

Running head: MEASURES OF EXTERNALIZING BEHAVIOR AND ACHIEVEMENT

CONVERGENT VALIDITY OF DIRECT AND INDIRECT MEASURES OF  
EXTERNALIZING BEHAVIOR SYMPTOMS ON ACADEMIC ACHIEVEMENT

A DISSERTATION

SUBMITTED TO THE FACULTY

OF

THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY

OF

RUTGERS,

THE STATE UNIVERSITY OF NEW JERSEY

BY

NATALIE MANNO

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY

AUGUST 2022

APPROVED:

\_\_\_\_\_  
Linda A. Reddy, Ph.D.

\_\_\_\_\_  
Adam Lekwa, Ph.D.

DEAN:

\_\_\_\_\_  
Ryan Kettler, Ph.D.

Copyright 2022 by Natalie Manno

### Abstract

The present study examined the convergent validity of observer ratings of externalizing behavior symptomatology using the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2004) and teacher ratings of externalizing behavior symptoms on the Behavior Assessment System for Children, Third Edition (BASC-III; Reynolds, Kamphaus, & Vannest, 2015) on student academic achievement measured by the Woodcock Johnson Tests of Achievement, Fourth Edition (WJ-IV ACH; Schrank, Mather, & McGrew, 2014). The study included a sample of 197 K-5th grade students with or at-risk for externalizing behavior disorders from 36 schools. Teacher ratings of student behavior symptomatology, namely aggression ( $r = -.16, p = .02$ ), and school problems ( $r = -.36, p < .001$ ), were significantly related to student achievement. Direct observation data did not significantly relate to student achievement. A series of multilevel models were conducted to examine the convergent validity of the BOSS and BASC-3 externalizing behavior scores on academic achievement. Variables were related, but not potent predictors of achievement for these children. Research limitations and implications of findings for school psychology practice are presented.

### Acknowledgments

I am fortunate to have such a long list of individuals to thank for my accomplishments in my dissertation and doctoral studies. First, I must thank my dissertation chair and faculty mentor, Dr. Linda Reddy, whose commitment to my academic and professional success has enabled me to complete this project. It has been a pleasure to study school psychology under someone as devoted to the field as Dr. Reddy. Thank you to Dr. Adam Lekwa, my committee member, who has been willing to offer his time and has challenged me to think critically about my study and apply the findings in ways that are meaningful to the field of research. I must also thank the Rutgers, GSAPP faculty and the close friends I have made it through this program with. Words cannot describe how transformative these four years have been for me, both professionally and personally, thanks to you.

To my family, your unwavering belief that I would achieve this milestone is the reason for my success. To the ‘Dr. Mannos’ that came before me, Uncle Chris and Uncle Andy, and my incredible grandparents who have always pushed me to achieve goals I once thought were impossible, thank you for your love and guidance. To my siblings, Nick and Julia, thank you for lending an ear and offering distraction from my stress when I needed it most. To my fiancé Carmine, thank you for being my partner throughout this long and arduous journey. These years have been so special, albeit challenging, and I am looking forward to moving into the next chapter of life with you. Finally, to my wonderful parents who have believed in me when I struggled to believe in myself and went above and beyond to meet every single one of my needs throughout my twenty yearlong educational journey - I am forever indebted to you. You have been by my side for every moment, small and large, and this accomplishment is truly for all of us to share.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	vi
INTRODUCTION .....	1
Overview of Externalizing Behavior .....	3
Hyperactivity .....	4
Aggression .....	5
Conduct Problems .....	5
Assessment of Externalizing Behavior .....	6
Systematic Direct Observations .....	6
Direct Behavior Ratings .....	7
Relations Between Externalizing Behavior and Achievement .....	7
Rationale for Current Study .....	12
Purpose .....	12
METHODOLOGY .....	14
Participants .....	14
Instrumentation .....	15
Procedures .....	17
Data Analysis Plan .....	19
RESULTS .....	20
Relation Between Student Behavior and Achievement .....	21

Prediction of Student Achievement.....25

DISCUSSION.....27

    Limitations.....33

    Future Directions .....34

    Implications for School Practice .....35

    Conclusions .....36

REFERENCES .....38

## LIST OF TABLES

Tables	Page #
1. Behavioral Observation of Students in Schools, Behavior Assessment System for Children, Third Edition, and Woodcock Johnson Tests of Achievement, Fourth Edition Descriptive Statistics.....	21
2. Correlations of Student Behavior and Achievement.....	24
3. Multilevel Modeling to Predict Student Academic Achievement ... ..	26

## **Convergent Validity of Direct and Indirect Measures of Student Externalizing Behavior Symptomatology on Academic Achievement**

### **Introduction**

Student externalizing behavior has been identified as a critical factor for academic achievement (e.g., Barker et al., 2010; Lewis et al., 2017; Nelson et al., 2004). Externalizing behavior refers to observable behaviors interpreted as students acting out of their environment, including hyperactivity, aggression, and conduct problems (Liu, 2004; Reynolds, Kamphaus, & Vannest, 2015). Over several decades, the association between student externalizing behavior and academic achievement has been explored. As researchers have gained an understanding of the association between externalizing behavior and achievement, research needs have shifted toward designing assessment measures and promoting behavioral interventions to promote student academic success.

Behavioral assessment represents an area of research that is principal to theory and intervention. Historically, studies demonstrating the link between behavior and achievement have employed a variety of behavior assessment methods (e.g., parent interview, teacher interview, rating scales, interviews, and systematic direct observation; Fergusson et al., 1993, Frick et al., 1991; Hinshaw, 1992b; Lane et al., 2007). Findings from these studies demonstrated the association of behavior and achievement and played an important role in informing theoretical conceptualizations of the link (Hinshaw, 1992a; Hinshaw, 1992b). Aside from research, practitioners in school contexts use similar methods of behavioral assessment to inform the selection of and evaluating the effectiveness of interventions (Christ, Riley-Tillman, & Chafouleas, 2009). Thus, behavioral assessment is a key tool for developing and evaluating interventions in the school setting.



Standardized behavioral assessment has been relied upon by researchers and school practitioners as a means for identifying and evaluating student externalizing behavior problems. In one survey, school clinicians reported employing direct observations and informant rating scales for 60 to 90 percent of special education assessment cases (Christ, Riley-Tillman, & Chafouleas, 2009). Further, use of a minimum of one structured behavioral observation has become a requirement for eligibility assessment under special education code (NJ 6A:14, 2016). Thus, understanding how these assessments relate to each other and to student achievement could have significant implications on early identification, prevention, and intervention for students with or at risk for externalizing behavior disorders.

Behavioral measurement includes broad-band assessments like systematic direct observation and direct behavior rating scales. Systematic direct observation (SDOs) measures like the Behavioral Observation of Students in Schools (BOSS) can be a useful assessment method in the school setting, as the procedures can be effectively standardized to reduce subjectivity and objectively quantify a student's behavior (Christ et al., 2009). However, the literature has raised considerable limitations associated with SDOs including accessibility for teachers to conduct the measure during instruction, and alteration of student behavior because of the presence of an outside observer (Riley-Tillman et al., 2008). Researchers and practitioners have emphasized the value of direct behavior rating measures like the Behavior Assessment System for Children, Third Edition (BASC-3; Reynolds, Kamphaus, & Vannest, 2015), that allows informants to report on students' behavior retrospectively, making them more feasible than other methods of behavioral assessment. With measures like the BASC-3, teachers can complete the assessment after instruction with a contextual understanding of behaviors. Further, direct behavior ratings provide detailed results that examine the presence of a broad variety of

behavioral symptoms consistent with theory (Christ et al., 2009; Reynolds, Kamphaus, & Vannest, 2015).

Behavioral measures like the BASC-3 and BOSS have been validated as effective measures of student externalizing behavior (Reynolds, Kamphaus, & Vannest, 2015; Volpe, DiPerna, Hintze, & Shapiro, 2004), however there are critical components of the measures that have not been explored. The current study examined the convergent validity of these two measures of student externalizing behavior. Further, the study will examine the convergent validity of each behavioral measure with the Woodcock-Johnson Measures of Achievement, *Fourth Edition* (WJ-IV ACH; Schrank, Mather, & McGrew, 2014), a standardized assessment of student achievement. It was expected that the findings from this study would highlight the utility of the BASC-3 and BOSS as measures for assessing student's externalizing behavior in the classroom setting and further validate their association with achievement.

### **Overview of Externalizing Behavior**

Externalizing behavior is defined as a cluster of observable behaviors that are often interpreted as the child acting out on their environment (Liu, 2004). A recent study from the 2016 National Survey of Children's Health found that 7.4% of children aged 3 to 17, had a current behavioral or conduct problem and were receiving professional mental health treatment (Ghandour et al., 2019). These externalizing behaviors include hyperactivity, aggression, and conduct problems (Reynolds, Kamphaus, & Vannest, 2015). Children who exhibit significant externalizing behavior and/or Attention-Deficit/ Hyperactivity Disorder (ADHD) may receive special education services through the Individuals with Disabilities Education Act (IDEA; 2004) under the classifications Emotional Regulation Impairment and Other Health Impairment (OHI). IDEA states that to qualify for special education, a student must

have a disability and evidence of educational impact caused by that disability (IDEA, 2004). According to special education code, this must be demonstrated using at least two methods behavioral assessment, including at least one structured observation (NJ 6A:14, 2016). Thus, DBRs and SDOs are often used on students who exhibit high levels of externalizing behavior at school.

