THE USE OF BEHAVIORAL SCREENERS IN ELEMENTARY SCHOOLS: CONCURRENT VALIDITY AND CONCORDANCE RATES

by

KATHLEEN R. KING-VOGEL

(Under the Direction of Amy L. Reschly)

ABSTRACT

Response to Intervention (RTI) is gaining momentum in school psychology as a means of identifying children in need of more specialized school services (Reschly & Bergstrom, 2009). RTI is associated with a tiered-model of service delivery in which most children (around 80% in Tier 1) receive general classroom instruction (Batsche et al., 2005). Those children who do not make adequate academic or behavioral progress (around 15% in Tier 2 and 5% in Tier 3) receive intensifying degrees of individualized instruction as they move up the tiers of intervention. Fundamental to this service delivery model is the need to identify those children in need of further intervention. Universal screening is a comprehensive yet concise method of evaluating every child within a school in order to determine those in need of further support. Academic screening has been accurately identifying children for years (Shinn, Shinn, Hamilton, & Clarke, 2002); however, behavior screening is less advanced. One reason for this is the absence of adequate behavioral screening measures. The purpose of the current studies is to evaluate two behavior screening instruments with a new population. The sample for both studies contained 496 elementary school children from the rural southeast. Study 1 examined the psychometric properties (including internal consistency, inter-rater reliability, convergent validity, and factor

structure) of the Teacher, Parent, and Student Forms of the Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007). Results indicated that all forms had adequate internal and external properties. Additionally, an exploratory factor analysis revealed multiple factors per form. Study 2 compared the BESS with the Behavioral Screening Checklist (BSC; Muskens, Marston, & Reschly, 2007). The psychometric properties of the screening instruments were evaluated in terms of inter-rater reliability and predictive validity of academic and behavioral outcomes. Results of this study revealed that both screening measures were highly correlated with behavioral and academic variables. A measure of social validity revealed that the BSC was somewhat more preferred by educators. Consumers of screening instruments are encouraged to select a screening instrument that has sound psychometric properties, such as those evaluated, and is practical for use in applied settings.

INDEX WORDS: Behavior, Behavior Screening, Universal Screening, Response to Intervention, Positive Behavior Support, Behavioral and Emotional Screening System, Behavior Screening Checklist

THE USE OF BEHAVIORAL SCREENERS IN ELEMENTARY SCHOOLS: CONCURRENT VALIDITY AND CONCORDANCE RATES

by

KATHLEEN R. KING-VOGEL

B. A., Kutztown University, 2003

M. A., Western Carolina University, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

THE USE OF BEHAVIORAL SCREENERS IN ELEMENTARY SCHOOLS: CONCURRENT VALIDITY AND CONCORDANCE RATES

by

KATHLEEN R. KING-VOGEL

Major Professor:	Amy L. Reschly
Committee:	Scott P. Ardoin Jonathan M. Campbell Stacey Neuharth-Pritchett

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2010

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

Response to Intervention (RTI) is characterized as a multi-tiered approach designed to match level of instruction to demonstrated student need (Batsche et al., 2005). As a method of resource allocation, RTI employs cost-effective universal interventions for the general school population. For about 80% of school children, these general interventions, which include evidence-based high quality instruction, are sufficient (Batsche et al., 2005). However, some children are in need of more intensive interventions to be successful in a school setting. For these children, less cost-effective yet more individualized interventions are employed. As the level of student need rises, so does the intensiveness of intervention. In this sense, RTI is a resource allocation model.

As a method of identification of children with disabilities, RTI is a source of much debate. Some school districts and researchers believe that information obtained from a traditional comprehensive evaluation, including standardized IQ data, is necessary in determining special education eligibility under the label of Specific Learning Disability (Hale, Naglieri, Kaufman, & Kavale, 2004; Kavale, Kaufman, Naglieri, & Hale, 2005). Other school districts and another group of researchers believe that RTI data are essential in determining eligibility, and that a comprehensive evaluation, while mandated by law, need not consist of the same standardized measures for every child (Fletcher & Reschly, 2005; Gresham et al., 2005).

RTI, albeit controversial, is gaining momentum. In fact, all states are currently developing RTI models (Reschly & Bergstrom, 2009), and some have been successfully using

RTI models for many years (Burns, Deno, & Jimerson, 2007). As a method of early intervention, RTI has promise, yet is underutilized. Arguably, the most effective and efficient way to identify children in need of early intervention services is through universal screening measures. Academic screening, particularly in the area of reading, is well validated and widely used as a means of identifying children at-risk for academic difficulties (Shinn, Shinn, Hamilton, & Clarke, 2002; Vellutino, Scanlon, & Zhang, 2007). The need for behavioral and emotional screening is well supported in the literature (Sprague, Walker, Stieber et al., 2001; Walker, Colvin, & Ramsey, 1995; Walker, Horner, Sugai, & Bullis, 1996); however, it is underrepresented in schools. In fact, only 2% of schools screen children in the area of mental health (Romer et al., 2005).

One reason for the lack of behavior screening on a school-wide basis is a dearth of behavior screening instruments. In order to be effective and efficient, universal behavior screening instruments must be valid, reliable, cost-effective, and able to be quickly administered (Lane, Parks, Kalberg, & Carter, 2007). In this current work, two studies are presented that sought to examine the psychometric properties and social validity of behavior screening instruments for use in a rural elementary school.

Chapter 2 includes a review of the literature in the area of behavioral difficulties and the effects of said difficulties within the public school system. The information included in this literature review is provided as support for the development of behavior screening instruments. This literature review is followed by an examination of the psychometric properties of a newly developed behavior screening instrument within a rural elementary school setting.

Chapter 3 provides a detailed discussion of the many factors leading to the development of RTI models. Included in this section are discussions of positive behavior support (PBS) and

universal behavior screening as they relate to the implementation of RTI. The information included in this chapter is provided as support for the use of universal behavior screening in schools. Following the literature review, a study is presented in which two recently developed behavior screening instruments were examined in terms of reliability, validity, factor structure, and acceptability.

References

Batsche, G. M., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *Response to Intervention: Policy Considerations and Implementation*: National Association of State Directions of Special Education, Inc.

Burns, M. K., Deno, S. L., & Jimerson, S. R. (2007). Toward a Unified Response-to-Intervention Model. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *Handbook of Response To Intervention: The Science and Practice of Assessment and Intervention* (pp. 428-440). New York: Springer.

Fletcher, J. M., & Reschly, D. J. (2005). Changing procedures for identifying learning disabilities: The danger of perpetuating old ideas. *The School Psychologist, 59*(1), 10-15.

Gresham, F. M., Reschly, D. L., Tilly, W. D., Fletcher, J., Burns, M., Christ, T., et al. (2005). Comprehensive evaluation of learning disabilities: A response to intervention perspective. *The School Psychologist, 59*(1), 26-30.

Hale, J. B., Naglieri, J. A., Kaufman, A. S., & Kavale, K. A. (2004). Specific learning disabilities classification in the new individuals with disabilities education act: The danger of good ideas. *The School Psychologist, 58*(1), 6-13, 29.

Kavale, K. A., Kaufman, A. S., Naglieri, J. A., & Hale, J. B. (2005). Changing procedures for identifying learning disabilities: The danger of poorly supported ideas. *The School Psychologist, 59*(1), 16-25.

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*(4), 209-222.

Reschly, D. J., & Bergstrom, M. K. (2009). Response to Intervention. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of School Psychology* (4th ed., pp. 434-460). Hoboken, NJ: John Wiley & Sons, Inc.

Romer, D., McIntosh, M., Evans, D. L., Foa, E. B., Gur, R. E., Hendin, H., et al. (2005). The Role of Primary Care Physicians in Detection and Treatment of Adolescent Mental Health Problems. In *Treating and preventing adolescent mental health disorders: What we know and what we don't know: A research agenda for improving the mental health of our youth.* (pp. 579-595). New York, NY US: Oxford University Press.

Shinn, M. R., Shinn, M. M., Hamilton, C., & Clarke, B. (2002). Using Curriculum-Based Measurement in General Education Classrooms to Promote Reading Success. In M. R. Shinn, H. M. Walker & G. Stoner (Eds.), *Interventions for Academic and Behavior Problems II: Preventive and Remedial Approaches* (pp. 113-142). Bethesda, MD: National Association of School Psychologists.

Sprague, J., Walker, H. M., Stieber, S., Simonsen, B., Nishioka, V., & Wagner, L. (2001). Exploring the relationship between school discipline referrals and delinquency. *Psychology in the Schools, 38*(2), 197-206.

Vellutino, F. R., Scanlon, D. M., & Zhang, H. (2007). Identifying Reading Disability Based on Response to Intervention: Evidence from Early Intervention Research. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *Handbook of Response To Intervention* (pp. 185-211). New York: Springer.

Walker, H. M., Colvin, G., & Ramsey, E. (1995). *Antisocial behavior in school: Strategies and best practices*. Belmont, CA US: Thomson Brooks/Cole Publishing Co.

Walker, H. M., Horner, R. H., Sugai, G., & Bullis, M. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4*(4), 194-209.

CHAPTER 2

VALIDATION OF THE BEHAVIORAL AND EMOTIONAL SCREENING SYSTEM IN A RURAL ELEMENTARY SCHOOL

It is estimated that between 2% and 10% of children in the United States suffers from emotional and behavioral disorders (EBD; Walker, Colvin, & Ramsey, 1995). However, only 1% of children in our public schools systems are receiving special education services for EBD (Wagner, Kutash, Duchnowski, & Epstein, 2005), and children with EBD made up 7% of the total special education population in the 2005-06 school year (U.S. Department of Education, n.d.).

The consequences of not intervening with children who are at-risk for behavioral problems can be costly to those children, their schools, and society in general. Walker and colleagues (Walker & Reid as cited in Walker et al., 1995; Walker, Shinn, O'Neill, & Ramsey, 1987) followed groups of children identified as either at-risk for or with diagnosed behavioral disorders. In this longitudinal study, they found that the children who were classified as having behavioral disorders had higher rates of alcohol, tobacco, and drug use than even those children identified as at-risk. Moreover, these diagnosed children were more than seven times more likely to have been arrested than the children identified as at-risk (Walker & Reid as cited in Walker et al., 1995). By the time children with EBD reach middle school, they are four times more likely to have been suspended or expelled than children in any other disability category (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005).

There are several well-documented risk factors for emotional and behavioral difficulties, including poverty, single parent household, parental unemployment, family history of mental

illness, and low levels of parent education (Wagner, Kutash, Duchnowski, Epstein et al., 2005; Walker et al., 1995). These risk factors, which can be identified early in development, have predictable and potentially devastating consequences for the individual. Walker et al. concluded, "Thus, the path to delinquency, criminality, and ultimately prison begins for many individuals very early in their lives" (Walker et al., 1995, p.19).

Children at-risk for and with EBD have demonstrated academic difficulties as well (Lane, Little, Menzies, Lambert, et al., 2010; Lane, Barton-Arwood, Nelson, & Wehby, 2008). Walker et al. (1995) suggested that some at-risk students may enter school with lower levels of academic skills than typical peers. Children with and at-risk for EBD have consistently scored lower on standardized achievement measures than their typical peers (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005; Walker et al., 1987). At-risk children have also been found to spend less time engaged in academic tasks within classroom setting (Walker et al., 1987), which may compound academic difficulties.

The academic achievement gap between children with and without behavioral difficulties widens as children progress through school (Walker & Reid as cited in Walker et al., 1995), which is one factor that may lead to the high percentage of school dropout among children with behavioral difficulties. In fact, children diagnosed with EBD routinely have the highest dropout rates (44.9%) of any disability category, and low graduation rates (43.4%; U.S. Department of Education, n.d.). After dropping out of high school, individuals with behavioral difficulties have higher rates of arrest for serious violent and nonviolent crime (Huizinga & Jakob-Chien, 1998).

Children with behavioral problems are often disruptive to the overall classroom environment (Walker et al., 1995). These children have difficulty interacting in socially acceptable ways with peers and teachers, which Walker et al. (1995) attributed to the immature

behavior demonstrated by children with behavior difficulties. In addition to demonstrating immature behavior, children with behavioral difficulties are also more likely to exhibit aggressive behavior, which has been found to increase with age. Further, bullying is common among children with behavior disorders and also intensifies with age (Walker et al., 1995). Thus, children with untreated behavior problems can threaten school safety (Walker, Horner, Sugai, & Bullis, 1996).

Traditional disciplinary practices for dealing with behavior disruptions at school often rely on the use of exclusionary methods. While these methods may immediately (although briefly) lead to a more positive environment in the school building, they are ineffective and potentially counterproductive in preventing future disruptions (Sugai & Horner, 2008; Walker et al., 1996). Furthermore, removing a disruptive child from a situation which causes him or her distress may be serving the very function of that behavior (escape/avoidance), making the disruption more likely to occur in the future.

In addition to being ultimately ineffective, exclusionary disciplinary methods can be time consuming. Teacher time spent responding to behavioral disruptions is time that is not spent on instruction. This loss of instructional time is not only harmful to the disruptive child, by excluding him/her from instruction, but to every child in the class who is losing instruction while the teacher is disciplining the disruptive student. In addition to the teacher time required to discipline children with behavioral difficulties, these children often demand more attention to engage in academic tasks (Walker et al., 1995). Walker and et al. (1995) reported that children with EBD made more teacher initiations, both negative and positive, than even those children at-risk for EBD. The authors concluded that those children are dependent on teacher assistance to

perform academically because they have lower academic skills, more difficulty staying on-task, and lower attendance rates than their peers (Walker et al., 1995).

In sum, there is evidence to suggest that children who enter school with or at-risk for EBD are likely to progress into having serious behaviors that are damaging to themselves, their schools, and society in general (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005; Walker et al., 1995; Walker et al., 1996; Walker et al., 1987). Further, research suggests that reliable risk factors of behavioral problems can be detected early in life (Loeber & Farrington, 1998; Walker et al., 1995; Walker, Shinn, & Stoner, 2002). Based on this evidence, it is clear that there is a need for proactive identification of children at-risk for serious behavioral difficulties prior to the onset of this disorder (O'Shaughnessy, Lane, Gresham, & Beebe-Frankenberger, 2002; Sprague, Walker, Stieber et al., 2001).

Sprague, Walker, and Stieber et al. (2001) advocated for early screening initiatives in schools, stating that "the children and youth who are likely to encounter serious negative outcomes later in their lives need supports and intervention services early on within school and community settings to reduce, buffer, and offset early risk factors" (p.199). Early intervention for children with behavioral difficulties is one rationale for universal behavioral screening. Kazdin (1987) noted that often children who are at-risk for behavioral disorders are identified too late in the development of the disorder to alter the developmental path of delinquency. Walker and colleagues (1998) summarized this argument as follows:

> A consensus has emerged in the past decade regarding the optimal timing of comprehensive interventions for diverting vulnerable, at-risk children from a developmental path that begins with antisocial behavior patterns and too often ends in school dropout, delinquency, and adult criminality. This consensus

> strongly suggests that the earlier intervention occurs, the more likely it is that positive outcomes will be achieved in successfully addressing this disorder. (p. 260).

Walker and colleagues recommended that schools screen children for behavioral risk status at point of school entry and provide prevention efforts in an attempt to divert these children from projected antisocial paths. Moreover, they advised schools to discontinue the use of exclusionary tactics and allow for alternative placements for children with severe behavioral problems (Walker et al., 1996). Quality intervention, applied early in the course of EBD, can prevent delinquency, increase school achievement, and even reduce teen pregnancy rates (Zigler, Taussig, & Black, 1992). While early intervention is considered essential to divert at-risk children from the path of anti-social behavior, interventions provided later in the development of behavior disorders have been found to be effective to some degree (Loeber & Farington, 1998; Walker et al., 1995)

There is a clear consensus among researchers that universal behavioral screening is in the best interests of children with behavior difficulties and schools in general (Walker et al., 1995; Walker et al., 1996). Traditional methods of detecting those children who are or may be at-risk for behavior problems include office discipline referrals (ODR), number of suspensions, report card ratings of behavior, classroom behavioral observations, and teacher referral (Muyskens, Marston, & Reschly, 2007). The use of ODR and suspension data to determine at-risk status may not be the most effective means by which to identify children at-risk because problem behaviors are already occurring at some level as to require significant disciplinary action. Moreover, these methods can be inconsistent within schools. Report card ratings of behavior are often quite broad in nature (i.e., satisfactory and unsatisfactory) and do little to inform interventions. Classroom

observations require the time of trained staff members, which makes them too time consuming and expensive to be used on a regular basis (Muyskens et al., 2007). Teacher referral to special education can be described as unsystematic, at best (Donovan & Cross, 2002). Additionally, each of the above methods is inconsistent with the universal screening practices proposed in the problem solving and positive behavior support literature (Gresham, 2004).

