August 2019

Examining the Classification Accuracy of the Social, Academic, Emotional Behavior Risk Screener and Its Relationship with Writing Performance

Narmene Hamsho
Syracuse University

Abstract

This study examined the relationship between two school-wide screening measures, one examining classroom behaviors (the Social, Academic, and Emotional Behavior Risk Screener; SAEBRS; Kilgus, Chafouleas, & Riley-Tillman, 2013) and another evaluating writing fluency (Curriculum-Based Measurement in Written Expression; CBM-WE). This study also evaluated the area under the receiver operating characteristic (ROC) curve of the SAEBRS and CBM-WE for identifying students at-risk for writing achievement deficits. A convenience sample of 147 third-grade general education students across two schools, who were determined to not have any significant impairment impacting their writing performance, participated in this study. The index tests (i.e., SAEBRS, CBM-WE) and reference standard (i.e., Wechsler Individual Achievement Test, III Essay Composition subtest; Pearson, 2009) were identified a priori. Of 147 participants, 18 students were identified as at-risk for writing achievement deficits and 129 students were identified as not at-risk. Results indicated a statistically significant relationship between the SAEBRS and CBM-WE for female and male students. In the identification of writing achievement deficits, the area under the ROC curve for CBM-WE revealed fair accuracy (AUC = .761; 95% Confidence Interval [.644, .878]) and poor accuracy was noted for the SAEBRS (AUC = .653; 95% Confidence Interval [.528, .778]). Although there was no statistically significant difference between the independent AUC values, the difference in the qualitative indicator suggest that CBM-WE is a superior screening measure for identifying at-risk students in comparison to the SAEBRS. The findings from this study highlight the contribution of classroom behaviors to the writing process of elementary-aged students and offers support for the use of CBM-WE to identify students at-risk for writing achievement deficits.

Keywords: classroom behaviors, writing performance, classification accuracy, sensitivity

EXAMINING THE CLASSIFICATION ACCURACY OF THE SOCIAL, ACADEMIC, EMOTIONAL BEHAVIOR RISK SCREENER AND ITS RELATIONSHIP WITH WRITING PERFORMANCE

by

Narmene Hamsho

B.A., SUNY University at Buffalo, 2014
M.S., Syracuse University, 2017

Dissertation
Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in *School Psychology.*

Syracuse University
August 2019

Acknowledgments

The completion of this incredible milestone would not be possible without the tremendous support I have received from many people in my life. First, my eternal cheerleaders, my parents. Dad, you have always pushed me to realize my fullest potential, and without your encouragement to apply to doctoral programs, I am not sure I would have ever achieved this accomplishment. Mom, there are not many mothers who would spend hours reading through their daughter's dissertation document to provide the perspective of an outside reader and I wholeheartedly share this success with you. I am so very appreciative of the constant emotional support you have both provided.

I would also like to express my sincere gratitude and deep respect for my advisor, Dr. Tanya Eckert, who provided her continuous support and generous time throughout my graduate career. As a result of your guidance, I have seen a vast improvement in my research writing skills. Thank you for your endless patience. You have inspired me to become an independent researcher and I hope I may one day emulate your skills.

I am extremely grateful for the unparalleled support I have received from my mentor, Dr. Kevin Antshel. I have learned so much from you and on behalf of all the students I have previously worked with and the future students I will one day encounter, I would like to thank you for lending me your clinical expertise. You are a professional model to which I aspire.

I must also express my deepest appreciation for my mentor, Dr. John Garruto. Your dedication to my professional development is apparent. I have thoroughly enjoyed our in-depth discussions as there are not many people in the world with whom I could hold a lively conversation about the fusiform gyrus. I have learned so much from you across my internship year. Thank you for instilling in me the pride of the profession.

I also want to provide an earnest thank you to my big sister, the double doctor, Sara Morelli, M.D., Ph.D. You have been my life-long role model and I am so proud of everything you have accomplished. From an early age, whenever I imagined my future, it always mimicked your path. The example you provided was a large reason I chose to pursue a doctoral degree. Thank you for leaving large shoes to fill, which has always pushed me to reach higher.

Last but certainly not least, words cannot even begin to communicate my sincere gratefulness to my beloved fiancé, David Fogel. David, you were by my side through every single step of this journey. I will never forget your desire to help me succeed in any way possible; so much so that you memorized Cronbach's alpha and at family gatherings you answered questions from relatives about my dissertation study before I even had the chance. I am forever indebted to your faithful support.

Narmene Hamsho

Syracuse University

May 2019

TABLE OF CONTENTS

Examining the Classification Accuracy of the Social, Academic, Emotional Behavior Risk Screener and Its Relationship with Writing Performance

Writing is an essential tool for effective functioning in daily life (MacArthur, Graham, & Fitzgerald, 2016). It is a necessary skill students must develop in order to achieve academic success (Duncan et al., 2007), and it serves as a prerequisite skill required for many occupations (Mikulecky, 1998). Early writing difficulties can result in greater long-term negative risks, including dropping out of high school (Lloyd, 1978) and incarceration (Morrisroe, 2014). As a result, research has focused on developing instructional interventions to remediate writing difficulties (Graham, 2006; Koster, Tribushinina, de Jong, & van den Bergh, 2015; Wanzek, Gatlin, Al Otaiba, & Kim, 2017); however, little attention has focused on examining classroom factors that may concurrently impact students' ability to develop writing skills.

One factor that impacts students' writing development is their behavior in the classroom. Numerous studies have demonstrated a significant relationship between classroom behaviors and students' academic performance in reading and mathematics (Alexander, Entwisle, & Dauber, 1993; Cobb, 1972; Lam & Beale, 1991; McKinney et al., 1975; Wentzel, 1993). In addition, some studies (Dowdy, Doane, Eklund, & Dever, 2013; Kamphaus, Distefano, Dowdy, Eklund, & Dunn, 2010; Kilgus, Bowman, Christ, & Taylor, 2017; Lane, Bruhn, Eisner, & Kalberg, 2010) have demonstrated this relationship, specifically within the areas of mathematics and reading, using school-wide screening measures examining classroom behavior, which are typically used within a multi-tiered system of support approach to identify and intervene with students at-risk for school difficulties. However, few studies examined the relationship that students' classroom behaviors share with their writing performance. Although a considerable number of studies provide support suggesting there is a relationship between students' writing performance and

specific executive functioning skills that are related to students' classroom behaviors, including attention, inhibitory control, and organization (Berninger, Abbott, Cook, & Nagy, 2016; DeBono et al., 2012; Decker, Roberts, Roberts, Stafford, & Eckert, 2016; Hamsho, Antshel, Eckert, & Kates, 2017; Hooper et al., 2011; Hooper, Swartz, Wakely, de Kruif, & Montgomery, 2002; Mayes, Calhoun, Bixler, & Zimmerman, 2009; Molitor, Langberg, & Evans, 2016), no study to date has specifically examined the relationship classroom behaviors shares with students' writing performance.

In this introduction, I review students' writing performance, as well as discuss outcomes in the United States, gender differences in students' writing performance, and current theoretical conceptualizations of writing. The empirical research examining the relationship classroom behaviors share with students' writing performance, as well as related academic areas, including reading and mathematics, are discussed. This introduction highlights previous studies that have examined this relationship by using school-wide screening measures to assess classroom behaviors. Gaps in our current understanding of the relationship between classroom behaviors and writing performance are reviewed.

The Condition of Writing Education

Given that writing difficulties can lead to negative long-term outcomes, there is great cause for concern when evaluating the writing performance of our nation's students. Only 36% of fourth-grade and 34% of eighth-grade students in the United States performed at or above the proficient level on national assessments of writing performance (National Center for Education Statistics, 2012). The proficient level denotes only partial mastery of a skill. This means that a large portion of students in the United States are not demonstrating mastery in the area of writing.

Given that a majority of our nation's students demonstrate deficient writing skills, it is important that educational professionals are accurately identifying students who may need additional support beyond what is provided within the general education classroom. The responsibility of schools to identify and serve students with disabilities is mandated by the Child Find clause of the Individuals with Disabilities Education Act ([IDEA], 2004). IDEA is a federal law that requires schools to provide students classified as having a disability with a free and appropriate public education in the least restrictive environment, which may include the provision of special education and/or related services. IDEA mandates schools to follow the criteria adopted by their state to identify and evaluate students between the ages of 3 and 21 years who are suspected of having a disability.

With regard to a specific learning disability within the area of written expression, the criteria used to determine eligibility varies by state. In New York, eligibility is based upon information from multiple sources (including but not limited to intelligence tests, achievement tests, parent report, social history, adaptive behavior, and teacher recommendations). The Committee on Special Education (CSE) reviews the information gathered from multiple sources regarding the student's writing performance and must conclude that the learning disability is not due to a lack of appropriate academic instruction or the student's limited proficiency in the English language. New York state education law indicates that either of the following two conditions can deem a student eligible for a specific learning disability within the area of written expression: 1) the student demonstrates a profile of academic, cognitive, or behavioral strengths and weaknesses when compared to what is expected given their chronological age, grade-level, or intellectual development, and 2) the student does not demonstrate sufficient academic progress after receiving an evidence-based intervention, also known as response to intervention.

Response to intervention is encompassed within a multi-tiered system of support approach to identify and intervene with students in need of academic and behavioral services. The majority of multi-tiered systems of support feature three tiers (Walker, & Shinn, 2010). The first tier emphasizes the use of school-wide screening procedures, which is an assessment process evaluating all students within a class, grade, or school to identify students at-risk for later difficulties. Tier one also consists of providing high quality, scientifically-based interventions to all students within a class, grade, or school and it requires continual data collection to make informed decisions regarding student progress. Students who are identified as at-risk on school-wide screening measures or students who do not benefit from tier one evidence-based intervention (i.e., a downward or stable slope is demonstrated throughout data collection) are moved to the second tier of the model. Tier two elevates the intensity of services by implementing evidence-based interventions in small group formats with frequent data collection to make informed decisions regarding student progress. Students at tier two who do not demonstrate improved performance are moved to the third and final tier of the model, which once again elevates the intensity of services by individually implementing evidence-based interventions. Students who do not respond to intervention provided at the third tier are then referred to the Committee on Special Education for a comprehensive evaluation to determine eligibility as a student with a disability, which would provide them with access to special education and/or related services.

The multi-tiered systems of support approach is especially beneficial in the identification of a specific learning disability in written expression as results from national assessments indicate a large portion of students are struggling with basic writing skills. By examining response to evidence-based interventions at varied intensities, a multi-tiered systems of support

approach allows for the immediate provision of support without requiring a comprehensive and time-consuming individual evaluation to be completed first. However, multi-tiered systems of support typically do not consider the influences of classroom behaviors as a way to identify at-risk students, explain initial academic struggles, and explain later nonresponse to evidence-based academic interventions. Instead, many school districts identify students at-risk for writing difficulties using school-wide screening measures that directly measure academic performance. In addition, students identified to be at-risk for writing difficulties are provided with academic interventions that aim to increase writing performance by focusing on improving a written product. This approach is partially supported by the current literature, as three recent meta-analyses evaluating the efficacy of writing interventions found that explicit instruction of writing strategies produced a large effect (range, $d = 0.96$ to 1.02) on students' writing performance (Graham, 2006; Graham, McKeown, Kiuhare, & Harris, 2012; Koster et al., 2015). Despite the large effects reported in these meta-analyses, there remain some students for whom the interventions are not effective. Studies, including randomized controlled trials (RCT), rarely examine or report the percentage of students that do not respond to writing interventions. However, in one recent RCT (Hier & Eckert, 2014) that reported a large effect size ($d = 0.89$) for students receiving a performance feedback writing intervention, 34% of the students continued to demonstrate post-intervention writing performance at or below the 25th percentile when compared to their same-aged peers. To improve the identification of students who are at-risk for writing difficulties and provide them with the necessary supports, future research must develop a more comprehensive understanding of factors, such as classroom behaviors, which can influence the writing process in classroom settings.

Theoretical Conceptualization of Writing

Theoretical models provide a basic understanding of the writing process from which future research can be developed. An early theoretical model, the Simple View of Writing (Juel, Griffith, & Gough, 1986), noted two component skills that are linked to students' writing performance: (a) transcription (i.e., motor output required to produce orthographic symbols) and (b) text generation (i.e., translating ideas into linguistic representation and written words). The Not So Simple View of Writing (Berninger & Winn, 2006) expanded upon this model in two ways: (1) by clearly defining the component skill of transcription (i.e., handwriting and spelling) and (2) by adding a third component skill to the model, executive functioning (defined as self-regulatory goal oriented behaviors, including the planning, reviewing, and revising of written work). The Not So Simple View of Writing model suggested that all component skills work together to support students' writing performance. As such, this model became the first theoretical model to recognize a behavioral component (executive functioning) associated with students' writing performance.

Recently, Kim and Schatschneider (2017) sought to extend upon the Not So Simple View of Writing in two ways: (1) to expand our current understanding of the nature of relations the three component skills (transcription, text generation, and executive functioning) share amongst themselves and with writing performance; and (2) to clearly define the component skills of text generation. Kim and Schatschneider evaluated the relationships among the three component skills with a sample of first-grade students and developed the Direct and Indirect Effects Model of Developmental Writing based on their findings. Unlike the Simple View of Writing, structural equation modeling revealed that the component skills of transcription and text generation were not significantly related to each other. Rather, these two component skills

provided independent contributions to first-grade students' writing performance. In other words, a weakness in transcription skills does not necessarily lead to a weakness in text generation skills, and vice versa. In addition, the results of this study confirmed that executive functioning was indirectly related to first-grade students' writing performance via text generation and transcription. Despite this indirect relationship, executive functioning had a substantial total effect ($d = 0.43$) on writing performance, which suggests that executive functioning serves an important role in students' text generation and transcription writing skills.

The aforementioned theoretical models of writing provide a framework for conceptualizing the developmental process of writing. However, only the Direct and Indirect Effects Model of Developmental Writing empirically evaluated the structural relationships between transcription, text generation, and executive functioning, which were assumed by the Simple View of Writing and Not So Simple View of Writing. However, the resulting findings are limited to emerging writers and may not encompass the structural relationships for students at different ages. Further, none of the theoretical models addressed the role of gender in students' writing development, which is an important variable given the extensive empirical evidence suggesting gender differences (Fearrington et al., 2014; Keller-Margulis, Mercer, Payan, & McGee, 2015; Malecki & Jewell, 2003; McMaster et al., 2017).

Gender Differences in Writing Performance

Across the span of several decades, results from national writing assessments have indicated significant gender differences in writing performance. Hedges and Nowell (1995) conducted a comprehensive examination of gender differences reported in national assessments across an eight-year period (1984 to 1992) and found that eleventh-grade female students scored at least a half a standard deviation higher in comparison to their male peers. The results from the

most recent national assessment of writing performance indicated a similar trend, such that eighth- and twelfth-grade female students performed higher in comparison to their same-aged male peers, with differences in standard scores ranging from 14 to 20 points (National Center for Education Statistics, 2012).

These gender differences in overall writing performance have also been demonstrated across curriculum-based measures of writing performance. For instance, statistically significant gender differences in an early study were found in a sample of first-, second-, and third-grade students, such that females demonstrated better performance on curriculum-based measures of letter and word writing fluency (Berninger & Fuller, 1992). These findings were corroborated by more recent studies with samples of first- through eighth-grade students, such that a statistically significant female advantage was demonstrated for both total words written (d = 0.10; 0.23; partial η^2 = .05) and writing fluency (d = 0.11; 0.31; partial η^2 = .05; Fearrington et al., 2014; Malecki & Jewell, 2003; McMaster et al., 2017). In addition to fluency and productivity, the results from Berninger and Fuller (1992) also indicated that first-, second-, and third-grade female students outperformed their male peers when evaluating the number of completed ideas in written text. Finally, gender differences in spelling performance have been demonstrated. In an early study of first- through sixth-grade students, results indicated a statistically significant female advantage on a standardized norm-referenced achievement test of spelling (Allred, 1990). Statistically significant gender differences favoring females within the area of spelling were also found by Malecki and Jewell's (2003) study which utilized a curriculum-based measure with first- through eighth-grade students (partial η^2 = .05).

Previous research examining gender differences with regard to writing quality presents a much less consistent pattern. In one study, a statistically significant female advantage in fourth-

through tenth-grade students was found on a curriculum-based measure using a trait scoring rubric (e.g., conventions, sentence fluency, number of ideas) to evaluate the quality of written stories (|*d*| = 0.42; Troia, Harbaugh, Shankland, Wolbers, & Lawrence, 2013). Similarly, studies utilizing a composite score of writing performance from standardized norm-referenced achievement tests indicated a statistically significant gender difference, such that females between the ages of 5 and 79 outperformed their male peers (|d| = 0.33 to 0.44; Camarata & Woodcock, 2006; Reynolds, Scheiber, Hajovsky, Schwartz, & Kaufman, 2015; Scheiber, Reynolds, Hajovsky, & Kaufman, 2015). However, two studies did not find gender differences in writing quality with a sample of students between the ages of 8 and 11 (Cameron et al., 1995; Williams & Larkin, 2013).

It is important that school professionals consider these gender differences when utilizing school-wide screening tools that measure writing performance with the understanding that female students may outperform their male peers. With regard to research, it is important that future studies consider gender differences when examining contributors to writing performance.

The Relationship between Classroom Behaviors and Academic Performance

In addition to the statistically significant gender differences, there can also be an impact of classroom behaviors on writing performance. Prior to reviewing the more recent studies that examined the relationship between classroom behaviors and writing performance, a large literature base spanning several decades examined the relation of classroom behaviors with academic performance (i.e., reading and/or mathematics). In one of the earliest studies examining this relation, researchers conducted direct observation of classroom behaviors with a sample of 90 second-grade students (McKinney et al., 1975). An overall achievement score from the reading (i.e., vocabulary and comprehension) and mathematics (i.e., concepts and

computation) subtests on the California Achievement Test (CAT; Tiegs & Clark, 1957) served as the outcome measure for this study. Results from multiple regression analysis identified the following classroom behaviors as statistically significant predictors accounting for 33% of the variance in students' reading and mathematics performance: distractibility ($F = 21.06$, $p < .001$), passive disengagement in academic tasks ($F = 7.94$, $p < .006$), dependence on the teacher for help ($F = 7.56$, $p < .001$), and playing or drawing at an inappropriate time ($F = 5.29$, $p < .001$).

Another study (Cobb, 1972) found similar results to McKinney et al. (1975), in that classroom behaviors were significantly related to academic performance in a sample of 103 fourth-grade students. Similar to McKinney et al. (1975), an observational coding system was utilized to assess classroom behaviors and the reading (i.e., comprehension and spelling) and mathematics (i.e., computation and application) subtests of the Stanford Achievement Test (SAT; Stake & Hastings, 1964) was utilized to assess academic performance. Results from stepwise regression analysis indicated that the following classroom behaviors were statistically significant predictors of a combined reading and spelling construct: talking to peers about the academic task ($r = .42$) and out-of-seat behavior ($r = -.25$). When examining mathematics performance, on-task behavior ($M r = .44$) was found to be the only significant predictor.