### *Hyperactivity*

According to the Diagnostic and Statistical Manual for Mental Disorders, Fifth Edition (DSM-5) criteria for ADHD, hyperactivity is characterized by fidgeting and squirming in seat, running or climbing in situations where it is not appropriate, leaving seat during class or another time when this is not appropriate, talking excessively, etc. (APA, 2013). The National Survey of Children's Health estimated that approximately 9.4% of children aged 2 to 17 have been diagnosed with ADHD in the United States (Danielson et al., 2018). These problems may impact student academic performance in several ways. Children who display hyperactivity may be disruptive in class, reducing the amount of instructional time they receive. They may also struggle to focus during instruction, impacting their learning. Further, it is estimated that the co-occurrence of ADHD and reading disabilities is between 25 and 48 percent (Froehlich et al., 2018; Willcutt & Pennington, 2000; Willcutt et al., 2010).

Researchers who have studied the relationship between externalizing behaviors and academic underachievement often highlight the methodological challenge of disentangling externalizing behaviors from attention problems in their findings (e.g., Barriga et al., 2002; Frick et al., 1991). This seems to be due, in part, to the common comorbidity of hyperactivity, inattention, and other externalizing problems. In some studies, researchers suggest attention problems are a mediator of the relationship between externalizing behavior and academic

underachievement, positing that the association can be explained by high rates of co-occurring behavior and attention problems seen in ADHD and other externalizing behavior disorders, though the nature of the relationship remains unclear (Carroll et al., 2005). In a study that examined data from a United Kingdom (UK) national survey of child mental health, Carroll (2005) stated that attention problems with or without hyperactivity represented the most salient risk factor for academic problems, though hyperactivity did not significantly contribute to academic underachievement on its own.

### ***Aggression***

Aggression is the tendency to act in a threatening or contentious manner toward others, and includes both physical and verbal actions (Reynolds, Kamphaus, & Vannest, 2015). Verbal aggression may include temper tantrums or verbal arguments or fights, while physical aggression may include hitting, kicking, or biting (APA, 2013). These behaviors have been identified as related to long-term academic outcomes in several studies.

### ***Conduct Problems***

Conduct problems refer to a tendency break rules, defy authority figures, and engage in property destruction (Reynolds, Kamphaus, & Vannest, 2015). Both aggression and conduct problems in childhood are related to poor outcomes in adulthood, including delinquency, school dropout, and antisocial behavior (e.g., Farrington, Loeber & Van Kammen, 1990; Hinshaw, 1992a).

Aggression and conduct problems externalizing behaviors have been associated with academic difficulties in several studies. Frick et al. (1999) used a clinic-referral sample of 177 7- to 12-year-old boys to examine the relationship between behavior problems and academic achievement. Results indicated that boys who met criteria for ADHD and/or Conduct Disorder

(CD) showed significantly higher rates of academic underachievement, measured by a standardized screener, than boys in the control group. Further, academic achievement appeared to be significantly associated with ADHD, even after controlling for boys who met criteria for both ADHD and CD (Frick et al., 1991). Farrington (1979) found that low vocabulary status at age eight to ten predicted self- and teacher- reported delinquency six years later. Similarly, Kupersmidt and Coie (1990) found that peer ratings of aggression predicted school dropout seven years later. Further, Carroll (2005) found that both CD and ADHD were associated with literacy difficulties. This research utilizes several methods of measurement for externalizing behavior including teacher, peer, and parent reports, direct observation, and clinical diagnostic rating scales for emotional and behavior disorders. Despite the various methods of assessment used, there is limited research related to differentiating different measures' relation to achievement.

### **Assessment of Externalizing Behavior**

Assessment of externalizing behavior serves a key function in research and practice. Methods of measurement allow clinicians and researchers to quantify and critically evaluate what is observed in a fashion that is standardized across observers (Christ et al., 2009). While several methodologies of measuring behavior are available, systematic direct observations and direct behavior ratings represent two predominant methods of measurement (Riley-Tillman et al., 2008).

#### ***Systematic Direct Observation***

Systematic direct observation measures like the BOSS enable trained observers to record student behavior in a systematic fashion that greatly reduces subjectivity. The BOSS is based on a set of clearly defined behavior categories that are tracked using fixed interval

recording (Shapiro, 2004). While subject to some degree of subjectivity and observer bias, this type of measure can yield valuable, quantitative data about student classroom behavior (Volpe, DiPerna, Hintze, & Shapiro, 2005).

### ***Direct Behavior Rating***

Direct behavior rating like the BASC-3 require an informant (e.g., teacher, parent) to rate the extent to which they have observed the student display a list of behaviors consistent with theoretical conceptualization of behavioral symptomatology (Reynolds, Kamphaus, & Vannest, 2015). Informants can complete rating scales with contextual knowledge of the student's behavior. While this can be seen as a strength, it is possible that the data generated from a direct behavior rating is prone to errors in reporting or bias due to the retrospective nature of the measure (Christ et al., 2009).

### **Relations Between Externalizing Behavior and Achievement**

Academic underachievement and externalizing behavior present major problems for school-aged children that are widespread and endure over time, having serious social, emotional, and financial implications on students and their families (e.g., Farrington, Loeber & Van Kammen, 1990; Hinshaw, 1992a; Lane et al., 2007). Externalizing behaviors (e.g., hyperactivity, aggression, conduct problems) are known to have lasting academic and social impacts, driving poor outcomes including school failure, substance abuse, risky behaviors, and antisocial behavior in adulthood (Farrington, 2004; Liu, 2004; Nelson et al., 2004). Academic underachievement and learning deficits require increased support from school professionals, special education evaluations, and often special education services, all of which incur additional expenses for schools. In addition to the significant financial and societal burden of academic underachievement and behavior problems, these are often associated with poor peer

and teacher outcomes including loss of instructional time and teacher burnout (e.g., Arnold, 1997; Hastings & Bham, 2003; Henry et al., 2012). For example, in a recent survey of 1,400 general education and special education teachers in elementary schools, teachers reported losing an average of two and a half hours of instructional time each week due to disruptive student behaviors (Education Advisory Board, 2019).

In high-poverty school districts, teacher burnout and attrition are prevalent, making the need for understanding such classroom stressors highly important (Boyd et al., 2007). Externalizing behaviors are more likely to endure over time compared to other types of childhood behavior problems, and they are not as amenable to intervention (e.g., Barker et al., 2010; Hinshaw, 1992a; Lewis et al., 2017). Thus, negative patterns of behavior and academic problems can become firmly established in the absence of effective prevention and intervention efforts, leading to long-term impacts in the aforementioned areas (Lane et al., 2007). The understanding of negative outcomes associated with both academic underachievement and externalizing behaviors (e.g., Barker et al., 2010; Lewis et al., 2017; Nelson et al., 2004), has led to a body of research examining the relationship between these factors. Early investigations asserted that academic underachievement and externalizing behavior co-occur at rates far higher than chance (Hinshaw, 1992b), and the relationship is often enduring (e.g., Fergusson et al., 1993; Frick et al., 1991; Lane et al., 2007). Several on this relationship examined clinical samples and relied on clinical assessment measures and parent reports in their research (e.g., Fergusson et al., 1993; Fergusson et al., 1997; Rapport et al., 1999). Further, studies that have used measures commonly used in schools focused primarily on either standardized academic achievement assessment or reliable behavior assessment methods, rather than both.

Understanding the relationship using measures that are accessible to and known by school professionals could have important implications for school-based practitioners. For example, data on the interrelation of the Woodcock Johnson IV Tests of Achievement (WJ-IV ACH), the BASC-3 and the BOSS may generate insights on possible individual and contextual factors that contribute to these important constructs. Likewise, information could be used to inform prevention, identification, and intervention efforts in schools. Given the BOSS and BASC-3 are regarded for their utility in informing intervention, the presence of an association with achievement would serve as a useful criterion for validating this relationship.