Lane et al. (2007) proposed a set of standards by which to evaluate screening measures. First, they called for high internal consistency (≥ .70) to ensure that the instrument is measuring the proposed construct. Second, an instrument should have demonstrated test-retest stability, or consistency over time. This type of reliability is arguably the most important to report for a screening instrument because scores can be used to determine students" eligibility for interventions. It is essential that the same students will be identified on repeated administrations of the same screener. Third, a screening instrument should have convergent validity with other established behavior rating instruments. Fourth, authors of screening instruments should report the positive predictive value and negative predictive value. This information is an important component of screening measures, as it demonstrates that those children who are above a selected cut score are truly those at-risk, and those children who are below that cut score are not at-risk. Finally, sensitivity and specificity, which go hand in hand with positive and negative predictive value, are essential to report. These terms are used to describe the proportion of at-risk students who are correctly identified and the proportion of not at-risk kids who are not identified, respectively (Lane et al., 2007). Consumers of screening instruments should carefully review screening literature to ensure that each of the above specifications is reported before deciding to use an instrument. In addition to having sound psychometric properties, screening instruments

must also be feasible to use on a school-wide basis. Screeners must be low in cost and require little time to administer, score, and analyze (Lane et al., 2007).

Finally, the current standards in the fields of educational and psychological testing state that tests should only be used with the populations in which their reliability and validity have been examined (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Further, test manuals should include detailed demographic information that describes the characteristics of the sample, in order to allow test users to determine the representativeness of their sample within the norm population. Thus, while it is the responsibility of the test developer to ensure that the norm sample is representative of the broader population, it is equally the responsibility of the test user to ensure that the prospective sample is similar to that used in the validation process. Before drawing conclusions based on test scores gathered with a dissimilar population, psychometric properties need to be examined to ensure that the test is in fact valid with this new population (American Educational Research Association et al., 1999).

The purpose of the current study is to examine the psychometric properties of a published screening instrument within a rural elementary school population. The Behavior and Emotional Screening System (BESS) was developed via Principal Components Analysis (PCA) of the longer, original Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992). Kamphaus et al. (2007) requested a one-component solution in the PCA with the goal of developing a shorter instrument more compatible with screening purposes. The individual items were chosen for the screener based on factor loadings (with a cut-off of .680 chosen by the authors), and the resulting screener met standards for acceptable internal consistency ($\alpha = .97$). It is noteworthy that the majority of items comprising the screener came from the Attention

Problems scale of the full length BASC, followed by Study Skills, Adaptability, Leadership, Aggression, Hyperactivity, and Learning Problems. Internalizing items did not factor significantly on to the first component and thus, may be underrepresented in the BESS. In reference to the items comprising the screener, the authors reported that the screener has the ability to "assess the behavioral requirements ideally suited to schooling" (Kamphaus et al., 2007, p. 351).

Kamphaus et al. (2007) used multiple outcome measures to assess the predictive validity of the BESS screener. They gathered screener data during one school year, and then correlated scores with outcome measures gathered during the next school year. The authors used zero-order partial correlations (controlling for age) to measure the relationship between screener scores and the other BASC scales that did not contribute items to the screener. Results indicated that the screener produced moderate correlations with Conduct Problems (r = 0.497), Atypicality (r = 0.479), and Social Skills (r = -0.471) subscales on the BASC. Acceptable correlations were noted for Depression (r = 0.370) and Withdrawal (r = 0.352), and a low correlation was found between screener scores and Anxiety scores (r = 0.195) on the BASC. Zero-order correlations were calculated between screener scores and the prediction of special education status, grades, and achievement scores. Results of this analysis indicated that there were acceptable significant correlations between screener scores and special education placement (r = 0.306), prereferral intervention (r = 0.308), and school maladjustment (maladjustment data missing from the published article). Moderate correlations were found between screener scores and achievement scores (r = -0.547 with math scores and r = -0.575 with reading scores), grades (r = -0.477 with math grades and r = -0.546 with reading grades), and work habits (r = -0.434). Finally, the screener was poor at predicting suspensions and attendance (r = 0.133 and r = 0.121

respectively), which the authors attributed to the unknown reliability and validity of those particular outcome measures. Interestingly, the screener showed stronger correlations with most outcome measures than did the overall Behavioral Symptoms Index on the full-length BASC.

The current study is an examination of the psychometric properties of the BESS within a rural elementary school population. The purpose of the current study is to explore the reliability and validity evidence of scores on the BESS for children in a rural elementary school. Moreover, the exploratory factor structure of the BESS will be examined, as the BESS was developed with a one-component PCA.

Method

Participants

The subjects of this study included 496 Kindergarten through 5th grade students in a public elementary school in rural northeast Georgia. The sample consisted of 48% male and 52% female students. A review of parent-reported ethnicity indicated that 65% of students were European American, 11% were African American, 19% were Hispanic American, less than 1% were Asian American, and 5% were defined by parents are multi-racial. Over 68% of the sample students were eligible for Free or Reduced Lunch. Every regular education teacher (n = 25) completed screeners in reference to the children in their classrooms. Additionally, 241 parents (48.6%) completed parent screeners. The data were examined to determine if any group differences were present between those parents who completed the Parent Form and those who did not. When controlling for home language, for reasons described below, the groups did not differ in terms of gender. However, they were found to differ in terms of grade (χ^2 = 20.06, df = 5, p < .01), ethnicity (χ^2 = 13.81, df = 4, p < .01), and Free or Reduced Lunch status (χ^2 = 12.16, df = 2, p < .01). Incidentally, the subsample of children included in the data set with Parent Form

data may not be representative of the school as a whole. The sample included a disproportionally large number of kindergartners and a disproportionally small number of Hispanic families, even when controlling for home language. Additionally, the sample is under representative of families eligible for Free or Reduced Lunch. Finally, 207 children in grades 3 through 5 completed the student screener (88.1%). These data were also examined for group differences between the children who completed the Student Form and those who did not. No such differences were found.

Measures

BASC-2 Behavioral and Emotional Screening System. The BASC-2 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007) was adapted from the well-validated and widely used Behavioral Assessment System for Children (BASC; Reynolds & Kamphaus, 1992). The BESS was developed to be a more efficient measure for screening purposes than the full length BASC forms.

The BESS system is comprised of three report forms (i.e., Parent, Teacher, and Student) for children in preschool through grade 12, which can be used individually or in combination. The BESS Child/Adolescent Teacher Form contains 27 items and is intended to be completed by an adult who has had extensive experience with the child in an educational setting. The Child/Adolescent Parent Form has 30 items and is intended to be completed by the child‟s caregiver. The Student Form, to be completed by children in grades 3 through 12, is comprised of 30 items. The BESS manual recommends obtaining ratings from multiple caregivers, if possible. Each form reportedly takes 5-10 minutes to complete. The BESS items are scored on a 4-point Likert-type scale, which result in an overall score that can be classified, using normative data, as having normal, elevated, or extremely elevated levels of risk. For copyright purposes, the

items printed in this study are truncated. They are intended to represent actual test items, but are not the test items themselves. Full test items can be found in the BESS Manual (Kamphaus & Reynolds, 2007).

Reliability and validity evidence is provided for each BESS form and is reported by age group. Given the diversity of item content on the varying BESS forms, split-half reliability is reported as a measure of internal-consistency, as opposed to the more common coefficient alpha. The median split-half reliability (across age groups) was .96, .94, and .92, for the teacher, parent, and student forms, respectively (Kamphaus & Reynolds, 2007). The test-retest reliability of the BESS, reported as an adjusted correlation coefficient, was .91, .84, and .80 for the teacher, parent, and student forms, respectively. Correlation values, adjusted for sample variability, between two raters with the same BESS form (i.e., two teachers or two parents) were reported as evidence of inter-rater reliability, and were .70 and .87 on the Teacher and Parent Forms, respectively.

Concurrent validity was reported for the total score on the BESS and other well-known behavior rating scales. The following adjusted correlation coefficients were obtained between the BESS Teacher Form and the Behavior Assessment Scale for Children, Second Edition (BASC-2; Reynolds & Kamphaus, 2004) Teacher Form composite scores: .79 (Externalizing Problems), .62 (Internalizing Problems), .89 (School Problems), -.85 (Adaptive Skills), and .90 (Behavioral Symptoms Index; Kamphaus & Reynolds, 2007). The correlation coefficients were found to be somewhat lower when the BESS Teacher Form and BASC-2 TRS were used on a new sample of children (Kamphaus, DiStefano, Dowdy, Eklund, & Dunn, 2010). The most recent correlation coefficients, obtained on the new sample of children attending school in an urban west coast

school district, were: .76 (Externalizing Problems), .52 (Internalizing Problems), .82 (School Problems), and -.82 (Adaptive Skills).

Concurrent validity correlation coefficients for the BESS Parent Form and BASC-2 Parent Form were: .79 (Externalizing Problems), .59 (Internalizing Problems), -.87 (Adaptive Skills), and .90 (Behavioral Symptoms Index). Similar coefficients are also reported for the BESS Student and BASC-2 Self-Report Form. Those adjusted coefficients are: .69 (School Problems), .84 (Internalizing Problems), .74 (Inattention/Hyperactivity), -.78 (Personal Adjustment), and .86 (Emotional Symptoms Index). There is a great degree of overlap of items on the BESS and on the BASC-2. In fact 24 of the 27 Teacher Form items, 26 of the 30 Parent Form items, and 29 of the 30 Student Form items on the BESS are also found on the corresponding BASC-2 forms. This overlap of items can partially explain the correlation coefficients presented above; however, the authors of the BESS stated that this comparison between forms is justified because the BASC-2 forms contain many more items than those in common with the BESS (Kamphaus & Reynolds, 2007).

Further, concurrent validity evidence was calculated between the BESS forms and the forms of the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001). The adjusted correlation coefficient between the BESS Teacher Form score and the ASEBA Teacher Rating Form (ASEBA TRF) Total Problems scale was .76. With the ASEBA TRF Externalizing scale, the adjusted correlation coefficient was .69, and with the ASEBA TRF Internalizing scale, it was .29. Similar statistics were reported for the BESS Parent Form and the ASEBA Child Behavior Checklist (ASEBA CBCL), which is a parent-completed rating scale. The adjusted correlation between the BESS Parent Form score and the ASEBA CBCL Total Problems scale was .76. The correlation was .66 and .64 for the ASEBA CBCL

Externalizing and Internalizing scales, respectively. The ASEBA system also has a self-report form called the Youth Self Report (ASEBA YSR). Correlations between the BESS Student Form and the ASEBA YSR were .77, .66, and .69 for the Total Problems, Externalizing, and Internalizing scales, respectively.

Student-level data. Student-level data were obtained from school records maintained with a computer program located in the central school office. These data included office discipline referrals (ODR), suspensions, and attendance records at mid-year. For the attendance data, tardies and absences were converted to a percent of days on-time relative to the number of days enrolled. As an indicator of mid-year reading achievement, scores from the November administration of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), 6th Edition, were used. Three Oral Reading Fluency (ORF) probes were administered according to standard administration procedures to students in grades 1 through 5. The passages were scored as the number of words read correctly in 1 minute. As is typical with this type of measurement, the median score of the three passages was used in analyses (Good, 2004). ORF and scores from similar reading probes (oral reading Curriculum Based Measurement) are moderately to highly correlated with other standardized measures of reading achievement (Reschly, Busch, Betts, Deno & Long, 2009). The DIBELS ORF multi-probe administration has shown strong test-retest reliability at every grade level, ranging from .96 (in one study of 1st grade ORF) to .99 (in another study of 1st grade ORF and a study of 2nd grade ORF; Dynamic Measurement Group, 2008). Data from the mid-point of the current school year were used as the best estimate of current student functioning because the data collection process occurred during the fall semester of the school year.

Procedures

BESS data were collected from parents/guardians, teachers, and students around 10 weeks into the school year. Parents were informed of the data collection via a letter sent home 1 week prior to fall parent-teacher conferences. Upon entry into the school building on the day of conferences, parents were approached and asked to complete the BESS Parent Form. After completing the form, they proceeded to the conference as scheduled. The parent-teacher conferences were chosen as an optimal time for data collection because of the school‟s traditionally high parent turnout for conferences. The BESS publishes a Spanish version of the Parent Form; however, due to limited availability of those forms at the time of data collection, they were not used in this study. Instead, Spanish speaking families (n = 89, 18%) were given the option to complete an English version. Most (n = 81, 91%) chose not to complete a BESS form. Those children were still included in the study, as teacher, and for many, student level data were available. The overall response rate, when controlling for home language, of the BESS Parent Form was 62% (n = 255).

The BESS Teacher Forms were distributed to teachers following the parent-teacher conferences. Teachers were given one week to complete the forms and were provided with a substitute teacher for 30 minutes to allow them time to complete forms during regular school hours. The response rate for BESS Teacher Form was 99%, meaning that a Teacher Form was completed for nearly every student enrolled in kindergarten through grade 5 at the time of data collection. There were two students for whom the BESS Teacher Form was not completed because those students moved into the district after the administration of the Teacher Form, but completed the Student Form, as described below.

The BESS Student Forms were distributed to 3rd, 4th, and 5th grade students during regular school hours. Group instructions were presented to children within each classroom and individual questions were answered. Eighty-eight percent of students in grades 3 through 5 completed the Student Form. Students who did not complete the Student Form either withdrew from school before the administration of that measure, were absent on the day of administration, or were involved in individualized instruction during the group administration of the Student Form. Participant data are presented in Table 1. To ensure accuracy of data entry, 15% of the data were checked for accuracy, and 99% of the data entered were found to be accurate. Errors were corrected in the data set.

Table 1

Characteristics of Participants by Grade

		Gender		Ethnicity/Race					Home Language		Lunch Status	
		Male	Female	EA	AfA	AsA	HA	M	English	Spanish	F/R	Full
Grade	*n*	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
K	85	46 (54.1)	39 (45.9)	66 (77.6)	7 (8.2)	0 (0)	10 (11.8)	2 (2.4)	75 (88.2)	10 (11.8)	57 (67.1)	28 (32.9)
1	85	44 (51.8)	41 (48.2)	47 (55.3)	8 (9.4)	0 (0)	26 (30.6)	4 (4.7)	57 (67.1)	28 (32.9)	65 (76.5)	20 (23.5)
2	91	42 (46.2)	49 (53.8)	56 (61.5)	10 (11.0)	2 (2.2)	19 (20.9)	4 (4.4)	74 (81.3)	17 (18.7)	63 (69.2)	28 (30.8)
3	70	32 (45.7)	38 (54.3)	42 (60.0)	9 (12.9)	0 (0)	14 (20.0)	5 (7.1)	58 (82.9)	12 (17.1)	52 (74.3)	18 (25.7)
4	93	39 (41.9)	54 (58.1)	61 (65.6)	14 (15.1)	0 (0)	12 (12.9)	6 (6.5)	82 (88.2)	11 (11.8)	60 (64.5)	33 (35.5)
5	72	33 (45.8)	39 (54.2)	49 (68.1)	8 (11.1)	0 (0)	11 (15.3)	4 (5.6)	61 (84.7)	11 (15.3)	42 (58.3)	30 (41.7)
Total	496	236 (47.6)	260 (52.4)	321 (64.7)	56 (11.3)	2 (0.4)	92 (18.5)	25 (5.0)	407 (82.1)	89 (17.9)	339 (68.3)	157 (31.7)

Note. Ethnicity/Race: EA=European American, AfA=African American, AsA=Asian American, HA=Hispanic American, M=Multiracial; Lunch Status: F/R=Free or reduced lunch, Full=full-priced lunch.

Data Analyses

The internal psychometric properties of the BESS forms were examined in terms of internal consistency reliability and factor structure. The first set of analyses examined the internal consistency of each of the BESS forms. As in previous studies of the BESS, split-half reliability was used because the screener is purported to measure heterogeneous test items across several content areas (Kamphaus & Reynolds, 2007). The exploratory factor structure of each of the BESS forms was examined with Principal Axis factoring and subjected to a Promax rotation with Kaiser normalization. This data-driven approach to factor analysis did not include an a priori number of factors, nor were there restrictions placed on the relationships between variables. The Principal Axis factoring extraction method was chosen a priori as the best choice of model fit because, with this method of extraction, data do not nave to meet distributional assumptions (Costello & Osborne, 2005; Fabrigar, Wegener, MacCallum, & Strahan, 1999). The Promax oblique rotation method was chosen because correlations were suspected to exist among scale items.

The external properties of the BESS forms were also examined. A nonparametric Independent Samples Kruskal-Wallis Test was used to determine the criterion-related external validity between each of the BESS forms and the academic (ORF) and behavioral (ODR, suspensions, attendance) measures. A comparison of group means with a MANOVA procedure was considered; however, it was determined that the data set violated a number of the assumptions associated with that test (including distributional normality). Additionally, the oral reading passages used to determine the ORF scores varied by grade, as do all CBM measures. Therefore, ORF scores were converted to *z*-scores for the analyses because the purpose was to compare groups (normal, elevated, extremely elevated) on an indicator of reading achievement

rather than to evaluate the ORF scores. When correlations were calculated among variables, Spearman‟s Rho coefficient was used with the ODR, suspensions, and attendance variables because of their skewed distributions.

The next set of analyses examined the agreement between each set of raters (e.g., parents, teachers, students). T-Scores, as opposed to raw scores, were used in these analyses because of the unequal number of items on the BESS forms (e.g., 27, 30, and 30 items on teacher, parent, and student forms, respectively). Pearson correlations coefficients were calculated as a measure of inter-rater reliability. Additionally, inter-rater agreement, or the extent to which raters identified the same children as "normal," "elevated," and "extremely elevated," on the BESS forms was calculated. Data were only included in this analysis if information was available from both raters. While Teacher Form data was available for 494 students, Parent Form data were only available for 255 students, and Student Form data were available for 207 students.

