A comparison of systematic screening tools for emotional and behavioral disorders: A replication. *Journal of Emotional and Behavioral Disorders, 18,* 100-112. doi:10.1177/1063426609341069

Lane, K. L., Little, M. A., Casey, A. M., Lambert, W., Wehby, J. H., Weisenbach, J. L., & Phillips, A. (2009). A comparison of systematic screening tools for emotional and

behavioral disorders: How do they compare? *Journal of Emotional and Behavioral Disorders, 17,* 93-105.

Lane, K. L., Menzies, H. M., Oakes, W. P., Lambert, W., Cox, M. L., & Hankins, K. (2012). A validation of the Student Risk Screening Scale for internalizing and externalizing behaviors: Patterns in rural and urban elementary schools. *Behavioral Disorders, 37,* 244-270.

Lane, K. L., Oakes, W. P., Swogger, E. D., Schatschneider, C., Menzies, M., & Sanchez, J. (2015). Student Risk Screening Scale for internalizing and externalizing behaviors: Preliminary cut scores to support data-informed decision making. *Behavioral Disorders, 40*.

Lane, K. L., Parks, R. J., Kalberg, J. R., & Carter, E. W. (2007). Systematic screening at the middle school level: Score reliability and validity of the Student Risk Screening Scale. *Journal of Emotional and Behavioral Disorders, 15*, 209-222.

Lasko, T. A., Bhagwat, J. G., Zou, H. H., & Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics, 38,* 404-415. doi:10.1016/j.jbi.2005.02.008

McGee, R., Prior, M., Williams, S., Smart, D., & Sanson, A (2002). The long-term significance of teacher-rated hyperactivity and reading ability in childhood: findings from two longitudinal studies. *The Journal of Child Psychology and Psychiatry*. *43*(3), 1004-1017.

McIntosh, K., Campbell, A. L., Russell, D., & Zumbo, B. D. (2009). Concurrent validity of office discipline referrals and cut points used in schoolwide positive behavior support. *Behavioral Disorders, 34*, 100-113.

McIntosh, K., Ty, S. V., & Miller, L. D. (2014). Effects of school-wide positive behavioral interventions and supports on internalizing problems: Current evidence and future directions. *Journal of Positive Behavior Interventions, 16,* 209–218. doi:10.1177/1098300713491980

McNemar, Q. (1947). Note on the sampling error of the differences between correlated proportions of percentages. *Psychometrika*, *12*, 153–157.

Menzies, H. M., & Lane, K. L. (2012). Validity of the student risk screening scale: Evidence of predictive validity in a diverse, suburban elementary setting. *Journal of Emotional and Behavioral Disorders, 20*, 82-91. doi:10.1177/1063426610389613

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., . . . Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the national comorbidity survey replication–adolescent supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry, 49,* 980-989. doi:10.1016/j.jaac.2010.05.017

Merrell, C., & Tymms, P. B. (2001). Inattention, hyperactivity and impulsiveness: Their impact on academic achievement and progress. *British Journal of Educational Psychology, 71*, 43–56. doi:10.1348/000709901158389

National Research Council and Institute of Medicine. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities.* Washington, DC: National Academies Press.

Oakes, W. P., Wilder, K., Lane, K. L., Powers, L., Yokoyama, L., O'Hare, M. E., & Jenkins, A. B. (2010). Psychometric properties of the Student Risk Screening Scale: An effective tool

for use in diverse urban elementary schools. *Assessment for Effective Intervention, 35*, 231-239. doi:10.1177/1534508410379796

Pappas, D. (2016). ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation. *Journal of Psychoeducational Assessment, 24*, 172-178. doi:10.1177/0734282905285792

President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families.* Retrieved from U.S. Department of Education: http://www.ed.gov/inits/commissionsboards/whspecialeducation/reports/pcesefinalreport.pdf

Rabiner, D. L., Coie, J. D., & Conduct Problems Prevention Research Group. (2000). Early attention problems and children's reading achievement: A longitudinal investigation. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*, 859-867.

Rabiner, D. L., Malone, P. S., & Conduct Problems Prevention Research Group. (2004). The impact of tutoring on early reading achievement for children with and without attention problems. *Journal of Abnormal Child Psychology, 32*, 273–284. doi:10.1023/B:JACP.0000026141.20174.17

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Raghubar, K., Cirino, P., Barnes, M., Ewing-Cobbs, L., Fletcher, J., & Fuchs, L. (2009). Errors in multi-digit arithmetic and behavioral inattention in children with math difficulties. *Journal of Learning Disabilities, 42*(4), 356-371. doi:10.1177/0022219409335211

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children—second edition (BASC-2).* Circle Pines, MN: AGS.

Ringel, J., & Sturm, R. (2001). National estimates of mental health utilization and expenditure for children in 1998. *Journal of Behavioral Health Services and Research, 28,* 319-322.

Robaey, P., Amre, D., Schachar, R., & Simard, L. (2007). French version of the strengths and weaknesses of ADHD symptoms and normal behaviors (SWAN-F) questionnaire. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 16,* 80-89.