In an international review of epidemiological reports, Hinshaw (1992b) asserted that co-occurrence of externalizing behavior problems and academic underachievement exists at a rate far higher than chance, estimating between 10% and 50% based on clinical samples of children and adolescents. Hinshaw (1992b) theorized four possible causal relationships between academic underachievement and externalizing behavior: underachievement leads to externalizing behavior, externalizing behavior leads to underachievement, both constructs lead to each other, or a third underlying variable results in problems in both domains. To explore these theorized mechanisms, Hinshaw's (1992b) meta-analysis examined the relationship between academic underachievement and externalizing behavior. Studies focused on elementary grades and secondary grades were examined and discussed separately. In one study of these relationships in early and middle childhood, teacher ratings on internalizing and externalizing predicted low IQ and poor reading status two years later (Kellam et al., 1975). Further, other studies of elementary aged children demonstrated that both teacher-rated antisocial behavior and early measures of inattention were correlated with poor reading readiness two years later (McMichael, 1979; Palfrey et al., 1985). In a review of studies of secondary grades,



Farrington (1979) demonstrated that low vocabulary status at age eight to ten predicted self- and teacher-reported delinquency six years later. Similarly, another study found that peer ratings of aggression predicted school dropout seven years later (Kupersmidt & Coie, 1990). While these findings provide compelling support for the presence of a relationship between behavior and academic underachievement, there are some methodological concerns. For one, significant findings were largely correlational and therefore interpretations regarding directionality should be made with caution. Further, externalizing behavior was measured using a range of methods in this review, including teacher ratings, parent ratings, semi-structured parent interviews, and psychologists' impressions based on clinical rating scales, some of which have limited reliability (Hinshaw, 1992b). Given these findings from early investigations, researchers continued to examine the relationship between student behavior and achievement using more reliable methodology.

Fergusson et al. (1993) used structural equation modelling and a longitudinal approach to examine the relationship between conduct disorder and attention deficit behaviors in middle childhood with juvenile delinquency and academic achievement. In a sample of 1,265 children in a single birth cohort in New Zealand, children were assessed for conduct disorder and attention deficit/hyperactivity behaviors at age 6, 8, and 10. Assessments used maternal and teacher behavior ratings, and then were administered the Test of Scholastic Abilities at age 13. Results suggested a significant relationship between middle childhood attention deficit/hyperactivity behaviors and later academic underachievement, while the relationship between conduct disorder and academic underachievement appeared to be due, almost entirely, to co-occurring attention problems. These findings added to existing knowledge on the relationship and supported the notion that early inattention and hyperactivity are related to

academic underachievement later (Fergusson et al., 1993). Findings from Fergusson et al. (1993) were replicated in several other studies, supporting the idea that the early behavior problems are related to academic underachievement years later (e.g., Fergusson et al., 1997; Rapport et al., 1999).

More recently, in a special education sample of 42 students with emotional and behavior disorders educated in a self-contained school, both elementary and secondary students scored well below the 25th percentile in math, reading, and written expression on the Woodcock-Johnson III Test of Achievement (WJ-III; Lane et al., 2008). Nelson et al. (2004) also found that in a special education sample of 155 K-12 students with EBD, 83 percent of students scored below the mean for their respective norm group on standardized measures of academic achievement overall and across specific academic content areas that include mathematics, reading, written expression. In this study, children and adolescents were grouped and their achievement scores were compared. Findings indicated a significant difference between children and adolescent math achievement scores in the sample, in which 56 percent of children and 83 percent of adolescents scored below the mean of the norm group (Nelson et al., 2004). This further supports the presence of a widening achievement gap for students with emotional or behavior disorders as they progress through school.

Taken together, these findings contribute to the body of research on the relationship between student behavior and academic achievement and confirmed the presence of a significant relationship in the school setting. However, further research is needed to understand the relationship in a non-special education and non-clinical sample. Children may exhibit externalizing behavior in the classroom, but not meet criteria for one of these disorders, thus underscoring the importance of measuring observable behaviors rather than

presence of a diagnosed behavior disorder. By using data from systematic direct observation and direct behavior rating, results may be more generalizable to students in general education who exhibit subclinical levels or are at-risk of externalizing behavior disorder. Thus understanding the relation between two methods of measuring behavior symptomatology and the role they play in relating to achievement is of importance.

### **Rationale for Current Study**

Historically, enhancing understanding of the nature of the relationship between externalizing behaviors and academic achievement has been the focus of researchers in this area. The existing body of literature demonstrates that elementary-aged children exhibit a range of behavioral and academic problems early in their schooling will experience poor educational, social, and behavioral outcomes and information may inform the selection and implementation of targeted prevention, and intervention approaches in schools. In these studies, behavioral measurements like SDOs and DBRs have served as a method to achieve the goal of examining the relationship. Therefore, examining the measures themselves to understand their association with each other and with academic achievement represents a gap in the research.

### **Purpose**

The purpose of the current study was to examine the convergent validity of two commonly used measures of student externalizing behavior on academic achievement in elementary schools. Student behavior was measured by the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2004) and teacher ratings on the Behavior Assessment System for Children, Third Edition (BASC-3; Reynolds, Kamphaus, & Vannest, 2015). Student academic achievement was measured using the Woodcock-Johnson Tests of Achievement, Fourth Edition (WJ-IV ACH; Schrank, Mather, & McGrew, 2014), Brief Achievement cluster. The study

addressed how the use of direct observation of student externalizing behavior and teacher ratings of student externalizing behavior symptomatology relate to each other and to academic achievement. The study specifically examined whether direct observation of student behavior using the BOSS and teacher ratings of student behavior on the BASC-3 separately relate to and predict student academic achievement. The presence of a relationship between externalizing behavior and academic underachievement is well-established (Barriga et al., 2002; Frick et al., 199; Lane et al., 2007), though there are very few studies that have examined this relationship using the BASC-3, BOSS and WJ-IV ACH.

The primary research questions (RQs) were: 1) To what extent are externalizing behaviors measured by the BOSS related to academic achievement as measured by the WJ-IV ACH among students in elementary schools? 2) To what extent are externalizing behaviors measured by the BASC-3 related to academic achievement as measured by the WJ-IV ACH among students in elementary schools? 3) Do teacher ratings of externalizing behavior, as measured by the BASC-3, predict academic achievement beyond what is explained by independent observer ratings on the BOSS?

Based on the larger scientific literature (e.g., Carroll et al., 2005; Conrad et al., 2013, Fergusson et al., 1997; Nelson et al., 2004), it was hypothesized that: 1) student externalizing behavior, as measured by the BOSS and BASC-3 will be negatively correlated with academic achievement, 2) student externalizing behavior, as measured by the BOSS and BASC-3, would predict academic achievement, and 3) teacher ratings of student behavior will predict student academic achievement and offer unique variance beyond direct observation of student behavior.

## Methods

## **Participants**

Data were obtained across 36 public school districts in New Jersey. The current study used data from the Paraprofessional Behavior Support Coaching Project, a randomized controlled trial (RCT) funded by the United States Department of Education Institute of Education Sciences (IES). Archival data used in the current study was collected at baseline from cohorts 1, 2, and 3 as part of the RCT. Participants included 197 students with or at risk for EBDs and 106 classroom teachers.

### ***Student Sample***

Students were mostly male (78%) and ranged from 5 to 12 years old. Students were enrolled in kindergarten (25.9%), first grade (15.5%), second grade (7.8%), third grade (15.0%), fourth grade (19.7%), and fifth grade (16.1%). According to school demographic information obtained from the New Jersey Department of Education, approximately 9.9% of students were classified as English Language Learners and 63.5% of students qualified for free or reduced-price lunch. 20.3% of students were White, 38.8% were Latino/a, 33.5% were Black, 5.2% were Asian, 0.2% were Native Hawaiian/ Pacific Islander, 0.3% were American Indian/Alaskan Native, and 1.7% were two or more races.

### ***Teacher Sample***

Teachers were mostly female (95.4%). Teachers self-identified as White (70.9%), Latino/a (5.1%), Black (15.2%), Asian (1.3%), Middle Eastern (1.3%) and multiple races (6.3%). Fifty-one percent of teachers had more than 10 years of teaching experience, 17% had five to ten years, 12% had two to five years, and 20% had less than two years of experience. The average teacher age was 39.4 ( $SD = 11.9$ ; Range = 24 to 62). Teachers held a bachelor's degree (51.3%),

master's degree (43.6%), two teachers held a graduate degree, one teacher held an associate's degree, and one teacher completed some college (no degree).

### **Instrumentation**

#### ***Behavior Assessment System for Children, Third Edition (BASC-3)***

The Behavior Assessment System for Children, Third Edition Teacher Rating Scales, Child Form (BASC-3 TRS-C; Reynolds, Kamphaus, & Vannest, 2015) is a broadband measure that assesses many areas of behavioral functioning in children. The Teacher Rating Scale was used to gather information about student behavioral functioning at school from the teachers' viewpoint. Respondents were instructed to complete the BASC-3 based on their observations of the students' behavior within the last four weeks. The BASC-3 TRS-C is intended for use with children aged 6 to 12 and consists of 156 items. The TRS-C includes a broad score (Behavioral Symptoms), four composite scores (Adaptive Skills, Externalizing Problems, Internalizing Problems, and School Problems), and 15 subscales. Items are rated on a Likert-type scale with four response options: 0 (*Never*), 1 (*Sometimes*), 2 (*Often*), or 3 (*Almost always*). Scores are reported as *T*-scores with a mean of 50 and a standard deviation of 10. Combined gender norms were used to calculate *T*-scores on each scale. The BASC-3 TRS-C has reliability coefficient alphas ranging from .92 to .97 for general, combined-gender norms, and acceptable test-retest reliability coefficients between .77 and .91 (Reynolds et al., 2015).