Results

Descriptive data and correlations among variables may be found in Tables 2 and 3. Data for item-level responses on each of the BESS forms is located in Tables 6, 9, and 12. There were low to moderate significant correlations between the Teacher Form and all outcome variables. The Parent Forms were significantly negatively correlated with attendance and ORF. The Student Forms were significantly correlated with ODR, attendance, and ORF. Correlations between screener forms will be discussed in a later section.

Internal Consistency

Internal consistency was examined with the Spearman-Brown split-half coefficient. Split-half reliability for the Teacher, Parent, and Student Forms was found to be 0.962, 0.939, and 0.904, respectively.

Table 2

Descriptives of Major Variables

Variable	*n*	Mean	*SD*	Min	Max
TF T-Score	494	52.75	11.66	4	86
PF T-Score	255	50.73	11.18	31	83
SF T-Score	212	55.50	10.94	33	88
ODR	496	0.43	1.29	0	13
Suspensions	496	0.05	0.29	0	3
Attendance	490	93.54	5.77	62.96	100.00
ORF					
Grade 1	80	26.99	22.38	2.00	133.00
Grade 2	85	87.36	35.19	9.00	168.00
Grade 3	68	96.60	38.65	9.00	186.00
Grade 4	88	110.38	34.28	6.00	194.00
Grade 5	71	120.73	36.31	9.00	237.00

Note. TF T-Score; PF T-Score; SF T-Score = T-Score on the Teacher Form, Parent Form, and Student Form , respectively. ODR = office discipline referrals. Suspensions = number of days of suspension. Attendance = percent of days on time. ORF = oral reading fluency.

Table 3

Correlations Between Major Variables

Variable	TF T-Score	PF T-Score	SF T-Score	ODR	Sus	Att	ORF
TF T-Score	1.00						
PF T-Score	.389**	1.00					
SF T-Score	.393**	.105	1.00				
ODR	.338**	.122	.281**	1.00			
Suspensions	.157**	.081	.096	.357**	1.00		
Attendance	-.112*	-.133*	-.171*	-.013	-.010	1.00	
ORF[1]	-.371**	-.148*	-.241**	-.140**	-.046	.095	1.00

Note. ** = Significant at 0.01 level. * = Significant at 0.05 level. TF T-Score; PF T-Score; SF T-Score = T-Score on the Teacher Form, Parent Form, and Student Form respectively. ODR = office discipline referrals; Sus = suspensions; Att = attendance; ORF = oral reading fluency. Spearman‟s Rho correlation coefficient was used for the correlations with the ODR, Suspensions, and Attendance variables. ORF[1] converted to z-score based on grade level means and standard deviations. Pearson‟s correlation coefficients used for all other correlations.

Factor Analysis

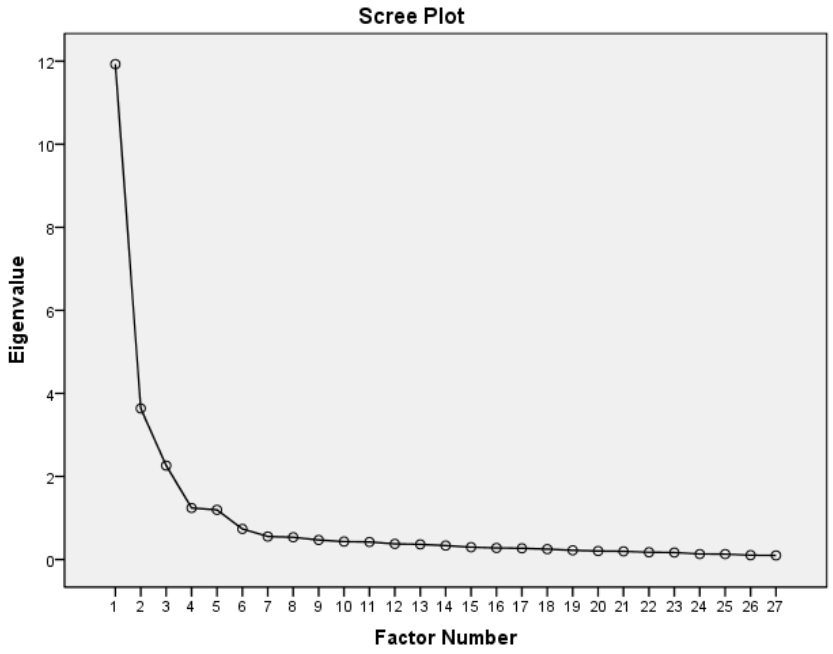
Factor analysis of each of the BESS forms was conducted with Principal Axis extraction and factors were rotated with promax with Kaiser normalization. Factor structures were examined in terms of eigenvalues, scree plots, and interpretability. While retaining factors with eigenvalues greater than 1 may be considered inadequate as the sole criteria for determining the number of factors to retain, many researchers still consider it to be an important element in the decision-making process. Similarly, many researchers report factor loadings of 0.4 or greater, while others justify the use of 0.3 or 0.5. Therefore, strength of factor loading is not to be considered the only criterion for determining factor structure. Costello and Osborne (2005) proposed that an examination of scree plots is the most valid indicator of true factor structure. For this reason, the factor matrices and item loadings presented below are arranged based on the number of factors suggested by the scree plot; however, alternate factor structures are discussed.

Teacher Form. Results presented in Tables 4 through 6. Examination of the eigenvalues of the Teacher Form revealed that five factors had eigenvalues greater than one (accounting for 69.25% of the variance). An examination of the scree plot suggested retaining five factors, as well (see Figure 1).

Finally, the interpretability of the factors was examined. With the exception of one item (item 20) each item loaded onto one and only one factor with a 0.4 or greater factor loading. A review of the items within each retained factor revealed that each could be interpreted. Factor 1 ($\alpha = 0.952$) consisted of externalizing behaviors, while Factor 2 ($\alpha = 0.956$) consisted of items that related to inattention and academic concerns. Factor 3 ($\alpha = 0.874$) contained items that were all internalizing in nature. The items on Factor 4 ($\alpha = 0.816$) were all positively worded adaptive skills. Factor 5 ($\alpha = 0.700$) contained two somatization items. When a four factor solution was

Figure 1

Teacher Form Scree Plot



requested, the two somatization items factored with the other internalizing items onto Factor 3 (new α = 0.869). With the internalizing items all on one factor, the four factor solution may be the most interpretable.

Table 4

Promax with Kaiser Normalization-Rotated Factor Pattern/Structure Coefficients for the BESS Teacher Form

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
2. Disobeys	**.885**	-.022	-.003	.029	.016
4. Breaks the rules	**.923**	-.010	-.073	.029	.046
6. Has poor self-control	**.686**	.203	.076	-.025	-.036
13. Annoys others on purpose	**.770**	.044	.063	.022	-.087
18. Gets into trouble	**.898**	.093	-.080	-.018	.022
21. Disrupts other children‟s activities	**.812**	.164	.013	-.059	-.035
1. Pays attention	.223	**.660**	-.061	.091	.056

5. Is well organized	.088	**.548**	.013	.248	.043
8. Completes assignments incorrectly because of not following instructions	.078	**.766**	.032	.022	.007
10. Has trouble keeping up in class	-.169	**.850**	.064	.117	-.050
16. Is easily distracted from class work	.169	**.853**	-.041	-.036	-.021
23. Has trouble concentrating	.147	**.894**	.011	-.145	.025
24. Has good study habits	-.045	**.592**	.030	.333	.031
26. Has a short attention span	.129	**.897**	-.011	-.092	-.013
3. Is sad	-.039	.096	**.638**	.047	.065
7. Is easily upset	.170	-.087	**.719**	-.014	.004
11. Worries about things that cannot be changed	-.054	.051	**.871**	-.052	-.060
12. Says, "Nobody likes me."	.234	-.091	**.502**	.030	-.079
14. Is fearful	-.203	.167	**.757**	-.040	-.021
20. Is negative about things	**.465**	-.231	**.497**	.093	.026
25. Worries	-.084	-.015	**.866**	-.008	.049
9. Is good at getting people to work together	.134	-.047	-.040	**.829**	.032
17. Is effective when presenting information to a group	-.153	.140	.088	**.759**	-.086
19. Gives good suggestions for solving problems	-.103	.281	-.020	**.684**	-.019
27. Encourages others to do their best	.112	-.109	-.056	**.891**	.026
15. Has headaches	-.035	-.003	-.077	-.016	**.802**
22. Complains about health	-.012	.021	.199	-.017	**.675**
N= 493					

Note. Pattern/structure coefficients greater than .4 are bolded. These items are copyrighted material of Pearson Assessments. Items originally published in Kamphaus et al. (2007).

Table 5

Teacher Form Factor Correlation Matrix

Factor	**1**	**2**	**3**	**4**	**5**
1	1.00				
2	.644	1.00			
3	.308	.344	1.00		
4	.462	.661	.233	1.00	
5	.140	.165	.470	.111	1.00

Table 6

Teacher Form Five Factor Solution – Factor Items and Properties

Factor 1	BASC Scale	Mean	*SD*	Alpha if Item Deleted
2. Disobeys	Conduct Problems	.73	.721	.943
4. Breaks the rules	Conduct Problems	.76	.733	.939
6. Has poor self-control	Hyperactivity	.84	.904	.949
13. Annoys others on purpose	Aggression	.52	.782	.948
18. Gets into trouble	Conduct Problems	.82	.828	.937
21. Disrupts other children‟s activities	Hyperactivity	.77	.840	.940
Total Factor		4.43	4.330	**Alpha .952**
Factor 2	BASC Scale	Mean	*SD*	Alpha if Item Deleted
1. Pays attention	Attention Problems	1.26	.945	.949
5. Is well organized	Study Skills	1.45	1.055	.954
8. Completes assignments incorrectly because of not following instructions	Learning Problems	.96	.894	.951
10. Has trouble keeping up in class	Learning Problems	.96	1.006	.953
16. Is easily distracted from class work	Attention Problems	1.24	.986	.946
23. Has trouble concentrating	Attention Problems	1.06	.944	.948
24. Has good study habits	Study Skills	1.51	.993	.953
26. Has a short attention span	Attention Problems	1.12	.973	.947
Total Factor		9.56	6.826	**Alpha .956**
Factor 3	BASC Scale	Mean	*SD*	Alpha if Item Deleted
3. Is sad	Depression	.59	.650	.854
7. Is easily upset	Depression	.68	.789	.851
11. Worries about things that cannot be changed	Anxiety	.47	.706	.840
12. Says, “Nobody likes me.”	Depression	.12	.363	.877*
14. Is fearful	Anxiety	.30	.562	.855
20. Is negative about things	Depression	.40	.612	.868
25. Worries	Anxiety	.52	.683	.838
Total Factor		3.07	3.357	**Alpha .874**
Factor 4	BASC Scale	Mean	*SD*	Alpha if Item

				Deleted
9. Is good at getting people to work together	Leadership	.76	.732	.858*
17. Is effective when presenting information to a group	Functional Communication	1.62	.975	.743
19. Gives good suggestions for solving problems	Leadership	1.70	.906	.712
27. Encourages others to do their best	Social Skills	1.77	.949	.727
Total Factor		5.85	2.875	**Alpha .816**
Factor 5	BASC Scale	Mean	*SD*	Alpha if Item Deleted
15. Has headaches	Somatization	.16	.440	---
22. Complains about health	Somatization	.36	.638	---
Total Factor		.52	.961	**Alpha .700**

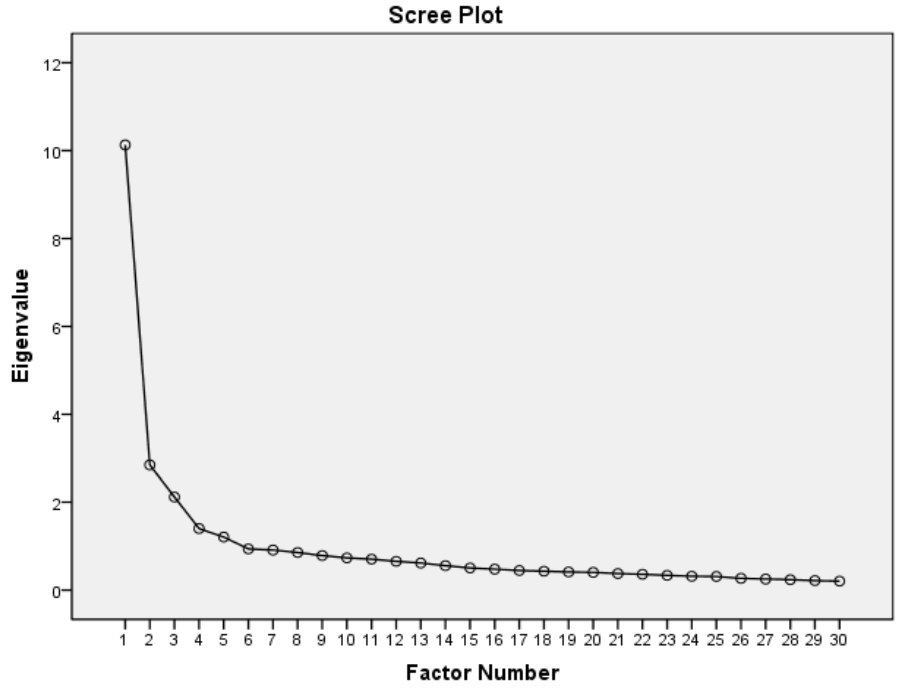
Note. * = Alpha increases if item is deleted. These items are copyrighted material of Pearson Assessments. Items originally published in Kamphaus et al. (2007).

Parent Form. Results of the exploratory factor analysis of the Parent Form are presented in Tables 7 through 9. An examination of the eigenvalues revealed that five factors on the Parent Form had eigenvalues greater than one and those five factors accounted for 51.23% of the variance. The scree plot suggested retaining five factors, as well (see Figure 2).

There were no items that loaded onto multiple factors (with coefficients of equal to or greater than 0.4); however, there were three items (7, 20, and 29) that did not load onto any factor with coefficient of at least 0.4. These items were considered to be part of the factor for which they had the highest loading (albeit less than 0.4) and the factors were interpreted. Factor 1 ($\alpha = 0.912$) consisted of items that were all positively worded adaptive skills and attention attributes. On the contrary, Factor 2 ($\alpha = 0.842$) contained negatively worded externalizing behaviors. Factor 3 ($\alpha = 0.802$) consisted of internalizing behaviors and Factor 4 ($\alpha = 0.733$) contained the two somatization items (similar to the five factor structure of the Teacher Form).

Figure 2

Parent Form Scree Plot



Finally, Factor 5 (α = 0.778) contained two negatively worded attention items. When a four factor solution was requested, the two inattention items that comprised Factor 5 loaded onto Factor 2, with the other negatively worded items (new α = 0.866). When a three factor solution was requested the two somatization items from Factor 4 loaded onto Factor 3 with the other internalizing items (new α = 0.801). With all items factoring together in expected ways, a three factor solution may be the most interpretable for the Parent Form.

Table 7

Promax-Rotated Factor Pattern/Structure Coefficients for the BESS Parent Form

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1. Attends	**.511**	.163	-.089	-.065	.273
3. Tracks down information	**.723**	-.030	-.104	-.003	.116

5. Brings out the best in others	**.801**	.132	-.172	-.034	-.188
9. Gets others to work together	**.809**	-.007	-.078	-.009	-.077
12. Organizes well	**.584**	-.082	-.005	-.010	.284
15. Communicates	**.620**	-.080	.063	.118	.128
17. Adjusts well	**.403**	-.062	.358	-.039	.012
19. Gives good suggestions	**.886**	-.163	-.065	.007	.033
22. Listens	**.616**	-.031	-.059	.061	.102
24. Gets along	**.604**	.159	-.039	.073	-.134
28. Recovers quickly	**.530**	-.173	.324	.083	.005
30. Realistic goals	**.728**	.042	.074	-.089	-.030
2. Does not obey	-.136	**.781**	-.183	.017	.084
4. Breaks rules	.035	**.603**	-.275	.118	.171
6. Out of control	.110	**.620**	.190	-.077	-.126
10. Defies	.127	**.510**	.109	.002	-.114
18. Trouble	-.087	**.850**	-.210	.112	.185
20. Disrupts	.230	.349	.100	-.044	.162
26. Loses temper	-.120	**.600**	.304	-.055	-.043
29. Disliked	.144	.366	.311	-.058	-.071
7. Lonely	.215	.096	.383	.057	-.207
11. Worries about things	-.154	-.118	**.769**	.004	.061
13. Frustrated	.016	.276	**.408**	-.062	.292
16. Nervous	-.064	-.088	**.682**	.109	.041
23. Easily upset	-.011	.189	**.532**	.059	.102
25. Worries	-.113	-.157	**.852**	.015	.101
14. Pain	.045	.030	.047	**.654**	-.004
21. Health	-.007	.087	.149	**.782**	-.093
8. Distracted	.096	.127	.076	-.035	**.616**
27. Trouble concentrating	.118	.197	.218	-.048	**.504**
N=255					

Note. Pattern/structure coefficients greater than .4 are bolded. Items are truncated because they are copyrighted material of Pearson Assessments.