The results from McKinney et al. (1975) and Cobb (1972) emphasized the important role of classroom behaviors with respect to students' mathematics, reading, and spelling performance. However, there are considerations in the classroom behavior assessment methods used in these two studies. First, although direct classroom observations are considered the gold standard for assessing students' classroom behavior (Wilson & Reschly, 1996), there is evidence to suggest that direct observation methods require extensive data collection and sampling of students' classroom behavior (i.e., one 15-minute observation per day across 3 days, totaling 45 minutes of

direct observation per student) in order to achieve acceptable levels of reliability (Hintze & Matthews, 2004). Due to practical concerns, teachers, school personnel, and researchers may not be able to achieve this threshold, which would ensure the reliability of the collected data. Both McKinney et al. and Cobb's studies did not achieve the prerequisite recommendations for conducting direct observation techniques, with McKinney et al. approaching the threshold (i.e., one 5-minute observation per day across 4 days, totaling 20 minutes of direct observation per student) and Cobb falling severely below threshold (i.e., one 10-second observation per day across 9 days, totaling 1 and a half minutes of direct observation per student). As a result, the findings from these two studies should be interpreted with some caution, as their observation methods may not have reliably measured the typical behaviors that students display in classrooms. In addition, the classroom behavior observation systems used in the prior studies were narrow in focus as they primarily evaluated classroom behaviors related to on-task academic engagement. Additional behaviors (e.g., aggressiveness, compliance) that can occur within the classroom environment and potentially contribute to academic performance were not assessed. Further, the feasibility of conducting direct classroom observations given the time, training, and additional resources needed limits the extent to which this assessment method can be routinely used in school settings within the context of a multi-tiered systems of support approach.

To address these concerns, more recent studies examined the relationship between classroom behaviors and academic performance by using informant reports (i.e., teacher rating scales) of classroom behavior. Unlike direct observation techniques, rating scales require limited resources and can be used as part of a school-wide screening approach to identify students in need of additional classroom supports (Nantais, Martin, & Barnes, 2014; Walker & Shinn, 2010).

In one of the first studies to use a teacher-report measure, Alexander et al. (1993) assessed the classroom behaviors of 790 first-grade students with a 13-item teacher-report measure adapted from the National Survey of Children (Zill, Furstenberg, Peterson, & Moore, 1992) to assess classroom behaviors, including (a) Interest-Participation (e.g., expresses ideas, cheerful, creative); (b) Cooperation-Compliance (e.g., teases peers, irritable, not considerate of others); and (c) Attention Span-Restlessness (e.g., demonstrates restlessness, high strung, doesn't concentrate). The psychometric evidence provided for this scale indicated acceptable coefficient alphas (α range: .74 to .82) for the overall measure, however no information regarding the validity of this scale was reported. The reading comprehension and mathematics concepts/reasoning subtests of the California Achievement Test (CAT; CTB Macmillan/McGraw-Hill, 1992) served as the outcome measure in this study.

The results from this study substantiate the findings from McKinney et al. (1975) and Cobb (1972), such that there was a statistically significant association between classroom behaviors and academic performance. More specifically, the results of the regression analysis revealed that 22% of the variance in reading performance was related to teacher ratings of Attention Span-Restlessness ($\beta = .30$, $p \leq .01$), teacher ratings of Interest-Participation ($\beta = .31$, $p \leq .01$), and teacher ratings of Cooperation-Compliance ($\beta = -.10$, $p \leq .05$). With regard to mathematics performance, 41% of the variance was explained by teacher ratings of Attention Span-Restlessness ($\beta = .19$, $p \leq .01$) and teacher ratings of Interest-Participation ($\beta = .28$, $p \leq .01$).

Another study (Finn, Pannozzo, & Voelkl, 1995) of 1,013 fourth-grade students using a different teacher-report measure provided further support for the findings of Alexander et al. (1993). The Student Participation Questionnaire (Finn, Folger, & Cox, 1991) is a 29-item

measure that assesses two factors: (a) Disruptive Behavior; and (b) Inattentive Behavior. High internal consistency across each of the two subscales has been reported (α range = .89 to .94; Finn et al., 1991), although there is currently no information regarding the validity of this scale. Reading and mathematics performance was assessed using a norm-referenced achievement measure (Comprehensive Test of Basic Skills, 1981).

The results of this study indicated that students with lower inattention scores performed approximately one standard deviation greater in reading and mathematics when compared to their same-aged peers with higher inattention scores. Additionally, students with lower disruptive behavior scores performed approximately a half of a standard deviation greater in reading and mathematics when compared to their same-aged peers with higher scores of disruptive behaviors. A multivariate analysis of variance (MANOVA) indicated that students who did not display any disruptive or inattentive behaviors performed over half a standard deviation greater in reading ($d = .69$, $p < .001$) and mathematics ($d = .72$, $p < .001$) when compared with students who were classified as either disruptive or inattentive. These findings further emphasize the significant relationship between classroom behaviors and academic performance.

Alexander et al. (1993) and Finn et al. (1995) were one of the earliest large-scale studies to examine the relationship between classroom behaviors and students' academic performance using teacher-report measures of classroom behaviors. Despite the differences in the procedures used to collect classroom behavioral information, the results from these two studies align with the findings reported by Cobb (1972) and McKinney and colleagues (1975) suggesting a statistically significant relationship between on-task classroom behaviors and students' academic performance, as well as extend the results to include other classroom behaviors such as prosocial

competence and compliance. Although Alexander et al. and Finn et al. evaluated a wider range of classroom behaviors than the direct observational systems utilized by Cobb and McKinney and colleagues, there was limited psychometric evidence reported for these rating scales. Information regarding the internal consistency of each scale was provided, however there is currently no research to support the validity of these scales. As a result, it is difficult to conclude whether the rating scales were accurately measuring the purported constructs of classroom behaviors.

To date, three studies have examined the relationship between classroom behaviors and academic performance using teacher-report measures with more extensive psychometric support. In the first study, Rabiner and Coie (2000) examined the relationship between classroom behavior and reading performance with 387 elementary-aged students by using the ADHD rating scale (DuPaul, 1991), which is a 14-item teacher-report measure assessing inattention-hyperactivity, impulsivity-hyperactivity, and total behavioral functioning. Psychometric support (DuPaul, 1991) for the scale includes evidence of high internal consistency across each of the two subscales (α range = .94 to .96) as well as support for the criterion validity with the following measures: direct observation of on-task behaviors (r = -.53, p < .001) and the Abbreviated Conner's Teacher Rating Scale (Goyette, Conners, & Ulrich, 1978; r = .90, p < .001). The Letter-Word Identification subtest from the Woodcock-Johnson Psychoeducational Battery-Revised (Woodcock & Johnson, 1989), a norm-referenced test of academic achievement, was administered to evaluate reading performance.

The results of this study did not find hyperactivity to be related to students' reading performance, however a statistically significant relationship was reported between attention and reading performance (β = -.29, p < .01), providing further support to the results from the

previously described studies suggesting a statistically significant relationship between classroom behaviors specifically related to on-task engagement and reading performance. Although the results corroborate prior findings reported between on-task engagement and academic performance using direct observation techniques (Cobb, 1972; McKinney et al., 1975) and teacher-report methods (Alexander et al., 1993; Finn et al., 1995), the measure used in this study was developed specifically to inform ADHD classification decisions and does not incorporate a broad assessment of behaviors that can occur within the classroom settings (e.g., interest in academic topics, relationships with peers, compliance).

In a second study, Lam and Beale (1991) investigated the relationship between classroom behaviors and reading performance with 190 elementary-aged students by using the inattention construct of the Conners' Teacher Rating Scale (CTRS; Werry & Hawthorne, 1976), a 39-item teacher-report measure. Previous studies evaluating the psychometric properties of the CTRS indicate high test-retest reliability across 8 days (r = .88 to .96; Edelbrock, Greenbaum, & Conover, 1985) and moderate to large correlation coefficients (r = .50 to .80, $p < .01$) with the Pittsburgh Adjustment Survey Scale (PASS; Ross, Lacey, & Parton, 1965), another teacher report measure of classroom behavior (Camp & Zimet, 1974). The vocabulary and comprehension subtests of the Progressive Achievement Test (PAT; Reid & Elley, 1991), a norm-referenced achievement measure, was administered to evaluate reading performance.

The results of this study found inattention to be a statistically significant predictor of reading performance (R^2 = .09), which aligns with the findings from previous studies utilizing both direct observation and informant reporting methods. However, similar to Rabiner and Coie, Lam and Beale utilized a classification clinical measure that not only narrowly focused upon inattentive behaviors, ignoring a wide-range of additional classroom behaviors, but also would

be infeasible to implement as a school-wide screening measure due to its lengthiness. Given the results of Alexander et al. (1993) and Finn et al. (1991), which emphasized the importance of on-task as well as social behaviors to academic performance, it is important to examine the relationship between a broad assessment of classroom behaviors and academic performance.

In the third study, Barriga et al. (2002) investigated the relationship between classroom behaviors and academic performance with a sample of 58 students (mean age = 15.02) using the Teacher's Report Form (TRF; Achenbach, 1991), a multidimensional assessment of behavior. This 112-item measure produces eight behavioral factors: Withdrawal, Somatic Complaints, Anxiety/Depression, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior. Psychometric support (Achenbach, 1991) for this scale includes evidence of test-retest reliability (r = .92) across 15 days as well as construct validity with the following measures: Conners' Teacher Rating Scale-Revised (C-TRS-R; Conners, Sitarenios, Parker, & Epstein, 1998; r = .71 to .85) and the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 2004; r = .40 to .87). An overall achievement score from the reading (i.e., letter and word recognition), spelling (i.e., writing dictated words), and mathematics (i.e., number recognition and computation) subtests on the Wide Range Achievement Test, third edition (WRAT3; Wilkinson, 1993), a norm-referenced achievement test, served as the outcome measure for this study.

The results found that Inattention (r = -.44, p < .001), Delinquent Behavior (r = -.28, p < .05) and Aggression (r = -.28, p < .05) were significantly correlated with an overall composite of academic performance. The results of multiple regression analyses indicated that Withdrawal (F = 5.33, R^2 = .16, p < .01), Somatic Complaints (F = 5.83, R^2 = .17, p < .01), Delinquent Behavior (F = 4.70, R^2 = .16, p < .05), and Aggressive Behavior (F = 5.34, R^2 = .16, p < .01) were

statistically significant predictors of academic performance. Similar to the limitations associated with the prior studies reviewed (Lam & Beale, 1991; Rabiner & Cole, 2000), the informant report measure used in this study was specifically developed to assess internalizing and externalizing behavior disorders to inform psychiatric diagnoses and the length makes it infeasible to be implemented as a school-wide screening measure within the context of a multi-tiered systems of support approach.

Despite the varying techniques utilized to assess classroom behaviors (e.g., direct observation versus teacher rating scales), the overall findings from all of the studies were consistent: academic classroom behaviors were significantly related to academic performance. Of the seven studies reviewed, only one study (Barriga et al., 2002) utilized a validated measure to broadly assess a wide range of classroom behaviors and the findings emphasized the importance of considering the contribution of classroom behaviors beyond on-task engagement (more specifically social behaviors and internalizing problems) to academic performance. However, the lengthiness of the C-TRS-R utilized by Barriga et al. (2002) would be prohibitive for school personnel especially if used within the context of school-wide screenings to identify at-risk students.

The Relationship between School-Wide Behavioral Screening Measures and Academic Performance

School-wide screening measures are brief assessments utilized within a multi-tiered systems of support approach that serves three purposes: 1) identify at-risk students who require evidence-based interventions administered at a higher level of intensity in addition to what is initially provided within the general education curriculum, 2) provide feedback to school administrative officials to identify teachers who are in need of additional support to enhance their

classroom management strategies and instructional practices, and 3) to prevent false negatives when identifying at-risk students by examining student performance approximately three times a year and immediately providing support to identified students (Johnson, Mellard, Fuchs, & McKnight, 2006). Until recently, no previous study utilized school-wide screening measures to examine the relationship between classroom behaviors and academic performance. Unlike the classification clinical measures utilized in previous research examining this relation, the efficiency of school-wide screening measures allows for a large amount of data to be collected for nearly all students within a class, grade, or school without requiring a significant amount of time. Due to this fact, school-wide screening measures can be implemented multiple times within an academic year allowing school personnel to provide immediate support to identified students as opposed to delaying services while waiting for teacher referrals, which can be inaccurate or biased (Eklund et al., 2009; Green, 1996; Lloyd, Kauffman, Landrum, & Roe, 1991). To date, four studies have examined the relationship between classroom behavior and academic performance using school-wide screening measures.

In the first study, Lane et al. (2010) investigated the relationship between classroom behaviors and academic performance with 534 middle school students by using the Student Risk Screening Scale (SRSS; Drummond, 1994), a 7-item teacher-report measure assessing antisocial behavior. Previous studies evaluating the psychometric properties of the SRSS indicate high internal consistency (α range = .78 to .85; Lane, 2007) and a strong correlation (r = .79; Walker, Ramsey, & Gresham, 2004) with the Aggressive Behavior factor of the Child Behavior Checklist (Achenbach, 1991). Support for the classification accuracy of the SRSS indicates high sensitivity (.94), high specificity (.95; Lane et al., 2009), and sufficient evidence of predicting externalizing behaviors as measured by the Systematic Screening for Behavior Disorders (SSBD;

Walker & Severson, 1992) by 45% (AUC = .952; Lane et al., 2012). Cumulative grade point average (GPA) for the school year served as a measure of academic performance.

The results from one-way analysis of variance (ANOVA) in this study revealed that students who demonstrated acceptable social classroom behaviors had a GPA two standard deviations higher than their peers who demonstrated antisocial classroom behaviors ($F(1, 112) = 82.48$, $p < .0001$, $d = -1.94$). Despite the different measures implemented to assess classroom behaviors (e.g., school-wide screening measures versus classification clinical measures) this study aligns with the findings from the previously reviewed articles (Alexander et al., 1993; Barriga et al., 2002; Finn et al., 1995) suggesting a significant relationship between social classroom behaviors and academic performance. However, GPA is not considered a reliable measure of academic performance as it may be influenced significantly by teacher bias and instructional practices (Lei, Bassiri, & Schulz, 2001). Additionally, the SRSS primarily focuses upon social behaviors and does not include an assessment of on-task academic behaviors, an area that multiple studies (Alexander et al., 1993; Barriga et al., 2002; Cobb, 1972; Finn et al., 1995; Lam & Beale, 1991; McKinney et al., 1975; Rabiner & Coie, 2000) have identified to be strongly correlated with academic performance. In addition, the SRSS solely evaluates maladaptive classroom behaviors without considering the contribution of adaptive classroom behaviors to academic performance. By excluding these classroom behaviors, the results of this study may only partially explain the contribution of classroom behavior to academic performance.

In the second study, Kamphaus et al. (2010) examined the relationship between classroom behaviors and reading performance with 309 elementary-aged students by using the Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), which is a

27-item teacher-report screening measure assessing behavioral and emotional strengths and weaknesses. Psychometric support for the scale includes evidence of high internal consistency (α = .939) as well as support for the criterion validity (Kamphaus & Reynolds, 2007; Kamphaus et al., 2010), which indicated a statistically significant correlation with the following measures: 1) across several factors of the full teacher report measure on the Behavior Assessment System for Children, second edition (BASC-2; Reynolds & Kamphaus, 2004; r = .523 to .820, p < .05), 2) across several factors of the Conners' Teacher Rating Scale Revised (Conners, 1997; r = .73 to .79), and 3) the total problems score from the Achenbach System of Empirically Based Assessment Teacher Report Form (Achenbach & Rescorla, 2001; r = .76). Research examining the classification accuracy of the BESS indicated high specificity (.95), high sensitivity (.80), a strong negative predictive value (.96; Kamphaus & Reynolds, 2007), and an ability to identify additional students (27%) who were not initially referred by teachers as needing services (Eklund et al., 2009). Academic performance was determined from the English/language arts and mathematics subtests of the California Standards Test (California State Board of Education, 2010a, 2010b), a criterion-referenced academic achievement test.

Results of correlational analyses indicated a negative relationship between behavioral functioning and both mathematics (r = -.447, p < .05) and reading performance (r = -.432, p < .05), suggesting that increased behavioral problems was associated with decreased academic performance. Despite the different measures utilized to assess classroom behaviors and academic performance, the results from Kamphaus et al. (2010) align with the findings from Lane et al. (2010), suggesting a statistically significant negative correlation between problem behaviors and academic performance.

In the third study, Dowdy et al., (2013) also used the BESS to examine the relationship between classroom behaviors and academic performance with a sample of 849 elementary and middle school students. Dowdy et al. (2013) utilized end of the academic year report card grades to examine academic performance. Results from a one-way, between-groups multivariate analysis of variance (MANOVA) indicated that students who demonstrated decreased behavioral problems received end of the year report card grades for mathematics that was approximately one standard deviation higher ($F(6, 804) = 37.93$, $p < .001$, partial $\eta^2 = .12$) and end of the year report card grades for reading that was approximately half a standard deviation higher ($F(6, 804) = 28.75$, $p < .001$, partial $\eta^2 = .10$) than peers who demonstrated increased behavioral problems. Similar to GPA, report card grades are an unstandardized assessment of academic performance that can impact the reliability of the findings reported in this study. Despite this fact, the results align with previous research (Kamphaus et al., 2010; Lane et al., 2010) suggesting a significant relationship between classroom behaviors and academic performance. Although the BESS is simple and assesses a broad array of behaviors that may occur within the classroom setting, it is cost-prohibitive and requires extensive teacher time.

The fourth study (Kilgus et al., 2017) examined the relationship between classroom behaviors and academic performance using the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS; Kilgus et al., 2013) in 1,058 elementary and middle school students. This 19-item criterion-referenced teacher rating scale assesses four classroom behavior factors: social behavior, academic behavior, emotional behavior, and a total factor. High internal consistency across all four factors ($\alpha = .83$ to $.93$) and strong criterion validity with the BESS ($r = -.75$ to $-.94$; Reynolds & Kamphaus, 2004) and the SRSS ($r = -0.84$; Drummond, 1994) has been reported (National Center on Intensive Intervention, 2018). Research examining the

classification accuracy of the SAEBRS across three time points during the academic year indicated high specificity (range = .95 to .99), high sensitivity (range = .72 to .95), a strong negative predictive value (range = .93 to .99), and excellent classification accuracy (range = .93 to .99; National Center on Intensive Intervention, 2018). Academic performance was evaluated through use of the Adaptive Reading (i.e., aReading, which assesses phonological awareness, vocabulary, and comprehension) and Adaptive Math (i.e., aMath, which assesses geometry, computation and concepts), which are norm- and criterion-referenced academic screening measures (Christ et al., 2014).