The Externalizing Problems composite was used to measure externalizing behavior. The Externalizing Problems composite is comprised of three different scales measuring hyperactivity, aggression, and conduct problems. Example items include "Speaks out of turn in class", "Acts out of control", and "Breaks the rules".

***The Behavior Observation of Students in Schools (BOSS)***

The Behavior Observation of Students in Schools (BOSS; Shapiro, 2004) is a systematic direct observation tool that measures student classroom behavior, specifically capturing academic engagement and disruptive behavior. Momentary time sampling was used to record student academic engagement at the beginning of 15-second intervals. Academic engagement was recorded when students were actively and passively attending to assigned work (e.g., writing, reading aloud, listening to the teacher's lecture, looking at a worksheet). During the remainder of each interval, partial interval time sampling was used to measure observed externalizing behaviors including inappropriate physical (e.g., hitting another person), inappropriate verbal (e.g., calling out, verbal teasing), noncompliance (e.g., not completing assigned task), and/or disruptive academic (e.g., off-task, daydreaming). The BOSS has consistently high inter-observer reliability (.90-1.0) and treatment sensitivity (Fredricks et al., 2011; Volpe et al., 2005). Discriminant validity evidence suggests that the BOSS can effectively differentiate children with ADHD from children without ADHD. In this study, Externalizing Behavior was measured using the inappropriate physical and inappropriate verbal scores on the BOSS. One study examined the relationship among the BOSS and various measures of ADHD in elementary students and found that academic engagement scores on the BOSS were incrementally predictive of ADHD symptomatology on the teacher-rated Children's Symptom Inventory (CSI). However, student behavior scores on the BOSS were not significantly related to teacher-rated behavior on the BASC-3 (Jiang et al., 2019).

***Woodcock-Johnson IV Test of Achievement (WJ-IV ACH)***

The Woodcock-Johnson Tests of Achievement, Fourth Edition (WJ-IV ACH; Schrank, Mather, & McGrew, 2014) is a norm-referenced, standardized test of academic achievement. The

WJ-IV ACH includes a total of 20 tests (11 standard battery, 9 extended battery) that measure academic achievement in various subject areas. The subject areas of interest in this study include reading, written language, and mathematics. Academic Achievement is measured using the Brief Achievement cluster, an academic proficiency cluster that consists of three widely used tests in the WJ-IV ACH standard battery: Letter-Word Identification, Applied Problems, and Spelling. These tests measure academic achievement in the areas of reading, mathematics, and written language, respectively. Internal consistency for cluster scores on the WJ-IV ACH are high (.92-.97) and test-retest correlations (.83-.95) are in the acceptable to excellent range (Villarreal, 2015).

Content of the measure was created to address core curriculum and achievement as described by federal legislation (Schrank et al., 2014). The WJ-IV ACH has demonstrated strong correlations with other widely used measures of academic achievement including the Kaufman Test of Educational Achievement–Second Edition (KTEA-II; Kaufman & Kaufman, 2004) and the Wechsler Individual Achievement Test–Third Edition (WIAT-III; Wechsler, 2009). The WJ-IV ACH was administered individually in pencil-and-paper format. Raw scores are converted into standard scores that have a mean of 100 and a standard deviation of 15.

## **Procedures**

### ***Recruitment***

Recruitment included several methods. Project staff first met with school district administrators and reviewed the purpose and protocol of the study. Following district-level meetings, project staff met with elementary school principals to review information about the study. Flyers and presentations about the project were shared with teachers and paraprofessionals



during elementary school meetings. Informed consent was obtained from all study participants in accordance with the Institutional Review Board at Rutgers University.

### ***Student Nomination and Eligibility***

A two-stage gating procedure was used to recruit students with disruptive classroom behavior. Participating teachers nominated and rank-ordered the top five students with disruptive behavior problems within their classroom. Students with intellectual disability, pervasive neurodevelopmental disorders (i.e., Autism Spectrum Disorder), or Bipolar Disorder were excluded from the study due to qualitative differences in the etiology of their disorders as well as the unique interventions needed to address them. Teachers completed the *BASC-3 Behavioral and Emotional Screening System (BESS-3; Kamphaus & Reynolds, 2015)* rating scale for each of the five ranked students. Students whose scores were in the “elevated” or “extremely elevated” risk for behavioral and emotional problems categories on the BESS, were observed in their respective classrooms by trained observers. Observers conducted two classroom observations of students’ behavior using the BOSS coding criteria described below. Students who were observed exhibiting disruptive behaviors in at least 15% of the intervals observed across two observations were considered eligible for participation in the study. Of those eligible, two to three students from each classroom were randomly selected to participate in the study.

### ***BOSS Observer Training***

Trained independent observers performed the BOSS observations. Independent observers were doctoral students and undergraduates who were trained by a licensed psychologist experienced in BOSS observation. Training consisted of a two-hour training session focused on the theory, design, constructs, procedure, and scoring of the measure. Observers were trained on taking frequency count data using both partial interval and momentary time sampling during

classroom lessons. Training also included group and individual practice completing the BOSS with classroom videos and were provided feedback from the licensed psychologist. Observers were required to independently complete three BOSS observations with proficiency prior to conducting observations for the current study. For the purpose of BOSS training, observers were considered proficient when they achieved 80% agreement with experienced trainers.

### ***WJ-IV ACH Administration Training***

The WJ-IV ACH was administered by doctoral students who attended a two-hour training session led by a licensed psychologist that focused on administration and scoring of the three subtests used in the current study. Additionally, all examiners completed doctoral level coursework on assessment of learning disabilities including administration and scoring of the protocol, the theory of the WJ-IV ACH, and the utility of the test. The examiners administered and scored at least three WJ-IV ACH protocols with proficiency prior to conducting evaluations for the current study. Assent was obtained from all children prior to test administration. Doctoral students administered and scored all test protocols. Raw scores were converted to standard scores in accordance with standard scoring procedures.

### **Data Analysis Plan**

Ratings of student behavior and academic achievement that were collected at baseline as part of a larger RCT were analyzed. Data was analyzed at the individual classroom (teacher) level. Descriptive statistics and correlations were computed on all BOSS scores, BASC-3 scales, and WJ-IV ACH Brief Achievement cluster scores. Spearman's rank-order correlations were used to examine the relation between the BOSS and other measures. Pearson's product-moment correlations were used for all remaining measures. Bivariate correlations with magnitudes in the

.00s were considered nonexistent, .10s and .20s small, .30s and .40s medium, .50s and .60s large, .70s and .80s very large, and .90s nearly perfect (Cohen, 1992).

A series of multilevel models (i.e., hierarchical linear models) were used to determine whether direct observation scores of student externalizing behavior (BOSS) and teacher ratings of externalizing behavior symptomatology (BASC-3) predicted student academic achievement. Multilevel modeling was used to account for nesting in the dataset, in which the 197 individual students were seated within 36 different classrooms (Peugh, 2009). Model 1 was the null model and did not include predictors to provide a standard that all other models were compared to. In Model 2, four student externalizing behavior scores from the BASC-3 were entered as predictors for student achievement. In Model 3, four student externalizing behavior scores from the BOSS were entered as predictors for student achievement. In Model 4, externalizing behavior scores on the BOSS and BASC-3 were entered as predictors in the model. Overall fit of the models was evaluated on the basis of examination of the Akaike information criterion (Akaike, 1974). Descriptive statistics and correlations were conducted in JASP Version 0.14.1 and multilevel models were conducted in SPSS Version 27.0.1.0.

## Results

Descriptive statistics for student BOSS, BASC-3, and WJ-IV ACH cluster and subtest scores are presented in Table 1. Prior to analyses, both outcome and predictor variables were assessed for distribution and analysis assumptions, such as normality, linearity, independence, homoscedasticity, and multicollinearity. A visual evaluation of data indicated no significant deviations from normality, as well as no violations of homoscedasticity of residual and random effect plots. Student Inappropriate Physical (IP) was coded on average of 25% of total intervals recorded. Inappropriate Verbal (IV) was coded on average of 18% of total intervals. Disruptive

Academic (DA) was coded on average of 3% of intervals, and Noncompliance on average of 15% of intervals recorded. Teacher ratings on the BASC-3 resulted in an average Behavioral Symptoms *T* score of 70.74 (*SD* = 12.02; Range = 40 to 101). The mean Externalizing Problems *T* score of 72.94 (*SD* = 15.91; Range = 41 to 115) suggests that on average, teachers rated students as having significantly high levels of externalizing behavior, compared to their norm group. Teacher ratings of School Problems resulted in an average *T* score of 64.06 (*SD* = 9.44; Range = 39 to 86), suggesting that teachers rated students as having significant school problems.