Table 8

Parent Form Factor Correlation Matrix

Factor	**1**	**2**	**3**	**4**	**5**
1	1.00				
2	.580	1.00			
3	.423	.456	1.00		
4	.187	.152	.274	1.00	
5	.476	.442	.250	.134	1.00

Table 9

Parent Form Five Factor Solution - Factor Items and Properties

Factor 1	BASC Scale	Mean	*SD*	Alpha if Item Deleted
1. Attends	Attention Problems	.94	.800	.904
3. Tracks down information	Functional Communication	1.13	.856	.903
5. Brings out the best in others	Social Skills	.93	.854	.905
9. Gets others to work together	Leadership	1.21	.899	.903
12. Organizes well	Activities of Daily Living	1.43	.907	.905
15. Communicates	Functional Communication	.60	.798	.904
17. Adjusts well	Adaptability	1.08	.916	.912
19. Gives good suggestions	Leadership	1.13	.842	.901
22. Listens	Attention Problems	.97	.822	.901
24. Gets along	Social Skills	.51	.687	.907
28. Recovers quickly	Adaptability	.97	.857	.908
30. Sets goals	Activities of Daily Living	1.10	.874	.902
Total Factor		12.00	52.202	**Alpha .912**
Factor 2	BASC Scale	Mean	*SD*	Alpha if Item Deleted
2. Does not obey	Conduct Problems	1.03	.571	.827
4. Breaks rules	Conduct Problems	.96	.500	.831
6. Out of control	Hyperactivity	.56	.660	.812
10. Defies	Aggression	.55	.662	.825
18. Trouble	Conduct Problems	.99	.687	.807
20. Disrupts	Hyperactivity	.71	.638	.825
26. Loses temper	Aggression	.80	.791	.825
29. Disliked	Depression	.46	.735	.832
Total Factor		6.06	3.674	**Alpha .842**
Factor 3	BASC Scale	Mean	*SD*	Alpha if Item Deleted
7. Lonely	Depression	.44	.690	.806*
11. Worries about things	Anxiety	.90	.780	.779
13. Frustrated	Depression	1.17	.798	.780
16. Nervous	Anxiety	.69	.722	.763
23. Easily upset	Depression	1.10	.762	.759
25. Worries	Anxiety	.83	743	.735

Total Factor		5.13	3.190	**Alpha .802**
Factor 4	BASC Scale	Mean	*SD*	Alpha if Item Deleted
14. Pain	Somatization	.60	.619	---
21. Health	Somatization	.41	.600	---
Total Factor		1.01	1.083	**Alpha .733**
Factor 5	BASC Scale	Mean	*SD*	Alpha if Item Deleted
8. Distracted	Attention Problems	1.28	.792	---
27. Trouble concentrating	Attention Problems	.99	.825	---
Total Factor		2.27	1.463	**Alpha .778**

Note. * = Alpha increases if item is deleted. Items are truncated because they are copyrighted material of Pearson Assessments.

Student Form. Results for the exploratory factor analysis of the Student Form are presented in Tables 10 through 12. Eigenvalues of seven factors on the Student Form were greater than one, and those seven factors accounted for 46.21% of the variance. The scree plot suggested retaining four factors (see Figure 3). When a seven-factor solution was examined, it was determined that no items loaded onto the seventh factor and only two loaded onto the sixth factor. Based on the scree plot, a four factor solution was examined and was determined to be interpretable. Factor 1 ($\alpha = 0.867$) contained items related to attitudes towards school and teachers. Factor 2 ($\alpha = 0.788$) consisted of items that related to feelings of self-worth and esteem. The items on Factor 3 ($\alpha = 0.800$) all related to inattention and hyperactivity. Factor 4 ($\alpha = 0.765$) contained items that related to internalizing/anxious behaviors.

Figure 3

Student Form Scree Plot

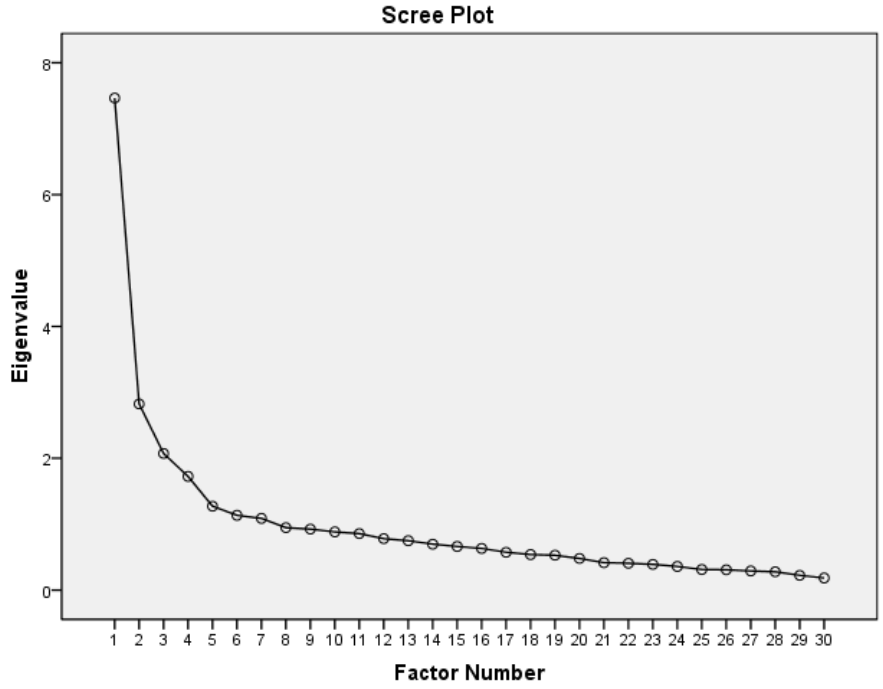


Table 10

Promax-Rotated Factor Pattern/Structure Coefficients for the BESS Student Form

Item	Factor 1	Factor 2	Factor 3	Factor 4
6. Want to quit school	**.595**	-.086	.077	.078
12. Boring	**.921**	-.060	-.036	-.119
17. Dislike	**.861**	-.073	.053	-.112
19. Unfair	**.643**	.010	.036	.114
22. Make me feel stupid	**.598**	.077	.080	.037
29. Good to me	**.627**	.271	-.019	-.115
1. Making decisions	-.153	**.522**	.322	-.206
4. Looks	.147	.373	-.043	-.070
9. Liked	-.211	**.662**	.042	.055
15. Trust	.017	**.604**	.089	-.029
18. Parents listen	.051	**.598**	-.048	-.061
21. Fun	-.098	**.572**	-.060	.078
26. Proud of me	.132	**.650**	-.041	-.103
27. Fail	.108	.313	-.018	.078

30. Respect	.064	**.632**	-.181	.125
2. Talk while others are	.205	-.094	**.509**	.025
8. Trouble paying attention	.080	.102	**.617**	-.013
11. Trouble sitting still	.089	-.116	**.593**	.060
24. Noisy	-.070	.159	**.469**	.086
25. Get into trouble for inattention	.187	.072	.575	.018
28. Trouble standing still	-.074	-.110	**.757**	.014
3. Worry	-.184	.010	.050	.344
5. Out of place around others	-.108	.167	.057	.373
7. Mad at me	.104	.105	-.124	**.532**
10. Life is getting worse	.137	.346	-.032	.383
13. People are out to get me	.104	-.091	-.014	**.687**
14. Worry about future	-.149	-.222	.076	**.667**
16. Left out	.043	.052	.015	**.542**
20. Want to do better	-.117	.243	.134	.300
23. Get blamed	.200	.043	.062	**.412**
N= 212				

Note. Pattern/structure coefficients greater than .4 are bolded. Items are truncated because they are copyrighted material of Pearson Assessments.

Table 11

Student Form Factor Correlation Matrix

Factor	**1**	**2**	**3**	**4**
1	1.00			
2	.413	1.00		
3	.484	.346	1.00	
4	.354	.502	.364	1.00

Table 12

Student Form Four Factor Solution - Factor Items and Properties

Factor 1	BASC Scale	Mean	*SD*	Alpha if Item Deleted
6. Want to quit school	Attitude to School	.82	1.047	.860
12. Boring	Attitude to School	1.11	1.100	.828
17. Dislike	Attitude to	.87	1.085	.825

	School			
19. Unfair	Attitude to Teachers	.64	.916	.848
22. Feel stupid	Attitude to Teachers	.55	.920	.853
29. Good to me	Attitude to School	.96	1.099	.849
Total Factor		4.95	4.792	**Alpha .867**
Factor 2	BASC Scale	Mean	*SD*	Alpha if Item Deleted
1. Making decisions	Self-Reliance	1.33	.857	.776
4. Looks	Self-Esteem	.97	.980	.784
9. Liked	Interpersonal Relations	1.10	1.030	.761
15. Trust	Relations with Parents	.98	1.077	.759
18. Parents listen	Relations with Parents	1.06	1.078	.763
21. Fun	Interpersonal Relations	1.17	1.065	.765
26. Proud of me	Relations with Parents	.71	.865	.758
27. Fail	Sense of Inadequacy	1.08	.853	.787
30. Respect	Interpersonal Relations	1.25	1.053	.755
Total Factor				**Alpha .788**
Factor 3	BASC Scale	Mean	*SD*	Alpha if Item Deleted
2. Talk while others are	Hyperactivity	1.08	.915	.774
8. Trouble paying attention	Attention Problems	.89	.915	.757
11. Trouble sitting still	Hyperactivity	1.22	1.110	.770
24. Noisy	Hyperactivity	1.15	1.038	.792
25. Get into trouble for inattention	Attention Problems	1.18	1.024	.753
28. Trouble standing still	Hyperactivity	.92	1.048	.766
Total Factor		6.44	4.286	**Alpha .800**
Factor 4	BASC Scale	Mean	*SD*	Alpha if Item Deleted
3. Worry	Anxiety	1.31	1.034	.773*
5. Out of place around others	Social Stress	1.08	1.018	.752
7. Mad at me	Locus of Control	1.41	1.087	.734
10. Life is getting worse	Depression	1.04	1.099	.733
13. People are out to get me	Atypicality	1.02	1.060	.725
14. Worry about future	Anxiety	1.44	1.049	.743
16. Left out	Social Stress	1.07	.939	.736
20. Want to do better	Sense of Inadequacy	1.35	1.090	.752
23. Get blamed	Locus of Control	1.24	1.080	.740
Total Factor		10.96	5.580	**Alpha .765**

Note. * = Alpha increases if item is deleted. Items are truncated because they are copyrighted material of Pearson Assessments.

External Validity

Based on the Box's Test of Equality of Covariance Matrices, the dependent variables violated the assumption of equality of covariance ($p = 0.000$). Additionally, Levene‟s Test revealed unequal variance across groups. Finally, the ODR, suspensions, and attendance variables were not normally distributed, as the vast majority of the children in the sample had never had an office discipline referral or suspension, and were present and on time for school most days. Based on the violations of MANOVA assumptions, nonparametric tests to compare means were required. Because of a low number of students in the extremely elevated category of both the Parent and Student Forms ($n = 11$ and 19 on the Parent and Student Forms, respectively), the elevated and extremely elevated groups were combined for these analyses.

Results of the Kruskal-Wallis Test comparing BESS Teacher Form groups of normal ($n = 353$), elevated ($n = 102$), and extremely elevated ($n = 37$) on academic (ORF) and behavioral (ODR, suspensions, and attendance) variables revealed significant effects. The tests of Teacher Form group differences resulted in rejecting the null hypotheses for ORF ($p = 0.000$), ODR ($p = 0.000$), and suspensions ($p = 0.000$). The Parent Form tests were significant in the areas of ODR ($p = 0.033$) and attendance ($p = 0.035$). The Student Form revealed significant differences in the areas of ORF ($p = 0.000$), ODR ($p = 0.000$), and attendance ($p = 0.036$), while suspensions approached significance ($p= 0.061$).

Inter-rater Reliability and Agreement

Agreement among raters was determined with inter-rater reliability and inter-rater agreement. Inter-rater reliability was calculated with a Pearson correlation between T-Score

ratings on the BESS forms (see Table 3). Results revealed that the Teacher and Parent forms were significantly correlated ($r = 0.389$). Likewise, the Teacher and Student forms were found to correlate ($r = 0.393$). Interestingly, there was not a significant correlation between Parent and Student ratings ($r = 0.105$). Inter-rater agreement is reported as the percent agreement between raters of a child"s risk status according to the BESS forms. Characteristics of students identified by level of risk on the Teacher, Parent, and Student Forms are presented in Tables 13 through 15. For this measurement, the BESS classifications of elevated and extremely elevated were combined. In other words, raters were considered to agree if the child was determined to be at either level of elevated risk above normal. Results revealed that teachers and parents agreed on 72.7% of children who were identified as normal, and 31.1% of children who were identified as at-risk ($n = 255$). Students and teachers agreed on 62.9% of children who were identified as normal, and 37.0% of those identified as at-risk ($n = 207$). Students and parents agreed on 65.6% of those identified as normal, and 15.4% of at-risk children ($n = 101$).

Table 13

Characteristics of Students by Risk Level on Teacher Form

	Normal			Elevated			Extremely Elevated		
Variable	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
ODR	353	.07	.36	102	.41	1.03	37	1.00	1.65
Suspensions	353	.00	.05	102	.01	.11	37	.11	.32
Attendance	350	93.71	5.90	101	93.42	5.08	36	92.73	4.84
ORF[1]	279	.21	.95	79	-.55	.88	29	-.60	.85

Note. ODR = office discipline referrals.ORF[1] = oral reading fluency z-score calculated with grade-level means and standard deviations.

Table 14

Characteristics of Students by Risk Level on Parent Form

	Normal			Elevated or Extremely Elevated		
Variable	*n*	*M*	*SD*	*n*	*M*	*SD*
ODR	203	.19	.83	51	.27	.60
Suspensions	203	.01	.12	51	.04	.20
Attendance	201	94.11	5.08	48	92.55	5.84
ORF[1]	152	.08	.89	35	-.19	1.19

Note. ODR = office discipline referrals. ORF[1] = oral reading fluency z-score calculated with grade-level means and standard deviations.

Table 15

Characteristics of Students by Risk Level on Student Form

	Normal			Elevated or Extremely Elevated		
Variable	*n*	*M*	*SD*	*n*	*M*	*SD*
ODR	145	.13	.48	67	.66	1.48
Suspensions	145	.01	.08	67	.04	.18
Attendance	145	95.02	5.15	65	94.27	4.02
ORF[1]	139	.12	1.02	65	-.25	.93

Note. ODR = office discipline referrals. ORF[1] = oral reading fluency z-score calculated with grade-level means and standard deviations.

Discussion

Behavior difficulties are increasingly the focus of school intervention efforts, as it is estimated that between 2% and 10% of our school children suffer from emotional and behavioral disorders (Walker, Colvin, & Ramsey, 1995). In addition, emotional and behavioral difficulties have been found to impact academic achievement (Wagner et al., 2005; Walker et al., 1987), school engagement (Walker et al., 1987), and involvement in the criminal justice system (Huizinga & Jakob-Chien, 1998). Traditional disciplinary methods for dealing with behavioral

difficulties are often exclusionary in nature (i.e., suspensions and expulsions) and result in less time in the classroom for the student (Sugai & Horner, 2008; Walker et al., 1996). Hence, recent research has been devoted to detecting the well-documented risk factors for emotional and behavioral difficulties, in an effort to provide early interventions for those children at-risk for behavioral difficulties (Wagner, Kutash, Duchnowski, Epstein et al., 2005; Walker et al., 1995). While the need for behavioral screeners is high, few behavioral screening instruments have been developed in comparison to the number of academic screening instruments currently on the market. One behavior screener, the Behavioral and Emotional Screening System (BESS) was developed from the widely used Behavioral Assessment System for Children (BASC; Reynolds & Kamphaus, 1992). The purpose of the current study was to examine the psychometric properties of this screening instrument when used in a rural elementary school.

The BESS is a relatively new instrument, as it was published in 2007, and this is the first study of the BESS conducted by researchers other than the BESS authors. The current study sought to compare the BESS to criteria that have been proposed for the evaluation of screening instruments (e.g., Glover & Albers, 2007; Lane et al., 2007). The internal consistency of the three BESS forms was well above the recommended standard for screening instruments of .70 and each was similar to those found in previous research (Kamphaus & Reynolds, 2007).

The factor structure of each of the BESS forms was examined. These analyses were undertaken because the factor structures of the BESS forms have not been reported. It may be beneficial to researchers and practitioners to see the underlying dimensions of adaptive and maladaptive behaviors of each BESS form. Results of the exploratory factor analyses revealed that each form contained multiple factors. The eigenvalues and scree plot of the Teacher Form suggested retaining five factors; however, a four factor solution was most interpretable and

parsimonious. The four factor solution of the Teacher Form resulted in factors representing externalizing behaviors, inattention and academic problems, internalizing behaviors, and adaptive skills. The scree plot and eigenvalues of the Parent Form indicated a five factor solution, while a three factor solution may be the most interpretable and parsimonious. The wording of the items on the Parent Form varied by factor and may represent a measurement artifact. The three factor solution of the Parent Form resulted in factors representing positively worded adaptive and attention skills, negatively worded externalizing behaviors and inattention, and internalizing behaviors. The Student Form was most interpretable with four factors. The scree plot further suggested retaining four factors, while eigenvalues greater than one were present for seven factors. A four factor solution of the Student Form retained factors representing attitude towards school and teachers, self-worth and esteem, inattention/hyperactivity, and internalizing behaviors.