The results of this study indicated that the total classroom behavior score on the SAEBRS was significantly related to both mathematics ($r = .335$, $p < .01$) and reading performance ($r = .373$, $p < .01$). Multiple regression analysis indicated a statistically significant model in which all three SAEBRS factors (i.e., social, academic, and emotional behaviors) accounted for 27% of the variance of the overall composite for academic performance ($R^2 = .271$, $F(3,1,054) = 130.75$, $p < .001$). However, only academic (e.g., on-task behavior and academic productivity; $\beta = .641$, $p < .001$) and social (e.g., interpersonal relations and compliance; $\beta = -.225$, $p < .001$) classroom behaviors were significantly related to academic performance. Emotional classroom behaviors were not a statistically significant predictor of academic performance ($\beta = .045$, $p > .05$).

When considering the findings of the earlier studies investigating the relationship between classroom behaviors and academic performance (Alexander et al., 1993; Barriga et al., 2002; Cobb, 1972; Finn et al., 1995; Lam & Beale, 1991; McKinney et al., 1975; Rabiner & Coie, 2000), the results from these four recent studies (Dowdy et al., 2013; Kamphaus et al., 2010; Kilgus et al., 2017; Lane et al., 2010) provide further support for the significant relationship between classroom behaviors and academic performance. Given the previously

stated advantages of using school-wide screening measures over clinical classification measures, the results from the most recent studies emphasize the benefits of using school-wide screening measures as part of a multi-tiered system of support approach to identify students at-risk for academic difficulties.

All three of the measures reviewed in this section (the SRSS, the BESS, and the SAEBRS) demonstrate strong psychometric support for their use as school-wide screening measures assessing classroom behaviors, however there are distinct advantages to using the SAEBRS to examine the relationship between classroom behaviors and academic performance. First, previous research (Kilgus et al., 2017) identified a statistically significant relationship between the SAEBRS and a standardized measure of academic performance, whereas the studies that have investigated this relationship using the SRSS (Dowdy et al., 2013; Lane et al., 2009) utilized an unstandardized measure that is highly susceptible to informant bias to assess academic performance. Although one study has identified a significant relationship between the BESS and a standardized measure of academic performance (Kamphaus et al., 2010), the BESS is cost-prohibitive and the lengthiness of this scale requires extensive teacher time. Therefore, the second distinct advantage of the SAEBRS is that unlike the BESS, the SAEBRS is feasible to implement as a school-wide screening measure because it does not require a significant amount of teacher time to complete and it is free. Third, unlike the SRSS, the SAEBRS provides a comprehensive assessment of classroom behavioral functioning by breaking down classroom behaviors into three domains related to students' academic performance (Kilgus et al., 2017).

All the studies reviewed thus far, including earlier research that utilized direct observation and clinical classification measures, identified a statistically significant relationship between classroom behaviors and academic performance. Of these studies, only three (Barriga

et al., 2002; Cobb, 1972; McKinney et al., 1975) incorporated an assessment of writing performance when evaluating academic performance. However, within the three studies, writing performance was incorporated within an overall composite of either reading performance or academic performance making it difficult to determine the specific relationship between classroom behaviors and writing performance. In addition, students' spelling performance was used as the primary indicator of writing performance for all three studies. This limits our understanding of the relationship between classroom behaviors and writing performance as spelling primarily focuses upon the component skill of transcription within the Not So Simple View of Writing (Berninger & Winn, 2006) and largely ignores the component skills of text generation. Given that early writing difficulties can result in greater long-term negative risks (Lloyd, 1978; Morrisroe, 2014), it is important that research studies evaluate factors, such as classroom behaviors, that may influence writing performance. Such research may inform multi-tiered systems of support to identify and intervene with at-risk students for writing difficulties. Until recently, little research has examined the relationship between classroom behaviors and writing performance.

The Relationship between Classroom Behaviors and Writing Performance

In comparison to the extensive literature base examining the relationship that classroom behaviors share with mathematics and reading performance, studies examining the relationship between classroom behaviors and writing performance are limited but have grown within recent years. To date, three studies have examined the relationship between classroom behaviors and writing performance. In one of the earliest studies, Kent, Wanzek, Petscher, Al Otaiba, and Kim (2014) longitudinally examined the relationship between classroom behaviors and writing performance across one-year with 265 kindergarten students by using the Strengths and

Weaknesses of ADHD Symptoms and Normal Behavior Rating Scale (SWAN; Swanson et al., 2006). The SWAN is a 30-item teacher-report measure assessing selective attention (Sáez, Folsom, Al Otaiba, & Schatschneider, 2012) and was completed when the students were in kindergarten. High internal consistency across all 30 items ($\alpha = .99$) of the SWAN has been reported (Kent et al., 2014). No study has evaluated the validity of the full 30-item teacher-report version of the SWAN. However, psychometric evidence (Lakes, Swanson, & Riggs, 2011) provides support for the convergent validity of the parent-report of the shortened versions of the SWAN (the first 18-items of the 30-item measure) with the following measures: the parent-report of the Disruptive Behavior Rating Scale ($r = .54$, $p < .01$; DuPaul et al., 1998) and the parent-report on the Strengths and Difficulties Questionnaire (Cramer's V = .53; Goodman & Scott, 1999). Within Kent et al.'s (2014) study, writing performance was assessed by scoring written responses to a story prompt from a curriculum-based measure (McMaster, Du, & Pétursdóttir, 2009) for writing productivity (i.e., total number of words written) in kindergarten and writing fluency (i.e., correct writing sequences; a measure of writing speed and accuracy) and quality (i.e., 6+1 traits rubric, which assesses organization and theme development) in the first-grade.

Results from structural equation modeling (CFI = .965, TLI = .954, RMSEA = .076, SRMR = .050) indicated a statistically significant relationship between classroom behaviors and writing productivity in kindergarten ($\gamma = .16$, $p = .001$) after controlling for early literacy skills (e.g., handwriting fluency, oral language, and word reading). Further analyses suggested that a model including classroom behaviors demonstrated a better fit when compared with a model only including early literacy skills ($\chi^2 = 73.5$, df = 4, $p < .001$), further emphasizing the contribution of classroom behavior to writing performance. An examination of the longitudinal

relationship between kindergarten classroom behaviors and first-grade writing performance using structural equation modeling (CFI = .964, TLI = .953, RMSEA = .061 SRMR = .047) indicated a statistically significant relationship between classroom behaviors and writing fluency ($\gamma = .23$, $p < .001$) as well as quality ($\gamma = .19$, $p = .001$), after accounting for early literacy skills.

In a second study, Kim, Al Otaiba et al. (2015) investigated the relationship between classroom behaviors and writing performance with a cross-sectional sample of 494 second- and third-grade students using a shortened version (the first 9-items of the original 30-item measure) of the same teacher-report measure (the SWAN) as Kent et al. (2014), which solely evaluates regulation of attention (Sáez et al., 2012). High internal consistency across all 9-items ($\alpha = .91$) of the SWAN has been reported (Kim, Al Otaiba et al., 2015). However, no research has evaluated the validity of only the first 9-items of the SWAN. Writing performance was assessed by scoring written responses to a story prompt from a curriculum-based measure (McMaster et al., 2009) and from the essay composition subtest of a standardized norm-referenced achievement test (Wechsler Individual Achievement Test-Third Edition, WIAT-III; Wechsler, 2009) for writing fluency, quality, and productivity.

Results from confirmatory factor analysis and multilevel modeling indicated that classroom behavior was a significant predictor of writing fluency ($\beta = .09$, $p = .02$) and writing quality ($\beta = .005$, $p = .03$) but was not a statistically significant predictor for writing productivity (CFI = .90, TLI = .88, RMSEA = .11, SRMR = .083). Further analyses to examine gender differences in writing performance indicated that female students performed significantly better (d range = .37 to .46) across all three writing outcome measures in comparison to their male peers. In a model that examined the relationship between gender and writing after accounting for classroom behaviors, females continued to demonstrate statistically and significantly greater

scores (*d* range = .22 to .34) when compared to their male peers for all three writing outcomes. However, in this second model that included gender, classroom behaviors were found to be statistically significant predictors for writing fluency alone (β = .10) and were no longer related to writing quality. Similar to Kent et al. (2014), the results from Kim, Al Otaiba et al. (2015) emphasized the important role of classroom behaviors, more specifically attention, with respect to students' writing performance. Kim, Al Otaiba et al. (2015) extended our understanding of this relationship by considering gender differences in students' writing performance.

In a third study of 80 third-grade students, Hamsho (2017) examined gender differences in the relationship between classroom behaviors and writing performance within the context of a performance feedback intervention aimed at increasing writing fluency. Similar to Kim, Al Otaiba et al. (2015), this study utilized the first nine-items of the original 30-item SWAN to assess attention. In addition, a second teacher-report measure, the Academic Performance Rating Scale (APRS; Dupaul, Rapport, & Perriello, 1991), was included to measure classroom behaviors. The APRS is a 22-item rating scale that assesses three factors: (a) Academic Success, (b) Impulse Control, and (c) Academic Productivity. Psychometric support for this scale includes evidence of acceptable to high internal consistency (α range = .72 to .94), with the Impulse Control factor demonstrating the lowest Cronbach's alpha (Dupaul et al., 1991). Additionally, the APRS evidenced variable criterion validity (DuPaul et al., 1991) when compared with the following measures: (a) The ADHD Rating Scale, teacher report (DuPaul, 1991; *r* = -.72), (b) direct observations of on-task behavior (*r* = .29), and (c) percentage of assignments completed accurately (*r* = .53). Writing performance was assessed by scoring written responses to a curriculum-based measure story prompt (McMaster et al., 2009) for writing fluency.

Results from stepwise linear regression analysis revealed that classroom behaviors assessed through the SWAN ($R^2 = .099$, $F_{(1,39)} = 4.29$, $p = .045$) and the APRS ($R^2 = .212$, $F_{(1,39)} = 10.50$, $p = .002$) were a statistically significant predictor of post-intervention writing fluency for female students but not for male students. The results from this study build upon the findings from Kent et al. (2014) suggesting that a statistically significant relationship continues to exist between classroom behaviors and writing performance within the context of a tier 1 writing intervention. Similar to Kim, Al Otaiba et al. (2015), the Hamsho (2017) study considered the impact of gender on writing performance and identified classroom behaviors to be a significant predictor of post-intervention writing performance for female students but not for male students. In contrast to the previously mentioned two studies (Kent et al., 2014; Kim, Al Otaiba et al., 2015), the Hamsho study incorporated a rating scale, the APRS, that was specifically developed for use within the school system and accounted for behavioral factors unique to the classroom environment, whereas the prior two studies relied solely upon a clinical measure, the SWAN, primarily used for classification purposes. However, similar to the SWAN, the APRS has a narrow focus upon classroom behaviors impacting on-task performance and ignores a wide-range of behaviors that may occur within the classroom environment, including behaviors (e.g., compliance, aggressiveness) that have been previously demonstrated to impact students' academic performance (Alexander et al., 1993; Barriga et al., 2002; Dowdy et al., 2013; Finn et al., 1995; Kamphaus et al., 2010; Kilgus et al., 2017; Lane et al., 2010).

In summary, the overall findings from all three studies were consistent: classroom behaviors were statistically significant contributors to writing performance. Furthermore, of the two studies that considered gender differences in writing performance (Hamsho, 2017; Kim, Al Otaiba et al., 2015), findings consistently suggested that the relationship between classroom

behaviors and writing performance was influenced by gender. Only one of these studies (Hamsho, 2017) incorporated a school-based measure utilized to evaluate classroom behaviors, the APRS, however this measure does not meet conventional psychometric standards to be used as a school-wide screening measure due to limited evidence regarding classification accuracy. No study to date has utilized a school-wide screening measure that broadly evaluates a wide range of classroom behaviors to examine gender differences in the relationship between classroom behaviors and writing performance.

Purpose of the Proposed Study

Several studies have demonstrated the important contribution of classroom behaviors to students' academic performance in reading and mathematics (Alexander et al., 1993; Barriga et al., 2002; Cobb, 1972; Finn et al., 1995; Kilgus et al., 2017; McKinney et al., 1975). Recent studies (Dowdy et al., 2013; Kamphaus et al., 2010; Kilgus et al., 2017; Lane et al., 2010) highlight the benefits of using school-wide screening measures as part of a multi-tiered system of support to identify at-risk students for school difficulties, including an emphasis on students' reading and mathematics performance. However, the empirical literature examining the relationship between classroom behaviors and writing performance is limited, with only three studies (Hamsho, 2017; Kent et al., 2014; Kim, Al Otaiba et al., 2015) examining this topic. Two of these three studies (Kent et al., 2014; Kim, Al Otaiba et al., 2015) narrowly defined classroom behaviors by using a unidimensional teacher rating scale of student attention developed for classification purposes. Although the third study (Hamsho, 2017) incorporated a school-based measure, it primarily examined on-task behaviors and does not meet conventional psychometric standards for school-wide screening measures. No study to date has evaluated the relationship between classroom behaviors and writing performance using school-wide screening

measures. Given that early writing difficulties can lead to negative long-term outcomes (Lloyd, 1978; Morrisroe, 2014), understanding the relationship between classroom behaviors and writing performance is important as it may inform multi-tiered systems of support to identify and intervene with at-risk students for writing difficulties and may ultimately lead to more positive long-term outcomes for these students.

The primary aim of the proposed study was to identify the relationship between the two domains addressed by a multi-tiered systems of support approach: classroom behaviors (as measured by the Social, Academic, and Emotional Behavior Risk Screener; SAEBRS; Kilgus et al., 2013) and academic performance, in particular writing performance (a Curriculum-Based Measure of Written Expression; CBM-WE), among female and male third-grade students. The SAEBRS was chosen over alternative measures because it is feasible to implement, it meets conventional psychometric standards for use as a school-wide screening measure (National Center on Intensive Intervention, 2018), and previous research has indicated a statistically significant relationship between the SAEBRS and standardized academic achievement measures (Kilgus et al., 2017). The CBM-WE was chosen over a standardized norm-referenced academic achievement measure of writing performance because it meets conventional psychometric standards for use as a school-wide screening measure.

A secondary aim of the proposed study was to evaluate the classification accuracy of these two screening measures (the SAEBRS and CBM-WE) for identifying students at-risk for writing achievement deficits. The WIAT-III was utilized as the reference standard to determine which students demonstrated at-risk writing achievement performance. The WIAT-III was utilized as the reference standard instead of a curriculum-based measure because it is the most

frequently used standardized norm-referenced academic achievement measure (Benson et al., 2019).

To address the aims of the proposed study, the following two research questions were proposed:

(1) To what extent are the factors of the SAEBRS related to CBM-WE among female and male third-grade students? Similar to the results from Kilgus et al. (2017), CBM-WE was expected to be related to two factors of the SAEBRS (i.e., social and academic); however, the SAEBRS total composite score including all three factors was hypothesized to be significantly related to CBM-WE. This hypothesis is based on previous research findings suggesting a significant relationship between classroom behaviors and writing performance (Kent et al., 2014; Kim, Al Otaiba, et al., 2015) and extrapolated from an earlier study that utilized the SAEBRS (Kilgus et al., 2017).

(2) How does the classification accuracy of the SAEBRS compare to the classification accuracy of CBM-WE when identifying students at-risk for writing achievement deficits? It was hypothesized that the SAEBRS would demonstrate better classification accuracy than the CBM-WE. Only one study has examined the classification accuracy of the CBM-WE in its ability to predict writing achievement (AUC = .74; Furey, Marcotte, Hintze, & Shackett, 2016), and no study has evaluated the classification accuracy of the SAEBRS when identifying students at-risk for academic performance difficulties. Therefore, this hypothesis was not based on any previous research examining the classification accuracy of the SAEBRS. Instead this hypothesis was derived from the results of earlier research suggesting a significant relationship between the SAEBRS and students' academic performance (Kilgus et al., 2017), as well as from recent theoretical conceptualizations of writing performance that have emphasized the importance of a

behavioral component to students' writing performance. It was expected that the SAEBRS would demonstrate higher levels of classification accuracy in comparison to CBM-WE as writing fluency narrowly assesses one area of writing achievement (i.e., speed and accuracy of written text) whereas classroom behaviors may broadly influence additional areas of writing achievement (i.e., text generation and organization of written text).

Method

Participants and Setting

Approval from the Institutional Review Board and from the participating school district was attained before commencement of the study. In addition, parent consent, student assent, and teacher consent were obtained. Students who received parental permission to participate in this study were required to sign an assent form. This form formalized their agreement to participate in the study. This was a convenience sample of students; the participating students were enrolled in the third-grade at the two elementary schools that granted permission to collect data.

After attaining all necessary approval, third-grade students were screened for eligibility. Students who fit the eligibility criteria: (a) did not have any serious motor deficits (e.g., neurological conditions) that impacted their writing performance; (b) did not have serious cognitive impairments (e.g., intellectual disability, traumatic brain injury, autism with accompanying intellectual impairment) which may have impacted their writing performance; (c) did not have any significant hearing or vision impairments; and (d) spoke and were able to write English at a proficient level (as determined by the general education teacher). Eligibility criteria was assessed for each student through review of student records and interviews with the general education teachers. Students who were determined to be ineligible to participate in this study

completed the writing tasks that were part of this study's procedures, however their results were excluded from all analyses.

A total of 171 third-grade students across two schools (105 from an urban public school located in a moderately-sized city and 66 from a public school located in a town classified as distant) were recruited for this study. Of these students, four moved prior to data collection. Of the 167 remaining students, 10 were identified by the general education teacher as not being able to write English at a proficient level and as a result were excluded from the current study. Of the 157 remaining students, 10 students were identified by the general education teacher as having a disability that significantly impacted their writing performance (4 students were identified as learning disabled; 1 student was identified as speech/language impaired; 5 students were identified as other health impairment). These identified students were excluded from the present study. Therefore, the total sample size for this study included 147 third-grade students (87 from the urban public school located in a moderately-sized city and 60 from the public school located in a town classified as distant; see Figure 1).

The mean age of the participants was 8 years and 4 months. A total of 70 students (47.6%; 37 from the urban public school and 33 from the school in a distant town) identified as female and 77 students (52.4%; 50 from the urban public school and 27 from the school in a distant town) identified as males. The majority of participants (55.8%; $n = 82$) identified as White, with a smaller percentage identified as Black or African American (28.6%; $n = 42$), Asian (6.8%; $n = 10$), Hispanic or Latino (5.4%; $n = 8$), Native Hawaiian/Other Pacific Islander (2%; $n = 3$), American Indian or Alaska Native (0.7%; $n = 1$), or two or more races (0.7%; $n = 1$). In addition, 6 of the 147 participants were eligible for special education services, however none of the participants had a Section 504 plan. Table 1 illustrates the student demographic information.

