Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Hayley Elia Date

Concurrent and Predictive Validity of the SDQ and the BASC-2 BESS in the Project to Learn About Youth- Mental Health

By

Hayley J. Elia

MPH

Behavioral Sciences and Health Education

Delia Lang, PhD

Committee Chair

Susanna Visser, DrPH

Committee Member

Colleen McBride, PhD

Department Chair

Concurrent and Predictive Validity of the SDQ and the BASC-2 BESS in the Project to Learn About Youth- Mental Health

By

Hayley J. Elia

B.S.

University of South Carolina

2013

Thesis Committee Chair: Delia Lang, PhD

Committee Member: Susanna Visser, DrPH

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Behavioral Sciences and Health Education

2016

Abstract

Concurrent and Predictive Validity of the SDQ and the BASC-2 BESS in the Project to Learn About Youth- Mental Health

By Hayley J. Elia

Introduction: The early identification of behavioral and emotional difficulties is imperative to guarantee that the mental health needs of youth are met. Limitations in traditional identification systems have led a large number of at-risk youth to remain unrecognized and without appropriate services. School-based screening improves upon conventional systems by demanding an assessment of all children potentially at-risk for mental disorders so that at-risk youth may be identified and linked to services more efficiently. However, comparative research about instruments used to conduct school-based screening is lacking. The goal of this paper is to assess the concurrent and predictive validity of two widely used screening instruments, the SDQ and BASC-2 BESS.

Methods: Data from Stage 1 of the Project to Learn About Youth- Mental Health were used to assess linear associations and levels of agreement between the teacher forms of the SDQ and BASC-2 BESS. Data from Stage 1 were then merged with in-depth mental health assessment data to evaluate each instrument's ability to accurately predict DSM-IV criteria for mental health disorders overall and by internalizing and externalizing subcategories as ascertained by the gold standard, DISC version-IV.

Results: The SDQ and BASC-2 BESS demonstrated sufficient concurrent validity overall [$\kappa = 0.64$ (95% CI: 0.60, 0.68); $r = 0.85$, $p < .0001$]. Levels of agreement varied according to students' grade level, with transition grades demonstrating the lowest levels of agreement. The SDQ demonstrated greater sensitivity, positive predictive power, and negative predictive power, while the BASC-2 BESS demonstrated greater specificity. Both screeners demonstrated low positive predictive power in predicting internalizing conditions.

Discussion: Both screeners demonstrated sufficient levels of concurrent validity. Agreement between instruments was lowest for grade levels typically viewed as transition periods in students' schooling. Agreement was also lowest for students performing in the top 25% of students, indicating that emotional and behavioral characteristics may be more difficult to ascertain for top-performing students. The SDQ demonstrated a greater ability to classify children according to risk level and to detect children truly at risk, while the BASC-2 BESS was better able to detect low-risk individuals. Neither screening tool demonstrated adequate predictive validity for internalizing disorders.

Concurrent and Predictive Validity of the SDQ and the BASC-2 BESS in the Project to Learn About Youth- Mental Health

By

Hayley J. Elia

B.S.

University of South Carolina

2013

Thesis Committee Chair: Delia Lang, PhD

Committee Member: Susanna Visser, DrPH

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Behavioral Sciences and Health Education

2016

Acknowledgments

I would like to thank the Child Development Studies Team for allowing me to engage in this PLAY-MH work in order to conduct the current analysis. I would like to particularly thank Dr. Susanna Visser for her abounding support and guidance in this project. Thank you to Joe Holbrook and Melissa Danielson for their assistance in data preparation. A special thank you to Dr. Delia Lang for providing me with the foundations to conduct this research.

Table of Contents

.

CHAPTER 1: INTRODUCTION

Emotional and behavioral wellbeing is a pivotal factor in positive child development that contributes to a child's ability to thrive mentally, socially, and academically (O'Connell, Boat, & Warner, 2009; Lippman, Moore, & McIntosh, 2011). Children affected by emotional and behavioral difficulties face serious challenges that threaten optimal child development and may result in critical impairment in home, school, and community environments (Hoagwood, Jensen, Petti, & Burns, 1996; Satcher, 2000; O'Connell, Boat, & Warner, 2009). Nationwide, over 20% of children aged 4-17 years experienced emotional or behavioral difficulties in 2014, and an estimated 13 to 20% of children endure diagnosed mental, emotional, or behavioral disorders on an annual basis (National Health Interview Study, 2014; Perou et al., 2013).

According to data from the 2007 National Survey of Children's Health, an estimated 4.6% of children aged 3-17 years had a history of an externalizing disorder, such as oppositional defiant disorder (ODD) or conduct disorder, and an estimated 8.9% of children met ADHD criteria in the past year. Data from the same survey indicated that 3.9% of children had a history of diagnosed depression and an estimated 2.1% of children experienced current depression. An estimated 4.7% of children aged 3–17 years reported a past history of anxiety, and 3% reported current anxiety. The prevalence of phobias or fears diagnosed in the past year among children aged 4–17 years was 2.6% (Perou et al., 2013).

Mental disorders that originate early in life have been associated with a number of adverse short- and long-term health outcomes. Mental disorders have been linked to greater disciplinary concerns, decreased academic success, and an increased probability of engaging in risky behaviors during childhood (Gilliam, 2005; Institute of Medicine, 2006; Teplin et al., 2002; Visser et al., 2011). Individuals who experience early onset mental disorders are at a higher risk for substance abuse issues, crime, and even premature death in adulthood (Institute of Medicine, 2006; O'Connell, Boat, & Warner, 2009; Dalsgaard et al., 2015; Nordentoft et al., 2013). Further, the economic costs of pediatric mental disorders are great. From 2006 to 2011, the United States spent approximately $11.6 million on hospital services for mental illness and in 2007 incurred an estimated $247 billion in mental health service costs overall. For children specifically, the United States spent an estimated $13.9 billion

to address mental disorders among children aged 0-17 years in 2012 (Torio et al., 2015; Eisenberg & Neighbors, 2007; Soni, 2015).

Limitations in traditional systems used to identify mental health disorders have led a large number of youth experiencing emotional and behavioral difficulties to remain unrecognized and without appropriate services to address their needs. In 2013, only 23% of children with serious behavioral difficulties received special education services, and only 55% of parents of children with difficulties contacted a mental health provider about the child's needs (U.S. Department of Health and Human Services, 2015). The receipt of services for mental health needs is disproportionately low among youth of uninsured, lower income, and minority groups, and boys are more likely to receive non-pharmacological services for serious mental health difficulties than girls (Howie, Pastor, & Lukacs, 2014; Jones et al., 2014).

The early identification of behavioral and emotional difficulties is imperative to guarantee that the mental health needs of at-risk youth are met and to ensure that the negative effects of childhood mental disorders are mitigated. Strategies used to achieve early identification involve proactively assessing children and addressing mental health issues before symptoms become severe and lead to impairment in daily functioning. Support for early identification has grown in recent years as its integral role in the prevention and treatment of mental health disorders has been recognized by public health practitioners, mental health providers, policymakers and educators (Conroy, 2004; Doll & Cummings, 2008; Mills, et al., 2006; New Freedom Commission on Mental Health, 2003; Satcher, 2000; Severson, et al., 2007). On a population level, identification is vital for gaining a clear understanding of the epidemiology of mental, behavioral, and emotional difficulties among youth. Specifically, identification is crucial for examining the prevalence and distribution of children vulnerable to mental disorders; for assessing the burden of mental disorders on individuals, families, and communities; and for detecting deficiencies in the delivery of mental health services. Early identification is also essential in the development and implementation of effective programs focused on mental health promotion, prevention, and intervention (Dowdy, Ritchey, & Kanmphaus, 2010).

To date, no single system is solely dedicated to the identification of children with mental health needs. Population-level information about children with emotional and behavioral difficulties is largely obtained through clinical research, administrative claims databases, and national surveys (Danielson, Visser Bitsko, & Holbrook, 2015). Data from these sources are often collected anonymously and on an infrequent basis, restricting their ability to contribute to individual-level service linkage or routine mental health monitoring. Prevalence estimates calculated from these data are often established using information about youth with clinical diagnoses of mental disorders. Therefore, children who experience mental health issues but lack access to clinical care remain unidentified as cases and are thus omitted from these estimates. Further, children at lower levels of risk for mental disorders are generally not captured in these estimates (Dowdy et al., 2010).

1.1 Community-Based Screening as a Means of Identification

Emerging approaches to early identification seek to improve upon the shortcomings of traditional systems. One method for achieving early identification is conducting systematic community-based screening. This approach necessitates that all youth within a defined community receive assessment for emotional and behavioral difficulties using mental health screening tools. Based on screening data, researchers and community stakeholders can begin to understand the unique epidemiology of emotional and behavioral difficulties affecting youth within the target community. Community-based screening improves upon conventional systems, as it demands an assessment of all children potentially at-risk for emotional and behavioral difficulties, rather than those exclusively at the highest risk (Dowdy et al., 2010). In this way, systematic screening provides a population-based lens through which communities can more broadly view mental health. Due to its comprehensive nature, the information gleamed from community-based screening is ideal for tailoring prevention and treatment measures to address the distinctive mental health needs of communities. Screening information can also be used to direct community resources to address mental health needs; to connect families and children to appropriate services; and to develop research projects to explore potential factors influencing mental health and treatment patterns within the community (Centers for Disease Control and Prevention, 2015; Dowdy et al., 2010; Severson et al., 2007).

Community-based screening systems frequently operate within multiple gating frameworks, which utilize sequential assessments to identify children at the highest risk of emotional and behavioral difficulties within a population. Typically, the first gate requires communities to conduct population-wide screenings to detect all children potentially at risk. Children classified as at-risk continue to a second gate consisting of a more in-depth psychological evaluation. Each subsequent gate employs increasingly comprehensive assessment measures to identify with a substantial level of specificity children at the highest risk of mental disorders. Much research has supported the ability of such systems to improve diagnostic precision in a cost-efficient manner (Severson et al., 2007; Miller et al., 2014; Dowdy, Chin, Twyford, & Dever, 2011).