An examination of the factor structure of the BESS Forms revealed multiple factors; however, in use, only the overall score of the BESS is calculated. Future research should examine the utility of using subscales of the BESS forms, based on the underlying factor structure, to identify children at-risk of behavioral difficulties. Items on the BESS Parent Form appeared to group by wording style (i.e., positive v. negative), which may indicate a measurement artifact. This could be a direction for future research.

Criterion-related external validity was examined with measures of reading achievement and behavior. The Teacher Form of the BESS showed significant group effects on the ORF, ODR and suspensions variables. Results of the Parent Form revealed the least significant group differences, as the results only varied by ODR and attendance. The Student Form results indicated that groups varied on ORF, ODR, and attendance. The criterion-related validity results

of this study are similar to those of Kamphaus et al. (2007); however, different statistical methods of comparison were used. Regardless of the method used, both studies revealed that the BESS Teacher Form is negatively related to academic variables (math and reading scores in the Kamphaus et al. study and oral reading fluency in the current study), and is not related to suspensions or attendance. Results of the current study are consistent with previous work showing a relationship between behavioral and academic difficulties (Lane et al., 2008; Lane et al., 2010; Wagner et al., 2005; Walker et al., 1987; Walker et al., 1995).

Agreement among raters was reported in terms of inter-rater reliability and agreement. Inter-rater reliability, while significant for the Teacher and Parent forms was low (Cohen, 1977). There was not a significant correlation between Student and Teacher forms. In a large meta-analysis of cross-informant behavioral ratings, Achenbach et al. (1987) found that small correlations between raters, and particularly between students and other raters (i.e., parents and teachers) was common. Achenbach et al. suggested that the low correlations may be due to actual changes in behavior by setting, and argued for the use of multiple raters, despite the low correlations, as additional sources of information. Future research should examine this with a regression model to determine what, if any, variance is explained with the addition of parent and student reports. Based on the findings of the current study, including reliability and validity analyses, practitioners should consider the cost of additional forms and the usefulness of information provided when deciding to include parent and student input in the screening process.

As suggested by Lane et al. (2007), screening instruments should adhere to and report certain psychometric properties. The current study sought to examine several of those properties, yet others have yet to be examined. The internal consistency and convergent validity of the BESS have now been examined in several studies (see Kamphaus & Reynolds, 2007; Kamphaus

et al., 2010; and the current study) and have been found to be adequate. The test-retest reliability, positive and negative predictive values, and sensitivity and specificity of the BESS have only been examined by the BESS authors (Kamphaus & Reynolds, 2007). Future research, conducted by independent researchers, should examine these principles again and with new samples.

Universal behavior screening is intended to identify children in need of behavioral intervention. Due to the scope of this task, Lane et al. (2007) suggested that screening measures be cost effective and efficient to administer. However, these considerations are likely not the most important to researchers focused on producing sound psychometric instruments. The authors of the BESS, for example, have set sensitivity and specificity to over identify children at-risk for behavioral difficulties, in an attempt to identify all children who are potentially at risk. This over identification significantly increases the cost and time required to complete the screening process. While it is important to identify all potential problems, the practical limitations found in applied settings need to be considered.

Finally, limitations of the current study must be discussed. At the time of data collection for the current study, only the English versions of the BESS Parent Form were available. A Spanish version of the BESS has since been published. Given the percentage of Spanish-speaking families in the school where data was collected, a Spanish version may have proved useful. Future research should examine the psychometric properties of the Spanish version of the BESS Parent Form. There were also significant group differences between parent responders and nonresponders in terms of home language, ethnicity, grade of student, and free and reduced lunch status. Future research on these populations should seek a more representative sample. The use of the Spanish version of the BESS Parent Form may have remedied these group differences.

The sample used in the current study was drawn from one rural elementary school. While nearly every child in the school was included in the sample, the sample is considered limited in that only one school was used. Additionally, the sample contained 496 students; however, when examined by grade-level, the sample size narrows considerably. Future research should seek to draw a more comprehensive sample of students.

References

Achenbach, T.M., MsConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213-232.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Bulrington, VT: University of Vermont, Research Center for Chidlren, Youth, and Families.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New\brk: Academic Press.

Costello, A.B. & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10 (7)*, 1-9.

Donovan, M. S., & Cross, C. T. (Eds.). (2002). *Minority students in special education*. Washington, DC: National Academy Press.

Dynamic Measurement Group (2008). DIBELS 6th Edition Technical Adequacy Information (Tech. Rep. No. 6). Eugene, OR: Author. Available: http://dibels.org/pubs.html.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272-299.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*(2), 117-135.

Gresham, F. M. (2004). Current Status and Future Directions of School-Based Behavioral Interventions. *School Psychology Review, 33*(3), 326-343.

Good, R. H., III. (2004). Using the Dynamic Indicators of Basic Early Literacy (DIBELS). Retrieved May 10th, 2004, from http://dibels.uoregon.edu/index.php.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures.* New York: Wiley.

Huizinga, D., & Jakob-Chien, C. (1998). The Contemporaneous Co-Occurrence of Serious and Violent Juvenile Offending and Other Problem Behaviors. In R. Loeber & D. P. Farrington (Eds.), *Serious and Violent Juvenile Offenders* (pp. 47-67). Thousand Oaks, CA: Sage Publications.

Kamphaus, R. W., DiStefano, C., Dowdy, E., Eklund, K., and Dunn, A. R. (2010). Determining the presence of a problem: Comparing two approaches for detecting youth behavioral risk. *School Psychology Review, 39 (3)*, 395-407.

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 Behavioral and Emotional Screening System Manual*: Pearson.

Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & VanDeventer, M. C. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement, 67*, 342-356.

Kazdin, A. E. (1987). *Conduct Disorders in Childhood and Adolescence*. London: Sage.

Lane, K. L., Barton-Arwood, S. M., Nelson, J. R., Wehby, J. (2008). Academic performance of students with emotional and behavioral disorders served in a self-contained setting. *Journal of Behavioral Education, 17*, 43-62.

Lane, K .L., Little, A., Menzies, H., Lambert, W., & Wehby, J. (2010). A comparison of students with behavior challenges educated in suburban and rural settings: Academic, social, and behavioral outcomes. *Journal of Emotional and Behavioral Disorders, 18*(3), 131-148.

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*(4), 209-222.

Loeber, R., & Farrington, D. P. (Eds.). (1998). *Serious and Violent Juvenile Offenders*. Thousand Oaks, CA: Sage Publications.

Muyskens, P., Marston, D., & Reschly, A. L. (2007). The use of response to intervention practices for behavior: An examination of the validity of a screening instrument. *The California School Psychologist, 12*, 31-45.

O'Shaughnessy, T. E., Lane, K. L., Gresham, F. M., & Beebe-Frankenberger, M. E. (2002). Students With or At-risk for Learning and Emotional-Behavioral Difficulties. In K. L. Lane, F. M. Gresham & T. E. O'Shaughnessy (Eds.), *Interventions for Children With or At-risk for Emotional and Behavioral Disorders*. Boston: Allyn & Bacon.

Reschly, A.L, Busch, T.W., Betts, J., Deno, S.L., Long, J.D. (2009). Curriculum-based measurement of oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*(6), 427-469.

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

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

Sprague, J., Walker, H. M., Stieber, S., Simonsen, B., Nishioka, V., & Wagner, L. (2001). Exploring the relationship between school discipline referrals and delinquency. *Psychology in the Schools, 38*(2), 197-206.

Sugai, G., & Horner, R. (2008). What we know and need to know about preventing problem behavior in schools. *Exceptionality, 16*, 67-77.

U.S. Department of Education (n.d.). Office of Special Education Programs, Data Analysis System, *Children with Disabilities Exiting Special Education, 2005–06* (OMB #1820-0521). Retrieved November 28, 2007, from https://www.ideadata.org/arc_toc8.asp#partbEX.

Wagner, M., Kutash, K., Duchnowski, A. J., & Epstein, M. H. (2005). The Special Education Elementary Longitudinal Study and the National Longitudinal Transition Study: Study Designs and Implications for Children and Youth With Emotional Disturbance. *Journal of Emotional and Behavioral Disorders, 13*(1), 25-41.

Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The Children and Youth We Serve: A National Picture of the Characteristics of Students With Emotional Disturbances Receiving Special Education. *Journal of Emotional and Behavioral Disorders, 13*(2), 79-96.

Walker, H. M., Colvin, G., & Ramsey, E. (1995). *Antisocial behavior in school: Strategies and best practices*. Belmont, CA US: Thomson Brooks/Cole Publishing Co.

Walker, H. M., Horner, R. H., Sugai, G., & Bullis, M. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4*(4), 194-209.

Walker, H. M., Shinn, M. R., O'Neill, R. E., & Ramsey, E. (1987). A longitudinal assessment of the development of antisocial behavior in boys: Rationale, methodology, and first year results. *Remedial and Special Education, 8*(4), 7-16.

Walker, H. M., Shinn, M. R., & Stoner, G. (2002). Structuring School-Based Interventions to Achieve Integrated Primary, Secondary, and Tertiary Prevention Goals for Safe and Effective Schools. In *Interventions for academic and behavior problems II: Preventive and remedial approaches.* (pp. 1-25). Washington, DC US: National Association of School Psychologists.

Zigler, E., Taussig, C., & Black, K. (1992). Early childhood intervention: A promising preventative for juvenile delinquency. *American Psychologist, 47*(8), 997-1006.

CHAPTER 3

A COMPARISON OF SCREENING INSTRUMENTS: PREDICTIVE AND SOCIAL VALIDITY OF THE BESS AND BCS

Response to intervention (RTI) is arguably one of the most innovative yet controversial ideas to arise in school psychology and special education. While some consider RTI to be an improvement in the field of school psychology and a force that will drive overall educational improvement (Fletcher & Reschly, 2005; Gresham et al., 2005) others fear that RTI lacks empirical evidence, particularly for special education eligibility determination (Hale, Naglieri, Kaufman, & Kavale, 2004; Kavale, Kaufman, Naglieri, & Hale, 2005). In this chapter, a brief history and essential components of RTI are presented, followed by a discussion of positive behavior support (PBS) and universal screening as part of RTI models.

Special Education Law

From the initial passage of P.L. 94-142 in 1975, federal special education law has had an enormous influence on states, educator practices, and students. In 1997, Congress reauthorized P.L. 94-142 as the Individuals with Disabilities Education Act (IDEA 1997). With this reauthorization, schools were given the flexibility to use federal funds to support school-wide programs to benefit students with and without disabilities. The goal of this change in law was to provide a more integrated special education system within public schools. The reauthorization of IDEA in 1997 further allowed for flexibility in special education eligibility determination, by including wording that permitted the use of multiple sources of information, including parent input, when making eligibility determinations (Batsche et al., 2005).

In 2004, IDEA was reauthorized again. One of the most significant, and controversial, changes included specific wording allowing for use of students‟ response to intervention data as part of the eligibility determination for the special education category of Specific Learning Disability (Batsche et al., 2005). The wording of IDEA 2004 has lead to much discussion and debate in the field of school psychology, as researchers, practitioners, and the American Academy of School Psychology disagree about the practical implications of using RTI for special education eligibility determination (Fletcher & Reschly, 2005; Gresham et al., 2005; Hale et al., 2004; Kavale et al., 2005; Schrank et al., 2005).

Although the reauthorization of IDEA allowed for the use of RTI, the impetus toward RTI as a large scale reform came from a number of places, including those concerned with poor outcomes in special education (Kavale, 2001), the overrepresentation of certain groups in special education (Donovan & Cross, 2002), bias in the referral and assessment process (Vaughn & Fuchs, 2003), the growing number of students receiving special education under the category of Specific Learning Disability (MacMillan, Gresham, & Bocian, 1998), and dissatisfaction with the disjointed programs across general and special education (Reschly & Bergstrom, 2009). In their 2002 report on behalf of the National Research Council Panel of Minority Overrepresentation, Donovan and Cross emphasized the lack of research regarding the benefits of special education for minority children. The authors recommended that the federal government change the way in which it views regular and special education, from segregated learning environments with separate funding, to more integrated services. Outlined in their report is a proposed method for determining special education eligibility, beginning with a student exhibiting large differences from peers. As proposed by Donovan and Cross (2002), if a child exhibits these differences, and does not respond to high-quality interventions aimed to remediate

those differences, he/she may be determined to be eligible for special education services. Furthermore, schools should use universal screening and multi-tiered interventions to promote early identification and intervention for children with difficulties (Donovan & Cross, 2002).

Response to Intervention

According to a seminal publication by the National Association of State Directors of Special Education (Batsche et al., 2005), RTI is defined by 1) high quality instruction that is matched to student needs, 2) the use of level of performance and learning rate over time as sources of information, and 3) the making of educational decisions regarding the intensity and duration of intervention based on a student‟s response to intervention. RTI is a system designed to support all children, including those with academic and behavioral difficulties.

An RTI model has three essential and necessary components: a multi-level/multi-tier intervention system (see Figure 4), a problem-solving model, and an integrated data collection system (Batsche et al., 2005). Multi-level or multi-tier models are practical for the uses of resource allocation and meeting the needs of individual students, by matching the level of support to individual student needs. At the lowest tier are school-wide modifications, including preventative and proactive measures designed to meet the educational and behavioral needs of most students. Within Tier 1, high-quality, research-based instruction is provided to all students. It is estimated that 80% to 90% of school children respond appropriately to Tier 1 and need no further intervention to succeed both academically and behaviorally (Batsche et al., 2005; Sugai et al., 2002; Sugai, Sprague, Horner, & Walker, 2000). Important in Tier 1 is the use of universal screening to identify those children who are in need of more intensive intervention. If children are identified via screening, additional interventions are provided in Tier 2.

Figure 4. Three-Tier Model of School Supports

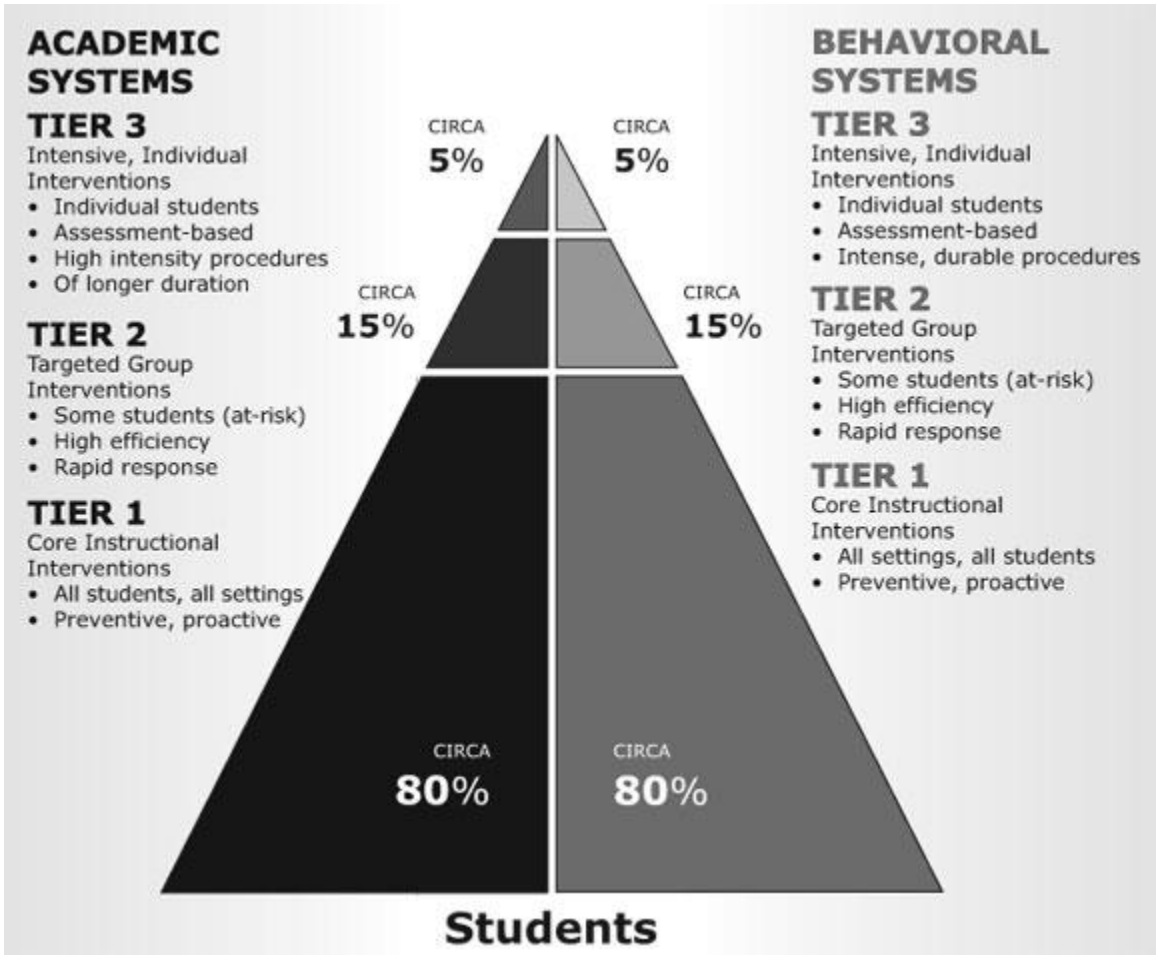


Figure 4. The three-tier model of academic and behavioral supports as used in RTI models (Batsche et al., 2005).