Data collection was conducted in seven third-grade general education classrooms across two elementary schools in the northeast. According to the most recent New York State School Report Card (2016-17), 924 kindergarten through eighth-grade students were enrolled in the first elementary school which was classified as an urban public school located in a moderately-sized city. A large percentage of the students (79%) in the first school were eligible for free or reduced-priced lunch. The majority of students enrolled in the first school were identified as either White (39%) or Black or African American (35%), with a smaller percentage identified as Asian or Native Hawaiian/Other Pacific Islander (9%), Hispanic or Latino (8%), two or more races (8%), and American Indian or Alaska Native (2%). According to the most recent New York State School Report Card (2016-2017), 444 kindergarten through sixth-grade students were enrolled in the second public elementary school which was located in a town classified as distant from an urbanized area. Approximately half of the students (51%) in the second school were eligible for free or reduced-priced lunch. The majority of students enrolled in the second school were identified as either White (90%) with a smaller percentage identified as Hispanic or Latino (5%), two or more races (3%), and Black or African American (1%).

Researcher Assistants

Doctoral students enrolled in a school psychology program served as the primary research assistants. Advanced undergraduate students served as secondary research assistants and provided support to the primary research assistants during data collection. All primary and secondary research assistants were required to complete formal training in research ethics through an online training program (e.g., Collaborative Institute Training Initiative; CITI) that emphasized the protection and ethical treatment of human research participants. Training

consisted of completing the following courses provided by CITI: (1) Social and Behavioral Focus and (2) Responsible Conduct of Research.

All research assistants received training on the research-related tasks they were responsible for completing. These included the following areas: (a) administration and scoring of writing performance measures, (b) administration of procedural scripts used during data collection, (c) conducting procedural integrity observations, and (d) data entry. Manuals that explain the scoring procedures for the writing performance measures administered during the course of the study were accessible to every research assistant. Before research assistants participated in data collection, they were required to demonstrate 100% proficiency with conducting procedural checks and scoring outcome measures.

Measures

Writing performance. To assess writing performance, the following two assessments were administered: (a) a Curriculum-Based Measurement in Written Expression probe (CBM-WE); and (b) the Essay Composition subtest from the Wechsler Individual Achievement Test – Third Edition (WIAT-III; Pearson, 2009).

Curriculum-Based Measurement in Written Expression (CBM-WE). Students written responses from a Curriculum-Based Measurement in Written Expression (CBM-WE) probe were scored for correct writing sequences, a measure of writing fluency (i.e., accuracy and speed), using the procedures outlined by Shapiro (2004). Previous research suggests that writing fluency is a valid and reliable indicator of writing performance as it simultaneously accounts for a number of factors including number of words written, grammar, spelling, and punctuation (Espin et al., 2000; Espin, Scierka, Skare, & Halverson, 1999; Videen, Deno, & Marston, 1982). A previous study (Videen, Deno, & Marston, 1982) evaluating the psychometric properties of

correct writing sequences indicates moderate validity with a standardized achievement measure (r = .69; the Test of Written Language; TOWL; Hammill & Larsen, 1978). To answer research question two, scores of 15 or greater, which corresponds with a percentile score of 25 or greater, represented "typical performance," whereas scores of 14 and lower represented "at-risk" performance, which corresponds with percentile scores of 24 or lower.

Previous studies evaluating the psychometric properties of CBM-WE indicate strong alternate-form reliability (r = .73 to .90; McMaster, Wayman, Deno, Espin, & Yeo, 2010), moderate to strong criterion validity(r = .57 to .80; Deno, Marston, & Mirkin, 1982) between three indices of writing performance (i.e., correct letter sequences, total words spelled correctly, and total words written) and the Test of Written Language (TOWL; Hammill & Larsen, 1978).

Wechsler Individual Achievement Test, Third Edition (WIAT-III) Essay Composition. The Wechsler Individual Achievement Test–Third Edition (WIAT–III; Pearson, 2009) is a standardized, norm-referenced achievement test that is used to measure the academic skills of children between the ages of 4 and 19. The Essay Composition subtest of the WIAT-III was administered to assess writing performance beyond writing fluency, including text structure and theme development. Within this subtest, participants were given 10 minutes to plan and compose an essay responding to a verbally and visually presented writing prompt. To answer research question two, a raw score of 180 or greater, which corresponds with a percentile scores of 25 or greater, represented "typical performance," whereas scores of 89 and lower represented "at-risk" performance, which corresponds with percentile scores of 24 or lower.

The psychometric properties of the WIAT-III were reported by the test maker (Pearson, 2009). Among children between the ages of 8 and 9, the Essay Composition subtest of the WIAT-III demonstrates a strong test-retest reliability (r = .88) with a test-retest interval that

averaged 13 days and ranged from two to 32 days. In addition, the WIAT-III reliably differentiates between students who are typically developing in the area of written expression from students who are classified with a Specific Learning Disability in writing. Previous research indicates a low positive correlation (r = .44) between the WIAT-III Essay Composition subtest and correct writing sequences on CBM-WE (Kim, Puranik, & Al Otaiba, 2015).

The Social, Academic, and Emotional Behavior Risk Screener (SAEBRS). Classroom behaviors were evaluated through use of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS; Kilgus et al., 2013; see Appendix A). This school-wide screening measure is a 19-item teacher-report questionnaire developed to assess behavioral and emotional risk within the classroom setting. Each item is rated on a 4-point Likert-type scale and lower values reflect areas of weakness while higher values reflect areas of strength. Eleven items on the measure are reverse-scored. In addition to a total score that measures overall behavioral functioning, confirmatory factor analysis revealed three domains (von der Embse, Pendergast, Kilgus, & Eklund, 2016). The Social Behavior domain (maximum raw score = 18) consists of six items that assesses a student's ability to engage in appropriate interpersonal relationships with peers and adults. The Academic Behavior domain (maximum raw score = 18) also consists of six items and measures a student's ability to benefit from instruction by assessing their preparedness and participation. Finally, the Emotional Behavior (maximum raw score = 21) domain includes seven items and measures emotion regulation, adaptability, and resiliency. The scale developers developed cut scores to determine classification as "at risk" within the following domains: Social behavioral risk scores between 0 and 12; Academic behavioral risk scores between 0 and 9; Emotional behavior risk scores between 0 and 17; and Total behavioral risk scores between 0 and 36.

Although this scale was developed recently, several studies have provided support for the psychometric properties of this measure. The scale has high internal consistency for the total score (α = .93), Social Behavior factor (α = .89), Academic Behavior factor (α = .92), and Emotional Behavior factor (α = .83; Kilgus, Eklund, Von Der Embse, Taylor, & Sims, 2016). The criterion validity of the SAEBRS was compared to another teacher-report screening measure assessing the behavioral and emotional functioning of students (the Behavior and Emotional Screening System; BESS; Reynolds & Kamphaus, 2004) indicating a strong and significant relationship with the total score (r = -.94, p < .001), the Social Behavior factor (r = -.88, p < .001), the Academic Behavior factor (r = -.75, p < .001), and the Emotional Behavior factor (r = -.75, p < .001). Research examining the classification accuracy of the SAEBRS across three time points during the academic year where the BESS served as the reference standard indicated high specificity (range = .95 to .99), high sensitivity (range = .72 to .95), a strong negative predictive value (range = .93 to .99), and excellent classification accuracy (range = .93 to .99; National Center on Intensive Intervention, 2018)

Procedures

All writing performance measures were group-administered to third-grade students in each classroom on one occasion. These measures were administered in the general education classroom setting. One SAEBRS rating scale was given to each student's assigned general education teacher. A packet of five rating scales was distributed to teachers on a weekly basis, such that data on the classroom behavior of five students from each classroom were collected each week across approximately five weeks. This strategy ensured that packets were received in a timely manner. Data from the urban public school located in a moderately sized city were collected in February of 2018 and data from the public school located in a distant town were

collected in November of 2018. Norms were applied based upon the appropriate data collection period. The index tests (i.e., SAEBRS, CBM-WE) and reference standard (i.e., WIAT-III Essay Composition subtest) were identified a priori.

Study Design

A correlational design was used to examine the association between the SAEBRS and CBM-WE among female and male students. An a priori power analysis was conducted to determine the minimum sample size needed to conduct this study. GPower (Erdfelder, Faul, & Buchner, 1996), a statistical software program, was utilized to determine the needed sample size to conduct the simple linear regression analyses to examine the association between the SAEBRS and CBM-WE in research question 1. The effect size was determined based on the results of a previous study (Hamsho, 2017), which was topically and statistically similar to the present study. Based on a medium effect size ($f^2 = .15$) and an alpha level of .05, a minimum sample size of 55 participants per analyses was suggested. As a result, simple linear regression analyses separated by gender required a total sample size of 110 students. A total of 147 third-grade students (77 males and 70 females) participated in this study. Analyses separated by gender exceeded requirements set by the power analysis.

To answer the second research question comparing the classification accuracy of the index tests (i.e., SAEBRS, CBM-WE) to identify students at-risk for writing achievement deficits on the reference standard (i.e. WIAT-III Essay Composition subtest), an a priori analysis using a web-based calculator (Goksuluk, Korkmaz, Zararsiz, & Karaağaoğlu, 2016) was computed to determine the minimum sample size needed. No study to date has evaluated the classification accuracy of any classroom behavior screening measure's ability to predict academic performance. Therefore, the estimated Area Under the Curve (AUC) value for both

analyses used in the present study was determined based on the results of a previous study examining the classification accuracy of CBM-WE in its identification of students at-risk for writing achievement deficits (Furey et al., 2016). Using an alpha level of $p = .05$, power of $b = .80$, and an estimated AUC value of .74, a minimum sample size of 32 participants was suggested. The total sample of 147 third-grade in this study exceeded the requirements set by the power analysis.

Procedural Integrity

To assess procedural integrity, the primary research assistant followed a procedural script when administering the writing performance measures utilized in this study. Each step successfully administered by the primary research assistant was marked as completed. A secondary research assistant followed along with the procedural script and checked off all the steps they observed the primary research assistants complete. Agreements between the primary and secondary research assistants were tallied up to calculate agreement. In order to measure procedural integrity, the total number of agreements was divided by the sum of agreements and disagreements. The mean procedural integrity was 100%.

Results

Data Preparation

Data input and consistency checks. Raw data for CBM-WE, the SAEBRS, and raw scores of the WIAT-III Essay Composition subtest were entered into Microsoft Excel by either a primary or secondary research assistant. A different trained research assistant re-entered the data. The accuracy of data entry was verified using a double data entry procedure. All discrepancies were compared with the original data and corrected. The data were transferred from Excel into SPSS 23 (SPSS Inc., 2015) for analysis.

Missing data. Ten students (four males and six females) were absent when the WIAT-III Essay Composition subtest was administered. The missing data for these 10 students were imputed using multiple imputation procedures in SPSS 23.

Data inspection. Prior to conducting the main analyses, all data were inspected for violations of the statistical assumptions underlying the statistical analyses. Tests of linearity, homoscedasticity, and independence of observation indicated that these underlying assumptions of multiple regression were met. However, the distribution of the unstandardized and standardized residuals for the male sample suggested a non-normal distribution of the CBM-WE. In addition, correlation analyses among the three SAEBRS composites indicated significant multicollinearity, which is described in further detail below.

Descriptive and Correlational Analyses

Descriptive statistics of the reference standard and index measures utilized in this study were calculated and are reported in Tables 2 and 3. The average WIAT-III Essay Composition score for the total sample (M = 212.04) corresponds with a standard score of 107 and a percentile score of 68, indicating that students were performing within the Average range. Similarly, the mean CBM-WE score for the total sample (M = 19.67; 35th percentile) indicated average performance. On the SAEBRS, the average total behavior composite (M = 44.50), social composite (M = 14.24), and academic composite (mean = 12.85) reflected behaviors assessed as falling within the average range. However, the average score on the emotional composite on the SAEBRS (M = 17.40) fell in the "at-risk" range.

Correlation analysis revealed that for the entire sample (n = 147), all variables measured within the study were significantly correlated with each other. More specifically, CBM-WE scores were significantly correlated with scores on the WIAT-III Essay Composition subtest (r =

.504, $p < .001$). Additionally, scores on the total SAEBRS composite were significantly correlated with CBM-WE ($r = .313$, $p < .001$) and scores on the WIAT-III Essay Composition subtest ($r = .336$, $p < .001$). A further examination of the SAEBRS composites indicated that the academic composite on the SAEBRS was significantly correlated with CBM-WE ($r = .370$, $p < .001$) and the scores on the WIAT-III Essay Composition subtest ($r = .356$, $p < .001$). Similarly, the emotional composite of the SAEBRS was significantly correlated with CBM-WE ($r = .267$, $p = .001$) and scores on the WIAT-III Essay Composition subtest ($r = .346$, $p < .001$). Finally, social behaviors on the SAEBRS was significantly correlated with CBM-WE ($r = .201$, $p = .014$) and the WIAT-III Essay Composition subtest ($r = .210$, $p = .011$). Significant multicollinearity between all the SAEBRS composites on the SAEBRS were demonstrated, with correlational values ranging from .688 to .734.

Due to the results of previous research identifying significant gender differences in writing performance (Fearrington et al., 2014), the main analyses in the current study examining the relationship between CBM-WE and the SAEBRS were conducted separately for each gender (see Table 4). Correlation analyses revealed that female CBM-WE scores were significantly correlated with scores on the WIAT-III Essay Composition subtest ($r = .526$, $p < .001$). Additionally, female scores on the total SAEBRS composite were significantly correlated with CBM-WE ($r = .338$, $p = .004$) and scores on the WIAT-III Essay Composition subtest ($r = .276$, $p = .021$). Further examination of the SAEBRS composites indicated that the academic composite on the SAEBRS was significantly correlated with CBM-WE ($r = .448$, $p < .001$) and the scores on the WIAT-III Essay Composition subtest ($r = .360$, $p = .002$) for female students. Similarly, the emotional composite of the SAEBRS was significantly correlated with CBM-WE ($r = .265$, $p = .026$) and scores on the WIAT-III Essay Composition subtest ($r = .262$, $p = .028$)

for female students. However, the social composite of the SAEBRS was not found to be significantly correlated with either measure evaluating writing performance for female students. For female students, all SAEBRS composite scores were correlated with each other, with correlation values ranging from .596 to .890.

A similar pattern of results was found for the male sample. That is, male CBM-WE scores were significantly correlated with scores on the WIAT-III Essay Composition subtest ($r = .433$, $p < .001$). Additionally, male scores on the total SAEBRS composite were significantly correlated with CBM-WE ($r = .228$, $p = .046$) and scores on the WIAT-III Essay Composition subtest ($r = .249$, $p = .029$). A further examination of the SAEBRS composite scores indicated that the academic composite was significantly correlated with CBM-WE ($r = .261$, $p = .022$) and the scores on the WIAT-III Essay Composition subtest ($r = .284$, $p = 0.12$). Although a statistically significant relationship was revealed between the emotional composite on the SAEBRS and the WIAT-III Essay Composition subtest ($r = .292$, $p = .010$), the emotional composite on the SAEBRS was not significantly correlated with CBM-WE ($r = .205$, $p = .073$). Similar to the results demonstrated for the female sample, the social composite on the SAEBRS was not found to be significantly correlated with either measure evaluating writing performance. For male students, all the SAEBRS composites were correlated with each other, with correlation values ranging from .721 to .910.

Each outcome variable was analyzed to determine whether gender differences existed (see Table 3). To control for Type 1 Error, an adjusted alpha value of .008 was applied. With regard to CBM-WE, there were no statistically significant difference between female ($M = 22.25$m $SD = 12.10$) and male ($M = 17.32$, $SD = 10.88$) students; $t\ (145) = 2.60$, $p = .010$). However, there was a statistically significant difference on the WIAT-III Essay Composition subtest with female

students demonstrating higher mean raw scores (M = 223.71, SD = 23.30) in comparison to male students (M = 201.42, SD = 29.25; t (145) = 5.07, p < .001). A statistically significant female advantage was also observed on the total SAEBRS composite scores (M = 47.71, SD = 9.11) when compared with their male peers (M = 41.58, SD = 12.03; t (140.52) = 3.50, p < .001). A similar pattern of findings was also discovered upon further examination of the social composite of the SAEBRS, such that a female advantage was observed (M = 15.64, SD = 3.31) in comparison to their male peers (M = 12.97, SD = 4.55; t (138.42) = 4.08, p < .001). A female advantage was also observed on the emotional composite of the SAEBRS (M = 18.34, SD = 3.11) in comparison to their male peers (M = 16.54, SD = 4.03; t (141.36) = 3.03, p = .003). However, no statistically significant differences between male and female students were observed on the academic composite of the SAEBRS (t (145) = 2.31, p = .022).

Relationship between the SAEBRS and CBM-WE

To answer the first research question, multiple linear regression analyses were conducted to examine the relationship between students' classroom behaviors on the three composite scores of the SAEBRS and their writing fluency on the CBM-WE (see Appendix B). Because this study was interested in evaluating whether there was a statistically significant relation between the two domains addressed in a multi-tiered systems of support approach (i.e., writing performance and classroom behaviors), CBM-WE, a school-wide screening measure, was included as the independent variable instead of a standardized academic achievement measure of writing performance. It is important to note that high levels of multicollinearity were observed between the three SAEBRS composite scores (r range, .688 to .734), which violated the underlying statistical assumptions of these analyses (see Table 4). Due to these violations, the first research question and accompanying analytical plan was altered. Specifically, two linear regression

analyses examining the relationship between male and female students' total SAEBRS composite and their performance on CBM-WE were conducted.

Results from these analyses indicated that the overall composite score on the SAEBRS was a statistically significant predictor of female (R^2 = .114, $F_{(1,68)}$ = 8.75, p = .004) and male (R^2 = .052, $F_{(1,75)}$ = 4.12, p = .046) students' CBM-WE scores. The proportion of variance in CBM-WE scores that was explained by the SAEBRS was greater among females (11%) in comparison to males (5%). Results indicated that an interaction term between gender and the SAEBRS did not account for a significant proportion of the variance CBM-WE, ΔR^2 =.011, $\Delta F_{(3, 143)}$ = 1.80, p = .181).