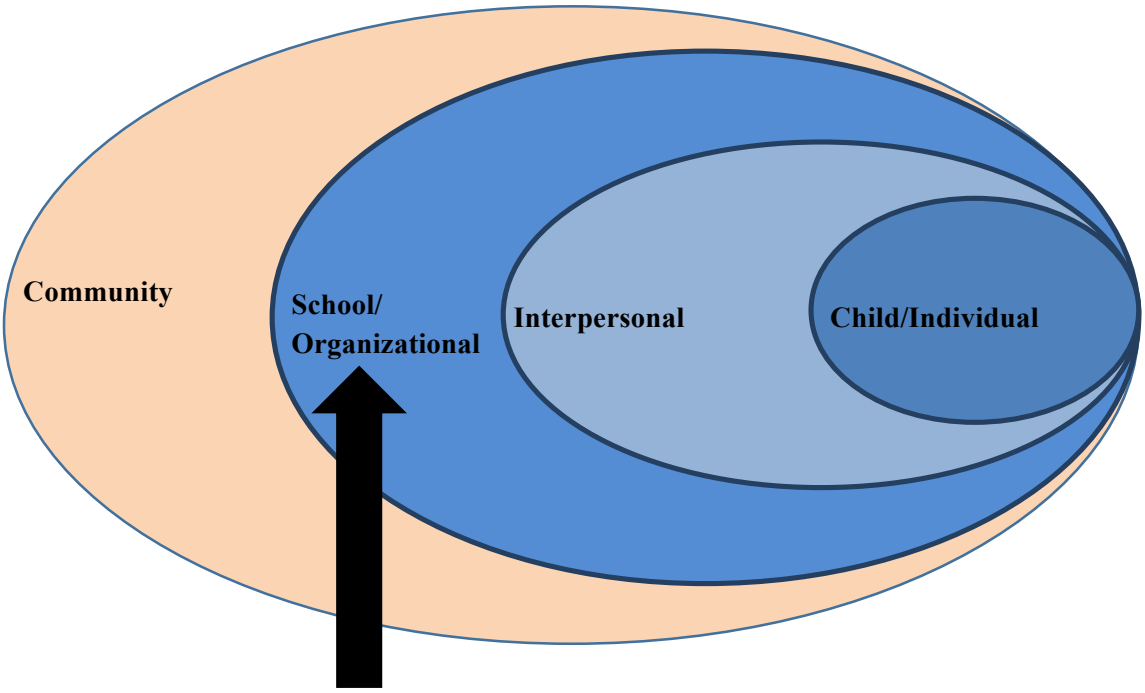
Although various environments can be leveraged, schools have been pinpointed as ideal settings for assessing the mental health needs of youth using community-based screening. Schools interact with students and families daily during the academic year, making them well situated to screen students. By screening students, schools are able to play an active role in preventing academic and social impairment that children may experience as a result of emotional and behavioral difficulties (Dowdy et al., 2010; Nemeroff et al., 2008; Satcher, 2000; Short, 2003). Further, since schools frequently serve as central providers of behavioral health resources for youth, screening provides a pathway to ensure that service linkage is accomplished in an efficient and timely manner (Dowdy et al., 2010).

1.2 Theoretical Underpinnings

Traditional identification systems operate within a referral-for-service framework, whereby youth at the greatest risk of experiencing mental disorders are identified on an individual level and are subsequently referred for clinical assessment and treatment (Dowdy et al., 2010). However, mental health issues arise from a culmination of a child's experiences that occur within multiple systems, including those beyond the individual level. Conversely, mental health issues impact the school, family, and local systems in which youth interact. From a theoretical standpoint, the socio-ecological model can be used to explain this reciprocal relationship between a child's mental health status and the environments in which a child interacts (Bronfenbrenner, 1979). As shown in Figure 1, in applying the socio-ecological model, each child represents the center of a succession of increasingly

large circles, each of which depicts progressively complex systems in which the child interacts. The mental health needs of the child must be addressed in a way that extends beyond the individual level if sustained change is to be achieved for both the child and for the systems within which the child lives. School-based screening initiatives target the mental health needs of children at the school level in order to create a lasting impact on the individual, interpersonal, and community levels. In this way, school-based screening programs take a preventative approach to addressing mental health needs on multiple levels rather than take the reactionary, individual-level approach that has been used in conventional referral-for-service approaches.

Figure 1. Socio-ecological model in the application of school-based mental health screening



1.3 Selection of High Quality Screening Tools

While the potential benefits of community-based screening are clear, the utility of this approach ultimately depends upon the quality of tools used to conduct screening. As described by Glover & Albers (2006),

three essential characteristics determine the quality of population-based screening tools: (a) usability of the tools, (b) appropriateness of the tools for their proposed use, and (c) technical adequacy. Each of these factors should be considered in identifying tools that are most suitable for use within a given screening context and population.

Usability refers to the degree at which a screening tool can be feasibly administered in a particular situation. First, the financial cost of utilizing the instrument should be commensurate to the benefits of its use. This criterion is particularly important in the selection of tools for school-based screening programs, as schools are often faced with competing needs and scarce resources. Additionally, the instrument should be practical to implement within the screening context at hand, another criterion that is particularly important in school-based screening systems that must account for the time constraints of teachers and school administrators. Overall, a tool that holds sufficient usability should be one that is user-friendly given the level of expertise of test administrators. Further, the screener should be realistic for use provided the resources available for implementation (Glover & Albers, 2006).

Appropriateness relates to whether the instrument is well aligned with the screening purpose and suitable for the population in which it is being administered. Specifically, the tool should consist of constructs that will properly measure the outcomes of interest within the screening context. As purported by Levitt, Saka, Romanelli, & Hoagwood (2007), screening instruments may be classified as broad, specialized, or targeted according to the purpose for which they were designed. Broad tools detect general emotional and behavioral difficulties that may be of concern. Specialized instruments, on the other hand, are used to identify symptoms that indicate when individuals may be at high risk for a range of mental health disorders. Targeted screeners are designed to distinguish individuals at risk for a small number of mental health disorders. The type of instrument chosen should be in agreement with the purpose for which it will be used. Additionally, evidence should support the applicability of measurement given the unique characteristics and makeup of the population.

Most importantly, technical adequacy describes the power of a screening tool to detect risk with a high degree of consistency and accuracy in the target population. Ideally, the instrument should have a proven level of reliability, or ability to consistently assess outcomes across test items, time points, and raters. The consistency of a

particular instrument is typically evaluated using three types of reliability measures: internal consistency, test-retest reliability, and inter-rater reliability. Internal consistency refers to the degree to which specific scale items measure the same construct. Test-retest reliability determines the level of consistency of an individual's responses to the same questionnaire across multiple time points. Inter-rater reliability gauges the consistency of responses when the same questionnaire is administered to the same individual by different interviewers (Glover & Albers, 2006; Elkin, 2012; Thompson, 2002).

In addition to reliability, a screening instrument must also have a high level of validity to be technically adequate. Validity refers to the capacity of a tool to accurately measure what it is designed to measure. Although it may be evaluated in many ways, validity is commonly assessed using the following measures: concurrent validity, positive predictive power, negative predictive power, specificity, and sensitivity. Concurrent validity evaluates how well measurement from a particular tool correlates with a similar tool. Positive predictive power measures the probability that individuals identified as at-risk by a screener are truly at risk. Negative predictive power evaluates the probability that individuals not detected as at-risk by a screener are truly not at risk. Sensitivity measures an instrument's ability to detect individuals who are truly at risk. Conversely, specificity assesses a tool's ability to correctly classify individuals not at risk (Cronbach, 1955; Elkin, 2012; Glover & Albers, 2006).

1.4 Statement of the Problem

Clearly, much thought and consideration is needed to identify tools that are most appropriate for a particular context. Empirical support is imperative in determining which instruments are most suitable for a specific population. Of particular value in this decision-making process is evidence demonstrating the comparative performance of multiple screeners within the same community. Through concurrent assessments, decision-makers can examine various screening options to ascertain which tool is most appropriate to address the community's needs.

Despite its importance, comparative information is lacking in the literature and within community-based programs aimed at conducting population-based screening. Such information is not only important to entities

implementing these systems; it is a valuable contribution to mental health surveillance efforts as a whole. This research seeks to add concurrent comparisons of broad mental health screening tools in order to fill to this gap in the literature.

1.5 Purpose

The purpose of this paper is to compare aspects of technical adequacy in two broad screening tools, the teacher form of the Strengths and Difficulties Questionnaire (SDQ) and the Behavior Assessment System for Children, Second Edition/Behavioral and Emotional Screening System (BASC-2 BESS). Data from the Stage 1 of the Project to Learn About Youth- Mental Health (PLAY-MH) will be used to directly compare the two screening instruments using measures of concurrent validity. Stage 2 data will then be merged with data from Stage 1 of the PLAY-MH in order to assess the predictive validity of the SDQ and BASC-2 BESS. Predictive validity will be assessed by comparing the performance of each instrument to the gold standard of meeting DSM-IV criteria for mental health disorders based on the Diagnostic Interview Schedule for Children (DISC) version-IV. Specifically, data will be analyzed to answer the following research questions:

1. Do the SDQ and BASC-2 BESS exhibit sufficient concurrent validity?

2. Do the SDQ and BASC equally predict mental health disorders as ascertained by the Diagnostic Interview Schedule for Children and the Mental Health Diagnosis, Treatment, and Satisfaction Questionnaire?

3. Do the SDQ and BASC-2 BESS equally predict externalizing and internalizing conditions?

1.6 Significance of the Research

Currently, a call to action from stakeholders and educators invested in the mental health of youths is growing louder in the movement to scale up school-based mental health screening programs (Doll & Cummings, 2008; Dowdy et al., 2010; Eklund et al., 2009; Levitt et al., 2007; Miller et al., 2015; Mills et al., 2006; Nemeroff, et al., 2008; Severson, et al., 2007; Short, 2003; Walker et al., 2000; Weist et al., 2007). At the same time,

literature comparing various options of instruments used to assess mental health is lacking in the area of mental health screening research. Without this evidence base, program developers are unable to weigh the strengths and weaknesses of various screening instruments. Further, their ability to properly select instruments that have been shown to perform accurately, consistently, and efficiently across school contexts is severely limited.

In particular, research involving direct comparisons of multiple screening instruments used to assess the same population is lacking. Specifically, little to no existing research has involved the direct comparison of the widely used SDQ and BASC-2 BESS instruments, particularly within the same population. This research will add to the literature on school-based mental health screening and to research focused on the psychometric properties of instruments used in the assessment of youth mental health. This research may be particularly valuable to communities that are considering the use of the SDQ and BASC-2 BESS. More narrowly, this research will help inform future decision-making in Stage 1 of the gating process in the PLAY-MH research study.

CHAPTER 2: LITERATURE REVIEW

The following review of the literature examines existing research related to the SDQ, BASC-2 BESS, and DISC questionnaires. The review focuses on literature that utilizes teacher versions of the SDQ and BASC-2 BESS, as data from these forms of the questionnaires will be analyzed in the proposed research. Specifically, studies evaluating measures of reliability and validity of the SDQ and BASC-2 BESS are emphasized in the current review in order to align with the outcome measures of interest in the proposed research.

2.1 BASC-2 BESS Overview

The BASC-2 BESS is a tool that was developed to measure the emotional and behavioral problems and adaptive behaviors for students aged 3 to 18. The instrument is available in parent, teacher, and self-report student versions, with the student and parent versions containing 30 items apiece and the teacher version containing 27 items. The instrument items are based on a four-factor model measuring externalizing behaviors, internalizing behaviors, adaptive skills, and inattention (Dowdy et al., 2011). Each item warrants a response on a 4-point ordinal scale (never, sometimes, often, almost always). Based on responses from all of the items, a raw summary

score is computed by summing the responses to items pertaining to behavior problems and summing reverse scores of items related to adaptive behaviors. All summary scores can be calculated by hand, computer-entry or through a scanning process. From the raw score, a standardized T-score is calculated (Dever, Mays, Kamphaus, & Dowdy, 2012). In terms of usability, the BASC-2 BESS can be completed within 5 minutes, and the instruments are proprietary, costing approximately $667 for online scoring software, $29 for a package of 25 teacher forms, and $75 for a BASC-2 BESS manual (www.pearsonclinical.com).