Tier 2 interventions are provided *in addition* to Tier 1 interventions, meaning that children who are receiving Tier 2 interventions are still receiving the school-wide services and high-quality instruction provided in Tier 1. In Tier 2, students are provided more specified and intense intervention in their determined area of need (i.e., reading, writing, behavior, etc.). It is estimated that 5% to 15% of children are in need of Tier 2 support (Batsche et al., 2005; Sugai et al., 2002; Sugai, Sprague et al., 2000).

There are two types of Tier 2 interventions: problem-solving interventions and standard treatment protocol interventions (Batsche et al., 2005). Problem-solving interventions are based

on a team approach, where a team of school professionals uses a problem-solving model to determine why a particular child is not responding to group intervention. These teams, which are sometimes called student support teams or student study teams, then develop an individualized intervention plan for that child based on his/her functional impairment.

Standard treatment protocol interventions are less individualized than problem-solving interventions, but tend to have strong empirical support (Batsche et al., 2005). The majority of research with standard protocol interventions has been conducted in the area of reading (Reschly & Bergstrom, 2009). Standard protocol interventions are typically a predetermined response to a particular type of problem. In other words, a school may have a standard protocol of reading interventions if a child presents with reading fluency difficulties. Because many children may be receiving the same intervention, standard treatment protocol interventions are amenable to a small group format (Batsche et al., 2005).

Dan Reschly described the debate over standard protocol and problem-solving interventions as a „false dichotomy" (Reschly & Bergstrom, 2009). That is, both standard protocol and problem-solving interventions are often used with the same child. When both are used in unison, standard protocol procedures are more often applied to academic concerns, while problem-solving procedures are used to remedy problem behavior (Reschly & Bergstrom, 2009). Furthermore, at their core, both standard protocol and problem-solving approaches use the problem-solving method to remediate difficulties (Jimerson, Burns, & VanDerHeyden, 2007). Other researchers have suggested that a combination of standard protocol and problem solving intervention plans is the most effective method (Batsche et al., 2005). The student support team should determine the function of the deficiency, as suggested in the problem-solving method, and then draw from a set of standard-protocol treatment interventions that is likely to be effective

or has shown to be effective for addressing that particular difficulty. This combination of intervention plans draws on the strength of the team approach, while ensuring that children are receiving empirically-supported interventions (Batsche et al., 2005).

Still other researchers suggest using standard protocol procedures as a first step, and then problem solving interventions as a second step for children who do not respond to standard protocol small group interventions (Burns, Deno, & Jimerson, 2007). Based on the assumption that both standard protocol and problem solving interventions are effective, efficiency can be used to differentiate between them. In other words, standard protocol interventions, which are typically delivered in a small group format, are more efficient than problem solving interventions, which are typically individually administered, in terms of teacher time. However, if a child does not respond sufficiently to standard protocol interventions, individual problem solving strategies should be used (Burns et al., 2007).

If a child does not adequately respond to Tier 2 interventions, then that child moves to Tier 3, where more intensive, individual instruction is provided. In practice, special education services are typically considered to be Tier 3 interventions, although it is important to note that a student may receive interventions at Tier 3 but not be identified as a student with a disability. In a 3-tier model, it is the progression from Tier 2 to Tier 3 that has been the subject of most debate. Some states and districts require a full psychological assessment and the use of traditional eligibility criteria between Tiers 2 and 3, while others have argued against this practice (e.g., Fletcher & Reschly, 2005; Gresham et al., 2005) and for the use of student response to intervention data as part of eligibility determination (Fletcher & Reschly, 2005; Gresham et al., 2005; Hale et al., 2004; Kavale et al., 2005; Schrank et al., 2005). It is estimated that 1% to 7%

of school children require Tier 3 interventions (Batsche et al., 2005; Sugai et al., 2002; Sugai, Sprague et al., 2000).

The second essential feature of RTI models is a problem-solving process. According to Marston and colleagues in the Minneapolis School District "Problem solving involves applying a logical sequence of steps to address an issue or difficulty" (Marston, Reschly, Lau, Muyskens, & Canter, 2007, p.265). The use of problem solving procedures with academic and behavioral problems can be traced to the behavioral consultation work of Bergan (1977) and the Curriculum-Based Measurement (CBM) work of Deno (Batsche et al., 2005; Deno, 1985; Marston et al., 2007). The defining features of problem solving models include progress monitoring and formative evaluation (Marston et al., 2007). Using this approach, when children are identified as having deficits, they are provided an intervention and their progress is continually monitored while receiving that intervention. This process is considered formative because a child"s own progress in meeting goals dictates how the instructional goals are altered. If a child is successful in meeting goals, then the goals are made more challenging. If a child is unable to meet goals, the intervention is altered (Marston et al., 2007).

When Bergan and Deno first proposed versions of the problem solving model, the intention was not to diagnose or identify children as disabled. However, due to the change in wording of special education law (as previously discussed) school districts have begun using the problem solving model as a way of identifying children in need of special education services (Marston et al., 2007). Minneapolis Public Schools, among others, have used the problem solving model successfully and for a number of years as a means of identifying students with disabilities and designing interventions (Marston et al., 2007).

The third essential component of RTI, an integrated data system, is used to determine Response to Intervention. Batsche and colleagues (2005) defined several characteristics of acceptable RTI assessment procedures, including the criteria that they are relevant to state performance standards and to the area of individual need, that they can be used to monitor even small amounts of progress over time, and that they can be administered repeatedly over short periods of time. The problem solving model based on formative evaluation is closely related to the idea of an integrated data system. Specifically, CBM methods (Deno, 1985) are often considered an excellent source of data on which to base intervention decisions. While a thorough discussion of CBM is beyond the scope of the current study (see Deno, 1985 for an excellent description), a brief discussion of the core components of CBM and how they fit within an RTI and problem solving model is warranted. CBM is a process in which school curriculum materials are used to validly and accurately measure student achievement (Deno, 1985). CBM is amenable to the process of an integrated data system because CBM instruments are sensitive to short-term growth and can be administered on a frequent basis (Marston et al., 2007).

A general 3-Tier RTI model was described in this chapter; however, many states have developed variations on this model, including Georgia, which has a four tier model. Kovaleski (2007) noted that the differences in tiers between and within states is a potential source of confusion, and subsequently chose to label the tiers the Benchmark, Targeted, and Ongoing Support Phases. Further, there are implementation issues that frequently arise in RTI models, including the selection of evidence-based curricula and effective instructional strategies amidst the preponderance of commercially available materials purporting to be evidence-based (Kovaleski, 2007). Additionally, school districts need to have clearly defined regulations regarding special education eligibility determination (Kovaleski, 2007).

Screening students is a firmly supported practice among researchers and educators alike (Batsche et al., 2005; Muyskens et al., 2007; O'Shaughnessy, Lane, Gresham, & Beebe-Frankenberger, 2002; Severson, Walker, Hope-Doolittle, Kratochwill, & Gresham, 2007; Sprague & Walker, 2000; Walker, 2000; Walker, Horner, Sugai, & Bullis, 1996); however, schools rarely use screening data to inform instruction, but rather to identify children in need of more intensive intervention. While identifying children via screening is of the highest priority, screening data delivered to school personnel in a timely manner has the potential to influence educational practices (Kovaleski, 2007).

Positive Behavior Support

Appropriate student behavior is crucial to maintaining a productive academic environment within schools. Children identified as serious violent and serious nonviolent offenders are more likely to be truant, have lower academic achievement scores, and are more likely to drop out than nondeliquent children (Huizinga & Jakob-Chien, 1998). Serious offenders aside, difficulties with student discipline are a primary reason for teacher burnout (President‟s Commission on Excellence in Special Education, 2002). Further, traditional school discipline practices are reactive, rather than proactive, and focus heavily on the use of punishment to manage student behavior (Sugai & Horner, 2008; Walker et al., 1996) Recent movements, both in legislation and recommended practice, have emphasized the responsibility of schools in promoting prosocial behavior (Sugai, Horner et al., 2000).

When IDEA was reauthorized in 1997, specific changes were made in regard to disruptive behavior. Specifically, the law was worded such that schools needed to consider the use of positive behavior supports for children whose behavior was impeding their own learning or the learning of others. Positive behavior support (PBS) was founded in reaction to aversive

behavioral intervention for children with severe and self-injurious behavior (Warren et al., 2006). Sugai and colleagues (2000) defined PBS as "a general term that refers to the application of positive behavioral interventions and systems to achieve socially important behavior change" (Sugai, Sprague et al., 2000; p.133).

Since its inception, PBS has grown to serve as the behavioral model upon which entire schools have based behavior goals. School-wide PBS has several core components, including the selection of a PBS team, the definition of school-wide behavior goals, directly teaching behavior goals to students, the development of a system to acknowledge positive behavior and discourage negative behavior, and a continual monitoring of the effectiveness of the program (Sugai, Horner et al., 2000; Warren et al., 2006). School-wide PBS has been implemented in nearly 6,000 schools nationwide (Sugai & Horner, 2008). All states are currently developing RTI models (Reschly & Bergstrom, 2009); however, only a handful of states and districts have large-scale RTI models in place (Burns et al., 2007). While PBS exists outside of the RTI model and debate, RTI applications that include behavior are largely based on the PBS model.

Essential to both RTI and PBS is universal screening. Within a tiered model of intervention, children who are at-risk for developing difficulties, either academic or behavioral, need to be identified in order to receive more intensive and appropriate interventions. Universal screening is recommended in both the academic and behavioral literature as an effective method of determining at-risk children (Harrison, 2009). Screening in academic areas, particularly reading, is widespread and well validated; however, screening in the area of behavior is far behind that of academics (Muyskens et al., 2007). Two likely reasons why behavioral screening technology has not been widely adopted is that there are few validated screening measures and many of the widely-used behavior ratings scales are not suited for screening in terms of money

and time. Despite the functional reasons for a lack of behavioral screening, researchers have overwhelmingly agreed that behavioral screening is not only necessary, but needs to occur as early as possible in a child‟s school career (Kazdin, 1987; Sprague, Walker, Golly et al., 2001; Walker, Colvin, & Ramsey, 1995).

The costs of delaying intervention for children with behavioral difficulties are well noted, and include higher rates of suspension, expulsion, and drop out (Walker, Shinn, O‟Neill, & Ramsey, 1987). Moreover, children with behavior difficulties have consistently scored lower on standardized achievement tests than typical peers (Wagner, Kutash, Duchnowski, Epstein & Sumi, 2005; Walker et al., 1987). However, if intervention occurs early enough, the path to antisocial and further disruptive behavior may be diverted (Sprague, Walker, Golly et al., 2001).

Lane and colleagues (2007) proposed a set of criteria for use in examining the psychometric properties of screening measures. In addition to the proposed psychometric standards for reliability and validity, they recommend evaluating the cost-effectiveness of screeners. An additional criterion in the evaluation of screeners is social validity. Educators must believe that behavior screening measures are worthwhile in order for them to be willing to devote the time necessary to complete them. Not only should they understand the role of screening in securing behavioral interventions, but they should feel that the screener is addressing issues pertinent to classroom behavior. It is important to consider the social validity of screening instruments, in addition to the psychometric properties, because one does not necessarily predict the other. For example, Walker et al. (1996) found that the measured effectiveness of an instrument and teacher acceptability of said instrument did not always coincide. An additional factor that influences acceptability is the amount of teacher time

required, with results indicating that less teacher time is viewed as more acceptable to teachers (Bergan & Kratochwill, 1990).

Comprehensive applications of RTI, i.e., those that include both academics and behavior, and the implementation of a school-wide model for discipline and behavior support rely on principles of screening, early intervention and progress monitoring. These principles provide the source of information that allows for the determination of response to intervention. Despite the importance of screening in these models, few instruments have been developed for behavioral screening purposes, and very few have promise as universal screening instruments. Most current behavior rating scales and assessment instruments are too long to be used as universal screening measures. For example, the Behavior Assessment System for Children – Second Edition (BASC-2; Reynolds & Kamphaus, 2004), a widely used behavior rating scale, contains over 100 items. Traditional methods of identifying children with behavior difficulties, including discipline records, observations, and teacher referral are also problematic. Specifically, office discipline referrals (ODR) and suspension data are gathered well after the at-risk phase of behavior disorders. Report card ratings of behavior provide little useful information to parents and educators about remedying problem behavior because they are often quite broad in nature (i.e., satisfactory and unsatisfactory). Classroom observations, while informative, are far too time consuming and expensive to be used at a universal level (Muyskens et al., 2007). Finally, teacher referral to special education has been criticized for being biased and unsystematic (Donovan & Cross, 2002; Vaughn & Fuchs, 2003).

The purpose of this study is to examine and compare two recently published instruments designed to be used for universal behavioral screening as part of a multi-tiered RTI model. Both

of the measures will be examined in terms of: 1) psychometric properties; 2) its ability to accurately predict school-based outcome measures; and 3) acceptability by school staff.

Method

Participants

The subjects of this study included 492 Kindergarten through 5th grade students in a public elementary school in rural northeast Georgia. The sample consisted of 47% male and 53% female students. A review of parent-reported ethnicity indicated that 65% of students were European American, 11% were African American, 19% were Hispanic American, less than 1% were Asian American, and 5% were defined by parents are multi-racial. Over 68% of the sample students were eligible for Free or Reduced lunch. Every regular education teacher (n = 23) completed screeners in reference to the children in their classrooms. Demographic information of the participants is presented in table 16.

Measures

BASC-2 Behavioral and Emotional Screening System. The BASC-2 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007) was adapted from the well-validated and widely used Behavioral Assessment System for Children (BASC; Reynolds & Kamphaus, 1992). The BESS was developed for use as a screening measure through the use of principal components analysis that reduced the number of items on the full length BASC forms.

Table 16

Characteristics of Participants by Grade

		Gender		Ethnicity/Race					Home Language		Lunch Status	
Grade	*n*	Male *n* (%)	Female *n* (%)	EA *n* (%)	AfA *n* (%)	AsA *n* (%)	HA *n* (%)	M *n* (%)	English *n* (%)	Spanish *n* (%)	F/R *n* (%)	Full *n* (%)
K	85	46 (54.1)	39 (45.9)	66 (77.6)	7 (8.2)	0 (0)	10 (11.8)	2 (2.4)	75 (88.2)	10 (11.8)	57 (67.1)	28 (32.9)
1	85	44 (51.8)	41 (48.2)	47 (55.3)	8 (9.4)	0 (0)	26 (30.6)	4 (4.7)	57 (67.1)	28 (32.9)	65 (76.5)	20 (23.5)
2	91	42 (46.2)	49 (53.8)	56 (61.5)	10 (11.0)	2 (2.2)	19 (20.9)	4 (4.4)	74 (81.3)	17 (18.7)	63 (69.2)	28 (30.8)
3	70	32 (45.7)	38 (54.3)	42 (60.0)	9 (12.9)	0 (0)	14 (20.0)	5 (7.1)	58 (82.9)	12 (17.1)	52 (74.3)	18 (25.7)
4	89	36 (40.4)	53 (59.6)	59 (66.3)	13 (14.6)	0 (0)	12 (13.5)	5 (5.6)	78 (87.6)	11 (12.4)	57 (64.0)	32 (36.0)
5	72	33 (45.8)	39 (54.2)	49 (68.1)	8 (11.1)	0 (0)	11 (15.3)	4 (5.6)	61 (84.7)	11 (15.3)	42 (58.3)	30 (41.7)
Total	492	233 (47.4)	259 (52.6)	319 (64.8)	55 (11.2)	2 (0.4)	92 (18.7)	24 (4.9)	403 (81.9)	89 (18.1)	336 (68.3)	156 (31.7)

Note.. Ethnicity/Race: Euro Am=European American, Af Am=African American, As Am=Asian American, Hisp Am=Hispanic American, Multi=Multiracial;

Lunch Status: F/R=Free or reduced lunch, Full=full-priced lunch.

Like the BASC system, the BESS system contains three report forms (i.e., Parent, Teacher, and Student) for children in preschool through grade 12. The three forms can be used individually or in combination. The current study used only the BESS Teacher Form, which reportedly takes 5-10 minutes to complete. The Teacher Form contains 27-items that are scored on a 4-point Likert-type scale. The resulting overall score can be classified, using normative data, as having normal, elevated, or extremely elevated levels of risk. BESS Teacher Form items are presented in Table 17.

Reliability and validity evidence of the Teacher Form is provided by age group. Split-half reliability, as opposed to coefficient alpha, is reported as a measure of internal-consistency due to the diversity of item content on the Teacher Form. The median split-half reliability (across age groups) was .96 (Kamphaus & Reynolds, 2007). The test-retest reliability of the BESS, after adjusting for sample variability, was .91. A correlation coefficient between teacher ratings was reported as evidence of inter-rater reliability, and was .70 after adjusting for sample variability.

A validation study of the BESS Teacher Form found moderate correlations between screener scores and achievement scores (r = -0.547 with math scores and r = -0.575 with reading scores), grades (r = -0.477 with math grades and r = -0.546 with reading grades), and work habits (r = -0.434). However, the screener was poor at predicting suspensions and attendance (r = 0.133 and r = 0.121 respectively; Kamphaus, Thorpe, Winsor, Kroncke, et al., 2007).