To examine whether differences existed based upon data collection period, linear regression analyses were conducted separately across the two schools. Gender differences were not examined in these follow-up analyses because the sample size did not meet the minimum power analysis requirements. The results from both analyses suggested that the SAEBRS continued to be a statistically significant predictor of students' CBM-WE scores for students in the urban public school (R^2 = .078, $F_{(1,85)}$ = 7.22, p = .009) and students in the public school located in a distant town (R^2 = .134, $F_{(1,58)}$ = 8.95, p = .004). A visual plot of this analysis revealed that an increase in the SAEBRS scores led to an increase in CBM-WE for both male and female students (see Figure 2).

Classification Accuracy of the SAEBRS and CBM-WE

To examine the classification accuracy of the SAEBRS and CBM-WE when the WIAT-III Essay Composition was used as the reference standard, cut scores were used. The WIAT-III was utilized as the reference standard because it is the most frequently used standardized norm-referenced academic achievement measure (Benson et al., 2019). Based on established

guidelines (Kilgus et al., 2013), a cut score of 36 or lower on the total SAEBRS composite was used to identify students at-risk for behavioral issues. For CBM-WE measure, a cut score at or below the 24th percentile was used to identify students at-risk for writing difficulties. For the WIAT-III Essay Composition subtest, a cut score at or below the 24th percentile was used to identify students at-risk for writing achievement deficits. Applying this threshold, 14% of the students in the sample were classified as at-risk based on the WIAT-III Essay Composition subtest score (see Table 5).

To answer the second research question, two separate receiver-operating characteristic (ROC) curves were estimated to provide a measure of classification accuracy. Due to the limited sample of female students identified as at-risk for writing achievement deficits on the WIAT-III Essay Composition subtest (n = 2), the ROC curve analyses were not separated by gender. The ROC analyses included an evaluation of the Area Under the Curve (AUC), which was calculated by plotting true positive rates (i.e., sensitivity) against false positive (1-specificity) rates. An AUC estimate of .69 or lower is considered to be poor, an AUC estimate between .70 and .79 is considered to be fair, and an AUC estimate of .80 and higher is considered good (Swets, 1996). Additional indicators of classification accuracy (i.e., sensitivity, specificity, positive predictive power, and negative predictive power) were conducted where values under 49% were considered unacceptable, values between 50% and 79% were considered acceptable, and values of 80% and higher were considered excellent (Daniels, Volpe, Fabiano, & Briesch, 2017).

Results of the first ROC curve analysis, which examined the classification accuracy of CBM-WE when the WIAT-III Essay Composition subtest was used as the reference standard, indicated fair overall classification accuracy for identifying writing achievement difficulties (AUC = .761; 95% CI [.644, .878]; see Figure 3 and Table 6). Additional results indicated that

CBM-WE demonstrated acceptable sensitivity (61.11%; 95% CI [35.75%, 82.70%]), acceptable specificity (75.19%; 95% CI [66.82%, 82.37%]), unacceptable positive predictive power (25.58%; 95% CI [17.61%, 35.61%]), and excellent negative predictive power (93.27%; 95% Confidence Interval [88.51%, 96.14%]; see Table 7).

Results of the second ROC curve analysis, which examined the classification accuracy of the SAEBRS when the WIAT-III Essay Composition subtest was used as the reference standard, indicated poor overall classification accuracy for identifying writing achievement difficulties (AUC = .653; 95% Confidence Interval [.528, .778]; see Figure 4 and Table 6). Additional results indicated the SAEBRS demonstrated unacceptable sensitivity (38.89%; 95% CI [17.30%, 64.25%]), acceptable specificity (76.74%; 95% CI [68.49%, 83.73%]), unacceptable positive predictive power (18.92%; 95% CI [10.78%, 31.07%]), and excellent negative predictive power (90%; 95% CI [86.02%, 92.94%]; see Table 7).

Although CBM-WE demonstrated higher levels of classification accuracy than the SAEBRS, further analyses did not reveal statistically significant differences between the areas under the two independent ROC curves (z = -1.07, standard error of the difference = .100, p = .283).

To examine whether differences existed based upon data collection period, the ROC curve analysis for each index test was conducted across the two schools. The results suggested that the area under the curve value for the SAEBRS was qualitatively higher for the public school in a distant town (AUC = .711; 95% Confidence Interval [.337, 1.00]) in comparison to the urban public school (AUC = .646; 95% Confidence Interval [.507, .785]). However, further analyses did not reveal statistically significant differences between the areas under the two independent ROC curves (z = -0.373, standard error of the difference = .173, p = .354). With regard to CBM-

WE, once again the public school in a distant town evidenced a qualitatively higher value (AUC = .815; 95% Confidence Interval [.668, .961]) in comparison to the urban public school (AUC = .746; 95% Confidence Interval [.612, .880]). Further analyses did not reveal statistically significant differences between the areas under the two independent ROC curves (z = -0.528, standard error of the difference = .130, p = .597). Additional ROC curve analyses excluding the data from the 10 participants of which the WIAT-III Essay Composition subtest was missing did not indicate a statistically significant difference in the AUC estimate for either CBM-WE (AUC = .775; 95% Confidence Interval [.658, .892]; z = 0.197, standard error of the difference = .071, p = .843) or the SAEBRS (AUC = .665; 95% Confidence Interval [.541, .790]; z = 0.135, standard error of the difference = .088, p = .892).

Discussion

The majority of our nation's students are unable to demonstrate proficient skills within the area of written expression (National Center for Education Statistics, 2012). It is particularly alarming that nearly three quarters of elementary-aged students are unable to write with grade-level proficiency as writing is an essential tool for effective functioning in daily life (MacArthur et al., 2016) that has been linked to long-term negative outcomes (Lloyd, 1978; Morrisroe, 2014). Recognizing the need to address concerns regarding students' academic performance, many states have enacted prevention efforts within school systems. Using a multi-tiered system of support framework, schools identify "at-risk" students using school-wide screening measures, provide evidence-based interventions to ameliorate academic difficulties, and evaluate student response to interventions.

The multi-tiered systems of support framework targets students' academic and behavioral functioning. Previous research has demonstrated the contribution of classroom behaviors to

students' academic performance in reading and mathematics (Alexander et al., 1993; Barriga et al., 2002; Cobb, 1972; Finn et al., 1995; Kilgus et al., 2017; McKinney et al., 1975). Although the research evaluating this relationship in writing is limited, two studies have demonstrated a statistically significant relationship between classroom behaviors and students' writing performance (Kent et al., 2014; Kim, Al Otaiba et al., 2015). However, no study has evaluated the relationship between classroom behaviors and writing performance by using a broad assessment of classroom behaviors that meets conventional psychometric standards for school-wide screening measures. As a result, the purpose of this study was two-fold: 1) to evaluate the relationship between the SAEBRS and CBM-WE among male and female students, and 2) to compare the classification accuracy of the SAEBRS to CBM-WE when identifying students at-risk for writing achievement deficits. Based upon the results from Kilgus et al., (2017), it was hypothesized that in addition to the total SAEBRS composite, the social and academic composite would also be related to CBM-WE. Derived from the results of earlier research (Kilgus et al., 2017), it was also hypothesized that the SAEBRS would demonstrate better classification accuracy than the CBM-WE.

Overall, the results from this study indicated a statistically significant relationship between the SAEBRS and CBM-WE for female and male students. In addition, the SAEBRS and CBM-WE demonstrated fairly comparable classification accuracy rates across the majority of indicators, with the exception of sensitivity and area under the curve. From an evidence-based practice perspective, CBM-WE was superior to the SAEBRS as CBM-WE demonstrated acceptable rates of sensitivity and overall classification accuracy while the SAEBRS indicated unacceptable rates of sensitivity and poor overall classification accuracy. Although further analyses did not reveal a statistically significant difference between the areas under the two

independent ROC curves, the differences in the qualitative indicators of sensitivity and classification accuracy suggest that it would not be justifiable to use the SAEBRS over CBM-WE as a school-wide screening tool to identify students at-risk for writing achievement difficulties.

Relationship between the SAEBRS and CBM-WE

Because previous research has demonstrated a significant relationship between classroom behaviors and students' writing performance (Kent et al., 2014; Kilgus et al., 2017; Kim, Al Otaiba, et al., 2015), it was expected that the SAEBRS total composite scores would be significantly related to female and male students' CBM-WE outcomes. This hypothesis was confirmed, suggesting that classroom behaviors are related to students' writing performance. This finding is in line with the previous literature identifying a relationship between various school-wide screening measures assessing classroom behaviors and academic performance (Dowdy et al., 2013; Kamphaus et al., 2010; Kilgus et al., 2017; Lane et al., 2010).

Additionally, the findings of the current study highlighted differences in the strength of the relationship between classroom behaviors and writing performance across female and male students. Specifically, for females, a larger percentage of the variance (11%) in writing performance was explained by classroom behaviors in comparison to the variance found for their male peers (5%). These results indicate that classroom behaviors play a larger role in the writing performance of third-grade female students in comparison to their male peers. Previous studies have attempted to understand gender differences with regard to writing performance and have identified a female advantage for letter and word writing fluency (Berninger & Fuller, 1992), spelling (Allred, 1990; Malecki & Jewell, 2003), and handwriting skills (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008). One possible explanation offered by these studies for the

observed gender differences relates to the development of writing skills across female and male students. Berninger et al. (2008) hypothesized that unlike their female peers, male students may not have automatized the lower level skills that promote successful writing performance (e.g., letter formation). Due to these differences, male students may focus heavily upon the motor transcription of individual letters when asked to write. In contrast, female students may focus more heavily upon idea development and planning when asked to write. This finding in combination with the results from the current study indicating that male students demonstrated levels of emotional and social behaviors on the SAEBRS that were classified to be in the at-risk range may help to explain why classroom behaviors demonstrated a larger contribution to the writing performance of female students in comparison to male students. It is suspected that maladaptive classroom behaviors may have impeded male students' ability to benefit from general education instruction aimed at developing lower-level writing skills. Male students who were noncompliant with teacher requests and struggled with emotion regulation may not have been able to learn and eventually master the lower-level skills involved in the writing process. When asked to write, these students may need to expend a large amount of their cognitive resources on lower-level skills, which may ultimately account for a larger percentage of variance in writing performance in comparison to classroom behaviors. In contrast, females within the current study demonstrated adaptive levels of social, academic, and emotional behaviors, which may have allowed them to develop lower-level writing skills without interference of at-risk classroom behaviors. As a result, female students may not need to expend as large of an amount of their cognitive resources on lower-level writing skills as their male peers when writing. Instead for female students, classroom behaviors may contribute a larger portion of variance to

the writing process in comparison to lower-level writing skills, which at this point are most likely well-developed.

Only one other study has examined the relationship between classroom behaviors and writing performance across genders, however it did so within the context of a tier 1 intervention aimed at increasing writing productivity (Hamsho, 2017). Although the current study evaluated pre-intervention writing performance and incorporated a broader measure of classroom behaviors, Hamsho (2017) also found a difference in the strength of the relationship between classroom behaviors and post-intervention writing performance across female and male students. More specifically, academic classroom behaviors and attention were identified to be statistically significant predictors of post-intervention writing performance for female students but were not statistically significant predictors of post-intervention writing performance for male students. In summary, the results from the current study and Hamsho's (2017) study indicated that classroom behaviors demonstrate a stronger contribution to the pre- and post-intervention writing performance of female students in comparison to their male peers. It is suspected that classroom behaviors may demonstrate a larger contribution to male students' writing performance once they have automatized lower level writing skills

When comparing the present study's results to prior research utilizing the SAEBRS (Kilgus et al., 2017), it was noted that the SAEBRS explained a larger percentage of variance (27%) on measures of reading and mathematics performance. A number of factors may account for these differences, including participant demographics, the decision to account for gender within the current study, and the procedures used to assess academic performance (curriculum-based measurement versus individualized adaptive computerized assessment). Because no other research to date has quantified the percentage of variance in pre-intervention writing

performance explained by classroom behaviors, it is unclear whether the results from this study accurately reflect the true population. However, the results of this study suggest that the SAEBRS is a better predictor of reading and mathematics performance in comparison to writing performance. Nevertheless, the statistically significant relationship identified in the current study between the SAEBRS and CBM-WE suggests that although the contribution may not be large, classroom behaviors can influence the writing performance of third-grade male and female students. Overall, the results from this study suggest that students, especially females, with better-developed classroom behaviors (i.e., higher scores on the SAEBRS) will perform better on CBM-WE in comparison to peers who demonstrate at-risk classroom behaviors.

Classification Accuracy

Because prior research has highlighted the benefits of using school-wide screening measures as part of a multi-tiered system of support to identify at-risk students for school difficulties (Dowdy et al., 2013; Kamphaus et al., 2010; Kilgus et al., 2017; Lane et al., 2010), one of the main purposes of this study was to compare the classification accuracy of the SAEBRS and CBM-WE to identify students at-risk for writing achievement deficits. It was hypothesized that the SAEBRS would demonstrate higher levels of classification accuracy in comparison to CBM-WE. Overall, the results of this study did not support the hypothesis. When compared with CBM-WE, the SAEBRS revealed comparable rates of specificity, positive predictive power, and negative predictive power. However, unlike CBM-WE, the SAEBRS did not demonstrate acceptable rates of classification accuracy and sensitivity. That is, when CBM-WE was used as the index measure, results indicated acceptable overall classification accuracy (AUC = .761) and sensitivity (61.11%), whereas when the SAEBRS was used as the index

measure, results indicated poor overall classification accuracy (AUC = .653) and sensitivity (38.89%).

Further analyses comparing the differences in qualitative indicators of classification accuracy between the SAEBRS and CBM-WE did not identify a statistically significant difference between the AUC estimates. This finding suggests that CBM-WE and the SAEBRS demonstrated comparable rates of identifying students who are at-risk for writing achievement deficits. However, a large difference was observed between the two index measures with regard to the value of sensitivity. The sensitivity of the SAEBRS fell well below acceptable levels, such that less than half of the students identified by the WIAT-III were also identified as at-risk for writing achievement deficits on the SAEBRS. In comparison, when CBM-WE was used as the index measure, over half of the students were identified as at-risk for writing achievement deficits on both the WIAT-III and CBM-WE. The qualitative indicators for classification accuracy and sensitivity indicate that CBM-WE is superior to the SAEBRS. This is most likely due to the fact that unlike the SAEBRS, CBM-WE is directly measuring a skill that is encompassed within the reference standard. In addition, the findings from the first research question revealed that the SAEBRS did not explain a large percentage of the variance in writing performance. As a result, the SAEBRS may be too broad of a measure to accurately identify students at-risk for writing achievement deficits.

Both index measures fell well below the desired levels for positive predictive power. Within the current sample less than a quarter of the students identified by the WIAT-III were also identified on both index measures to be at-risk for writing achievement deficits. No study to date has investigated the classification accuracy of the SAEBRS to identify students at-risk for academic deficits; however, the positive predictive power of CBM-WE from the current study

was noted to fall well below what has been found in previous research (Furey et al., 2016). In the current study, less than 15% (n = 18) of the participants were identified as at-risk across both the reference standard and CBM-WE. In contrast, Furey et al. (2016) identified 53% of students as at-risk on CBM-WE and their reference standard (English Language Arts Composition subtest of Massachusetts Comprehensive Assessment System; MCAS; Massachusetts Department of Elementary and Secondary Education, 2014). The extremely low positive predictive values observed in the current study across both criterion measures may have been influenced by the small sample of students identified as "at-risk" for writing achievement deficits as base rate is crucial to the calculation of this indicator (Pepe, 2003).

Across other indicators of classification accuracy, the area under the curve value of CBM-WE obtained in the current study was comparable to prior research (AUC = .74; Furey et al., 2016). CBM-WE of the current study demonstrated higher rates of accurately identifying students who were not at-risk for writing achievement deficits (Specificity = 75.19%, NPP = 93.27%) in comparison to the results obtained by Furey et al. (Specificity = 40%, NPP = 79%). However, the current study indicated lower rates of accurately identifying students at-risk for writing achievement deficits (Sensitivity = 65.11%, PPP = 25.58%) when compared with Furey et al. (Sensitivity = 87%, PPP = 53%). The differences in the values for sensitivity, positive predictive power, specificity, and negative predictive power between the current study and Furey et al. may be due to the different reference standard utilized in the ROC curve analyses across the two studies. While the current study utilized a measure specifically evaluating theme development, organization, and text productivity, Furey et al., incorporated a measure that also considered textual conventions (i.e., spelling, grammar, punctuation, and usage), idea development, level of detail, and the ability to engage reader's interest. The reference standard

utilized by Furey et al. encompassed a comprehensive evaluation of writing performance that extended beyond the scope of the reference standard used within the current study. It is also important to note that very limited evidence regarding the reliability and validity of state assessment measures are available, whereas considerable evidence exists for the WIAT-III in terms of psychometrics and standardization. As it is unclear whether the reference standard utilized by Furey et al. is an accurate measure of writing performance, the results between Furey et al. and the current study may not be directly comparable.

The greater advantage CBM-WE demonstrated on two indictors (i.e., area under the curve and sensitivity) revealed that CBM-WE possessed a higher rate of accurately identifying students who may be at-risk for writing achievement deficits in comparison to the SAEBRS. Therefore, the findings from the current study support the use of CBM-WE instead of the SAEBRS as a school-wide screening measure to identify students who may be at-risk for writing achievement deficits. At this time, no additional research is available to provide support for the findings of the current study with regard to the classification accuracy of the SAEBRS in writing. As a result, it is not recommended that school districts use the SAEBRS as a school-wide screening measure to identify students at-risk for writing performance deficits.

Limitations

Several limitations to the current study are noted. First, the use of rating scales to gather information regarding student classroom behavior may limit the confidence in the reliability of the results. Due to the subjective nature inherent to rating scales, it is possible that the SAEBRS was susceptible to errors like the halo effect (i.e., ratings reflect impression rather than actual performance) and/or error of central tendency (i.e., scores tending to hover in the middle of the

Likert scale and do not accurately reflect student behavior). As such, the results of the current study should be interpreted with some degree of caution.

Second, only a small percentage of students were identified as "at-risk" for writing achievement deficits on the reference standard (the WIAT-III Essay Composition subtest). Base rate is related to the overall accuracy of classifications including the positive and negative predictive powers of the criterion measures (Pepe, 2003). As a result, it is possible that the limited sample size of students who demonstrated "at-risk" writing skills may have influenced the results of this study.

A third limitation relates to the exclusionary criteria that was utilized to select participants. Because students who were identified by the school district and their teacher as having a significant disability or diminished language skills that impaired their writing performance were excluded, support for CBM-WE and SAEBRS as a school-wide screening measure, which typically involves all general education students regardless of classification status, may be limited. Furthermore, the exclusion of these students may have ultimately influenced the base rate of the current study.