2.2 BASC-2 BESS Psychometric Properties & Reliability/Validity

Several studies have evaluated the level of reliability and validity of the BASC-2 BESS across research and community settings. The BASC-2 BESS technical manual developed by Kamphaus & Reynolds (2007) documented a 2-year longitudinal study, which yielded evidence of the predictive value of the BASC-2 BESS in predicting academic and behavioral outcomes in school. The manual demonstrated sufficient concurrent validity with the Child Behavior Checklist with a correlation coefficient of 0.76. Values of sensitivity from the same manual for the teacher form of the instrument ranged from 0.53-0.80 and values of specificity ranged from 0.90-0.95. Positive predictive values ranged from 0.47-0.77, while negative predictive values ranged from 0.92-0.96 (Kamphaus & Reynolds, 2007; Jenkins et al., 2014). Further, for internalizing problems, the BASC-2 BESS indicated a fairly low sensitivity of 0.30 (Kamphaus & Reynolds, 2007). Another study by Furlong (2009) reported satisfactory levels of test-retest reliability with values ranging from 0.80-0.91.

Other studies have examined the ability of BASC-2 BESS to predict school-based outcomes. Renshaw et al. (2009) detected a significant relationship between level of risk as indicated by a student's score on the BASC-2 BESS and mean report card scores in a sample of elementary school children. In a study by King, Reschly, and Appleton (2012), however, the teacher version of the BASC-2 BESS found low values of positive predictive power and sensitivity in the relationship between BASC-2 BESS risk level and school office referrals. Finally, in research conducted by Eklund, Tanner, Stoll, and Anway (2014) to assess emotional and behavioral risk in gifted versus non-gifted students, scores from the BASC-2 BESS indicated that gifted children had elevated internalizing behaviors in comparison to non-gifted students.

2.3 Strengths and Difficulties Questionnaire (SDQ) Overview

The SDQ is a 25-item instrument that assesses strengths as well as challenge areas in order to evaluate the emotional and behavioral health of a child. The SDQ was adapted from the Rutter Questionnaires, which it expanded upon using criteria from the Diagnostic and Statistical Manual of mental disorders (Goodman, 1997; Stone et al., 2010). The questionnaire includes both a parent and teacher version for 3- to 16-year-olds and a self-report questionnaire for individuals 11-16 years of age. The SDQ contains five subscales: the emotional symptoms scale, conduct problems scale, hyperactivity scale, peer problems scale, and prosocial scale. Each subscale consists of five items, which contribute to individual subscale scores. Additionally, the scores from all subscales other than the prosocial scale are combined to produce a total difficulties score, which ranges from 0 to 40. Extended versions of the questionnaire include an "Impact Supplement" to assess overall impairment (Goodman, 2001). In terms of usability, the SDQ is highly feasible for implementation in many contexts, as it is user-friendly and freely accessible online (www.sdqinfo.com). Additionally, the SDQ requires little to no training prior to use and can be scored electronically or by hand (Stone et al., 2010).

2.4 SDQ Psychometric Properties & Reliability/Validity

The SDQ was originally developed in Britain and was extensively evaluated in various clinical and community settings within the United Kingdom. In one of the earliest validation studies, the comparative validity of the SDQ teacher screener was tested against the Rutter Questionnaires from which the instrument was derived in a sample of 403 children aged 4-16 seen at a dental and psychiatric clinic in the U.K. Receiver operating characteristic curves were examined to determine how well the two questionnaires distinguished between high and low risk samples (children seen in the psychiatric clinic versus children seen in the dental clinic), and the correlation between the SDQ and Rutter Questionnaires were calculated. Similar areas under the curve were detected for both questionnaires (0.85 for the SDQ versus 0.84 for the Rutter Questionnaires), indicating a similar ability to distinguish between high and low risk samples. Agreement between the two questionnaires was also high with a Pearson correlation coefficient of 0.92 (Goodman, 1997). Similarly, a later study conducted in 1999 by Goodman & Scott indicated comparative validity between the SDQ and the Child Behavior Checklist in a

sample of 132 children aged 4-7 in Britain. In addition to obtaining similar results for the two questionnaires in a receiver operating curve analysis, the study indicated that the two questionnaires performed similarly in measuring externalizing behaviors and internalizing behaviors (Goodman, 1999). Other studies conducted in Britain have yielded similar results (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000; Goodman & Goodman, 2011).

Further, studies conducted in Britain have pointed to the predictive power of the SDQ in detecting mental health diagnoses in youth. In a study conducted in Britain with 1,025 children living residentially or in foster care, Goodman, Ford, Corbin, and Meltzer (2004) measured sensitivities of 82.7% and 97.7% for residential and foster care children, respectively, when comparing SDQ subscale and total difficulties scores with the Development and Well-Being Assessment. On the other hand, some studies conducted in Britain have yielded lower levels of validity. In a study by Goodman (2001), sensitivity was fairly low at 43% with a similarly low positive predictive value of 44%.

Since its original use in Britain, the SDQ has been widely used and evaluated in studies that were conducted in diverse communities and clinical settings across the world. In a review by Achenbach (2008), reliability and validity measures were provided from global studies conducted in Australia, Finland, Germany, Russia, Sweden, and a multi-European sample. The mean alpha coefficients across studies ranged from the 0.50s to 0.70s for individual subscales and from the 0.70s to the 0.80s for the Total Difficulties score. Test-retest correlations for the same studies averaged at 0.73 for teacher SDQ Total Difficulties scores. In the area of predictive power, one study conducted in Germany compared SDQ subscale scores to ICD-10 diagnoses using data from children seen in inpatient and outpatient settings. Receiver operating curve analyses indicated that areas under the curve averaged 0.72 for self-report questionnaires and 0.77 for parent questionnaires (Becker, Hagenberg, Roessner, Woerner, & Rothenberger, 2004). Another global review of studies conducted outside of Europe supported the sound psychometric properties of the SDQ in diverse settings (Woerner et al., 2004).

In addition to studies conducted in Europe and the broader global community, research conducted in the United States has pointed to its applicability in this country. In a study by Hill and Hughes (2007), the SDQ

yielded sufficient convergent validity in a confirmatory factor analysis that examined the fit of SDQ subscales using data from a longitudinal study of 784 children attending school in Texas. Results from a study conducted by the Violence Institute of New Jersey at UMDNJ (2013) indicated good test-retest reliability ranging from 0.70-0.85. Finally, a study implemented by Bourdon, Goodman, Rae, Simpson, and Koretz (2005) found good levels of internal consistency.

2.5 Diagnostic Interview Schedule for Children (DISC)

The National Institute of Mental Health Diagnostic Interview Schedule for Children (DISC) was developed for the purpose of obtaining epidemiological information about the prevalence of mental disorders among youth in the United States. The instrument is designed so that administers without formal clinical training could evaluate mental health diagnoses in children. DISC items assess criteria from the Diagnostic and Statistical Manual of Mental Disorders and the WHO International Classification of Diseases. Items ask for information regarding symptoms exhibited during the past year. The instrument is divided into six portions based on mental health disorder categories: Anxiety Disorders, Disruptive Disorders, Substance-Use Disorders, Schizophrenia and Miscellaneous Disorders (National Institute of Mental Health, 2006). Results from the survey can then be scored to obtain assessments for over 30 mental health diagnoses that do not require clinical observation or specialized testing. The DISC has been validated and has been used annually in the cross-sectional National Health and Nutrition Examination Survey (NHANES), as well as in numerous other epidemiological and clinical studies (Shaffer et al., 2000; Perou et al., 2013; Lucas et al., 2001; Costello, Edelbrock, & Costello, 1985).

CHAPTER 3: METHODOLOGY

3.1 PLAY-MH Study Design

PLAY-MH is a community-based longitudinal study implemented by the Center for Disease Control and Prevention's Child Development Studies Team to study mental disorders and related symptoms and impairment among youth in schools. The study is specifically aimed at gaining epidemiological information related to the prevalence of mental disorders, related risks, and treatment patterns. Currently, the study is being implemented at

four research sites based at the University of South Carolina, University of Colorado-Denver, University of Florida-Jacksonville, and Ohio University. Each research site is responsible for coordinating research activities at school districts within the surrounding community. Since data collection and processing had not been completed at other study sites at the time that these analyses were conducted, the current analyses are based on data collected by the University of South Carolina site, which included data for two school districts in South Carolina. While data were collected for all youths in grades K-12, an error in data processing for grades K-5 rendered the data for these grades unusable for the current analyses. Therefore, the current analyses are based on data collected for grades 6-12. (Danielson et al., 2015).

This study employed a multi-staged, stratified random sampling design to select youth for participation. Students within two school districts in South Carolina were assigned as primary sampling units, and stratification was used to increase the representativeness of particular subgroups within the community to allow for the estimation of strata of interest for the specific aims of the study (Wolraich et al., 2014).

PLAY-MH utilizes a multiple-gating system to identify children in schools who are at-risk for mental health disorders. This multiple-gating system consists of a 2-stage design, in which a broad assessment of mental health risk for all students is completed in Stage 1, and an in-depth assessment of a subsample of students is completed in Stage 2. Specifically, in Stage 1, teachers in participating schools used the teacher versions of the SDQ and BASC-2 BESS to assess all children in the classrooms in which they taught for emotional and behavioral difficulties. In developing the research design for this study, it was determined that the two questionnaires would be used in conjunction in this stage based on the hypothesis that the two instruments have differential levels of sensitivity in detecting internalizing and externalizing behavioral problems (Wolraich et al., 2014; Danielson et al., 2015). Based on results from the questionnaires, students were classified as “low” or “high” screeners. In this study, “low” screeners were defined as students receiving a BASC-2 BESS score of less than 61 or a SDQ score of less than 12 and “high” screeners were defined as students receiving a BASC-2 BESS score of greater than or equal to 61 or a SDQ score of greater than or equal to 12. In addition to information obtained from the SDQ and BASC-2 BESS, all individuals provided basic demographic information in Stage 1.

Demographic data included information about students' current grade in school, gender, and race/ethnicity. Additionally, teachers were asked to estimate each student's level of school performance in comparison to other students in the classroom. In order to rate students for this variable, teachers were asked to classify students as performing in the bottom 25% of students; below the top 50% of students but above the bottom 25% of students; below the top 25% of students but above the bottom 50% of students; or in the top 25% of students.