Table 1  
*BOSS, BASC-3 TRS, and WJ-IV ACH Descriptive Statistics*

	Mean	SD	Range
<b>BOSS (<i>n</i> = 197)</b>			
Inappropriate Physical	0.25	0.15	0.07 - 0.95
Inappropriate Verbal	0.18	0.14	0.00 - 0.81
Disruptive Academic	0.15	0.12	0.00 - 0.59
Noncompliance	0.03	0.04	0.00 - 0.24
<b>BASC-3 TRS (<i>n</i> = 197)</b>			
Behavioral Symptoms	72.94	15.91	41.00 - 115.00
Hyperactivity	69.14	11.94	40.00 - 94.00
Aggression	73.01	19.26	43.00 - 119.00
Conduct Problems	71.05	16.86	41.00 - 118.00
<b>WJ-IV ACH (<i>n</i> = 197)</b>			
Brief Academic	80.25	16.56	39.00 - 112.00
Letter-Word Identification	82.23	18.14	40.00 - 118.00
Applied Problems	77.54	16.00	40.00 - 114.00
Spelling	85.95	15.58	40.00 - 123.00

### **Relations Between Student Behavior and Achievement**

To address research questions (RQ) 1 and 2, relationships among variables were examined using Pearson product-moment and Spearman rank-order correlations (see Table 2). The relationships between the BASC-3 scores and the WJ-IV ACH scores were in the expected directions. Small to medium negative correlations were found between School Problems on the BASC-3 and Brief Academic (BA) ( $r = -.36, p < .001$ ), Letter-Word Identification ( $r = -.36, p <$

.001), Applied Problems ( $r = -.25, p < .001$ ), and Spelling ( $r = -.34, p < .001$ ), indicating that students who were rated by teachers as demonstrating higher levels of School Problems symptomatology (i.e., learning problems and attention problems) scored lower in academic achievement. A small negative correlation was found between Aggression and BA ( $r = -.16, p = .02$ ), Letter-Word Identification ( $r = -.15, p = .04$ ), and Spelling ( $r = -.14, p = .04$ ), suggesting that students who were rated by teachers as exhibiting more aggressive behaviors scored lower in academic achievement. Small to medium positive correlation was found between Adaptive Skills and Brief Academic ( $r = .20, p = .006$ ), Letter-Word Identification ( $r = .16, p = .023$ ), Applied Problems ( $r = .22, p = .002$ ), and Spelling ( $r = .18, p = .013$ ). This indicates that students who were rated by teachers as demonstrating more adaptive skills (i.e., social skills, leadership, study skills, functional communication) scored higher in academic achievement. Student externalizing symptomatology on the BOSS was found to have no significant relationship to achievement scores on the WJ-IV ACH.

Table 2  
*Correlations of Student Behavior and Achievement*

Variable	BOSS						BASC-3							WJ-IV ACH			
	AE	PE	IP	IV	DA	NC	BS	EP	SP	IP	AS	HY	AG	CP	BA	LW	AP
BOSS																	
Active Engagement																	
Passive Engagement	-.22**																
Inappropriate Physical	-.21**	-.26**															
Inappropriate Verbal	-.19**	-.32**	.55**														
Disruptive Academic	-.27**	-.04	.12	-.23**													
Noncompliance	-.20**	-.15*	.30**	.30**	.10												
BASC-3 Composite																	
Behavioral Symptoms	-.22**	-.20**	.23**	.25**	-.01	.23**											
Externalizing Problems	-.18*	-.21**	.25**	.29**	-.01	.21**	.78**										
School Problems	-.21**	-.18*	.25**	.10	.16	.16*	.52**	.42**									
Internalizing Problems	-.05	-.15*	.03	.10	-0.08	.09	.64**	.27**	.17**								
Adaptive Skills	.21**	.10	-.25**	-.09	-0.13	-.11	-.54**	-.40**	-.64**	-.20**							
BASC-3 Subtest																	
Hyperactivity	-.22**	-.11	.30**	.31**	-.01	.19**	.71**	.84**	.48**	.23**	-.34**						
Aggression	-.16**	-.23**	.23**	.31**	-.03**	.20**	.73**	.94**	.32**	.30**	-.36**	.66**					
Conduct Problems	-.12	-.20**	.19**	.21**	.01	.17**	.69**	.94**	.39**	.20**	-.39**	.72**	.82**				
WJ-IV ACH																	
Brief Academic	.09	.05	-.03	-.04	.00	-.01	-.11	-.13	-.36**	-.06	.20**	-.05	-.16**	-.11			
Letter-Word Identification	.07	.08	-.07	-.04	-.03	-.00	-.09	-.11	-.36**	-.04	.16**	-.04	-.15**	-.08	.93**		
Applied Problems	.17*	-.04	.02	-.03	-.00	-.03	-.10	-.10	-.25**	-.02	.22**	-.04	-.11**	-.11	.76**	.54**	
Spelling	.03	-.05	.01	-.02	-.02	.01	-.12	-.11	-.34**	-.11	.18**	-.06	-.14**	-.09	.90**	.84**	.56**

*Note.* Spearman's rank-order correlations were computed for BOSS scores and Pearson product-moment correlations were used on BASC-3 and WJ-IV Ach scores. AE = Active Engagement; PE = Passive Engagement; IP = Inappropriate Physical; IV = Inappropriate Verbal; DA = Disruptive Academic; NC = Noncompliance; BS = Behavioral Symptoms; EP = Externalizing Problems; SP = School Problems; IP = Internalizing Problems; AS = Adaptive Skills; HY = Hyperactivity; AG = Aggression; CP = Conduct Problems; BA = Brief Academic; LW = Letter-Word Identification; AP = Applied Problems; SP = Spelling.

\* $p < .05$ . \*\* $p < .01$ .

### **Prediction of Student Achievement**

A series of multilevel models (see Table 3) were conducted to address research questions 2 and 3. The variables (predictors) were entered simultaneously in each of the four regression models. The model assumptions were met through visual examination of residual and random effect plots. For all models the outcome variable was Brief Achievement (BA) as measured by the WJ-IV ACH. Model 1 was the null model and did not include predictors to provide a standard that all other models were compared to. For Model 2, four student externalizing behavior scores from the BASC-3 were entered as predictors (fixed effects) for student achievement. BASC-3 externalizing scores did not significantly predict student BA scores. Model 3 included the BOSS externalizing scores as predictors of student BA. BOSS externalizing scores did not significantly predict student achievement. Model 4 included eight externalizing scores from both BASC-3 and BOSS as predictors of BA. This model did not significantly predict student achievement. While examination of unstandardized beta (*b*) scores for the fixed effects in the four models did not yield significant results, assessment of model fit through examination of Akaike information criteria (AIC) was conducted to further evaluate the results.

A series of fit indices were calculated to evaluate the fit of the data to the final models. As presented in Table 3, the multilevel analyses resulted in variation in AIC across models. For Model 2, which included four BASC-3 externalizing measures as predictors of achievement, there was a slightly improved fit over the null model ( $\Delta\text{AIC} = 2.73$ ;  $\text{AIC} = 1639.12$ ). Model 3, which included four BOSS externalizing measures as predictors of achievement, had lower AIC, indicating better fit ( $\text{AIC} = 1612.88$ ). The fourth model which included externalizing scores from the BASC-3 and the BOSS demonstrated lower AIC (indicating better fit) over the other three

models. Thus, entering both the BOSS and BASC-3 externalizing scores as predictors of BA resulted in a more credible model, as compared with models entering BOSS or BASC-3 scores separately (Vrieze, 2012). No meaningful changes to model fit were observed when individual subtests of academic achievement (i.e., Applied Problems, Spelling, Letter-Word Identification) were entered as outcome variables individually.

Table 3  
*Multilevel Modeling to Predict Student Achievement*

Parameter Estimates	WJ-IV Brief Achievement			
	<i>b</i> (Std. Error)			
	Model 1	Model 2	Model 3	Model 4
<b>Fixed Effects</b>				
Intercept	80.06*** (1.47)	61.75** (17.86)	81.89*** (3.11)	65.47** (18.36)
<b>BASC-3</b>				
Externalizing Problems		-4.43 (3.85)		-4.01 (3.90)
Aggression		1.44 (1.39)		1.27 (1.42)
Conduct Problems		1.57 (1.42)		1.43 (1.43)
Hyperactivity		1.80 (1.39)		1.64 (1.41)
<b>BOSS</b>				
Inappropriate Physical			.28 (9.89)	-1.28 (10.01)
Inappropriate Verbal			.24 (10.65)	3.02 (10.85)
Disruptive Academic			-12.42 (10.99)	6.63 (28.43)
Noncompliance			-1.87 (28.15)	-10.36 (10.99)
<b>Random Effects</b>				
Teacher (Intercept)	122.25*** (32.51)	115.27** (31.57)	131.34*** (34.66)	122.65** (33.46)
Residual	154.89*** (21.81)	155.80** (22.06)	153.27*** (22.08)	155.03** (22.43)
<b>Additional Information</b>				
ICC	0.44	0.43	0.46	0.44
AIC	1641.85	1639.12	1612.88	1610.07
BIC	1648.41	1645.63	1619.40	1616.54
Log Likelihood	1637.85	1635.12	1608.88	1606.07

Note. ICC = Intraclass Correlation Coefficient; AIC = Akaike Information Criterion; BIC = Bayesian Information Criteria.