Another aspect of this study that may limit the generalizability of findings related to the population that was sampled. The primary aim of this study was to examine the relationship between classroom behavior and writing performance among elementary-aged students and to evaluate the classification accuracy of two school-wide screening measures. This study attempted to recruit a diverse group of students across two different school settings (i.e., urban versus distant town) in order to increase the generalizability of the findings to school districts across the nation. However, because this study specifically focused upon third-grade students the result may not generalize to students in other grade levels. Writing skills vary widely across

grade levels (Abbott & Berninger, 1993). Students between the grade levels of kindergarten and second grade are utilizing lower-level processes in their writing development, which is characterized by establishing proficiency in motor production of written text. As students' writing abilities progress, they begin to automatize the motor skills involved with the production of written language and focus on generating sentences. Once these processes are automatized, students can fully engage in planning and revising their written work (cognitive constraints). Given this information, the generalizability of the results from the current study is restricted to third-grade students.

Finally, this study incorporated a relatively small sample when compared with previous research. For example, Kilgus et al. (2017) included over 1,000 students across four elementary schools and one middle school. Additionally, the current study recruited a smaller number of teachers when compared with the number of teachers participating in the study by Kilgus and colleagues. As a result, the findings may have been influenced by individual differences across teacher ratings.

Directions for Future Research

The results from the current study and previous research (Kent et al., 2014; Kim, Al Otaiba, et al., 2015) highlight the importance of classroom behaviors to female and male elementary-aged students writing performance. However, the current literature examining this topic is limited. This was the first study to consider the classification accuracy of the SAEBRS for identifying students at-risk for academic deficits, specifically within the area of writing. However, given the previously stated limitations, it is recommended that future studies replicate the current study with a sample containing a larger base rate to ensure that the results from the current study accurately represent the classification accuracy of these measures. Additionally,

given the differences in skills that are required to achieve successful writing performance across grade levels, it is suggested that that this study be replicated with older and younger participants.

There was quite a large difference in the percentage of variance explained in writing performance by the SAEBRS across female and male students. The potential to explain gender differences in the relationship between classroom behaviors and writing performance has yet to be fully explored. Future research examining variables (e.g., language and cognitive factors identified by Kim et al., 2015) influencing the relationship between classroom behaviors and writing performance is warranted.

Results of the current study established a significant relationship between classroom behaviors and writing performance. Despite this relationship, the results from the current study did not support the use of the SAEBRS (a measure assessing classroom behaviors) to identify students at-risk for writing achievement deficits. In addition to school-wide screening, tier 1 of the multi-tiered systems of support approach also encourages implementation of evidence-based intervention. However, no study to date has considered the relationship between classroom behaviors and writing performance within the context of an evidence-based intervention aimed at improving behavioral functioning. The results from the current study suggest that third-grade males may benefit greatly from an evidence-based intervention targeting behavior, as the male students in the current sample demonstrated a total SAEBRS composite that was significantly lower than their female peers. It is suggested that future research consider whether the implementation of a behavioral intervention could contribute to improved writing performance of male elementary-aged students.

The current study relied upon subjective reports from teachers in order to collect information regarding each student's behavioral functioning. As previously discussed,

subjective reporting measures are susceptible to errors that may call into question the validity and reliability of the data collected. To determine if these errors significantly impacted the study's results, it is recommended that future research obtain information regarding classroom behaviors from multiple informants to evaluate to the classification accuracy of the SAEBRS for identifying students at-risk for writing achievement deficits.

Conclusions

Writing is an important skill utilized across many different academic areas (The National Commission on Writing, 2004) and early writing deficiencies have been linked to negative long-term consequences (Lloyd, 1978; Morrisroe, 2014). Given its critical role to effective daily functioning, writing is an important area to target with interventions for struggling writers. As a result, it is important for researchers and practitioners to develop a comprehensive understanding of factors that predict writing performance and utilize effective screening practices to identify "at-risk" students.

The current study sought to examine the relationship between writing performance and classroom behaviors across male and female students. It also compared the classification accuracy of two school-wide screening measures (one assessing classroom behaviors and the other assessing writing performance) to identify students at-risk for writing achievement deficits. This study was unique from the previous research in three distinct ways. First, it examined the relationship between writing performance and a wide array of classroom behaviors, including social behaviors and emotional behaviors, areas that had not been examined in the previous literature. Second, it considered gender differences in the relationship between classroom behaviors and students' writing performance. Third, it was one of the first studies to evaluate the

classification accuracy of the SAEBRS in identifying students "at-risk" for writing achievement deficits.

In relation to the study aims, the total SAEBRS composite, which included social and emotional behaviors, was identified to be related to CBM-WE outcomes for male and female students. These results offer some guidance on the underlying contributors to the writing performance of elementary-aged students. With regard to the classification accuracy analyses, the results indicated an advantage when using CBM-WE to identify students at-risk for writing achievement deficits across some indicators of classification accuracy in comparison to the SAEBRS. Given the large discrepancy in the sensitivity values and differences in qualitative indicator for area under the curve between CBM-WE and the SAEBRS, it is recommended that schools use CBM-WE instead of the SAEBRS as a school-wide screening tool for identifying students as at-risk for writing achievement deficits. Table 1

Table 1

Student Demographic Characteristics (*n* = 147)

	Female (*n* = 70)		Male (*n* = 77)	
	M / %	(SD) / *n*	M / %	(SD) / *n*
Race				
American Indian or Alaska Native	0%	0	1.3%	1
Asian	8.6%	6	5.2%	4
Black or African American	25.7%	18	31.2%	24
Decline to Self-Identify	2.9%	2	7.8%	6
Native Hawaiian or Other Pacific Islander	1.4%	1	2.6%	2
Two or More Races	0%	0	1.3%	1
White	61.4%	43	50.6%	39
Ethnicity				
Arab	2.9%	2	1.3%	1
Chin	1.4%	1	0%	0
Hispanic or Latino	2.9%	2	7.8%	6
Hutu	0%	0	1.3%	1
Karen	1.4%	1	1.3%	1
Mandinka/Malinke	0%	0	1.3%	1
Masalit	0%	0	1.3%	1
Nepali	1.4%	1	2.6%	2
Not Hispanic or Latino	82.9%	58	76.6%	59
Oromo	2.9%	2	0%	0

Somali	2.9%	2	5.2%	4
Swahili/Waswahili	1.4%	1	0%	0
Vietnamese	0%	0	1.3%	1
Special Education Status				
Learning Disability	0%	0	2.6%	2
None	98.6%	69	93.5%	72
Other Health Impairment	1.4%	1	1.3%	1
Speech/Language Impairment	0%	0	2.6%	2
English as a Second Language (ESL) Status				
ESL	12.9%	9	13%	10
No ESL	87.1%	61	87%	67
	M	(SD)	M	(SD)
Age	8.04	(.03)	8.04	(.04)

Table 2

Descriptive Statistics of the Reference Standard and Index Measures for the Total Sample

Variable	*N*	Range	Mean	SD	Skew	Kurtosis
Reference Standard						
WIAT-III Essay Composition subtest	147	150.00	212.04	28.75	-0.706	1.04
Index Measures						
CBM-WE	147	53.00	19.67	11.70	0.189	-0.331
SAEBRS, Total Behavior Composite	147	43.00	44.50	11.13	-0.799	-0.344
SAEBRS, Social Composite	147	15.00	14.24	4.21	-1.15	0.379
SAEBRS, Academic Composite	147	18.00	12.85	4.41	-0.453	-0.920
SAEBRS, Emotional Composite	147	14.00	17.40	3.72	-0.946	-0.104

Table 3

Descriptive Statistics of the Reference Standard and Index Measure across Female and Male Participants

	Females (n = 70)				Males (n = 77)					
	Range	*M* (*SD*)	Skew	Kurtosis	Range	*M* (*SD*)	*Skew*	*Kurtosis*	*t*	*p*
Reference Standard										
WIAT-III Essay Composition	93.00	223.71 (23.30)	-.080	-.674	140.00	201.42 (29.25)	-.869	.867	5.07	.000
Index Measure										
CBM-WE	53.00	22.25 (12.10)	.048	-.478	50.00	17.32 (10.88)	.240	-.053	2.60	.010
SAEBRS Total Behavior Composite	35.00	47.71 (9.11)	-1.03	.179	43.00	41.58 (12.03)	-.509	-.816	3.50	.001
SAEBRS, Social Composite	15.00	15.64 (3.31)	-1.80	3.06	15.00	12.97 (4.55)	-.753	-.536	4.08	.000
SAEBRS, Academic Composite	12.00	13.72 (3.96)	-.561	-1.10	18.00	12.06 (4.67)	-.287	-.963	2.31	.022
SAEBRS, Emotional Composite	12.00	18.34 (3.11)	-1.16	.543	14.00	16.54 (4.03)	-.699	-.651	3.03	.003

Table 4

Intercorrelations between the Reference Standard and Index Measures

Total Sample (n = 147)

Measure	1	2	3	4	5	6
1 WIAT-III Essay Composition	-					
2 CBM-WE	.504***	-				
3 SAEBRS, Total Composite	.336***	.313***	-			
4 SAEBRS, Social Composite	.210*	.201*	.897***	-		
5 SAEBRS, Academic Composite	.356***	.370***	.903***	.688***	_	
6 SAEBRS, Emotional Composite	.346***	.267***	.904***	.733***	.734***	_

*p< .05. **p< .01. ***p<.001

Females (n = 70)

Measure	1	2	3	4	5	6
1 WIAT-III Essay Composition	-					
2 CBM-WE	.526***	-				
3 SAEBRS, Total Composite	.276*	.338**	-			
4 SAEBRS, Social Composite	.083	.143	.851***	-		
5 SAEBRS, Academic Composite	.360**	.448***	.890***	.596***	_	
6 SAEBRS, Emotional Composite	.262*	.265*	.887***	.665***	.696***	_

*p< .05. **p< .01. ***p<.001

Males (n = 77)

Measure	1	2	3	4	5	6
1 WIAT-III Essay Composition	-					
2 CBM-WE	.433***	-				
3 SAEBRS, Total Composite	.249*	.228*	-			
4 SAEBRS, Social Composite	.108	.152	.907***	-		
5 SAEBRS, Academic Composite	.284*	.261*	.910***	.721***	-	
6 SAEBRS, Emotional Composite	.292**	.205	.903***	.738***	.740***	-

*p< .05. **p< .01. ***p<.001

Table 5

Cross Tabulation of the SAEBRS and CBM-WE by the WIAT-III Essay Composition Subtest

	WIAT-III Essay Composition Subtest		
SAEBRS	**At-Risk**	**Not At-Risk**	**Total**
At-Risk	7	30	37
Not At-Risk	11	99	110
Total	18	129	147

	WIAT-III Essay Composition Subtest		
CBM-WE	**At-Risk**	**Not At-Risk**	**Total**
At-Risk	11	32	43
Not At-Risk	7	97	104
Total	18	129	147

Table 6

Area Under the Curve from Receiver Operating Characteristic Analyses

	Area Under the Curve	Standard Error	*p* value	95% Confidence Interval	
				Lower	Upper
CBM-WE	.761	.060	<.001	.644	.878
SAEBRS, Total Behavior Composite	.653	.064	.036	.528	.778

Table 7

Classification Accuracy of CBM-WE and the SAEBRS Predicting Writing Achievement Deficits on the WIAT-III Essay Composition Measure

Index	Sensitivity	Specificity	Positive Predictive Power	Negative Predictive Power
CBM-WE	61.11%	75.19%	25.58%	93.27%
SAEBRS, Total Behavior Composite	38.89%	76.74%	18.92%	90.00%