Representative sampling procedures were used to select all children classified as high screeners and a sample of individuals classified as low screeners to participate in the second gate, Stage 2. School staff contacted the families of eligible students to invite their participation in Stage 2 of the study. Consenting families were asked to complete the DISC and the CDC-developed Mental Health Diagnosis, Treatment, and Satisfaction Questionnaire (MHDTSQ). Parents completed DISC modules to elicit DSM-IV criteria for 18 mental health diagnoses: social phobia, separation anxiety, panic disorder, agoraphobia, generalized anxiety disorder, obsessive-compulsive disorder, post-traumatic stress disorder, eating disorders, trichotillomania, major depression/dysthymia, mania/hypomania, attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder, alcohol abuse/dependence, nicotine dependence, marijuana abuse/dependence, other substance abuse/dependence. The MHDTSQ is a short questionnaire that elicits information about past diagnoses and treatment received for mental health care. Parents are asked, among other things, to indicate whether their children have been diagnosed with any of 18 specific mental health diagnoses that align with the 18 mental health diagnoses that the DISC modules assess children for.

3.2 Statistical Analysis

SAS version 9.2 was used for all data management and data analyses. Baseline statistics were computed for all participants and included demographic information and scores on both the SDQ and BASC-2 BESS. To calculate concurrent validity, measures of agreement were calculated, including kappa statistics, Pearson correlation coefficients, and 95% confidence intervals. To assess the predictive power of each screener, results from the SDQ and BASC-2 BESS were compared with diagnostic information from the DISC and parent self-report information provided on the MHDTSQ to compute estimates for sensitivity, specificity, positive predictive

values, and negative predictive values. Results for individuals classified as high risk on both the SDQ and BASC-2 BESS or low risk on both the SDQ and BASC-2 BESS were also compared to DISC and MHDTSQ results in order to examine the ability of the screeners to predict diagnoses when combined in comparison to each screener individually. To assess the ability of each screener to predict externalizing and internalizing disorders specifically, subsets of DISC diagnoses were created and estimates of sensitivity, specificity, positive predictive values, and negative predictive values were calculated for externalizing and internalizing disorders specifically. Internalizing disorders included generalized anxiety, separation anxiety, and depression, while externalizing disorders included attention deficit/hyperactivity disorder, oppositional defiant disorder, and conduct disorder. Finally, logistic regression was used to model the ability of the BASC and SDQ to accurately predict whether students were classified as meeting no DSM-IV criteria for mental disorders on the DISC versus meeting criteria for at least one mental disorder on the DISC. Modeling was used to examine the Receiver Operating Curve characteristics of each instrument in order to compare the sensitivities of each instrument.

CHAPTER 4: RESULTS

The most recent wave of Stage 1 of the study included 3622 student participants in grades 6-12 overall, with completed data for 3602 students on the SDQ and 3621 students on the BASC-2 BESS. The most recent wave of the study included 94 students in grades 6-12 with completed data for both the SDQ and BASC-2 BESS in Stage 1 and for the DISC in Stage 2. Table 1 lists demographic information for students completing Stage 1 of the study. Information about gender, race/ethnicity, grade in school, and level of school performance are listed. Table 2 lists mean SDQ Total Difficulties Scores and BASC-2 BESS T-Scores for students overall and by categories of gender, race/ethnicity, grade in school, and level of school performance.

Table 1. Baseline demographic information for Stage 1 students

	Total Population of Children	
	n	**%**
Total	3622	**-**
Gender		
Boys	1865	51.5
Girls	1757	48.5
Race/Ethnicity		
American Indian/Alaskan Native	2	0.1
Native Hawaiian/Pacific Islander	18	0.5
Black or African American	984	27.2
Hispanic	164	4.5
White	2304	63.6
Other	60	1.7
Grade		
6	529	14.6
7	624	17.2
8	650	18.0
9	551	15.2
10	475	13.1
11	445	12.3
12	348	9.6
School Performance		
In the bottom 25% of students	551	15.2
Below the top 50% of students but above the bottom 25% of students	582	16.1
Below the top 25% of students but above the bottom 50% of students	1004	27.7
In the top 25% of students	1375	38.0
Don’t know	90	2.5

Table 2. Mean SDQ Total Difficulties Scores and BASC-2 BESS T-Scores for Stage 1 students

	SDQ				BASC-2 BESS			
	n	**Mean**	**Std Deviation**	**Median**	**n**	**Mean**	**Std Deviation**	**Median**
Total	3602	5.4	5.6	4	3621	48.0	10.3	46
Gender								
Boys	1851	6.5	6.0	5	1864	48.5	10.2	47
Girls	1751	4.2	4.9	3	1757	47.3	10.3	45
Race/Ethnicity								
American Indian/Alaskan Native	2	6.5	0.7	6.5	2	48	2.8	48
Asian/Pacific Islander	18	1.3	1.8	1	18	39	4.5	38
Black/Not Hispanic	975	6.5	5.8	5	984	50.7	10.4	50
Hispanic	164	4.7	5.0	3	164	46.7	9.2	45
White/Not Hispanic	2295	5.0	5.5	3	2303	47.0	10.1	45
Other	58	3.1	3.4	2	60	43.7	7.6	41.5
Teacher-reported Level of School Performance								
In the bottom 25% of students	543	10.4	6.5	10	551	59.4	9.9	59
Below the top 50% of students but above the bottom 25% of students	580	7.5	5.5	6	582	53.1	8.3	52
Below the top 25% of students but above the bottom 50% of students	1000	4.9	4.8	4	1003	47.3	8.1	46
In the top 25% of students	1369	2.6	3.4	1	1375	41.1	6.3	40
Don't know	110	9.2	6.4	8	110	53.8	10.2	54

Table 3 displays the number distribution of low versus high screeners on the SDQ and BASC-2 BESS by gender and related measures of agreement between the two screeners by gender. There was moderate agreement between the SDQ and BASC-2 BESS on whether individuals were classified as low or high screeners, $\kappa = 0.62$ (95% CI = 0.57, 0.67) for boys and $\kappa = 0.67$ (95% CI = 0.61, 0.73) for girls. There was a positive correlation between BASC T-Scores and SDQ Total Difficulties scores for boys, $r=0.86$, $p<.0001$, and for girls, $r=0.86$, $p<.0001$.

Table 3. Stage 1 agreement of SDQ and BASC-2 BESS by gender

	Overall	**Boys**	**Girls**
n	3,601	1,865	1,757
Kappa (95% CI)	0.64 (0.60, 0.68)	0.62 (0.57, 0.67)	0.67 (0.61, 0.73)
Pearson correlation coefficient (BASC T score and SDQ Total Difficulties)	0.85	0.86	0.86
Low screen on both	2974 (82.6%)	1,461 (79.0%)	1,513 (86.4%)
High screen on both	329 (9.1%)	199 (10.8%)	130 (7.4%)
High screen on SDQ, low screen on BASC-2 BESS	183 (5.1%)	148 (8.0%)	35 (2.0%)
High screen on BASC-2 BESS, low screen on SDQ	115 (3.2%)	42 (2.3%)	73 (4.2%)

frequency missing = 21
**where "high" is defined as a SDQ score of 12 or greater or a BASC score of 61 or greater*

Table 4 displays the number distribution of low versus high screeners on the SDQ and BASC-2 BESS by race/ethnicity and related measures of agreement between the two screeners by race/ethnicity. There was moderate agreement between the SDQ and BASC-2 BESS on whether individuals were classified as low or high screeners for students identified as Black, $\kappa = 0.66$ (95% CI = 0.60, 0.72); students identified as Hispanic, $\kappa = 0.67$ (95% CI = 0.45, 0.88); and students identified as White, $\kappa = 0.63$ (95% CI = 0.58, 0.68). There was a positive correlation between BASC T-Scores and SDQ Total Difficulties scores for students identified as Black, $r=0.85$, $p<.0001$; students identified as Hispanic, $r=0.83$, $p<.0001$; and students identified as White, $r=0.84$, $p<.0001$.

Table 4. Stage 1 agreement of SDQ and BASC-2 BESS by race/ethnicity

		Race/Ethnicity						
	Overall	**American Indian or Alaskan Native**	**Asian or Pacific Islander**	**Black, Not of Hispanic Origin**	**Hispanic**	**White, Not Hispanic**	**Other**	**Don't Know**
n	3,601	2	18	984	164	2304	60	90
Kappa (95% CI)	0.64 (0.60, 0.68)	N/A	N/A	0.66 (0.60, 0.72)	0.67 (0.45, 0.88)	0.63 (0.58, 0.68)	N/A	N/A
r (BASC T score and SDQ Total Difficulties)	0.85	N/A	N/A	0.85	0.83	0.84	N/A	N/A
Low screen on both	2974 (82.59%)	2 (100%)	18 (100%)	746 (76.5%)	147 (89.6%)	1937 (84.4%)	55 (94.8%)	69 (76.7%)
High* screen on both	329 (9.14%)	0	0	129 (13.2%)	9 (5.5%)	180 (7.9%)	1 (1.7%)	10 (11.1%)
High* screen on SDQ, low screen on BASC-2 BESS	183 (5.08%)	0	0	54 (5.5%)	6 (3.7%)	113 (4.9%)	0	10 (11.1%)
High* screen on BASC-2 BESS, low screen on SDQ	115 (3.19%)	0	0	46 (4.7%)	2 (1.2%)	64 (2.8%)	2 (3.5%)	1 (1.1%)

frequency missing = 21
**where "high" is defined as a SDQ score of 12 or greater or a BASC score of 61 or greater*

Table 5 displays the number distribution of low versus high screeners on the SDQ and BASC-2 BESS and related measures of agreement between the two screeners by grade in school. Agreement between the SDQ and BASC-2 BESS on whether individuals were classified as low or high screeners ranged from $\kappa = 0.54$ (95% CI = 0.42, 0.66) for children in grade 6 to $\kappa = 0.70$ (95% CI = 0.60, 0.81) for children in grade 11. There were varying levels of positive correlation between BASC T-Scores and SDQ Total Difficulties scores for students in all grades, with correlation ranging from $r = 0.75$ ($p<.0001$) for children in grade 12 to $r = 0.89$ ($p<.0001$) for children in grade 8. Figure 2 displays variations in values of kappa statistics by grade level in school.