\*\*\* $p < .0001$ , \*\* $p < .001$ , \* $p < .05$



## Discussion

The purpose of the current study was to examine the convergent validity of the Behavior Assessment Scale for Children, Third Edition (BASC-3) and the Behavior Observation of Students in Schools (BOSS), with the Woodcock-Johnson Tests of Achievement, Fourth Edition (WJ-IV ACH). While researchers have historically examined the relationship between externalizing behavior and achievement, this study focused specifically on understanding the association between these commonly used assessments. One primary research aim was to examine the relationship between teacher ratings of student externalizing behavior symptomatology measured by the BASC-3 and academic achievement measured by the WJ-IV ACH. It was hypothesized that significant relationships would emerge among student externalizing behavior symptomatology measured by the BASC-3, and academic achievement. Significant findings indicating correlations among these measures would provide confirmation of the relationship between externalizing behavior and academic achievement using these specific measures. More specifically, significant correlations could indicate the relationship is measurable using three common school-based tools. This type of finding could provide support for the use of the BOSS and BASC-3 as indicators of risk for academic underachievement in addition to their measurement of behavioral symptoms.

Correlational findings revealed small to medium negative relationships between teacher-reported School Problems, including observed learning and attention problems, and a brief measure of student achievement. Students who were rated by teachers as demonstrating higher levels of School Problems symptomatology (i.e., learning problems and attention problems) scored lower in academic achievement. Consistent with prior research, the combined presence of attention problems and learning problems represent a significant negative relationship to

academic achievement (e.g., Barriga et al., 2002; Carroll et al., 2005; Frick et al., 1991). Carroll (2005) stated that attention problems with or without hyperactivity represented the most salient risk factor for academic problems in a study from the UK national survey of child mental health. Like Carroll's (2005) findings, hyperactivity did not offer any significant relationship to achievement on its own.

A significant, negative relationship also emerged among students who were rated by teachers to demonstrate higher levels of aggression in the classroom scored lower in academic achievement. These findings are consistent with a previous study of the BASC-3 and WJ-III ACH which indicated that higher ratings of externalizing problems scales on the BASC-3 were negatively associated with achievement on the WJ-III ACH (Conrad et al., 2013). Further, these results are in line with previous findings by Carroll (2005), which indicated that both Conduct Disorder and Attention-Deficit/ Hyperactivity Disorder were significantly associated with literacy difficulties. Similarly, Farrington (1979) found that low vocabulary status at age eight to ten predicted self- and teacher- reported delinquency six years later, and Kupersmidt and Coie (1990) found that peer ratings of aggression predicted school dropout seven years later. While previous findings provide evidence of the relationship between constructs of aggression and achievement, these findings from the current study represent some of the first to identify a significant relationship between aggression on the BASC-3 and achievement on the WJ-ACH.

While a few significant relations emerged among School Problems and Aggression on the BASC-3 and the WJ-IV ACH, correlations between the other scales on the measures were largely nonsignificant. Conrad (2013) found that externalizing problems on the BASC-3 were significantly, negatively related to achievement on the WJ-III ACH. Yet, results from the current

study showed no significant relationship for most of the externalizing behavior scales, such as hyperactivity and conduct problems, on the BASC-3.

Another goal of the study was to examine the relationship between independent observer ratings of student externalizing behaviors measured by the BOSS and academic achievement measured by the WJ-IV ACH. Given the literature citing the relationship between externalizing behavior and achievement, it was hypothesized that student externalizing behavior, as measured by the BOSS would be negatively correlated with academic achievement. However, correlation results from the current study indicated no significant relationship between these variables. According to these findings, independent observer ratings of student behavior did not significantly relate to student academic achievement. While this contradicts previous findings, which have shown that student behavior and academic achievement are closely related, no existing research has used SDOs and the WJ-ACH to examine the relationship. Thus, the results suggest that while there is a well-documented relationship between externalizing behavior and achievement, a direct observation measure that represents a single, brief, point in time may not provide an accurate representation of this link. These nonsignificant findings may be due, in part, to measurement error associated with SDOs, as results from the BOSS are fundamentally unstable, as data represent a single, brief, point in time (Christ et al., 2009). Thus, a number of conciliatory factors impacting BOSS data may have contributed to behavioral screening data that may not accurately portray a student's level of symptomatology in the current study (Riley-Tillman et al., 2008).

The third research question hypothesized that student externalizing behavior, as measured by the BOSS and BASC-3, would predict academic achievement. Further, teacher ratings of student behavior would predict student academic achievement and offer unique variance beyond

direct observation of student behavior. This study is one of the first to examine the effect that this group of predictors (BOSS and BASC-3 symptomatology measures) jointly has on student academic achievement and aims to ascertain evidence indicating convergent validity among the measures. Through examination of fixed effects across four models, outcomes from the multilevel model analyses indicated that student behavior symptomatology did not significantly predict academic achievement. Variables were related, but not potent predictors of achievement for these children. Given that student behavior has been cited to have significant impact on academic achievement (e.g., Nelson et al., 2004; Lane et al., 2008), it was hypothesized that a relationship among student behavior on the BASC-3 and BOSS and academic achievement would emerge in the current study. Thus, it was unexpected that student externalizing behavior symptomatology did not significantly predict academic achievement using these school-based measures.

While examination of fixed effects did not yield significant evidence of a relationship among these variables, evaluation of change in Akaike's information criteria across the four models provided useful information related to model fit. Using AIC to evaluate model fit has become increasingly relied upon as a means for model selection in HLM because of its focus on higher order properties of data, rather than means and covariances using in factor analysis (Vrieze, 2012). Thus, examination of incremental change to AIC was relied upon as a credible method for model selection in the current study. AIC values in the current study indicated increased fit to the WJ-IV ACH with each subsequent model. This was observed through consistent decreases in each AIC from Models 1 through 4. Interestingly, Model 3 which included the BOSS externalizing variables as predictors of achievement indicated smaller AIC, and therefore better model fit than Model 2, which included BASC-3 externalizing variables as

predictors. This was an unexpected result based on prior research that indicated significant strengths in DBRs such as the BASC-3 at measuring externalizing behavior (Chafouleas et al., 2009). It is possible that while the BASC-3 is a strong measure for externalizing behavior symptomatology, certain characteristics of the BOSS lend themselves to stronger measurement of behavioral symptoms that relate to academic achievement.

Furthermore, examination of AIC provided support for the third hypothesis, which posited that teacher ratings of student behavior using the BASC-3 would offer unique variance beyond direct observation of student behavior. While this was not indicated through examination of the factor analysis or through observation that the BOSS alone (Model 3) produced a smaller AIC value than the BASC-3 alone (Model 2) did, examination of the incremental change in AIC when the BASC-3 was included in the model in addition to the BOSS (Model 4) provided important information. Model 4 which included both the BOSS and BASC-3 externalizing scores as predictors of achievement produced the smallest AIC of the four models (AIC = 1610.07). Thus, the addition of the BASC-3 to the BOSS model contributed to improved model fit over the BOSS model alone.

This finding provides support for the convergent validity of the three measures. The BOSS and BASC-3 are regarded as reliable behavioral measures that are relied upon to inform intervention in the schools. Given research has consistently cited the relationship between externalizing behavior and achievement, one would expect that two of the most used school measures would detect such a relationship. However, minimal research exists regarding their association to achievement. The current results indicate the BOSS more closely relates to achievement compared to the BASC-3 on its own. Further, it provides support that the two measures together more closely relate to achievement than either measure individually.

Comparison of AIC values suggest that this evidence provides some support for the convergent validity of the three measures. It is, however, important to remember that unstandardized beta ( $b$ ) scores for the fixed effects in the four models did not yield significant results independently, which may weaken the strength of this claim.

Findings from the current study offer valuable perspective on the understanding of the relationship between measures of student behavior and achievement. Researchers in the field have long asserted that student behavior is closely related to achievement. Some have even posited that the relationship exists clearly in both directions, such that students who demonstrate externalizing behavior will exhibit poorer achievement, while student identified as having poorer achievement will exhibit higher levels of externalizing behavior (Fergusson et al., 1993; Hinshaw, 1992b). While certain methods of measurement may have demonstrated this outcome in previous research, the results of the current study only partially confirm this claim. Examination of  $b$  values in the HLM did not yield significant associations among the two externalizing behavior measures and the achievement measure. However, comparison of the AIC values in each of the four models provided valuable information regarding the fit of each model, compared to each other. Thus, while the results of the fixed effects analyses were nonsignificant, AIC values indicate the BOSS is a stronger predictor of achievement than the BASC-3, and further the two measures together are stronger predictors of achievement than either measure independently. With the widespread use of behavioral measurement in schools, it is favorable that the current study suggests that the measures do, in fact, relate to achievement. Results like this are promising and suggest that school practitioners would be able to create meaningful change in student behavior based on academic interventions, or academic improvement based on

behavior interventions. Yet, these results are not definitive and therefore important questions regarding measurement of the relationship between behavior and achievement remain.