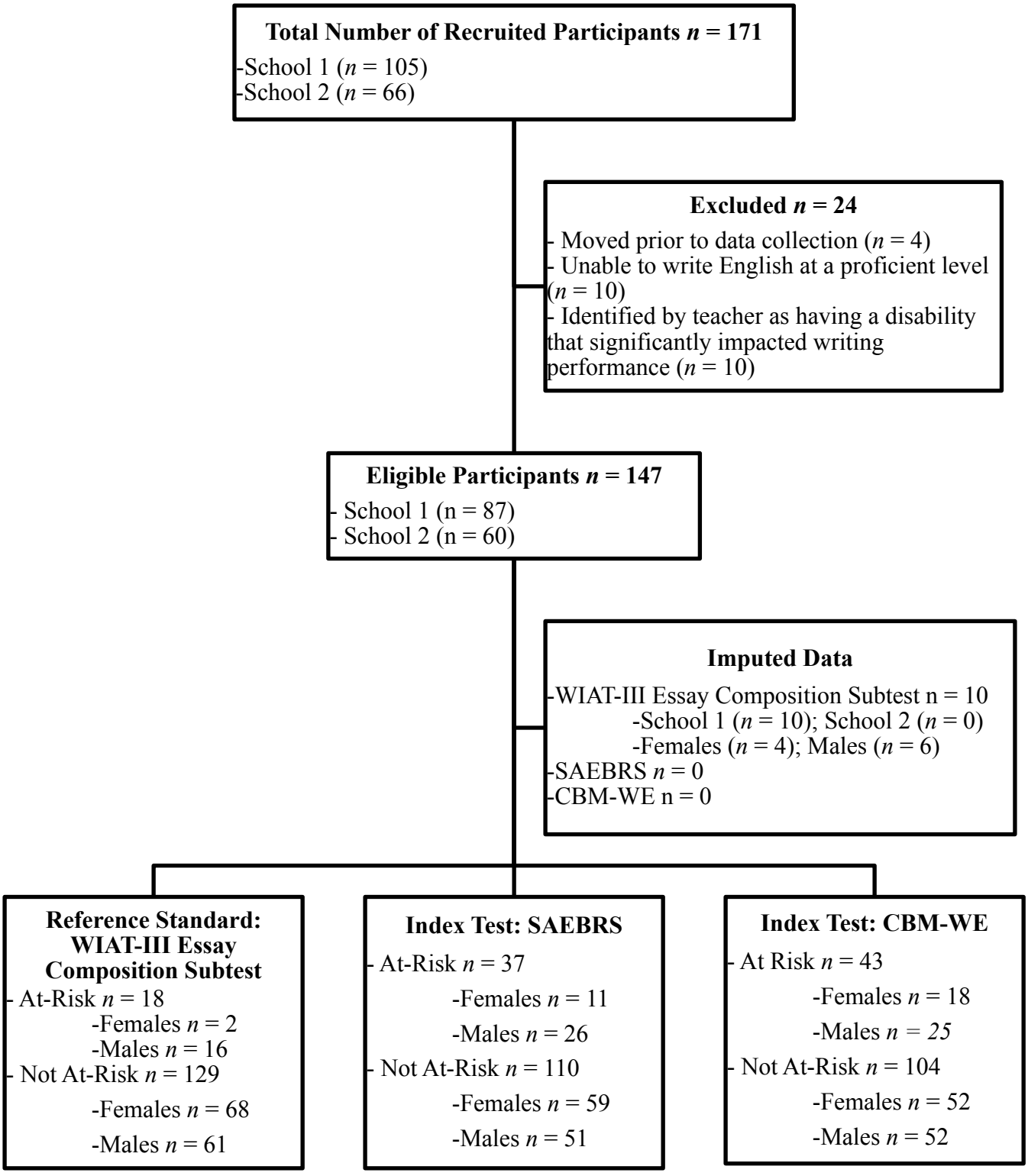

Figure 1. Flowchart of Participant Selection

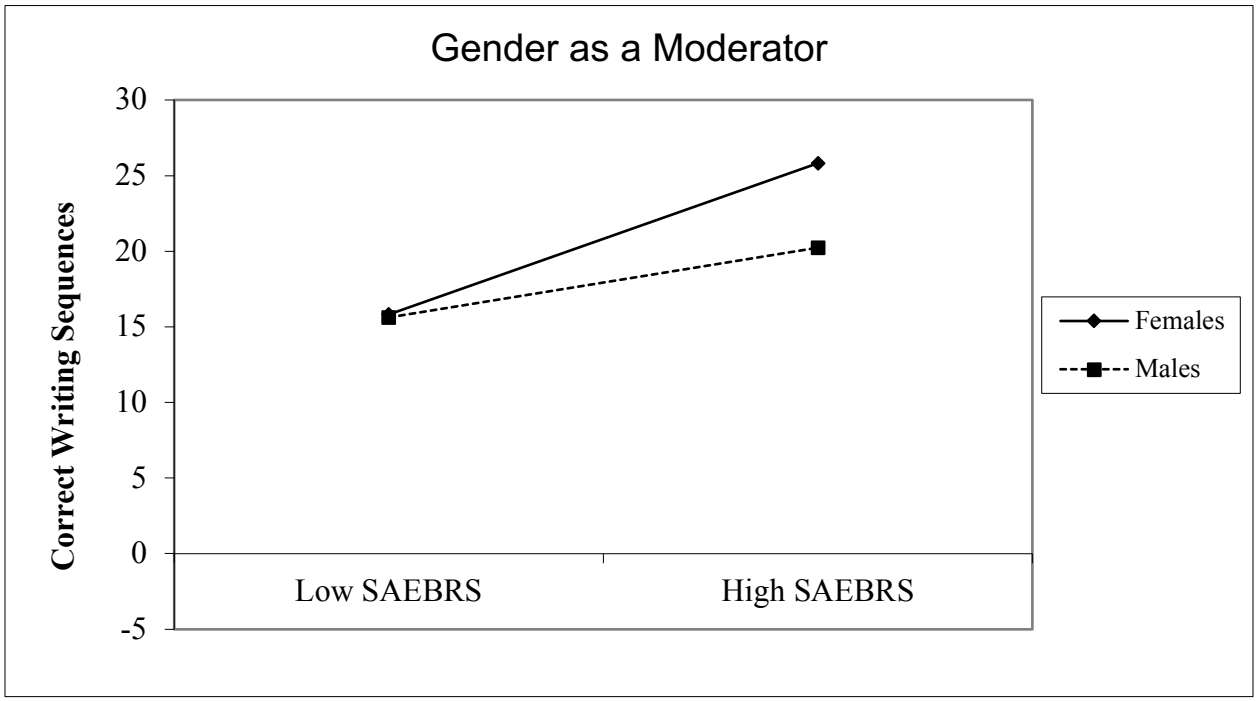

Figure 2. Moderation Analyses of the Relation Between the SAEBRS and CBM-WE

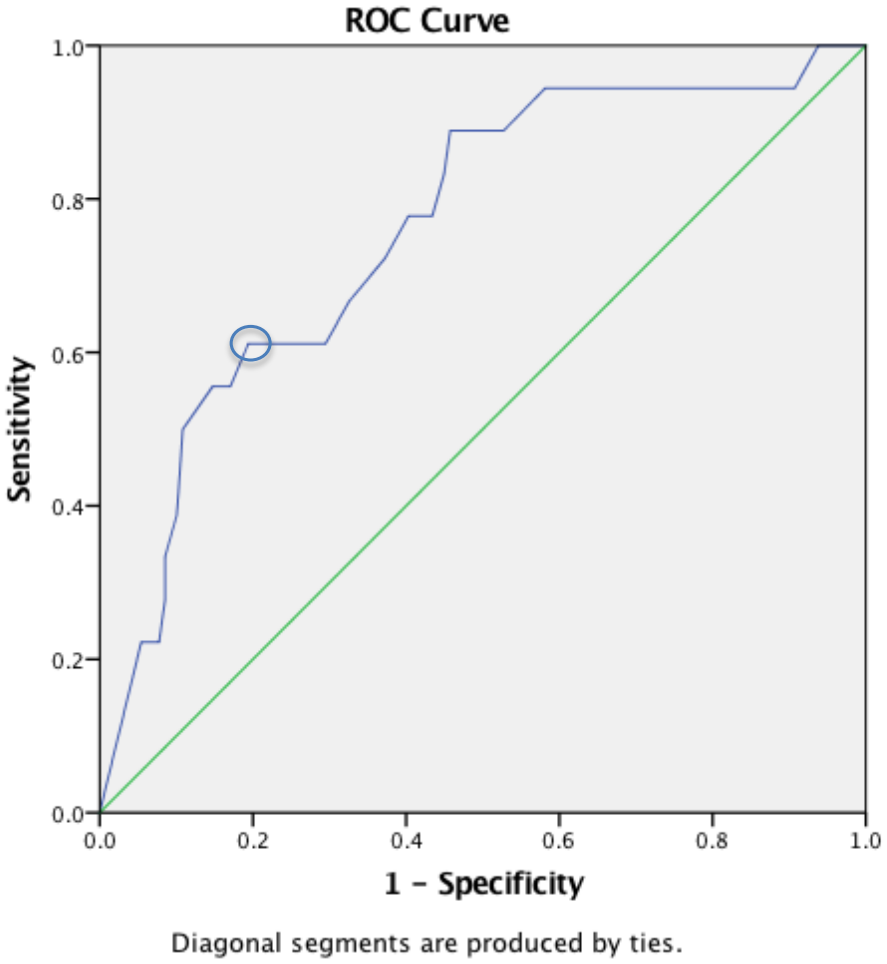

Figure 2. Classification Accuracy of CBM-WE

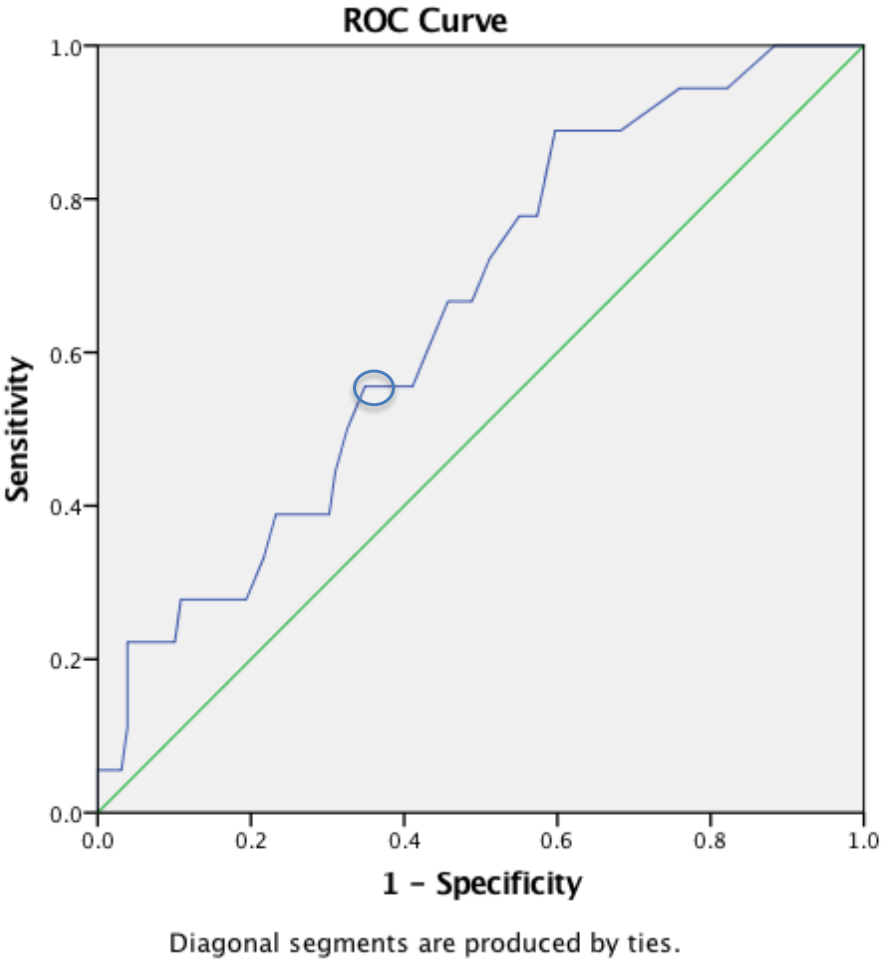

Figure 3. Classification Accuracy of the SAEBRS

Appendix A

SAEBRS Teacher Report of Classroom Behaviors

Student Name: _____

Directions: Using the following scale, identify how frequently the student has displayed each of the following behaviors during the previous month. Circle only one number for each behavior.

Social Behavior	**Never**	**Sometimes**	**Often**	**Almost Always**
Arguing	0	1	2	3
Cooperation with peers	0	1	2	3
Temper outbursts	0	1	2	3
Disruptive behavior	0	1	2	3
Polite and socially appropriate response to others	0	1	2	3
Impulsiveness	0	1	2	3
Academic Behavior	**Never**	**Sometimes**	**Often**	**Almost Always**
Interest in academic topics	0	1	2	3
Preparedness for instruction	0	1	2	3
Production of acceptable work	0	1	2	3
Difficulty working independently	0	1	2	3
Distractedness	0	1	2	3
Academic engagement	0	1	2	3
Emotional Behavior	**Never**	**Sometimes**	**Often**	**Almost Always**
Sadness	0	1	2	3
Fearfulness	0	1	2	3
Adaptable to change	0	1	2	3
Positive attitude	0	1	2	3
Worry	0	1	2	3
Difficulty rebounding from setbacks	0	1	2	3
Withdrawal	0	1	2	3

Appendix B

Multiple Regression Analysis Summary for the SAEBRS Composites Predicting Writing Fluency

Predictor	Females			Males		
	β	95% Confidence Interval		β	95% Confidence Interval	
		Lower	Upper		Lower	Upper
SAEBRS, Social Composite	-.195	-1.80	.382	-.107	-1.11	.605
SAEBRS, Academic Composite	.561**	.761	2.66	.283	-.183	1.50
SAEBRS, Emotional Composite	.005	-1.28	1.32	.075	-.800	1.20

Note. * $p \leq 0.05$, ** $p \leq 0.001$

References

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary- and intermediate-grade writers. *Journal of Educational Psychology*, *85*(3), 478–508. doi:10.1037/0022-0663.85.3.478

Achenbach, T. (1991). *Manual for the Teacher's Report Form and 1991 Profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington, VT.

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1993). First-grade classroom behavior: Its short- and long-term consequences for school performance. *Child Development*, *64*, 801–814. doi:10.2307/1131219

Allred, R. A. (1990). Gender differences in spelling achievement in grades 1 through 6. *Journal of Educational Research*, *83*, 187–193. doi:http://www.jstor.org/stable/27540383

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*, 233–240. doi:10.1177/10634266020100040501

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology, 72,* 29-48. doi:10/1016/j.jsp.2018.12.004

Berninger, V., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Gender differences in severity of writing and reading disabilities. *Journal of School Psychology*, *46*,

151–172. doi:10.1016/j.jsp.2007.02.007

Berninger, V. W., Abbott, R., Cook, C. R., & Nagy, W. (2016). Relationships of attention and executive functions to oral language, reading, and writing skills and systems in middle childhood and early adolescence. *Journal of Learning Disabilities*, *1*, 434–449. doi:10.1177/0022219415617167

Berninger, V. W., & Fuller, F. (1992). Gender differences in orthographic, verbal, and compositional fluency: Implications for assessing writing disabilities in primary grade children. *Journal of School Psychology*, *30*, 363–382. doi:10.1016/0022-4405(92)90004-O

Berninger, V. W., & Winn, W. D. (2006). Implications of advancements in brain research and technology for writing development, writing instruction, and educational evolution. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research.* (pp. 96–114). New York: Guilford Press.

California State Board of Education. (2010a). *California Common Core Standards: Mathematics*. Sacramento, CA.

California State Board of Education. (2010b). *California Common Core State Standards: English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects*. Sacramento, CA.

Camarata, S., & Woodcock, R. (2006). Sex differences in processing speed: Developmental effects in males and females. *Intelligence*, *34*, 231–252. doi:10.1016/j.intell.2005.12.001

Cameron, C. A., Lee, K., Webster, S., Munro, K., Hunt, A. K., & Linton, M. J. (1995). Text cohesion in children's narrative writing. *Applied Psycholinguistics*, *16*, 257. doi:10.1017/S0142716400007293

Camp, B. W., & Zimet, S. G. (1974). The relationship of teacher rating scales to behavior

observations and reading achievement of first-grade children. *Journal of Special Education*, *8*, 353–359. doi:10.1177/002246697400800408

Christ, T. J., Ara˜nas, Y. A., Kember, J. M., Kiss, A. J., McCarthy-Trentman, A., & White, M. J. (2014). *Formative assessment system for teachers technical manual: EarlyReading, CBMReading, aReading, aMath, and earlyMath*. Minneapolis, MN.

Cobb, J. A. (1972). Relationship of discrete classroom behaviors to fourth-grade academic achievement. *Journal of Educational Psychology*, *63*, 74–80.

Comprehensive Test of Basic Skills. (1981). *Examiner's manual*. Monterey, CA.

Conners, C. K. (1997). *Conners' Rating Scales–Revised*. Toronto, ON, Canada.

Conners, C. K., Sitarenios, G., Parker, J. D. A., & Epstein, J. N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology*, *26*(4), 279–291. doi:10.1023/A:1022606501530

CTB Macmillan/McGraw-Hill. (1992). *CAT/S, California Achievement Tests: Form A*. Monterey, CA.

Daniels, B., Volpe, R. J., Fabiano, G. A., & Briesch, A. M. (2017). Classification accuracy and acceptability of the integrated screening and intervention system teacher rating form. *School Psychology Quarterly*. doi:10.1037/spq0000147

DeBono, T., Hosseini, A., Cairo, C., Ghelani, K., Tannock, R., & Toplak, M. E. (2012). Written expression performance in adolescents with attention-deficit/hyperactivity disorder (ADHD). *Reading and Writing*, *25*, 1403–1426. doi:10.1007/s11145-011-9325-8

Decker, S. L., Roberts, A. M., Roberts, K. L., Stafford, A. L., & Eckert, M. A. (2016). Cognitive components of developmental writing skill. *Psychology in the Schools*, *53*, 617–625.

doi:10.1002/pits.21933

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children*, *48*, 368–371. doi:10.1177/001440298204800417

Dowdy, E., Doane, K., Eklund, K., & Dever, B. V. (2013). A Comparison of Teacher Nomination and Screening to Identify Behavioral and Emotional Risk Within a Sample of Underrepresented Students. *Journal of Emotional and Behavioral Disorders*, *21*, 127–137. doi:10.1177/1063426611417627

Drummond, T. (1994). *The student risk screening scale (SRSS)*. Grants Pass, OR.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446. doi:10.1037/0012-1649.43.6.1428

Dupaul, G. J. (1991). Parent and teacher ratings of ADHD symptoms: Psychometric properties in a community-based sample. *Journal of Clinical Child Psychology*, *20*, 245–253. doi:10.1207/s15374424jccp2003

DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M. J., & McGoey, K. E. (1998). Parent ratings of attention deficit/hyperactivity disorder symptoms: Factor structure and normative data. *Journal of Psychopathology and Behavioral Assessment*, *20*, 83–102. doi:10.1023/a:1023087410712

Dupaul, G. J., Rapport, M. D., & Perriello, L. M. (1991). Teacher ratings of academic skills: The development of the academic performance rating scale. *School Psychology Review*, *20*, 284–300.

Edelbrock, C., Greenbaum, R., & Conover, N. C. (1985). Reliability and concurrent relations

between the teacher version of the child behavior profile and the conners revised teacher rating scale. *Journal of Abnormal Child Psychology*, *13*, 295–303.

Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. *The California School Psychologist*, *14*, 89–95. doi:10.1007/BF03340954

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPower: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, *28*, 1–11.

Espin, C. A., Scierka, B. J., Skare, S., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary school students. *Reading and Writing Quarterly*, *15*, 5–27. doi:10.1080/105735699278279

Espin, C., Shin, J., Deno, S. L., Skare, S., Robinson, S., & Benner, B. (2000). Identifying indicators of written expression proficiency for middle school students. *The Journal of Special Education*, *34*, 140–153. doi:10.1177/002246690003400303

Fearrington, J. Y., Parker, P. D., Kidder-Ashley, P., Gagnon, S. G., Mccane-Bowling, S., & Sorrell, C. A. (2014). Gender differences in written expression curriculum-based measurement in third- through eighth-grade students. *Psychology in the Schools*, *51*, 85–96. doi:10.1002/pits.21733

Finn, J. D., Folger, J., & Cox, D. (1991). Measuring participation among elementary grade students. *Educational and Psychological Measurement*, *51*, 393–402. doi:10.1177/0013164491512013

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *The Elementary School Journal*, *95*, 421–

434. doi:10.1086/461853

Furey, W. M. W. M., Marcotte, A. M. A. M., Hintze, J. M. J. M., & Shackett, C. M. C. M. (2015). Concurrent Validity and Classification Accuracy of Curriculum-based Measurement for Written Expression. *School Psychology Quarterly*, *31*, 369–382. doi:10.1037/spq0000138

Goksuluk, D., Korkmaz, S., Zararsiz, G., & Karaağaoğlu, A. (2016). EasyROC: An interactive web-tool for ROC curve analysis using R language environment. *The R Journal*, *8*, 213–230.

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, *27*, 17–24. doi:10.1023/A:1022658222914

Goyette, C. H., Conners, C. K., & Ulrich, R. F. (1978). Normative data on Revised Conners Parent and Teacher Rating Scales. *Journal of Abnormal Child Psychology*, *6*, 221–236. doi:10.1007/BF00919127

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). New York: Guilford Press.

Graham, S., McKeown, D., Kiuhare, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, *104*, 879–896. doi:10.1037/a0029185; 10.1037/a0029185.supp (Supplemental)

Green, M. T. (1996). Understanding gender differences in referral of children to mental health services. *Journal of Emotional and Behavioral Disorders*, *4*, 182–190. doi:10.1177/106342669600400305

Hammill, D. D., & Larsen, S. C. (1978). *The Test of Written Language*. Austin, TX.

Hamsho, N. (2017). *The impact of classroom behaviors and student attention on written expression*. (Master's thesis) Available from ProQuest Dissertations and Theses database. (UMI No. 10256005).

Hamsho, N., Antshel, K. M., Eckert, T. L., & Kates, W. R. (2017). Childhood predictors of written expression in late adolescents with 22q11.2 deletion syndrome: A longitudinal study. *Journal of Intellectual Disability Research*, *61*, 501–511. doi:10.1111/jir.12370

Hedges, L., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41–45. doi:10.1126/science.7604277

Hier, B. O., & Eckert, T. L. (2014). Evaluating elementary-aged students' abilities to generalize and maintain fluency gains of a performance feedback writing intervention. *School Psychology Quarterly*, *29*, 488–502. doi:10.1037/spq0000040

Hintze, J. M., & Matthews, W. J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review*. ProQuest Central pg. doi:Article

Hooper, S. R., Costa, L. J., McBee, M., Anderson, K. L., Yerby, D. C., Knuth, S. B., & Childress, A. (2011). Concurrent and longitudinal neuropsychological contributors to written language expression in first and second grade students. *Reading and Writing*, *24*, 221–252. doi:10.1007/s11145-010-9263-x

Hooper, S. R., Swartz, C. W., Wakely, M. B., de Kruif, R. E. L., & Montgomery, J. W. (2002). Executive functions in elementary school children with and without problems in written expression. *Journal of Learning Disabilities*, *35*, 57–68. doi:10.1177/002221940203500105

Individuals with Disabilities Education Act, Pub. L. No. 1400 (2004). U.S.C.

Johnson, E., Mellard, D. F., Fuchs, D., & McKnight, M. (2006). *Responsiveness to intervention (RTI): How to do it*. Lawrence, KS.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*, *78*, 243–255. doi:10.1037/0022-0663.78.4.243

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

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

Keller-Margulis, M. A., Mercer, S. H., Payan, A., & McGee, W. (2015). Measuring annual growth using written expression curriculum-based measurement: An examination of seasonal and gender differences. *School Psychology Quarterly*, *30*, 276–288. doi:10.1037/spq0000086

Kent, S., Wanzek, J., Petscher, Y., Al Otaiba, S., & Kim, Y.-S. (2014). Writing fluency and quality in kindergarten and first grade: The role of attention, reading, transcription, and oral language. *Reading and Writing*, *27*, 1163–1188. doi:10.1007/s11145-013-9480-1

Kilgus, S. P., Bowman, N. A., Christ, T. J., & Taylor, C. N. (2017). Predicting academics via behavior within an elementary sample: An evaluation of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS). *Psychology in the Schools*, *54*, 246–260. doi:10.1002/pits.21995

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

School Psychology Quarterly, *28*, 210–226. doi:10.1037/spq0000024

Kilgus, S. P., Eklund, K., von der Embse, N. P., Taylor, C. N., & Sims, W. A. (2016). Psychometric defensibility of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) Teacher Rating Scale and multiple gating procedure within elementary and middle school samples. *Journal of School Psychology*, *58*, 21–39. doi:10.1016/j.jsp.2016.07.001

Kim, Y.-S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of dimensions, predictors, and the gender gap in written composition. *Journal of Educational Psychology*, *107*, 79–95. doi:10.1037/a0037210

Kim, Y.-S., Puranik, C., & Al Otaiba, S. (2015). Developmental trajectories of writing skills in first grade: Examining the effects of SES and language and/or speech impairments. *The Elementary School Journal*, *115*, 593–613. doi:10.1086/681971

Kim, Y.-S., & Schatschneider, C. (2017). Expanding the developmental models of writing: A direct and indirect effects model of developmental writing (DIEW). *Journal of Educational Psychology*, *109*(1), 35–50. doi:10.1037/edu0000129

Koster, M., Tribushinina, E., de Jong, P. F., & van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research*, *7*, 249–274. doi:10.17239/jowr-2015.07.02.2

Lakes, K. D., Swanson, J. M., & Riggs, M. (2011). The reliability and validity of the English and Spanish Strengths and Weaknesses of ADHD and Normal Behavior Rating Scales in a preschool sample: Continuum measures of hyperactivity and inattention. *Journal of Attention Disorders*, *16*, 510–516. doi:10.1177/1087054711413550

Lam, C. M., & Beale, I. L. (1991). Relations among sustained attention, reading performance,

and teachers' ratings of behavior problems. *Remedial and Special Education*, *12*, 40–47. doi:10.1177/074193259101200208

Lane, K. L. (2007). Identifying and supporting students at risk for emotional and behavioral disorders within multi-level models: Data driven approaches to conducting secondary interventions with an academic emphasis. *Education and Treatment of Children*, *30*, 135–164. doi:10.1353/etc.2007.0026

Lane, K. L., Bruhn, L. A., Eisner, S. L., & Kalberg, J. R. (2010). Score reliability and validity of the Student Risk Screening Scale: A psychometrically sound, feasible tool for use in urban middle schools. *Journal of Emotional and Behavioral Disorders*, *18*, 211–224. doi:10.1177/1063426611400082

Lane, K. L., Little, M. a., Casey, a. M., Lambert, W., Wehby, J., Weisenbach, J. L., & Phillips, A. (2009). A comparison of systematic screening tools for emotional and behavioral disorders: How do they compare? *Journal of Emotional and Behavioral Disorders*, *17*, 93–105. doi:10.1177/1063426608326203

Lane, K. L., Oakes, W. P., Harris, P. J., Menzies, H. M., Cox, M., & Lambert, W. (2012). Initial evidence for the reliability and validity of the Student Risk Screening Scale for internalizing and externalizing Behaviors at the elementary level. *Behavioral Disorders*, *37*, 99–122.