Table 5. Stage 1 agreement of SDQ and BASC-2 BESS by grade in school

		Grade in School						
	Overall	**6**	**7**	**8**	**9**	**10**	**11**	**12**
n	3,601	529	624	650	551	475	445	348
Kappa (95% CI)	0.64 (0.60, 0.68)	0.54 (0.42, 0.66)	0.60 (0.52, 0.69)	0.73 (0.65, 0.80)	0.58 (0.47, 0.69)	0.66 (0.57, 0.75)	0.70 (0.60, 0.81)	0.60 (0.45, 0.75)
r (BASC T score and SDQ Total Difficulties)	0.85	0.86	0.87	0.89	0.85	0.84	0.80	0.75
Low screen on both	2974 (82.6%)	450 (85.1%)	494 (79.7%)	518 (80.2%)	458 (84.4%)	369 (78.2%)	381 (85.8%)	304 (87.6%)
High* screen on both	329 (9.1%)	33 (6.2%)	62 (10%)	80 (12.4%)	39 (7.2%)	58 (12.3%)	37 (8.3%)	20 (5.8%)
High* screen on SDQ, low screen on BASC-2 BESS	183 (5.1%)	27 (5.1%)	54 (8.7%)	34 (5.3%)	25 (4.6%)	17 (3.6%)	16 (3.6%)	10 (2.9%)
High* screen on BASC-2 BESS, low screen on SDQ	115 (3.2%)	19 (3.6%)	10 (1.6%)	14 (2.2%)	21 (3.9%)	28 (5.9%)	10 (2.3%)	13 (3.8%)

frequency missing = 21

**where "high" is defined as a SDQ score of 12 or greater or a BASC score of 61 or greater*

Figure 2. Stage 1 agreement of SDQ and BASC-2 BESS by grade in school

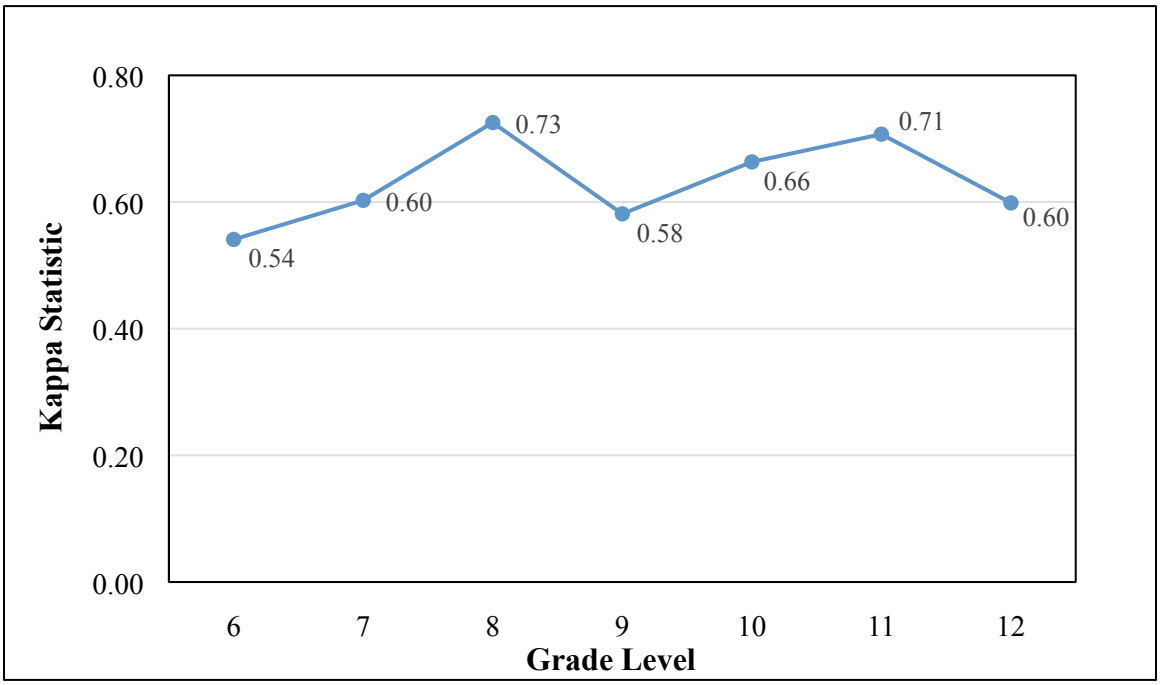


Table 6 displays the number distribution of low versus high screeners on the SDQ and BASC-2 BESS and related measures of agreement between the two screeners by teacher-reported level of school performance. Agreement between the SDQ and BASC-2 BESS on whether individuals were classified as low or high screeners was lowest for students classified as performing in the top 25% of students [$\kappa = 0.54$ (95% CI = 0.42, 0.66)], while it was highest for students classified as performing in the bottom 25% of students [$\kappa = 0.61$ (95% CI = 0.54, 0.67)] and for students classified as performing Below the top 25% of students but above the bottom 50% of students [$\kappa = 0.61$ (95% CI = 0.52, 0.70)]. Values of positive correlation between BASC T-Scores and SDQ Total Difficulties scores ranged from $r = 0.75$ ($p<.0001$) for children performing in the top 25% of students to $r = 0.84$ ($p<.0001$) for students performing in the bottom 25% of students. Figure 3 displays variations in values of kappa statistics by teacher-reported level of school performance, and Figure 4 displays the percent of students classified as "high" versus "low" screeners by teacher-reported level of school performance.

Table 6. Stage 1 agreement of SDQ and BASC-2 BESS by teacher-reported level of school performance

		Teacher-reported level of school performance			
	Overall	**In the bottom 25% of students**	**Below the top 50% of students but above the bottom 25% of students**	**Below the top 25% of students but above the bottom 50% of students**	**In the top 25% of students**
n	3,601	551	582	1004	1375
Kappa (95% CI)	0.64 (0.60, 0.68)	0.61 (0.54, 0.67)	0.54 (0.45, 0.63)	0.61 (0.52, 0.70)	0.46 (0.30, 0.61)
r (BASC T score and SDQ Total Difficulties)	0.85	0.84	0.79	0.79	0.75
Low screen on both	2974 (82.6%)	272 (50.1%)	425 (73.3%)	885 (88.6%)	1323 (96.6%)
High screen on both	329 (9.1%)	168 (30.9%)	71 (12.2%)	54 (5.4%)	14 (1.0%)
High* screen on SDQ, low screen on BASC-2 BESS	183 (5.1%)	40 (7.4%)	54 (9.3%)	47 (4.7%)	30 (2.2%)
High* screen on BASC-2 BESS, low screen on SDQ	115 (3.2%)	63 (11.6%)	30 (5.2%)	13 (1.3%)	2 (0.2%)

frequency missing = 21

**where "high" is defined as a SDQ score of 12 or greater or a BASC-2 BESS score of 61 or greater*

Figure 3. Stage 1 agreement of SDQ and BASC-2 BESS by teacher-reported level of school performance

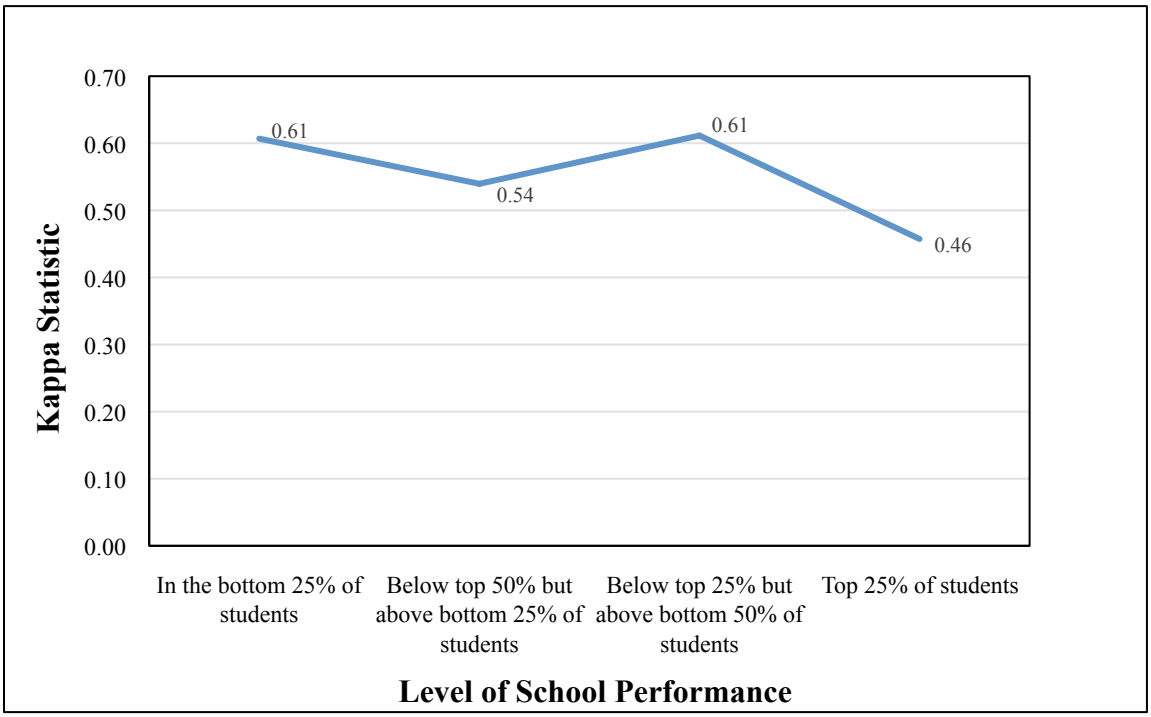


Figure 4. Stage 1 agreement of SDQ and BASC-2 BESS by teacher-reported level of school performance

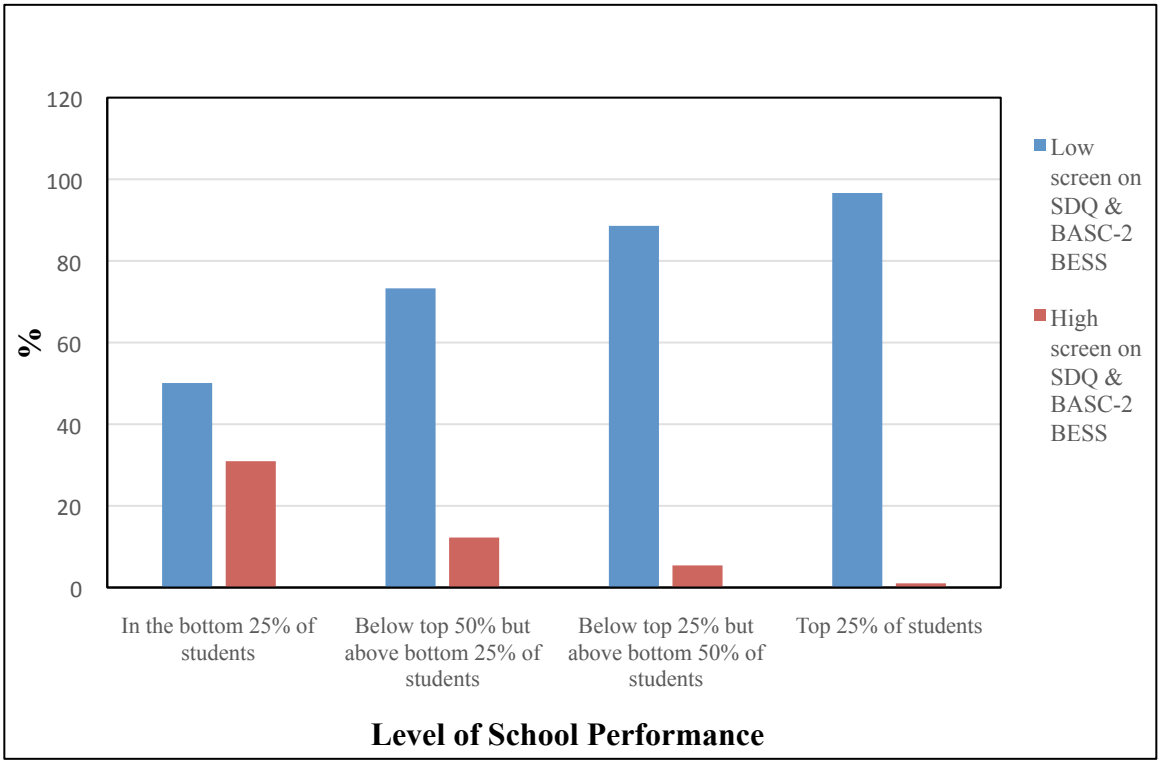


Table 7 displays the number of children meeting past-year diagnostic criteria via parent-report on the DISC during Stage 2 of the study. As shown, 70.2% of children did not meet diagnostic criteria for any mental disorder according to the DISC, while 29.8% of children met diagnostic criteria for at least one mental disorder.