Rodriguez, A., Järvelin, M., Obel, C., Taanila, A., Miettunen, J., Moilanen, I., ...Olsen, J. (2007). Do inattention and hyperactivity symptoms equal scholastic impairment? Evidence from three European cohorts. *BMC Public Health, 7,* 327. doi:10.1186/1471 2458-7-327

Severson, H. H., Walker, H. M., Hope-Doolittle, J., Kratochwill, T. R., & Gresham, F. M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology, 45,* 193-223. doi:10.1016/j.jsp.2006.11.003

Sims, D. M., & Lonigan, C. J. (2012). Inattention, hyperactivity, and emergent literacy: different facets of inattention relate uniquely to preschoolers' reading-related skills. *Journal of Child Clinical and Adolescent Psychology, 42,* 208-219. doi:10.1080/1537441 6.2012.738453

Smalley, S. L., McGough, J. J., Moilanen, I. K., Loo, S. K., Taanila, A., Ebeling, H., . . . Jarvelin, M. R. (2007). Prevalence and psychiatric comorbidity of attention-deficit/hyperactivity disorder in an adolescent Finnish population. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 1575-1583. doi: 10.1097/ chi.0b013e3181573137

State Board of Education. (2016). *Data & Reports - Student Accounting*. Retrieved March 20http://www.ncpublicschools.org/fbs/accounting/data/, 2017, from Public Schools of North Carolina.

Streiner, D. L. (2003). Diagnosing tests: using and misusing diagnostic and screening tests. *Journal of Personality Assessment, 81*, 209-219.

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *Canadian Journal of Psychiatry, 52*, 121-128.

Sugai, G., & Horner, R. H. (2006). A promising approach for expanding and sustaining school wide positive behavior support. *School Psychology Review, 35*, 245–259.

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior support: Integration of multi-tiered system approaches. *Exceptionality, 17*, 223-237. doi:10.1080/09362830903235375

Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders, 8*, 94-101.

Swanson, J., Kraemer, H., Hinshaw, S, Arnold, L., Conners, C., Abikoff, H., . . . Wu, M. (2001a). Clinical relevance of the primary findings of the MTA: success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 168-179.

Swets, J. A. (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology, 14*, 109-121.

Tilly, W. (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. *Best Practices in School Psychology V, 1*, 17-36.

United States Public Health Service. (2000). *Report of the Surgeon General's Conference on Children's Mental Health: A National Action Agenda.* Washington, D.C.: Department of Health and Human Services.

Walcott, C. M., Scheemaker, A., & Bielski, K. (2010). A longitudinal investigation of inattention and preliteracy development. *Journal of Attention Disorders*, *14*(1), 79-85. doi:10.1177/1087054709333330

Walker, H. M., Severson, H. H., Todis, B. J., Block-Pedego, A. E., Williams, G. J., Harring, N. G., & Barckley, M. (1990). Systematic Screening for Behavior Disorders (SSBD): further validation, replication, and normative data. *Remedial and Special Education, 11*, 32-46.

Westerlund, J., Ek, U., Holmberg, K., Naswall, K., & Fernell, E. (2009). The Conners' 10-item scale: Findings in a total population of Swedish 10–11-year-old children. *Acta Pediatrica, 98*, 828-833. doi:10.1111/j.1651-2227.2008.01214.x

Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgalarrondo, P., . . . Tannock, R. (2004). The strengths and difficulties questionnaire overseas: Evaluations and applications of the SDQ beyond Europe. *European Child and Adolescent Psychiatry, 3*(2), 47-54. doi:10.1007/s00787-004-2008-0

APPENDIX A: IRB APPROVAL

EAST CAROLINA UNIVERSITY
University & Medical Center Institutional Review Board Office
4N-70 Brody Medical Sciences Building· Mail Stop 682
600 Moye Boulevard · Greenville, NC 27834
Office **252-744-2914** · Fax **252-744-2284** · www.ecu.edu/irb

Notification of Initial Approval: Expedited

From: Social/Behavioral IRB
To: Christy Walcott
CC:

Erin Ezell
Date: 1/8/2016
Re: UMCIRB 15-002025
Effectiveness of a Computer-Assisted Intervention for Young Children with Inattention and Reading Delays

I am pleased to inform you that your Expedited Application was approved. Approval of the study and any consent form(s) is for the period of 1/7/2016 to 1/6/2017. The research study is eligible for review under expedited category # 7. The Chairperson (or designee) deemed this study no more than minimal risk.

Changes to this approved research may not be initiated without UMCIRB review except when necessary to eliminate an apparent immediate hazard to the participant. All unanticipated problems involving risks to participants and others must be promptly reported to the UMCIRB. The investigator must submit a continuing review/closure application to the UMCIRB prior to the date of study expiration. The Investigator must adhere to all reporting requirements for this study.

Approved consent documents with the IRB approval date stamped on the document should be used to consent participants (consent documents with the IRB approval date stamp are found under the Documents tab in the study workspace).

The approval includes the following items:

Name	Description
ADHD-IV Attention Problems Screener- School Version	Surveys and Questionnaires
Coding Sheet for CAI Intervention Study.docx	Data Collection Sheet
DanaEmailOptOut.jpg	Information Sheet
Parental-OptOut-Form-CAIstudy.docx	Information Sheet
SAEBRS Teacher Rating Scale- 3.3.14	Surveys and Questionnaires
Sample Intervention Progress Report.jpg	Information Sheet

The Chairperson (or designee) does not have a potential for conflict of interest on this study.