Lei, P. W., Bassiri, D., & Schulz, E. M. (2001). *Alternatives to the grade point average as a measure of academic achievement in college*. *ACT Research Report Series*. Iowa City, IA.

Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement*, *38*, 1193–1200.

Lloyd, J. W., Kauffman, J. M., Landrum, T. J., & Roe, D. L. (1991). Why do teachers refer pupils for special education? An analysis of referral records. *Exeptionality*, *2*(May 2015),

115–126. doi:10.1080/09362839109524774

MacArthur, C. A., Graham, S., & Fitzgerald, J. (2016). Introduction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *The Handbook of Writing Research, Second Edition* (2nd ed., pp. 1–7). New York, NY: The Guilford Press.

Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools*, *40*, 379–390. doi:10.1002/pits.10096

Mayes, S. D., Calhoun, S. L., Bixler, E. O., & Zimmerman, D. N. (2009). IQ and neuropsychological predictors of academic achievement. *Learning and Individual Differences*, *19*, 238–241. doi:10.1016/j.lindif.2008.09.001

McKinney, J. D., Mason, J., Perkerson, K., Clifford, M., Porter, F., Child, G., & Carolina, N. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology*, *87*, 198–203.

McMaster, K. L., Du, X., & Pétursdóttir, A.-L. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities*, *42*, 41–60. doi:10.1177/0022219408326212

McMaster, K. L., Shin, J., Espin, C. A., Jung, P. G., Wayman, M. M., & Deno, S. L. (2017). Monitoring elementary students' writing progress using curriculum-based measures: grade and gender differences. *Reading and Writing*, *30*, 2069–2091. doi:10.1007/s11145-017-9766-9

McMaster, K. L., Wayman, M. M., Deno, S. L., Espin, C. A., & Yeo, S. (2010). *Examining technical features of progress monitoring measures across grade levels in writing*. (Research Report No. 38). Retrieved from Research Institute on Progress Monitoring

Mikulecky, L. (1998). Adjusting school writing curricula to reflect expanded workplace writing. In M. S. Garay & S. A. Bernhardt (Eds.), *Expanding Literacies: English Teaching and the New Workplace* (pp. 201–223). Albany, NY: State University of New York Press.

Molitor, S. J., Langberg, J. M., & Evans, S. W. (2016). The written expression abilities of adolescents with Attention-Deficit/Hyperactivity Disorder. *Research in Developmental Disabilities*, *51*, 49–59. doi:10.1016/j.ridd.2016.01.005

Morrisroe, J. (2014). *Literacy Changes Lives 2014: A New Perspective on Health, Employment and Crime*. London.

Nantais, M., Martin, K., & Barnes, A. (2014). Best practices in facilitating professional development of school personnel in delivering multitiered services. In P. Harrison & A. Thomas (Eds.), *Best Practices in School Psychology: Systems-Level Services* (pp. 71–82). Bethesda, MD: National Association of School Psychologists.

National Center for Education Statistics. (2012). *The nation's report card: Writing 2011*. (NCES 2012-470). Washington DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

National Center on Intensive Intervention. (2018). *FAST: Social, Academic, & Emotional Behavior Risk Screener (SAEBRS)*. Washington DC.

Pearson, P. D. (2009). *Wechsler individual achievement test – third edition*. San Antonio, TX.

Rabiner, D. L., & Coie, J. D. (2000). Early attention problems and children's reading achievement: A longitudinal investigation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*, 859–867. doi:10.1097/00004583-200007000-00014

Reid, N. A., & Elley, W. B. (1991). *Progressive achievement tests of reading*. Wellington, N.Z.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children - Second*

Edition (BASC-2). Circle Pines, MN: AGS.

Reynolds, M. R., Scheiber, C., Hajovsky, D. B., Schwartz, B., & Kaufman, A. S. (2015). Gender Differences in Academic Achievement: Is Writing an Exception to the Gender Similarities Hypothesis? *Journal of Genetic Psychology*, *176*, 211–234. doi:10.1080/00221325.2015.1036833

Rogers, L. A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology*, *100*, 879–906. doi:10.1037/0022-0663.100.4.879

Ross, A. O., Lacey, H. M., & Parton, D. A. (1965). The development of a behavior checklist for boys. *Child Development*, *36*, 1013–1027.

Sáez, L., Folsom, J. S., Al Otaiba, S., & Schatschneider, C. (2012). Relations among student attention behaviors, teacher practices, and Beginning Word Reading Skill. *Journal of Learning Disabilities*, *45*(5), 418–432. doi:10.1177/0022219411431243

Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychology in the Schools*, *52*, 335–348. doi:10.1002/pits.21827

Shapiro, E. S. (2004). *Academic skill problems: Direct assessment and intervention*. New York: Guilford.

SPSS Inc. (2015). *SPSS Base 23.0 for Windows User's Guide*. Chicago, IL.

Stake, R. E., & Hastings, T. A. (1964). Stanford Achievement Battery. *Personnel & Guidance Journal*, *43*, 178–184. doi:10.1002/j.2164-4918.1964.tb02761.x

Swanson, J., Shuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., & McCleary, R. (2006). *Categorical and dimensional definitions and evaluations of symptoms of ADHD:*

The SNAP and SWAN Rating Scales. Unpublished manuscript. Irvine, CA.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnosis: Collected papers*. Mahwah, New Jersey: Erlbaum.

The National Commission on Writing. (2004). *Writing: A Ticket to Work... Or a Ticket Out: A Survey of Business Leaders*. *Writing*. Retrieved from http://www.collegeboard.com/prod_downloads/writingcom/writing-ticket-to-work.pdf

Tiegs, E., & Clark, W. W. (1957). *California Achievement Tests*. *California Achievement Tests*. Retrieved from http://login.proxy.library.vanderbilt.edu/login?url=http://search.proquest.com/docview/615351882?accountid=14816%0Ahttp://sfx.library.vanderbilt.edu/vu?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:book&genre=book&sid=ProQ:PsycINFO&atitle=&title=Ca

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing*, *26*, 17–44. doi:10.1007/s11145-012-9379-2

Videen, J., Deno, S., & Marston, D. (1982). *Correct word sequences: A valid indicator of proficiency in written expression*. Minneapolis, MN: Minnesota University Institution for research on Learning Disabilities.

von der Embse, N. P., Pendergast, L. L., Kilgus, S. P., & Eklund, K. R. (2016). Evaluating the applied use of a mental health screener: Structural validity of the social, academic, and emotional behavior risk screener. *Psychological Assessment*, *28*, 1265–1275. doi:10.1037/pas0000253

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices*. Belmont, CA: Wadsworth.

Walker, H. M., & Severson, H. (1992). *Systematic Screening for Behavior Disorders: Technical manual*. Longmont, CO.

Walker, H. M., & Shinn, M. R. (2010). Systemic, evidence-based approaches for promoting positive student outcomes within a multitier framework: Moving from efficacy to effectiveness. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for Achievement and Behavior Problems in a Three-Tier Model Including RTI* (pp. 1–26). Bethesda, MD: National Association of School Psychologists.

Wanzek, J., Gatlin, B., Al Otaiba, S., & Kim, Y. S. G. (2017). The impact of transcription writing interventions for first-grade students. *Reading and Writing Quarterly*, *33*, 484–499. doi:10.1080/10573569.2016.1250142

Wechsler, D. (2009). *Wechsler individual achievement test – third edition*. San Antonio, TX.

Wentzel, K. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, *85*(2), 357–364. doi:10.1037/0022-0663.85.2.357

Werry, J. S., & Hawthorne, D. (1976). Conners teacher questionnaire: Norms and validity. *Australian and New Zealand Journal of Psychiatry*, *10*, 257–262. doi:10.3109/00048677609159508

Wilkinson, G. S. (1993). *Wide Range Achievement Test 3 (wrat3)*. *Wide Range*. Wilmington, DE. https://doi.org/10.1007/978-1-4615-1185-4_13

Williams, G. J., & Larkin, R. F. (2013). Narrative writing, reading and cognitive processes in middle childhood: What are the links? *Learning and Individual Differences*, *28*, 142–150.

doi:10.1016/j.lind.2012.08.003

Wilson, M. S., & Reschly, D. J. (1996). Assessment in school psychology training and practice. *School Psychology Review*, *25*, 9–23.

Woodcock, R. W., & Johnson, M. (1989). *Woodcock-Johnson Psychoeducational Battery-Revised*. Allen, TX.

Zill, N., Furstenberg, F. F., Peterson, J., & Moore, K. (1992). *National Survey of Children: Wave I, 1976, Wave II, 1981, and Wave III, 1987*. Ann Arbor, MI.

Vita

Narmene Hamsho
Syracuse University
Department of Psychology
430 Huntington Hall
Syracuse, NY 13244

EDUCATION

Doctor of Philosophy, School Psychology **(Anticipated) August 2019**
Syracuse University
Advisor: Tanya Eckert, Ph.D.
Dissertation: *Examining the Classification Accuracy of the Social, Academic, Emotional Behavior Risk Screener and Its Relationship with Writing Performance*
Committee Members: Kevin Antshel, Ph.D.; Whitney Wood, Ph.D.; Leonard Newman, Ph.D. Shannon Sweeney, Ph.D.

Master of Science, Psychology **May 2017**
Syracuse University
Advisor: Tanya Eckert, Ph.D.
Master's Thesis: *The Impact of Classroom Behavior and Student Attention on Written Expression*
Committee Members: Kevin Antshel, Ph.D.; Whitney Wood, Ph.D.; Joshua Felver, Ph.D.

Bachelor of Arts, Psychology and Health and Human Services **May 2014**
Minor, Education
University at Buffalo, The State University of New York (SUNY)
Advisor: Gregory Fabiano, Ph.D.
Honors Thesis: *An Examination of the Influence of Parental ADHD Symptoms on Parenting Behaviors of Parents of Adolescents*

CLINICAL EXPERIENCES

School Psychology Intern **September 2018 – Present**
Oswego City School District
Supervisors: Larry Lewandowski, Ph.D. and John Garruto, Ed.D.

School Psychology Practicum Student **September 2017 –May 2018**
Syracuse City School District, Frazer K-8
Supervisors: Larry Lewandowski, Ph.D., Joshua Felver, Ph.D., and Kristi Cleary, Ph.D.

Consultation Practicum Student **February 2017 – May 2017**
Syracuse City School District, Frazer K-8
Supervisors: Brian Martens, Ph.D. and Kristi Cleary, Ph.D.

Neuropsychology Intern **July 2016 – July 2017**
SUNY Upstate University Hospital Cancer Center & Concussion Clinic
Supervisor: Brian Rieger, Ph.D. and Laura Jenkins, M.S.SpEd

ADHD Diagnostic Team Member **August 2015-August 2016**
SUNY Upstate University Hospital, Department of Psychiatry and Behavioral Sciences

Supervisor: Kevin Antshel, Ph.D. and George Starr, M.D.

ADHD Group Intervention Co-Facilitator **March 2016 – May 2016**
Syracuse University Psychological Services Center
Supervisor: Kevin Antshel, Ph.D.

Graduate Student Clinician **August 2015 – June 2016**
Syracuse University Psychological Services Center
Supervisor: Kevin Antshel, Ph.D. and Sarah Felver, Ph.D.

Graduate Assistant **March 2015 – March 2016**
Syracuse University Program for Refugee Assistance

Social Skills Group Facilitator **August 2014 – June 2016**
SUNY Upstate University Hospital Department of Psychiatry and Behavioral Sciences
Supervisor: Kevin Antshel, Ph.D.

Clinical Intern **June 2012 – August 2012**
Child Mind Institute
Supervisor: Steven Kurtz, Ph.D.

Advocate **May 2013 – May 2014**
Crisis Services

TEACHING EXPERIENCE

Undergraduate Course Instructor
Psychology of Adolescence, Syracuse University **Summer 2018**
Psychology of Adult Life and Aging, Syracuse University **Summer 2016**
Behavior Disorders in Children, Syracuse University **Summer 2015; 2017-2018**

Guest Lecturer **April 17, 2018**
Syracuse University's National Alliance on Mental Illness
Entitled: *Why is there Controversy Surrounding "13 Reasons Why"?*

Teaching Assistant
Foundations of Human Behavior, Syracuse University **2014- 2015**
Introductory Psychology **Summer 2013**

Student Teacher
Lorraine Elementary, Buffalo City Public School **Sept. 2012–January 2013**
University at Buffalo's Early Childhood Research Center **January 2012 – May 2012**
P.S. 54 and P.S. 58 New York City Public School **August 2008 – March 2010**

POSTER PRESENTATION

Williams, N.L., Eckert, T.L., Malandrino, R.M., Eggleston, B.N., **Hamsho, N.**, & Circe, J. (2019, February). *Improving students' writing and spelling skills through an integrative intervention.* Poster session presented at the annual convention of the National Association for School Psychologists, Atlanta, GA.

Hamsho, N. & Williams, N.L. (2018, October). *Literacy performance of urban students with ADHD symptomatology*. Poster session presented at the annual convention of the New York Association of School Psychologists, Lake Placid, NY.

Hamsho, N., Eckert, T.L., Williams, N.L., Eggleston, B.N., and Malandrino, R.D. (2018, February). *Why writing interventions might fail: An examination of student behaviors*. Poster session presented at the annual convention of the National Association for School Psychologists, Chicago, IL.

Eggleston, B.N., Eckert, T.L., Malandrino, R.D., **Hamsho, N.**, & Williams, N.L. (2017, February). *Relationship between writing self-Efficacy and fluency in an academic intervention.* Poster session presented at the annual convention of the National Association for School Psychologists, San Antonio, TX.

Hamsho, N., Eckert, T. L., Malandrino, R. D. (2016, February). *Behavioral predictors of writing performance: Third grade students*. Poster session presented at the annual convention of the National Association of School Psychologists, New Orleans, LA.

Eckert, T. L., Malandrino, R. D., Eggleston, B. N., Kim, Y., **Hamsho, N.,** & Williams, N. L. (2016, February). *Enhancing performance feedback interventions to increase elementary students' writing proficiency.* Poster session presented at the annual convention of the National Association of School Psychologists, New Orleans, LA.

PUBLICATIONS

Eckert, T. L., Hier, B. O., **Hamsho, N. F**., & Malandrino, R. D. (2017). Assessing children's perceptions of academic interventions: The Kids Intervention Profile. *School Psychology Quarterly, 32*, 268-281. doi:10.1037/spq0000200

Hamsho, N., Antshel, K. M., Eckert, T. L., & Kates, W. R. (2017). Childhood predictors of written expression in late adolescents with 22q11.2 deletion syndrome: A longitudinal study. *Journal of Intellectual Disability Research, 61*, 501-511. doi:10.1111/jir.12370

Haboush, K. L., Meltzer, A., Wang, R., & **Hamsho, N.** (2016). Cultural Competence and Child Interviewing: Understanding Religious Factors in Child Sexual Abuse Interviewing. In Parsons, O. (Ed.). *Sexual Abuse: Intervention, Coping Strategies and Psychological Impact*. Hauppauge, NY: Nova Publishers.

CERTIFICATIONS AWARDS AND HONORS

- *Ted Bernstein Winner* (Fall 2018), New York Association of School Psychologists (NYASP): Awarded to graduate students who demonstrate high quality work, leadership potential, and a strong interest in professional development.
- *Safer People, Safer Places Training Certificate* (Fall 2018), Syracuse University LGBT Resource Center: Completed a 3-hour seminar discussing issues related to the LGBTQ+ community and considered campus climates.
- *Certificate in University Teaching* (Spring 2017), Syracuse University: Reflects a strong commitment to professional development as a university instructor and documents preparedness to hold faculty appointments.
- *Outstanding Teaching Assistant Award* (Spring 2017), Syracuse University: Recognizes instructors "who have made distinguished contributions to Syracuse University by demonstrating excellence in significant instructional capacities".
- *Graduate Travel Award* (Spring 2016, 2018), Graduate Student Organization (GSO), Syracuse University
- *Graduate Travel Award* (Spring 2016; Spring 2018), Department of Psychology at Syracuse University
- *Graduate Tuition Scholarship* (2014 – present), Syracuse University
- *Psi Chi induction* (Spring 2013), International honor society in psychology
- *Honors Award* (Spring 2014), University at Buffalo, SUNY
- *Dean's Lists* (Fall 2011, Fall 2012, Spring 2013), University at Buffalo, SUNY

LEADERSHIP

- *Mentor* to an undergraduate psychology student (Spring 2016), Psychology Mentorship Program within the Department of Psychology at Syracuse University
- *Liaison* between the graduate students and the school psychology program faculty (Fall 2016 – Fall 2017), Department of Psychology at Syracuse University
- *Peer Mentor* to an incoming graduate student (Fall 2015 – present), Department of Psychology at Syracuse University
- *Member of the Future Professoriate Program* (Fall 2015 – Spring 2018), The Graduate School at Syracuse University
- *Outreach committee member* (Fall 2014- Spring 2015), Graduate Student Organization (GSO) at Syracuse University
- *Secretary* (Fall 2014 – Spring 2015), Psychology Action Committee (PAC) within the Psychology Department at Syracuse University