Table 7. Stage 2 children meeting past-year DSM-IV criteria based on the DISC

	Children meeting past-year diagnostic criteria via parent report DISC	
Children by # DISC diagnoses	**n**	**%**
None	66	70.2
1≥ DISC diagnosis	28	29.8
1	11	11.7
2	7	7.5
3	5	5.3
4	3	3.2
5	1	1.1
6	1	1.1

Table 8 displays values of sensitivity, specificity, positive predictive value, and negative predictive value to compare the predictive validity of the SDQ and BASC-2 BESS in the prediction of DSM-IV criteria based on the DISC. As shown, sensitivity was highest for the SDQ (78.57%); specificity was highest for the BASC-2 BESS (65.08%); positive predictive value was highest for the SDQ (56.92%); and negative predictive value was highest for the BASC_SDQ combined (90.24%).

Table 8. Predictive validity in the prediction of DSM-IV criteria based on the DISC

	BASC-2 BESS	**SDQ**	**BASC_SDQ Combined**
n	93	93	94
Low screen, no DISC diagnosis	41	37	37
Low screen, at least one DISC diagnosis	13	6	4
High screen*, no DISC diagnosis	22	28	21
High screen*, at least one DISC diagnosis	15	22	13
Sensitivity	53.57%	78.57%	76.47%
Specificity	65.08%	56.92%	64.79%
PPV	40.54%	44%	38.24%
NPV	75.93%	86.05%	90.24%

**where "high" is defined as a SDQ score of 12 or greater or a BASC score of 61 or greater*

Figure 4 displays receiver operating characteristics curves for the SDQ and BASC-2 BESS in the prediction of mental disorders as ascertained by the DISC. As shown, the area under the curve is greater for the SDQ than the BASC-2 BESS.

Figure 4. Receiver Operating Characteristics Curves for the SDQ and BASC-2 BESS in the prediction of DSM-IV diagnoses as indicated by the DISC

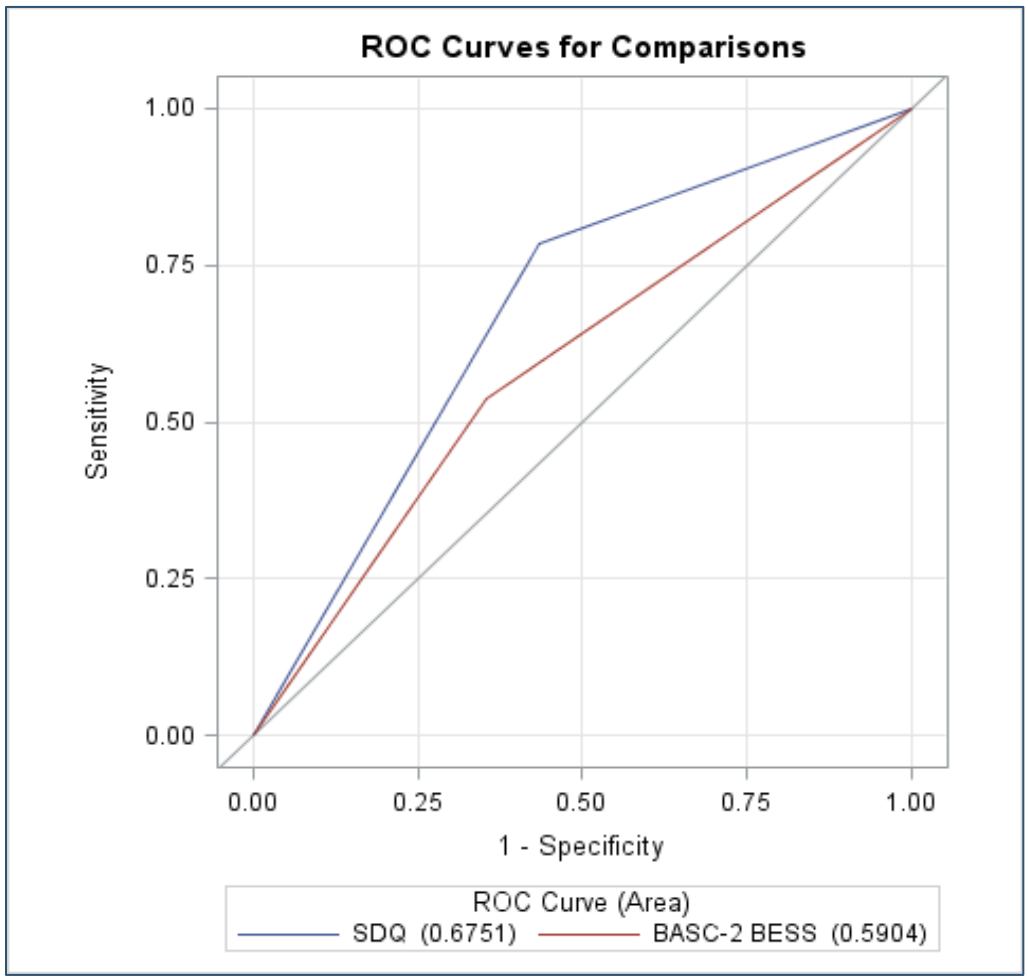


Table 9 displays values of sensitivity, specificity, positive predictive value, and negative predictive value to compare the predictive validity of the SDQ and BASC-2 BESS in the prediction of DSM-IV criteria for internalizing mental disorders based on the DISC. As shown, sensitivity was highest for the SDQ (70%); specificity was highest for the BASC-2 BESS (59.26%); positive predictive value was highest for the SDQ (14%); and negative predictive value was highest for the BASC_SDQ combined (92.86%).

Table 9. Predictive in the prediction of DSM-IV criteria for internalizing disorders based on the DISC

	BASC-2 BESS	SDQ	BASC_SDQ Combined
n	91	93	75
Low screen, no DISC internalizing diagnosis**	48	40	38
Low screen, at least one DISC internalizing diagnosis**	6	3	3
High screen*, no DISC internalizing diagnosis**	33	43	30
High screen, at least one DISC internalizing diagnosis****	4	7	4
Sensitivity	40.0%	70%	57.1%
Specificity	59.3%	48.2%	55.9%
PPV	10.8%	14.0%	11.8%
NPV	88.9%	93.0%	92.9%

**where "high" is defined as a SDQ score of 12 or greater or a BASC score of 61 or greater*
***where internalizing diagnoses include separation anxiety, generalized anxiety, major depression*

Table 10 displays values of sensitivity, specificity, positive predictive value, and negative predictive value to compare the predictive validity of the SDQ and BASC-2 BESS in the prediction of DSM-IV criteria for externalizing mental disorders based on the DISC. As shown, sensitivity was highest for the BASC_SDQ combined (70%); specificity was highest for the BASC-2 BESS (63.77%); positive predictive value was highest for the SDQ (36%); and negative predictive value was highest for the BASC_SDQ combined (95.12%).

Table 10. Predictive validity in the prediction of DSM-IV criteria for externalizing disorders based on the DISC

	BASC	SDQ	BASC_SDQ Combined
n	91	93	94
Low screen*, no DISC externalizing diagnosis**	48	39	39
Low screen*, at least one DISC externalizing diagnosis**	6	4	2
High screen, no DISC externalizing diagnosis****	33	32	24
High screen*, at least one DISC externalizing diagnosis**	4	18	10
Sensitivity	54.5%	81.8%	83.3%
Specificity	63.8%	54.9%	61.9%
PPV	32.4%	36.0%	29.4%
NPV	81.5%	90.7%	95.1%

**where "high" is defined as a SDQ score of 12 or greater or a BASC score of 61 or greater*
***where externalizing diagnoses include attention-deficit hyperactivity disorder (ADHD), opposition defiant disorder (ODD), and conduct disorder*

Table 11 indicates the number of children by the number of past mental health diagnoses as indicated by the MHDTSQ. As shown, 49 children in grades 6-12 had at least one parent-reported mental health diagnosis.

Table 11. Stage 2 children with history of mental diagnoses as indicated by the MHDTSQ

	Children with history of mental diagnoses as indicated by the MHDTSQ	
Children by # of parent-reported diagnoses	**n**	**%**
None	66	57.4
1≥ parent-reported diagnosis	49	29.6
1	34	7.0
2	8	3.5
3	4	0.9
4	1	0.9
5	1	0.9
6	1	0.9

Table 12 displays values of sensitivity, specificity, positive predictive value, and negative predictive value for the SDQ and BASC-2 BESS as compared to MHDTSQ parent-reported diagnoses. As the table indicates, the SDQ had higher values for sensitivity, positive predictive value, and negative predictive value, while the BASC-2 BESS had higher specificity.

Table 12. Predictive validity of SDQ & BASC-2 BESS in the prediction of parent-reported MHDTSQ diagnostic history

	BASC-2 BESS	SDQ
n	111	114
Low screen*, no parent-reported past diagnosis* **	43	37
Low screen*, at least one parent-reported past diagnosis* **	27	19
High screen, no parent-reported past diagnosis*** **	21	29
High screen, at least one parent-reported past diagnosis*** **	20	29
Sensitivity	42.55%	60.42%
Specificity	67.19%	56.06%
PPV	48.78%	50%
NPV	61.43%	66.07%

CHAPTER 5: DISCUSSION

5.1 Implications

The SDQ and BASC-2 BESS demonstrated sufficient concurrent validity as demonstrated by the overall kappa statistic of 0.64 (95% CI: 0.60, 0.68). Agreement by gender and race/ethnicity was fairly consistent, with kappa statistics ranging from 0.62-0.68 and correlation coefficients ranging from 0.83-0.86 across all gender/race classifications. Agreement varied according to grade levels in school and teacher-reported levels of school performance. By grade, the lowest kappa values were for grades 6 [$\kappa = 0.54$ (0.42, 0.66)], 7 [$\kappa = 0.60$ (0.52, 0.69)], 9 [$\kappa = 0.58$ (0.47, 0.69)], and 12 [$\kappa = 0.60$ (0.45, 0.75)]. In particular, agreement was lowest for grades in which students are typically adjusting or are in a transition period of their schooling. Therefore, it is possibile that teachers are less able to accurately ascertain aspects of emotional and behavioral health for students in these grades, perhaps because they do not know the students as well or because students tend to display different behaviors during these transition periods in comparison to other points in their schooling. By teacher-

reported level of school performance, agreement was lowest for students classified as performing in the top 25% of students [$\kappa = 0.46$ (0.30, 0.61)]. This research finding fits interestingly with past research conducted by Eklund, Tanner, Stoll, and Anway (2014) that compared gifted versus non-gifted students and determined that gifted students displayed greater internalizing behaviors than non-gifted students. Since the current study demonstrated a low ability to correctly ascertain internalizing behaviors in students, it is possible that this could be contributing to lower levels of agreement among top-performing students.