The total BESS Teacher Form score was compared to scores of other well-known behavior rating scales as a measure of concurrent validity. Correlation coefficients between the BESS Teacher Form and BASC-2 Teacher Rating Scale (BASC-2; Reynolds & Kamphaus, 2004), which were adjusted for restriction of range, were as follows: .79 (Externalizing

Problems), .62 (Internalizing Problems), .89 (School Problems), -.85 (Adaptive Skills), and .90 (Behavioral Symptoms Index). These correlation coefficients should be interpreted with caution, however, because there is a great degree of overlap of items on the BESS and on the BASC-2. In fact 24 of the 27 Teacher Form items on the BESS are also found BASC-2 Teacher Rating Scale. Despite the degree of overlapping items, Kamphaus and Reynolds (2007) argued that the comparison of forms is justified because the BASC-2 form contains many more items than only those it shares with the BESS.

Finally, concurrent validity evidence was calculated between the BESS Teacher Form and the Teacher Rating Form of the Achenbach System of Empirically Based Assessment (ASEBA TRF; (Achenbach & Rescorla, 2001). The adjusted correlation coefficients between the BESS Teacher Form score and the ASEBA TRF scores were as follows: .76 (TRF Total Problems), .69 (TRF Externalizing scale), and .29 (TRF Internalizing scale).

Behavior Screening Checklist. The Behavior Screening Checklist (BSC; Muyskens et al., 2007) is a teacher-report measure of children"s behavior. The full length version of the BSC has 12 items that are grouped into 3 categories: Classroom Behaviors, Externalizing Behaviors, and Socialization (See Table 18). Behavior is scored on a 5-point Likert-type scale. Teachers rated each child on a scale from 1 to 5, with 1 representing appropriate levels of the particular behavior and 5 representing problematic levels of behavior.

A cut-score of 36 was developed by the authors to identify children who are in need of further intervention. This cut-score was chosen because it was with this score that 5% of the students were identified as needing further intervention. Five percent of the student body was chosen because it is consistent with previous research findings (see (Sugai & Horner, 2002) that about 5% of a typical student body is in need of third tier behavior interventions, and it identified

a manageable number of students for the school system to accommodate for more intensive intervention (Muyskens et al., 2007).

The BSC was normed on over 22,000 children in Minneapolis Public Schools. Inter-rater reliability, which was reported as correlation coefficients between sets of co-teachers, was calculated with the original 10-item version of the BSC. The reliability coefficients ranged from .659 to .965 and had a mean rating of .825 (Muyskens et al., 2007).

Internal-consistency reliability, represented with Cronbach‟s alpha, was calculated for the 10-item and 12-item versions of the BSC. For the 10-item version, the alpha coefficient was .93. The alpha coefficients are provided by grade level for the 12-item version of the BSC, and range from .92 to .95 (Muyskens et al., 2007).

Predictive validity was calculated between BCS scores (separated into two groups: kindergarten through grade 5, and grades 6 through 8) and standardized achievement scores, office discipline referrals (ODR), and attendance data. All correlations were significant (p<0.001). The correlation between BSC scores and number of suspensions was .28 for the K-5 sample and .51 for the 6-8 sample. There was a small to moderate relationship between BCS scores and attendance (r = 0.18 and 0.46, for the K-5 and 6-8 samples, respectively). For the kindergarten through grade 5 sample, BSC scores and achievement scores correlated at -.39 and -.42 (for reading and math, respectively). For the grade 6 through 8 sample, correlation coefficients were -.45 and -.48 (for reading and math respectively).

An error occurred in the publication of the BSC that resulted in the printing of only 11 items. The missing item, which is grouped into the Socialization category, was excluded from print. Therefore, only 11 items were used in this study. Steps were taken to develop a new cut score based on the 11-item version used in the current study that corresponded to the top 5% of

the distribution. The utility of a cut score that identified the top 20% of students identified in teachers‟ ratings of behavior, which corresponds with Tier II of the RTI model, was also explored.

Table 17

Items on the BESS Teacher Form

BESS Teacher Form
1. Pays attention
2. Disobeys
3. Is sad
4. Breaks the rules
5. Is well organized
6. Has poor self-control
7. Is easily upset
8. Completes assignments incorrectly because of not following instructions
9. Is good at getting people to work together
10. Has trouble keeping up in class
11. Worries about things that cannot be changed
12. Says, “Nobody likes me.”
13. Annoys others on purpose
14. Is fearful
15. Has headaches
16. Is easily distracted from class work
17. Is effective when presenting information to a group
18. Gets into trouble
19. Gives good suggestions for solving problems
20. Is negative about things
21. Disrupts other children‟s activities
22. Complains about health
23. Has trouble concentrating
24. Has good study habits
25. Worries
26. Has a short attention span
27. Encourages others to do their best

Note. The BESS items are copyrighted material of Pearson Assessments. Items originally published in Kamphaus et al. (2007).

Table 18

Items on the BSC

Classroom Behaviors				
1. Attention:				
1	2	3	4	5
Consistently attends to classroom activities		Sometimes follows along with classroom activities		Rarely follows along with classroom activities
2. Follows Directions:				
1	2	3	4	5
Consistently follows rules		Sometimes follows rules		Rarely follows rules
3. Completing work:				
1	2	3	4	5
Consistently completes work independently		Sometimes completes work independently		Rarely completes work independently
4. Class Involvement:				
1	2	3	4	5
Participates well		Sometimes participates		Rarely participates
Externalizing Behaviors				
5. Physical Behavior Toward Others:				
1	2	3	4	5
Physically appropriate		Sometimes participates		Rarely participates
6. Verbal Behavior:				
1	2	3	4	5
Uses appropriate verbal behavior		Sometimes uses appropriate behavior		Rarely uses appropriate behavior
7. Physical Behavior Toward Materials or Property:				
1	2	3	4	5
Is consistently respectful of materials or property		Is sometimes respectful of materials or property		Is rarely respectful of materials or property
8. Out of Place:				
1	2	3	4	5
Remains in assigned area		Sometimes remains in assigned area		Rarely is in assigned area
Socialization				
9. Coping with Change:				
1	2	3	4	5
Handles change appropriately		Occasionally handles change appropriately		Rarely handles change appropriately
10. Adult Interactions:				
1	2	3	4	5
Seeks positive relationships		Sometimes seeks positive relationships		Rarely seeks positive relationships
11. Peer Interactions:				
1	2	3	4	5
Seeks positive relationships		Sometimes seeks positive relationships		Rarely seeks positive relationships
12. Projected Self Image:				

1	2	3	4	5
Speaks positively about self		Sometimes speaks positively about self		Rarely speaks positively about self

Note. BSC items originally published in Muyskens et al. (2007). Item 12, which is printed here for convenience, was excluded from the current study.

Student-level data. Student-level data were obtained from school records maintained with a computer program located in the central school office. These data included ODR, suspensions, and attendance records at the end of the school year of the screener administration. For the attendance data, tardies and absences were converted to a percent of days on-time relative to the number of days enrolled. As an indicator of end of year reading achievement, scores from the May administration of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), 6th Edition, were used. According to standard benchmark administration procedures, three grade-level Oral Reading Fluency (ORF) probes were administered to students in grades 1 through 5. Students received an ORF score for the number of words read correctly in 1 minute for each passage, and the median score of the three passages was used in analyses (Good, 2004). A meta-analysis of oral reading Curriculum Based Measurement (which includes the DIBELS ORF measure) found that R-CBM scores were moderately to highly correlated with other standardized measures of reading achievement (Reschly, Busch, Betts, Deno & Long, 2009). Across research studies, the DIBELS ORF multi-probe administration has been found to have strong test-retest reliability at every grade level, ranging from a low of .96 (in one study of 1st grade ORF) to high of .99 (in another study of 1st grade ORF and a study of 2nd grade ORF; Dynamic Measurement Group, 2008).

The instrument used as a measure of standardized achievement was the Criterion-Referenced Competency Test (CRCT). School children in grades 1 through 8 in the state of

Georgia take the CRCT in the spring of each school year. The CRCT was developed to measure Georgia Performance Standards (GPS) in the areas of reading, English/language arts, and mathematics. Additionally, children in grades 3 through 8 are tested in the areas of science and social studies. Internal consistency reliability coefficients (Cronbach‟s Alpha) were high for each subject area and across all grades included in the current study. Internal consistency for the Reading CRCT ranged from .86 (2nd and 5th grade) to .89 (3rd and 4th grade). The English Language Arts test revealed coefficient alpha‟s ranging from .89 (5th grade) to .90 (1st through 4th grade). Coefficients for the Mathematics test ranged from .91 (1st, 2nd, and 4th) to .93 (3rd grade). The Science test, which is only administered to grades 3 through 8, produced alpha‟s ranging from .90 (5th grade) to .92 (4th grade). Finally, the internal consistency of the Social Studies test was found to range from .91(4th grade) to .92 (3rd and 5th grade; Georgia Department of Education, 2008).

Teacher Questionnaire. The Teacher Questionnaire (Appendix A) was developed for use in the current study to examine teacher acceptability of the rating scales. Specific items inquire about the efficacy of the screening instruments, time to complete, and preference of screening instrument.

Procedure

The BESS Teacher Form and BSC were distributed to teachers around 10 weeks into the school year, after the first grading period in the fall of the school year. Teachers were given one week to complete the screeners and were provided with a substitute teacher for 30 minutes to allow them time to complete forms during regular school hours. The response rate for BESS and BSC was 99%, meaning that both screeners were completed for nearly every student enrolled in kindergarten through grade 5 at the time of data collection. Missing data were the result teachers

unintentionally completing only one or the other screener in reference to some children, as opposed to both. In addition to the two screening measures, each teacher was asked to complete a Teacher Questionnaire as a measure of social validity. Teachers were able to submit the questionnaire anonymously, which was considered essential to true opinions; however, it made following up with nonresponders impossible. The response rate for the Teacher Questionnaire was 40% (n = 10). To ensure accuracy of data entry, 15% of the data were checked for accuracy, and 99% of the data entered were found to be accurate. Errors were corrected in the data set.

Data Analyses

The current study examined the inter-rater agreement of the BESS and BSC. Inter-rater agreement differs from inter-rater reliability because the correlational relationship among scores of raters is not the variable of interest, but rather the placement of children as either above or below the cut score. In other words, the current study sought to identify which children are placed "at-risk" by each measure, and the variable was calculated as a percent agreement between the BSC and the BESS. For the purposes of this particular calculation, children identified as either elevated or extremely elevated on the BESS were grouped and compared to those children who surpassed the 5% and 20% cut-scores on the BSC. Only children for whom both screeners were completed were included in this analysis.

Individual children‟s scores on the BESS and BSC were analyzed to determine the concurrent validity of the two measures. Pearson product- moment correlations were calculated among the BSC and BESS scores.

The predictive validity of each screener was calculated using outcome measures gathered at the end of the school year in which the data collection took place. Specific outcome measures included number of ODR, days of suspension, ORF, and standardized achievement test scores as

measured by the CRCT. ORF probes and CRCT items varied by grade level, thus direct comparison of scores across grades was not possible. Therefore, ORF and CRCT scores were converted to *z*-scores, based on grade-level means and standard deviations. An alternative would have been to conduct each analysis by grade level; however, the *z*-score conversion was chosen to reduce the number of analyses and potential inflation of Type I error. The predictive validity of the screening instruments was calculated with Spearman"s rho correlations among screener scores and the behavioral outcome measures. As with previous screening studies, the Spearman"s rho was used because of the skewed distribution of screening scores and the ordinal nature of the scale (Muyskens et al., 2007). The academic measures (ORF and CRCT) were normally distributed in the sample, thus Pearson correlations were used with those variables.

The social validity of the BESS and BCS was examined by comparing the means of the ratings on the Teacher Questionnaire to determine if one measure was more preferred by teachers than the other.

Results

Descriptive data and correlations among variables may be found in Tables 19, 20, and 21. All achievement measures (CRCT areas and ORF) were moderately to highly correlated with each other. Significant but low negative correlations were found between ODR and suspensions and all areas of the CRCT. Attendance was found to have low significant correlations with the achievement measures, but was not correlated with the behavioral variables. Correlations between screeners and outcomes will be discussed in a later section.

Table 19

Descriptives of Major Variables

Variable	N	Mean	SD	Min	Max
BESS Total	492	24.52	14.61	0	66
BSC Total	492	19.82	8.56	11	49
ODR	492	0.43	1.30	0	13
Suspensions	492	0.05	0.29	0	3
Attendance	492	93.66	5.42	62.96	100.00
CRCT ELA					
Grade 1	82	814.00	20.72	774.00	858.00
Grade 2	84	831.85	31.16	773.00	910.00
Grade 3	67	829.49	35.80	752.00	930.00
Grade 4	86	827.10	26.25	761.00	900.00
Grade 5	69	837.45	22.30	770.00	884.00
CRCT Read					
Grade 1	82	826.16	25.79	774.00	920.00
Grade 2	84	849.02	32.10	776.00	920.00
Grade 3	67	830.24	31.96	755.00	920.00
Grade 4	86	824.45	25.02	776.00	870.00
Grade 5	69	826.55	19.33	785.00	867.00
CRCT Math					
Grade 1	83	821.87	26.88	770.00	883.00
Grade 2	85	840.08	28.74	784.00	930.00
Grade 3	67	826.25	41.82	732.00	953.00
Grade 4	86	830.58	41.33	740.00	936.00
Grade 5	69	853.72	45.86	767.00	990.00
CRCT Sci					
Grade 3	67	822.34	37.48	733.00	923.00
Grade 4	86	827.43	34.95	759.00	912.00
Grade 5	69	838.99	34.73	729.00	916.00
CRCT SS					
Grade 3	67	806.88	22.19	763.00	858.00
Grade 4	86	817.81	27.61	762.00	889.00
Grade 5	69	823.23	24.98	755.00	880.00
ORF					
Grade 1	80	50.86	28.75	9.00	150.00
Grade 2	80	100.04	35.98	14.00	202.00
Grade 3	66	110.96	38.47	22.00	208.00
Grade 4	84	123.02	39.48	12.11	214.00
Grade 5	68	129.18	32.69	17.00	194.00

Note. BESS and BSC = total raw score. ODR = office discipline referrals. Suspensions = number of days of suspension. Attendance = percent of days on time. ORF = oral reading fluency. CRCT = Criterion Referenced Competency Test scores.

Table 20

Correlations Between Major Variables

Variable	CRCT ELA[1]	CRCT Math[1]	CRCT Read[1]	CRCT Sci[1]	CRCT SS[1]	ODR	Susp	Att
CRCT ELA[1]	1.00							
CRCT Math[1]	.791**	1.00						
CRCT Read[1]	.786**	.745**	1.00					
CRCT Sci[1]	.734**	.790**	.747**	1.00				
CRCT SS[1]	.754**	.726**	.730**	.829**	1.00			
ODR	-.259**	-.218**	-.217**	-.359**	-.356**	1.00		
Susp	-.165**	-.125**	-.148**	-.215**	-.175**	.404**	1.00	
Att	.115*	.126*	.119*	.169*	.210**	-.065	-.054	1.00
ORF[1]	.675**	.571**	.681**	.608**	.600**	-.176**	-.143**	.086

Note. ** = Significant at 0.01 level. * = Significant at 0.05 level. ODR = office discipline referrals; Susp = Days of suspension; Att = percent of days on time; ORF = oral reading fluency. Variables indicated by [1] are z-scores calculated with grade level means and standard deviations. Correlations with ODR, Susp and Att were calculated with Spearman‟s Rho correlation coefficient. All other coefficients are Pearsons.

Reliability

Before examining the properties of the BSC and the BESS, appropriate cut-scores for the BSC were developed. The published cut-score for the BSC, which was designed to capture the top 5% of the students, is 36; however, that cut-score was developed with all 12 items and this study only used the first 11. Therefore, a new 5% BSC cut-score was developed for use in this study following the same guidelines as the original authors (Muyskens, Marston, & Reschly, 2007). The resulting 5% cut score for this study was determined to be 39. Additionally, a 20% cut score was calculated with this sample and was determined to be 27. The utility of this new "at-risk" category was evaluated.

Inter-rater agreement, or the degree to which the two screener scores agreed with at-risk placement, was calculated as percent agreement. Inter-rater agreement between the 5% category

of the BSC and the extremely elevated category of the BESS was found to be 27.1%. Further, the number of students identified in the extremely elevated risk category was 37 (7.5%), which is similar to the number of students identified by the BSC (*n* = 24; 5%). Inter-rater agreement was 34.2% between the 20% BSC and BESS elevated categories. This percentage is considerably lower than expected given that both were completed by the same teacher in reference to the same children. Moreover, both the BESS elevated and 20% BSC categories identified exactly 102 children; however, they only identified 52 of the same children.

Validity

The external properties of the BSC and BESS were examined with concurrent and predictive validity measures. Concurrent validity of the BESS and BSC was calculated with a Pearson Correlation, and revealed a strong relationship between teacher ratings on the BSC and teacher ratings on the BESS (*r* = 0.854). Predictive validity (presented in Table 21) between the screeners (BSC and BESS) and the behavioral outcome measures (ODR, suspensions, and attendance) was calculated with the Spearman‟s rho. Pearson product-moment correlation coefficients were used for the academic variables (ORF and CRCT). Results indicated that the BSC and BESS each correlated with every outcome measure. The BESS was the most predictive of the achievement variables, including ORF (*r* = -0.393), CRCT English/Language Arts (*r* = -0.525), CRCT Math (r = -0.485), CRCT Reading (*r* = -0.509), CRCT Science (*r* = -0.480), CRCT Social Studies (*r* = -0.498), and attendance (*r* = -0.154). The BSC was most predictive of the behavioral measures, which included ODR (*r* = 0.400) and suspensions (*r* = 0.221).