### **Limitations**

Several limitations should be considered when interpreting results of this study and generalizing findings to other contexts and populations. Sample characteristics may limit generalizability to other states, school contexts, and student populations. Data were collected from kindergarten through fifth grade from suburban and urban school districts in New Jersey. Students were predominately male (78%), and identified as Latino/a (33.5%), Black (33.5%), and White (20.3%). Further, teachers were mostly female (95.4%), identified as White or European American (70.9%), and reported having more than 10 years of professional experience (51%). Thus, the generalizability of the findings to other samples of students and teachers of varied gender, race/ethnicity, and years of professional experience may be limited. Additionally, selection bias may present another limitation in this study. Student participants were selected using teacher identification of students with perceived highest rates of externalizing behavior. Thus, the participants likely exhibited disproportionately high rates of externalizing behavior symptomatology compared to the population.

Methods of data collection present additional limitations. The study aimed to examine the relationship between measures of student externalizing behavior and achievement on the WJ-IV ACH, Brief Achievement Cluster. While the Brief Achievement cluster scores are meaningful for representing achievement, it is not comprehensive in nature and therefore assumptions about the impact of behavior on broad academic achievement should be made with this consideration.

Finally, methods of behavioral measurement should be interpreted with caution. Given that teachers rated student behavioral symptoms on the BASC-3, which may be susceptible to

social desirability bias. Further, independent observations were used as a measure of externalizing behavior symptomatology, and the presence of outside of observers may have impacted student behavior. Furthermore, data collected from the BOSS represents only a single, 15-minute, point in time. As such, this method of data collection using a brief observation period is fundamentally unstable, and therefore cannot be relied upon to accurately capture student externalizing behavior. Even further, the lack of generalizability of these one-time observations challenges the validity of the resulting relationship to academic achievement, as it may not be accurately captured.

### **Future Directions**

This study presents opportunities for future research. Future studies should examine larger, more race/ethnicity and gender inclusive samples of elementary, middle and/or high school students with or at-risk for externalizing behavior disorders to strengthen generalizability. Future research should collect behavior symptomatology data using a variety of informants including teachers, students, and parents. Further, achievement data should include a more comprehensive assessment of broad academic achievement. Given the lack of relationships found between externalizing behavioral dimensions on the BOSS and BASC-3 and achievement of the WJ-ACH, it would be beneficial to explore this relationship among additional measures of student behavior and achievement. Introducing a broad-band rating scale such as the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS; Kilgus et al., 2014) that examines both behavioral and academic domains together may provide valuable information relating to student externalizing behavior and academic achievement that may produce results more useful to school practitioners. Further, given the findings of a positive relationship between adaptive



skills and student achievement, it may be useful to examine more of these student strengths such as protective factors, adaptive skills, and prosocial behaviors, in relation to achievement.

### **Implications for School Practice**

Given existing research that demonstrates a close relationship between externalizing behavior and achievement such that intervening in one area may provide significant outcomes in the other area, this study aimed to examine the convergent validity of two brief, school-based measures to determine the relevance of this relationship to school-based practice. Unexpectedly, while previous research demonstrates that rating scales like the BASC-3 significantly relate to the WJ-ACH, this relationship was not significant in the current study. Although nonsignificant, the association among the BOSS, BASC-3, and WJ-IV ACH was observed through examination of AIC values in the four models. Thus, these results from this study offer valuable information for school practice.

First, results of this study provide support for the use of the BOSS and BASC-3 as measures of student behavior that relate to achievement. While this study was not able to clearly determine whether SDOs or DBRs are objectively more accurate in predicting achievement, it suggests the importance in considering both sources of information. Although SDOs have been regarded in the literature as less practical and more subjective than DBRs (Chafouleas et al., 2009), they represent a stronger predictor of achievement than DBRs alone in these results. Therefore, it is important for school practitioners to consider both SDOs and DBRs for student behavior in order to more clearly understand student behavioral symptomatology and develop plans for intervention.

Further, the evidence that the BOSS and BASC-3 are related to the WJ-IV ACH from this study represent a potentially positive outlook for students who demonstrate significant

behavioral symptoms on the BASC-3, as we found that elevated BASC-3 scores do not significantly relate to poor achievement. This suggests that there is room to intervene on behavior before the relationship to poorer achievement becomes intertwined. Therefore, school practitioners may be successful in focusing on early behavior interventions for young students to prevent future academic challenges.

Finally, we must think critically about the utility of school-based behavioral measurement methodologies. Many methods of behavioral measurement, including SDOs and DBRs, are used by school psychologists to inform classification and intervention planning (Christ et al., 2009). However, SDOs and DBRs are less often used as a method of screening for behavioral and academic problems outside of special education (Christ et al., 2009). The current results provide support for the use of the BOSS or the BASC-3 as behavioral and academic problems screeners, and we found that BOSS to be somewhat more effective at gathering relevant information than the BASC-3. As such, while the BASC-3 may provide comprehensive, reliable data about classroom behavior, some characteristics of the BOSS contributed to a stronger model fit for achievement. Thus, while the BOSS and BASC-3 provide some evidence of predictions of academic achievement in the areas of school problems and aggression, the question of whether another form of behavioral measurement may demonstrate stronger convergent validity to the WJ-IV ACH remains.

### **Conclusion**

It has long been a goal of researchers to understand the complex and interrelated factors leading to student academic achievement. Thus, the current study aimed to understand how direct and indirect measures of student externalizing behavior symptomatology (BASC-3 and BOSS) related to academic achievement. While the relation between student externalizing

behavior symptomatology and academic achievement has been documented in previous research (e.g., Lane et al., 2008; Nelson et al., 2004), findings from the current study demonstrated behavior's significant relationship to achievement only among learning problems, attention problems, and aggression on the BASC-3, suggesting that the relationship may not be as clearly defined as was once asserted. Further, using multilevel modelling analyses, student externalizing behavior symptomatology on the BASC-3 and the BOSS did not significantly predict academic achievement. However, examination of Akaike's information criteria indicated the BOSS is a stronger predictor of achievement than the BASC-3, and further the BOSS and BASC-3 together provide incrementally stronger predictor than either measure independently. While this finding does not clearly determine which measure is objectively more accurate, it provides a strong case for consideration of both measures rather than just one. This provides initial support for the use of these behavioral measures as screeners for achievement, though additional evidence is needed. Overall, the mixed findings from this study suggest that methods of measurement clearly impact the extent to which the relationship among externalizing behavior and academic achievement is detected. Therefore, for children with or at risk of EBDs, the influence of context and methods of measurement is relevant to their academic achievement in elementary schools.

## References

- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>.
- Arnold, D. H. (1997). Co-occurrence of externalizing behavior problems and emergent academic difficulties in young high-risk boys: A preliminary evaluation of patterns and mechanisms. *Journal of Applied Developmental Psychology*, *18*, 317-330.
- Ayllon, T. & Roberts, M. D. (1974). Eliminating discipline problems by strengthening academic performance. *Journal of Applied Behavior Analysis*, *7*(1), 71-76.
- Barker, E. D., Oliver, B. R., & Maughan, B. (2010). Co-occurring problems of early onset persistent, childhood limited, and adolescent onset conduct problem youth. *Journal of Child Psychology and Psychiatry*, *51*(11), 1217–1226.
- Barriga, A. Q., Doran, J. W., Newell, S. B., Morrison, E. M., Barbetti, V., & Dean Robbins, B. (2002). Relationships between problem behaviors and academic achievement in adolescents: the unique role of attention problems. *Journal of Emotional and Behavioral Disorders*, *10*(4), 233-240.
- Boyd, D., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. *Educational Evaluation and Policy Analysis*, *31*(4), 416-440.
- Carroll, J. M., Maughan, B., Goodman, R., & Meltzer, H. (2005). Literacy difficulties and psychiatric disorders: evidence for comorbidity. *Journal of Child Psychology and Psychiatry*, *46*(5), 524–532. <https://doi.org/10.1111/j.1469-7610.2004.00366.x>

- Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct Behavior Rating (DBR): An Emerging Method for Assessing Social Behavior Within a Tiered Intervention System. *Assessment for Effective Intervention, 34*(4), 195–200. <https://doi.org/10.1177/1534508409340391>
- Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Foundation for the Development and Use of Direct Behavior Rating (DBR) to Assess and Evaluate Student Behavior. *Assessment for Effective Intervention, 34*(4), 201–213. <https://doi.org/10.1177/1534508409340390>
- Coie, J., & Krehbiel, G. (1984). Effects of Academic Tutoring on the Social Status of Low-Achieving, Socially Rejected Children. *Child Development, 55*(4), 1465-1478. doi:10.2307/1130016
- Connor, D. F., Doerfler, L. A. (2008). ADHD with comorbid oppositional defiant disorder or conduct disorder: Discrete or nondistinct disruptive behavior disorders? *Journal of Attention Disorders, 12*, 126-134. doi:10.1177/1087054707308486
- Conrad, L., Jones, P., Loe, S., Paul, M., & Raines, T. (2013). Examining the relationships between internalizing and externalizing problems and academic achievement (ProQuest Dissertations Publishing). Retrieved from <http://search.proquest.com/docview/1513232480/>
- Danielson, M.L., Bitsko, R.H., Ghandour, R.M., Holbrook, J.R., Kogan, M.D., & Blumberg, S.J. (2018) Prevalence of Parent-Reported ADHD Diagnosis and Associated Treatment Among U.S. Children and Adolescents, 2016. *Journal of Clinical Child & Adolescent Psychology, 47*(2), 199-212, DOI: 10.1080/15374416.2017.1417860

- DiPerna, J., & Elliott, S. N. (2000). *Academic Competence Evaluation Scales*. San Antonio, TX: Harcourt Assessment.
- Education Advisory Board. (2019). *Breaking bad behavior: The rise of classroom disruptions in early grades and how districts are responding*. <http://pages.eab.com/rs/732-GKV-655/images/BreakingBadBehaviorStudy.pdf>
- Farrington, D. P. (2004). Conduct disorder, aggression, and delinquency. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (p. 627–664). John Wiley & Sons Inc.
- Farrington, D. P., Loeber, R., & Van Kammen, W. B. (1990). *Long-term criminal outcomes of hyperactivity-impulsivity-attention deficit and conduct problems in childhood*. In L. N. Robins & M. Rutter (Eds.), *Straight and devious pathways from childhood to adulthood* (p. 62–81). Cambridge University Press.
- Fergusson, D. M., Horwood, L. J., Lynskey, M. T. (1997). Attentional difficulties in middle childhood and psychosocial outcomes in young adulthood. *Journal of Child Psychology and Psychiatry*, 38(6), 633-644.
- Fergusson, D. M., Horwood, L. J., Lynskey, M. T. (1993). The effects of conduct disorder and attention deficit in middle childhood on offending and scholastic ability at age 13. *Journal of Child Psychology and Psychiatry*, 34(6), 899-916.
- Frick, P. J., Kamphaus, R. W., Lahey, B. B., Loeber, R., Christ, M. A. G., Hart, E. L., & Tannenbaum, L. E. (1991). Academic underachievement and the disruptive behavior disorders. *Journal of Consulting and Clinical Psychology*, 59, 289-294.

Froehlich, T. E., Fogler, J., Barbaresi, W. J., Elsayed, N. A., Evans, S. W., & Chan, E. (2018).

Using ADHD Medications to Treat Coexisting ADHD and Reading Disorders: A systematic review. *Clinical Pharmacology and Therapeutics*, *104*(4), 619–637.

Ghandour, R. M., Sherman, L. J., Vladutiu, C. J., Ali, M. M., Lynch, S. E., Bitsko, R. H., &

Blumberg, S. J. (2019). Prevalence and treatment of depression, anxiety, and conduct problems in US children. *The Journal of Pediatrics*, *206*, 256-267.

Hastings R.P., & Bham M.S. (2003). The relationship between student behaviour patterns and teacher burnout. *School Psychology International*. *24*(1):115-127.

doi:10.1177/0143034303024001905.

Henry, K. L., Knight, K. E., & Thornberry, T. P. (2012). School Disengagement as a Predictor of

Dropout, Delinquency, and Problem Substance Use During Adolescence and Early Adulthood. *Journal of Youth and Adolescence*, *41*(2).

Hinshaw, S. P. (1992a). Academic underachievement, attention deficits, and aggression:

Comorbidity and implications for intervention. *Journal of Consulting and Clinical Psychology*, *60*(6), 893-903.

Hinshaw, S. P. (1992b). Externalizing behavior problems and academic underachievement in

childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin*, *111*(1), 127-155. doi:10.1037/0033-2909.111.1.127

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Jiang, Capriotti, M., Beaulieu, A., Rooney, M., McBurnett, K., & Piffner, L. J. (2019).

Contribution of the Behavioral Observation of Students in Schools to ADHD

Assessment. *School Mental Health*, *11*(3), 464–475. <https://doi.org/10.1007/s12310-019-09313-5>

- Kellam, S.G., Branch, J.D., Agrawal, K.C., & Ensminger, M.E. (1975). Mental health and going to school. *American Journal of Sociology*, *81*(4). University of Chicago Press.
- Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & von der Embse, N. P. (2014). Social, Academic, and Emotional Behavior Risk Screener (SAEBRS). Minneapolis, MN: Theodore J. Christ & Colleagues.
- Kupersmidt, J. B., & Coie, J. D. (1990). Preadolescent peer status, aggression, and school adjustment as predictors of externalizing problems in adolescence. *Child Development*, *61*, 1350-1362.
- Kwon, K., Kim, E., & Sheridan, S. (2012). Behavioral Competence and Academic Functioning Among Early Elementary Children with Externalizing Problems. *School Psychology Review*, *41*(2), 123–140.
- Lane, K. L., Barton-Arwood, M., Nelson, J. R., & Wehby, J. (2007, S). Academic Performance of Students with Emotional and Behavioral Disorders Served in a Self-Contained Setting. *Journal of Behavioral Education*, *17*(1), 43-62. doi:10.1007/s10864-007-9050-1
- Lewis, G. J., Asbury, K., & Plomin, R. (2017). Externalizing problems in childhood and adolescence predict subsequent educational achievement but for different genetic and environmental reasons. *Journal of Child Psychology and Psychiatry*, *58*(3), 292–304.
- Liu J. (2004). Childhood externalizing behavior: theory and implications. *Journal of Child and Adolescent Psychiatric Nursing*. *17*(3), 93–103. <https://doi.org/10.1111/j.1744-6171.2004.tb00003.x>
- McMichael, P. (1979). The hen or the egg? Which comes first- antisocial emotional disorders or reading disability? *British Journal of Educational Psychology*, *49*, 226-238.



- Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic Achievement of K-12 Students with Emotional and Behavioral Disorders. *Exceptional Children, 71*(1), 59-73.  
doi:10.1177/001440290407100104
- New Jersey Administrative Code for Special Education, Title 6A, Chapter 14- Special Education § Section 3.4- Evaluation (2016)
- Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and correlates of DSM-IV conduct disorder in the National Comorbidity Survey Replication. *Psychological medicine, 36*(5), 699–710.  
<https://doi.org/10.1017/S0033291706007082>
- Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. *The Journal of Special Education, 32*(4), 194–206.
- Palfrey, J. S., Levine, M. D., Walker, D. K., & Sullivan, M. (1985). The emergence of attention deficits in early childhood: A prospective study. *Developmental and Behavioral Pediatrics, 6*, 339-348.
- Rapport, M. D., Scanlan, S. W., & Denney, C. B. (1999). Attention-deficit/hyperactivity disorder and scholastic achievement: A model of dual developmental pathways. *Journal of Child Psychology and Psychiatry, 40*(8), 1169-1183.
- Reynolds, C.R., Kamphaus, R.W. (2015). Behavior assessment for children: Third edition. (BASC-3). Bloomington, MN: Pearson.
- Riley-Tillman, T. C., Chafouleas, S. M., Sassu, K. A., & Flazer, A. D. (2008). Examining the agreement of direct behavior ratings and systematic direct observation data of on-task and disruptive behavior. *Journal of Positive Behavior Interventions, 10*(2), 136-143.

- Schrank, F. A., Mather, N., & McGrew, K. S. (2014). Woodcock-Johnson IV Tests of Achievement. *Journal of Psychoeducational Assessment, 33*(4), 391–398.
- Shapiro, E. S. (2004). Academic skills problems workbook (rev.). *New York: Guilford.*
- Villarreal, V. (2015). Test Review: Schrank, F. A., Mather, N., & McGrew, K. S. (2014). Woodcock-Johnson IV Tests of Achievement. *Journal of Psychoeducational Assessment, 33*(4), 391–398.
- Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing student in classroom settings: A review of seven coding schemes. *School Psychology Review, 34*(4), 454-474.
- Vrieze, S.I. (2012) Model Selection and Psychological Theory: A Discussion of the Differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods, 17*, 228-243. <http://dx.doi.org/10.1037/a0027127>
- Willcutt, E. G., & Pennington, B. F. (2000). Comorbidity of Reading Disability and Attention-Deficit/ Hyperactivity Disorder: Differences by Gender and Subtype. *Journal of Learning Disabilities, 33*(2), 179–191. <https://doi.org/10.1177/002221940003300206>
- Willcutt, E. G., Betjemann, R. S., McGrath, L. M., Chhabildas, N. A., Olson, R. K., DeFries, J. C., & Pennington, B. F. (2010). Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. *Cortex, 46*(10), 1345–1361.