The SDQ demonstrated greater sensitivity, positive predictive power, and negative predictive power in comparison to the BASC-2 BESS overall and for externalizing and internalizing disorders. This points to the instrument's ability to accurately classify students according to risk level and its ability to detect individuals who are truly at risk according to gold standard criteria. The BASC-2 BESS, on the other hand, demonstrated higher specificity overall and for externalizing and internalizing disorders. This points to the instrument's ability to detect children who are truly not at risk for mental disorders. However, the BASC-2 BESS had lower values of specificity in this study in comparison to levels of specificity found in the literature. Both instruments demonstrated very low positive predictive power for internalizing disorders, although this power was slightly greater for the SDQ. Combining the BASC and SDQ did not provide major improvements in psychometric performance other than leading to slight increases in negative predictive power. Psychometric properties were generally lower when comparisons were made between the instruments and the MHTDSQ versus the DISC. This could be anticipated, as parents might not always be aware of their children's mental disorders, which would be brought to light by the DISC.

5.2 Strengths and Limitations

This research was strong in that it compared two widely used screening instruments within the same population to assess concurrent validity. Research that elicits concurrent validity within the same population is limited; therefore, this study contributes important information to this area of mental health research. Additionally, this study utilized the widely accepted DISC as the gold standard for comparison in assessing predictive validity. This is important, as the DISC is often used in epidemiological and school-based mental

health screening programs that attempt to ascertain actual DSM-IV diagnoses. Limitations of this study included its small sample size and its exclusion of lower grade levels in school. Additionally, the study was limited to one geographic area, which might differ from other areas of the United States. Further, other versions of these instruments, namely the self-report and parent versions, were not assessed in the current analysis.

5.3 Recommendations

Future studies could benefit from conducting similar analyses comparing the SDQ and BASC-2 BESS using larger sample sizes. With larger sample sizes, the screening capabilities of each instrument could be modeled, taking into account potential covariates, in order to gain more in-depth information about the predictive validity of each instrument and the factors that impact validity. Additional analyses could be conducted using all grade levels in school and using multiple study sites in the PLAY-MH study in order to compare the performance of the instruments across study sites and a wider range of grade levels. Also, similar analyses could be conducted at the same site to examine trends in screening patterns over time. Finally, future research could further examine the relationship between grade level in school and agreement between instruments.

References

Achenbach, T. M., Becker, A., Dopfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., et al. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. Journal of Child Psychology and Psychiatry, 49(3), 251–275.

Becker, A., Hagenberg, N., Roessner, V., Woerner, W., & Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? European Child and Adolescent Psychiatry, 13(Suppl. 2), 17–24.

Becker A, Woerner W, Hasselhorn M, Banaschewski T, Rothenberger A (2004) Validation of the parent and teacher SDQ in a clinical sample. *Eur Child Adolesc Psychiatry*, **13 Suppl 2, II11-6**.

Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. J Am Acad Child Adolesc Psychiatry 44(6):557–64. 2005.

Boyle, C. A., Boulet, S., Schieve, L. A., Cohen, R. A., Blumberg, S. J., Yeargin-Allsopp, M., ... & Kogan, M. D. (2011). Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics*, peds-2010.

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Cambridge, MA: Harvard University Press.

Centers for Disease Control and Prevention. (2015, October 9). Attention-Deficit/Hyperactivity

Disorder. Retrieved from National Center on Birth Defects and Developmental Disabilities website: http://www.cdc.gov/ncbddd/adhd/research.html

Conroy, M. A., & Brown, W. H. (2004). Early identification, prevention, and early intervention with young children at-risk for emotional/behavioral disorders: Issues, trends, and a call for action. Behavioral Disorders, 29(3), 224–236.

Costello, E. J., Edelbrock, C. S., & Costello, A. J. (1985). Validity of the NIMH Diagnostic Interview Schedule for Children: A comparison between psychiatric and pediatric referrals. Journal of Abnormal Child Psychology, 13(4), 579-595.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological bulletin, 52(4), 281.
http://psychclassics.yorku.ca/Cronbach/construct.htm?viewType=Print&viewClass=Print

Dalsgaard, S., Østergaard, S. D., Leckman, J. F., Mortensen, P. B., & Pedersen, M. G. (2015). Mortality in children, adolescents, and adults with attention deficit hyperactivity disorder: a nationwide cohort study. The Lancet.

Danielson, M.L., Visser, S.N., Bitsko, R.H., Holbrook, J.R. (2015). Community-based surveillance of children's mental health: A methodological introduction to CDC's Project to Learn About Youth- Mental Health (PLAY-MH).

Deighton J, Wolpert M (editors). Mental health outcome measures for children and young people. CAMHS EBPU, University College London (UCL) and the Anna Freud Centre. 2009. Available from: https://www.ucl.ac.uk/ebpu/docs/publication_files/mental_health_outcome_measures_for_children_yp

Doll B, Cummings J. Best practices in population-based school mental health services. In: Thomas A, Grimes J, editors. Best practices in school psychology V. Bethesda, MD: National Association of School Psychologists; 2008. pp. 1333–1347.

Doll B, Cummings J. Why population-based services are essential for school mental health, and how to make them happen in your school. In: Doll B, Cummings J, editors. Transforming school mental health services: Population-based approaches to promoting the competency and wellness of children. Thousand Oaks, CA: Corwin Press; 2008.

Dowdy, E., Chin, J. K., Twyford, J. M., & Dever, B. V. (2011). A factor analytic investigation of the BASC-2 Behavioral and Emotional Screening System Parent Form: Psychometric properties, practical implications, and future directions. Journal of School Psychology, 49(3), 265-280.

Dowdy, E., Ritchey, K., & Kamphaus, R. W. (2010). School-Based Screening: A Population-Based Approach to Inform and Monitor Children's Mental Health Needs. *School Mental Health*, *2*(4), 166–176. http://doi.org/10.1007/s12310-010-9036-3.

Eisenberg, D., & Neighbors, K. (2007). Economics of preventing mental disorders and substance abuse among young people. *Paper commissioned by the Committee on Prevention of Mental Disorders and Substance Abuse among Children, Youth, and Young Adults: Research Advances and Promising Interventions, Board on Children, Youth and Families, National Research Council and the Institute of Medicine, Washington, DC*.

Elkin, E. (2012). Are you in need of validation? Psychometric evaluation of questionnaires using SAS. In ICON Late Phase & Outcomes Research (Author), SAS Global Forum 2012. Retrieved from http://support.sas.com/resources/papers/proceedings12/426-2012.pdf

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*(1), 89-95.

Eklund, K., Tanner, N., Stoll, K., & Anway, L. (2015). Identifying emotional and behavioral risk among gifted and nongifted children: A multi-gate, multi-informant approach. School Psychology Quarterly, 30(2), 197.

Fazel, S., Doll, H., & Långström, N. (2008). Mental disorders among adolescents in juvenile detention and correctional facilities: a systematic review and metaregression analysis of 25 surveys. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(9), 1010-1019. http://www.sciencedirect.com/science/article/pii/S0890856708600775

Fergusson DM, Horwood LJ, Ridder EM. Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. J Child Psychol Psychiatry 46(8):837–49. 2005.

Furlong, M. J., O'Brennan, L., & Johnson, K. (2009). Test review of the BASC-2 Behavior and Emotional Screening System. In K. F. Geisinger, R. A. Spies, & J. F. Carlson (Eds.), The eighteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Gilliam, W. S. (2005). Prekindergarteners Left Behind: Expulsion Rates in State Prekindergarten Systems. http://www.bridges4kids.org/PBS/NationalPreKExpulsionPaper03.02_new.pdf

Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. Journal of clinical psychology, 44(6), 1013-1023.

Glover, T., & Albers, C. (2006). Considerations for evaluating universal screening assessments. Journal of School Psychology, 45(2), 117–135.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. Journal of child psychology and psychiatry, 38(5), 581-586.

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(1), 17–24.

Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. Journal of child psychology and psychiatry, 40(05), 791-799

Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European child & adolescent psychiatry*, *9*(2), 129-134.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. British Journal of Psychiatry, 177, 534–539.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child & Adolescent Psychiatry, 40(11), 1337-1345. http://www.sciencedirect.com/science/article/pii/S0890856709605438

Goodman, R., Gledhill, J., & Ford, T. (2003). Child psychiatric disorder and relative age within school year: cross sectional survey of large population sample. *Bmj*, *327*(7413), 472. http://www.bmj.com/content/327/7413/472

Goodman A, Goodman R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. J Am Acad Child Adolesc Psychiatry 2009;48:400–3.

Goodman A, Lamping DL, Ploubidis GB (2010) When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. Journal of Abnormal Child Psychology, 38, 1179-1191.

Goodman A, Goodman R (2011). Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. J Child Psychol Psychiatry, 52, 100-8.

Goodman A, Heiervang E, Fleitlich-Bilyk B, Alyahri A, Patel V, Mullick MS, Slobodskaya H, Dos Santos DN, Goodman R (2011) Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence. *Soc Psychiatry Psychiatr Epidemiol.*

Greer, F. W., Wilson, B. S., DiStefano, C., & Liu, J. (2012). Considering social validity in the context of emotional and behavioral screening. School Psychology Forum, 6(4), 148–159.

Harrison, J. R., Vannest, K. J., & Reynolds, C. R. (2013). Social acceptability of five screening instruments for social, emotional, and behavioral challenges. *Behavioral Disorders*, 171-189. http://www.researchgate.net/profile/Kimberly_Vannest/publication/259295804_Social_Acceptability_of_Five_Screening_Instruments_for_Social_Emotional_and_Behavioral_Challenges/links/54b1999d0cf2318f0f93edd0.pdf

Henderson, J., & Strain, P. (2009). Screening for social emotional concerns: Considerations in the selection of instruments. Technical Assistance on Social Emotional Intervention. Retrieved from http://www.challengingbehavior.org/do/resources/documents/roadmap_1.pdf

Hill, L. G., Lochman, J. E., Coie, J. D., Greenberg, M. T., The Conduct Problems Prevention Research

Group. (2004). Effectiveness of early screening for externalizing problems: Issues of screening accuracy and utility. Journal of Consulting and Clinical Psychology, 72(5), 809–820.

Hoagwood, K., Jensen, P. S., Petti, T., & Burns, B. J. (1996). Outcomes of mental health care for children and adolescents: I. A comprehensive conceptual model. Journal of the American Academy of Child & Adolescent Psychiatry, 35(8), 1055-1063.