Table 21
Correlations Among Screeners and Outcome Measures

Variable	BSC	BESS TF
ORF[1]	-.314**	-.393**
CRCT ELA[1]	-.442**	-.525**
CRCT Reading[1]	-.425**	-.509**
CRCT Math[1]	-.402**	-.485**
CRCT Science[1]	-.440**	-.480**
CRCT SS[1]	-.459**	-.498**
ODR	.400**	.378**
Suspensions	.221**	.217**
Attendance	-.104*	-.154**

Note. ** = Significant at 0.01 level. BESS TF = Bess Teacher Form; ORF = oral reading fluency as measured by DIBELS; CRCT ELA = CRCT English/Language Arts; CRCT Social Stu = CRCT Social Studies; ODR = office discipline referrals. Variables indicated by [1] are z-scores calculated with grade level means and standard deviations. Correlations with ODR, Susp and Att were calculated with Spearman‟s Rho correlation coefficient. All other coefficients are Pearson‟s.

The social validity of each instrument was measured with the Teacher Questionnaire. Results indicated that each was well-liked by the teachers (see Table 22). When asked which measure they would choose to use again, half of the teachers chose the BSC and half chose the BESS (n = 10). The BSC had slightly higher averages than the BESS for each item. In regard to the screeners identifying problem behavior (Item 1), 72.7% of teachers felt that the BSC was appropriate, and 63.6% felt that the BESS was appropriate. Both the BSC and BESS were rated by 72.7% of teachers as being effective measures of the problem behaviors they see in their classrooms. When asked if they would recommend the measure to other educators, 72.7% agreed that they would recommend the BSC, and 63.6% would recommend the BESS. Both the BSC and the BESS were rated by 72.7% as being efficient. The “time to complete” item seemed to have been misinterpreted by some teachers. While most replied that each took them between 1

and 5 minutes to complete, one teacher wrote “60” and others wrote “10 “ and “15.” Hence, this item was not interpreted.

Table 22
Teacher Questionnaire Responses

	BSC		BESS	
Item	M	SD	M	SD
Identifying problem behavior	3.64	1.43	3.45	1.04
Behavior in my classroom	3.73	0.79	3.64	0.92
Would suggest this Screener	3.55	1.13	3.45	0.82
Is Efficient	3.64	1.43	3.45	1.04

Note. Full items can be found in Appendix A. Means are based on a 5-point Likert-type scale with higher numbers representing more positive thoughts towards the screening instruments.

Discussion

Response to intervention (RTI) is a “hot topic” in school psychology and special education, with well-respected proponents (Fletcher & Reschly, 2005; Gresham et al., 205) and opponents (Hale, Naglieri, Kaufman, & Kavale, 2004). Some opponents of RTI feel that the process lacks empirical evidence as a component of special education eligibility determination. Others, who promote the use of RTI, view it as protecting against overrepresentation in special education and bias in the referral process. Central to RTI is the tenet of a multi-tiered approach of resource allocation, with increasing interventions matched to children with increasing needs. Tier 1 of the model consists of school-wide initiatives to provide academic and behavioral supports. A key component of Tier 1 is the use of universal screening to determine which children are in need of more individualized or intensive supports (Tiers 2 and 3). While academic universal screening, for reading in particular, has advanced steadily over the past decade, behavioral screening is still in the beginning stages. Consequently, there are few behavioral

screening instruments currently used in schools. Two such instruments, the Behavioral Screening Checklist (BSC) and the Behavioral and Emotional Screening System (BESS), were reviewed in the current study.

The review process of the measures began with an investigation of the inter-rater agreement. Rather than a correlation of scores, the measures were compared in terms of the actual children identified by each measure as being at-risk for behavioral difficulties. Results indicated that 34.2% of the same children were identified by the „at-risk" categories of the measures, with the both identifying 102 children. When the extremely elevated category of the BESS was compared to the traditional cut score of the BSC (5%), there was 27.1% agreement in the children identified. Based on a three-tier RTI model, about 5% of children are typically receiving Tier 3 level of supports and 15% typically receive Tier 2 supports. When combined, the elevated and extremely elevated categories of the BESS identified a total of 28% of the students as potentially needing further intervention. The traditional 5% BSC cut-score coincides with the children likely in need of Tier 3 levels of supports, whereas the BESS over identifies even those children in need of Tier 2 supports. As stated in the BESS Manual:

> Cut scores were developed to maximize the likelihood of identifying children and adolescents who truly have behavioral and emotional problems, while accepting a somewhat greater (but still low) risk of referring children and adolescents who do not have behavioral and emotional problems (Kamphaus & Reynolds, 2007, p.48).

The authors of the BESS proposed a screening system in which the BESS is used to identify (and potentially over identify) children at risk for emotional and behavioral difficulties, and a follow-up broad-band behavioral rating scale, like the BASC-2 is then used for the children who were identified on the BESS. The BESS system has an optional intervention manual, to accompany

the BESS screening forms, that recommends prescribed interventions based on BESS scores. While this system of identification may be comprehensive and systematic, it has the potential to be time-consuming (completing a lengthy rating scale for many children who do not have emotional and behavioral difficulties) and expensive (the BESS and BASC-2 forms are proprietary). Alternatively, the cut score of the BSC is set to only identify the number of children coinciding with those typically needing Tier 3 supports, which may be an underrepresentation of the children in need of Tier 2 supports. A more comprehensive option, as examined in the current study, may be to adjust the cut score of the BSC to include the top 20% of children, which is the number of children that typically need Tier 3 *and* Tier 2 behavioral interventions. The BSC is non-proprietary; however, requires skill in the planning of appropriate interventions. Regardless of the screening measure chosen, intervention integrity is critical to student success. That is, the identification of children in need of intervention is only one step in the process of improving behavior. Implementing interventions with integrity is essential to improving behavior, and is not dependent on the screening measure used to identify children (Lochman & Gresham, 2009).

When a new cut-score was developed to accommodate the printing error of the BSC, a higher cut score was found than the one previous established (see Muskens et al. 2007). In other words, the cut score developed within the rural southern school district used for the current study was three points higher than the cut score developed within the Minneapolis Public School District. One explanation for the difference in cut scores across settings may be due to the difference in severity of behaviors exhibited across those settings. A less speculative investigation of the differences may be a direction for future research. The current study underscores the importance of local norm setting, a concept typically associated with CBM. Like

reading and math, behavior screening cut-scores or benchmarks may also be best determined with local norms.

In addition to having a strong relationship with each other ($r = 0.867$), the screeners each correlated with every outcome measure (ODR, suspensions, attendance, every CRCT area, and ORF). Each correlation was in the expected direction (i.e., behavior was negatively correlated with achievement and positively correlated with the behavioral outcomes). The BESS was slightly more highly correlated with attendance and the achievement measures (CRCT and ORF), while the BSC was somewhat more highly correlated with the behavioral outcomes (ODR and suspensions). Additionally, the performance of the screeners in the current study is similar to the findings in previous validation studies of the screeners (e.g., Kamphaus & Reynolds, 2007; Kamphaus et al., 2007; Muyskens et al., 2007).

The BSC performed somewhat better than the BESS on the Teacher Questionnaire, with higher means on each item. However, both screeners were equally recommended by the teachers for future use. The one item that may have shown a difference between the screeners, the time of completion, was not interpretable. Based on the number of items on the screeners (11 on the BSC and 27 on the BESS), one may surmise that the BSC takes less time to complete. One limitation of the current study is that only 10 teachers completed and returned the Teacher Questionnaire. Because anonymity was considered important, we were unable to identify which teachers did and did not complete the questionnaire to follow-up with those who had not. Future studies should strive to gain social validity from all parties involved.

Finally, office discipline referrals, suspensions, and attendance are problematic because they are greatly skewed variables. The majority of school children receive no office discipline referrals or suspensions, and are present and on time for school most days. However, even

though the nature of the variables makes statistical analyses difficult, ODR and suspensions are usually the only behavior data available to researchers. Future research may examine classroom behavioral differences via observation to further examine the validity of groupings attained from the BESS or BSC screeners.

Future research should seek to include a broader sample. The current sample included nearly every child in a rural elementary school; however, when examining the grade-level data, this sample size is small. Grade-level examinations would alleviate the need to convert ORF and CRCT to z-scores. Future researcher should include multiple schools to increase the grade-level sample size.

When deciding on a behavior screener, schools should consider the psychometric properties of the screener, including the population on which the screener was normed. As demonstrated in this study, norms may vary considerably by region. In addition to the psychometric properties, cost of administration and time to administer, score, and interpret (Lane et al., 2007). Preference should be given to a screener with sound psychometric properties that is of low cost and is quickly administered.

References

Batsche, G. M., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *Response to Intervention: Policy Considerations and Implementation*: National Association of State Directions of Special Education, Inc.

Bergan, J. R. (1977). *Behavioral Consultation*. Columbus, OH: Charles E. Merrill.

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral Consultation and Therapy*. New York: Plenum Press.

Burns, M. K., Deno, S. L., & Jimerson, S. R. (2007). Toward a Unified Response-to-Intervention Model. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *Handbook of Response To Intervention: The Science and Practice of Assessment and Intervention* (pp. 428-440). New York: Springer.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Donovan, M. S., & Cross, C. T. (Eds.). (2002). *Minority students in special education*. Washington, DC: National Academy Press.

Fletcher, J. M., & Reschly, D. J. (2005). Changing procedures for identifying learning disabilities: The danger of perpetuating old ideas. *The School Psycholgist, 59*(1), 10-15.

Georgia Department of Education. (2008). *Validity and reliability for the 2008 CRCT*. Atlanta, GA: Assessment Research and Development Division.

Gresham, F. M., Reschly, D. L., Tilly, W. D., Fletcher, J., Burns, M., Christ, T., et al. (2005). Comprehensive evaluation of learning disabilities: A response to intervention perspective. *The School Psychologist, 59*(1), 26-30.

Hale, J. B., Naglieri, J. A., Kaufman, A. S., & Kavale, K. A. (2004). Specific learning disabilities classification in the new individuals with disabilities education act: The danger of good ideas. *The School Psychologist, 58*(1), 6-13, 29.

Harrison, P. L. (2009). Preschool Assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The Handbook of School Psychology* (pp. 247-268). Hoboken, NJ: John Wiley and Sons, Inc.

Huizinga, D., & Jakob-Chien, C. (1998). The Contemporaneous Co-Occurrence of Serious and Violent Juvenile Offending and Other Problem Behaviors. In R. Loeber & D. P. Farrington (Eds.), *Serious and Violent Juvenile Offenders* (pp. 47-67). Thousand Oaks, CA: Sage Publications.

Jimerson, S. R., Burns, M. K., & VanDerHeyden, A. M. (2007). Response to Intervention at School: The Science and Practice of Assessment and Intervention. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *Handbook of Response To Intervention: The Science and Practice of Assessment and Intervention* (pp. 3-9). New York: Springer.

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 Behavioral and Emotional Screening System Manual*: Pearson.

Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & VanDeventer, M. C. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement, 67*, 342-356.

Kavale, K. A. (2001). Decision making in special education: The function of meta-analysis. *Exceptionality, 9*(4), 245-268.

Kavale, K. A., Kaufman, A. S., Naglieri, J. A., & Hale, J. B. (2005). Changing procedures for identifying learning disabilities: The danger of poorly supported ideas. *The School Psycholgist, 59*(1), 16-25.

Kovaleski, J. F. (2007). Potential Pitfalls of Response to Intervention. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *Handbook of Response To Intervention* (pp. 80-89). New York: Springer.

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*(4), 209-222.

Lochman, J. E. & Gresham, F. M. (2009). Intervention development, assessment, planning, and adaptation: The importance of developmental models. In M. J. Mayer, J. E. Lochman, & R. Van Acker (Eds.), *Cognitive-behavioral interventions for emotional and behavioral disorders: School-based practice* (pp. 29-57). New York, NY, US: Guilford Press.

MacMillan, D. L., Gresham, F. M., & Bocian, K. M. (1998). Curing mental retardation and causing learning disabilities: Consequences of using various WISC-III IQs to estimate aptitude of Hispanic students. *Journal of Psychoeducational Assessment, 16*(1), 36-54.

Marston, D., Reschly, A. L., Lau, M. Y., Muyskens, P., & Canter, A. (2007). Historical Perspectives and Current Trends in Problem Solving: The Minneapolis Story. In D. Hagar, J. Klinger & S. Vaughn (Eds.), *Evidence-Based Reading Practices for Response to Intervention* (pp. 265-285). Baltimore: Brookes.

Muyskens, P., Marston, D., & Reschly, A. L. (2007). The use of response to intervention practices for behavior: An examination of the validity of a screening instrument. *The California School Psychologist, 12*, 31-45.

O'Shaughnessy, T. E., Lane, K. L., Gresham, F. M., & Beebe-Frankenberger, M. E. (2002). Students With or At-risk for Learning and Emotional-Behavioral Difficulties. In K. L. Lane, F. M. Gresham & T. E. O'Shaughnessy (Eds.), *Interventions for Children With or At-risk for Emotional and Behavioral Disorders*. Boston: Allyn & Bacon.

President‟s Commission on Excellence in Special Education (2002). *A new era: Revitalizing special education for children and their families.* Washington, DC: US Department of Education.

Reschly, D. J., & Bergstrom, M. K. (2009). Response to Intervention. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of School Psychology* (4th ed., pp. 434-460). Hoboken, NJ: John Wiley & Sons, Inc.

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

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

Schrank, F. A., Teglasi, H., Wolf, I. L., Miller, J. A., Caterino, L. C., & Reynolds, C. R. (2005). American academy of school psychology reply to response-to-intervention perspective. *The School Psycholgist, 59*(1), 30-33.

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*(2), 193-223.

Sprague, J., Walker, H., Golly, A., White, K., Myers, D. R., & Shannon, T. (2001). Translating research into effective practice: The effects of a universal staff and student intervention

on indicators of discipline and school safety. *Education & Treatment of Children, 24*(4), 495-511.

Sprague, J. R., & Walker, H. (2000). Early identification and intervention for youth with antisocial and violent behavior. *Exceptional Children, 66*(3), 367-379.

Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy, 24*(1), 23-50.

Sugai, G., & Horner, R. (2008). What we know and need to know about preventing problem behavior in schools. *Exceptionality, 16*, 67-77.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., et al. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2*, 131-143.

Sugai, G., Horner, R. H., Gresham, F. M., Shinn, M. R., Walker, H. M., & Stoner, G. (2002). Behaviorally Effective School Environments. In *Interventions for academic and behavior problems II: Preventive and remedial approaches.* (pp. 315-350). Washington, DC US: National Association of School Psychologists.

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 & Behavioral Disorders, 8*(2), 94.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18*(3), 137-146.

Wagner, M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, W. C. (2005). The Children and Youth We Serve: A National Picture of the Characteristics of Students With

Emotional Disturbances Receiving Special Education. *Journal of Emotional and Behavioral Disorders, 13*(2), 79-96.

Walker, H. M. (2000). Investigating school-related behavior disorders: Lessons learned from a thirty-year research career. *Exceptional Children, 66*(2), 151-161.

Walker, H. M., Colvin, G., & Ramsey, E. (1995). *Antisocial behavior in school: Strategies and best practices*. Belmont, CA US: Thomson Brooks/Cole Publishing Co.

Walker, H. M., Horner, R. H., Sugai, G., & Bullis, M. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4*(4), 194-209.

Walker, H. M., Shinn, M. R., O'Neill, R. E., & Ramsey, E. (1987). A longitudinal assessment of the development of antisocial behavior in boys: Rationale, methodology, and first year results. *Remedial and Special Education, 8*(4), 7-16.

Warren, J. S., Bohanon-Edmonson, H. M., Turnbull, A. P., Sailor, W., Wickham, D., Griggs, P., et al. (2006). School-wide positive behavior support: Addressing behavior problems that impede student learning. *Educational Psychology Review, 18*, 187-198.

Appendix A

Teacher Questionnaire

Recently you were asked to complete behavioral screening measures of each child in your classroom. Now we would like to know what you thought about each measure you completed. For your convenience, each measure has been attached and labeled either A or B. Please refer to them as necessary to reply each of the following statements.

Please circle the number corresponding to your choice for each measure (A **and** *B)*

Item		Measure	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree
1	This screener is an effective measure for identifying a child"s problem behavior	**A**	1	2	3	4	5
		B	1	2	3	4	5
2	This screener is an effective measure for identifying the type of behaviors I see in my classroom	**A**	1	2	3	4	5
		B	1	2	3	4	5
3	I would suggest the use of this measure to other educators	**A**	1	2	3	4	5
		B	1	2	3	4	5
4	This measure is efficient	**A**	1	2	3	4	5
		B	1	2	3	4	5

5. On average, how much time per child did it take you to complete Measure A? _____
Measure B? _____

6. If you were able to choose one of the measures to use, which would you choose?
Please circle: **A** or **B**

7. Any additional comments that you would like to add?