Holbrook, J. R., Cuffe, S. P., Cai, B., Visser, S. N., Forthofer, M. S., Bottai, M., ... & McKeown, R. E. (2014). Persistence of parent-reported ADHD symptoms from childhood through adolescence in a community sample. *Journal of attention disorders*, 1087054714539997. Available from: https://www.researchgate.net/profile/Joseph_Holbrook2/publication/263709751_Persistence_of_Parent-Reported_ADHD_Symptoms_From_Childhood_Through_Adolescence_in_a_Community_Sample/links/5475f2020cf29afed612ea7f.pdf

Howie, L. D., Pastor, P. N., & Lukacs, S. L. (2014). *Use of medication prescribed for emotional or behavioral difficulties among children aged 6-17 years in the United States, 2011-2012*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Available from: http://www.nevadatrends.com/articles/2014-05-05--medication/db148.pdf

Institute of Medicine (US). Committee on Crossing the Quality Chasm: Adaptation to Mental Health and Addictive Disorders. (2006). Improving the quality of health care for mental and substance-use conditions. National Academy Press.

Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., ... & Tennant, J. (2014). A critical review of five commonly used social-emotional and behavioral screeners for elementary or secondary schools. Contemporary School Psychology, 18(4), 241-254.

Jones, L. I., Pastor, P. N., Simon, A. E., & Reuben, C. A. (2014). Use of Selected Nonmedication Mental Health Services by Adolescent Boys and Girls With Serious Emotional or Behavioral Difficulties: United States, 2010–2012. NCHS data brief, (163), 1-8. Available from: http://198.246.102.49/nchs/data/databriefs/db163.pdf

Kamphaus, R. W., Thorpe, J., Winsor, A. P., Kroncke, A., Dowdy, E. T., & VanDeventer, M. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. Educational and Psychological Measurement, 67(2), 342–356.

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*(3), 395. http://search.proquest.com/docview/758657279?pq-origsite=gscholar

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry*. Available from: http://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.159.9.1548

King, K., Reschly, A. L., & Appleton, J. J. (2012). An examination of the validity of the Behavioral and Emotional Screening System in a rural elementary school validity of the BESS. Journal of Psychoeducational Assessment, 30(6), 527-538.

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. Journal of Emotional and Behavioral Disorders, 71(2), 93–105.

Lane, K. L., Menzies, H. M., Ennis, R. P., & Bezdek, J. (2013). School-wide systems to promote positive behaviors and facilitate instruction. Journal of Curriculum and Instruction, 7(1), 6-31.

Lane,K. L., Menzies, H. M, Oakes, W. P., & Kalberg, J. R. (2012). Systematic screenings of behavior to support instruction: From preschool to high school. New York: Guilford Press

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. Journal of School Psychology, 45(2), 163-191. Available from: http://www.sciencedirect.com/science/article/pii/S002244050600104X

Lewis, T., Schwebel, D. C., Elliott, M. N., Visser, S. N., Toomey, S. L., McLaughlin, K. A., ... & Schuster, M. A. (2015). The association between youth violence exposure and attention-deficit/hyperactivity disorder (ADHD) symptoms in a sample of fifth-graders. *American journal of orthopsychiatry*, *85*(5), 504.

Lippman LH, Moore KA, McIntosh H. Positive indicators of child well-being: a conceptual framework, measures, and methodological issues. Appl Res Qual Life 2011;6:425–49. Available from: http://www.unicef-irc.org/publications/pdf/iwp_2009_21.pdf

Lucas, C. P., Zhang, H., Fisher, P. W., Shaffer, D., Regier, D. A., Narrow, W. E., ... & Lahey, B. B. (2001). The DISC Predictive Scales (DPS): efficiently screening for diagnoses. Journal of the American Academy of Child & Adolescent Psychiatry, 40(4), 443-449.

Mathai J, Anderson P, Bourne A (2004) Comparing psychiatric diagnoses generated by the Strengths and Difficulties Questionnaire with diagnoses made by clinicians. *Australian and New Zealand Journal of Psychiatry*, **38**, 639-643.

Mathai J, Anderson P, Bourne A (2003) Use of the Strengths and Difficulties Questionnaire as an outcome measure in a child and adolescent mental health service. *Australasian Psychiatry*, **11**, 334-337

McKeown, R. E., Holbrook, J. R., Danielson, M. L., Cuffe, S. P., Wolraich, M. L., & Visser, S. N. (2015). The Impact of Case Definition on Attention-Deficit/Hyperactivity Disorder Prevalence Estimates in Community-Based Samples of School-Aged Children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 53-61. Available from: http://www.jaacap.com/article/S0890-8567(14)00767-9/fulltext

McKeown, R.E., James, L., Cuffe, S.P., Teng, Y., and Holbrook, J.R. (2008). Prevalence of comorbid disorders and risky behaviors with ADHD in elementary-aged children. Paper presented at the Children and Adults with Attention-Deficit/Hyperactivity Disorder Conference, Anaheim, California.

Merikangas, K. R., Nakamura, E. F., & Kessler, R. C. (2009). Epidemiology of mental disorders in children and adolescents. Dialogues in Clinical Neuroscience, 11(1), 7–20.

Merikangas, K. R., He, J. P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. Pediatrics, 125(1), 75-81. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938794/

Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication—Adolescent Supplement (NCS–A). J Am Acad Child Adolesc Psychiatry 49(10):980–89. 2010.

Miller, F. G., Cohen, D., Chafouleas, S. M., Riley-Tillman, T. C., Welsh, M. E., & Fabiano, G. A. (2015). A comparison of measures to screen for social, emotional, and behavioral risk. School Psychology Quarterly, 30(2), 184.

Mills C, Stephan SH, Moore E, Weist MD, Daly BP, Edwards M. The President's New Freedom Commission: Capitalizing on opportunities to advance school-based mental health services. Clinical Child and Family Psychology Review. 2006;9(3):149–161. doi: 10.1007/s10567-006-0003-3.

Muris, P., Meesters, C., & van den Berg, F. (2003). The strengths and difficulties questionnaire (SDQ). European child & adolescent psychiatry, 12(1), 1-8.

National Institute of Mental Health. (2006). National Institute of Mental Health Diagnostic Interview Schedule for Children (DISC). Retrieved from http://www.cdc.gov/nchs/data/nhanes/limited_access/ydq.pdf

National Health Interview Survey. (2014). http://childhealthdata.org/browse/survey/results?q=3375&r=1

Nemeroff, R., Levitt, J. M., Faul, L., Wonpat-Borja, A., Bufferd, S., Setterberg, S., & Jensen, P. S. (2008). Establishing ongoing, early identification programs for mental health problems in our schools: a feasibility study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(3), 328-338. Available from: https://www.researchgate.net/profile/Sara_Bufferd/publication/5634463_Establishing_ongoing_early_identification_programs_for_mental_health_problems_in_our_schools_a_feasibility_study/links/0fcfd5105996a5bf34000000.pdf

Nordentoft, M., Wahlbeck, K., Hallgren, J., Westman, J., Osby, U., Alinaghizadeh, H., ... & Laursen, T. M. (2013). Excess mortality, causes of death and life expectancy in 270,770 patients with recent onset of mental disorders in Denmark, Finland and Sweden. *PloS one*, *8*(1), e55176.

O'Connell, M. E., Boat, T., & Warner, K. E. (Eds.). (2009). *Preventing Mental, Emotional, and Behavioral Disorders Among Young People:: Progress and Possibilities*. National Academies Press. http://www.nap.edu/download.php?record_id=12480

Pastor PN, Reuben CA, Duran CR. Identifying emotional and behavioral problems in children aged 4–17 years: United States, 2001–2007. National health statistics reports; no 48. Hyattsville, MD: National Center for Health Statistics. 2012. http://files.eric.ed.gov/fulltext/ED530166.pdf

Pearson Publishing. (n.d.). BASC-2 Behavioral and Emotional Screening System. Retrieved from http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid5PAaBASC2bess

Perou, R., Bitsko, R. H., Blumberg, S. J., Pastor, P., Ghandour, R. M., Gfroerer, J. C., ... & Parks, S. E. (2013). Mental health surveillance among children—United States, 2005–2011. *MMWR Surveill Summ*, *62*(Suppl 2), 1-35. http://www.cdc.gov/mmwr/preview/mmwrhtml/su6202a1.htm?s_cid=su6202a1_w

Renshaw, T. L., Eklund, K., Dowdy, E., Jimerson, S. R., Hart, S. R., Earhart Jr, J., & Jones, C. N. (2009). Examining the relationship between scores on the Behavioral and Emotional Screening System and student academic, behavioral, and engagement outcomes: An investigation of concurrent validity in elementary school. The California School Psychologist, 14(1), 81-88.

Reynolds, C. R., & Kamphaus, R. W. (1992). Behavior assessment system for children (BASC). Circle Pines, MN: AGS Publishing.

Reynolds, C. R., & Kamphaus, R. W. (2004). BASC-2: Behavior Assessment System for Children, Second Edition. Shoreview, MN: AGS Publishing.

Reynolds, C. R., & Livingston, R. A. (2012). Mastering modern psychological testing. Boston, MA: Allyn & Bacon.

Richardson, M. J., Caldarella, P., Young, B. J., Young, E. L., & Young, K. R. (2009). Further validation of the systematic screening for behavior disorders in middle and junior high school. Psychology in the Schools, 46(7), 605–615.

Satcher, D. (2000). Report of the surgeon general's conference on children's mental health: A national action agenda. http://www.ncbi.nlm.nih.gov/books/NBK44233/

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

Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. Journal of the American Academy of Child & Adolescent Psychiatry, 39(1), 28-38.

Short RJ. Commentary: School psychology, context, and population-based practice. School Psychology Review. 2003;32:181–184.

Silva, T. B., Osório, F. L., & Loureiro, S. R. (2015). SDQ: Discriminative validity and diagnostic potential. Name: Frontiers in Psychology, 6, 811.

Soni, A. *Top Five Most Costly Conditions among Children, Ages 0–17, 2012: Estimates for the U.S. Civilian Noninstitutionalized Population*. Statistical Brief #472. April 2015. Agency for Healthcare Research and Quality, Rockville, MD. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st472/stat472.pdf

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: a review. *Clinical child and family psychology review*, *13*(3), 254-274.

Stormont, M. (2002). Externalizing behavior problems in young children: Contributing factors and early intervention. Psychology in the Schools, 39(2), 127–138.

Teplin, L. A., Abram, K. M., McClelland, G. M., Dulcan, M. K., & Mericle, A. A. (2002). Psychiatric disorders in youth in juvenile detention. Archives of general psychiatry, 59(12), 1133-1143. http://archpsyc.jamanetwork.com/article.aspx?articleid=206941&resultclick=1

Thompson, B. (Ed.). (2002). Score reliability: Contemporary thinking on reliability issues. Newbury Park, CA: Sage.

Torio, C. M., Encinosa, W., Berdahl, T., McCormick, M. C., & Simpson, L. A. (2015). Annual report on health care for children and youth in the United States: national estimates of cost, utilization and expenditures for children with mental health conditions. Academic pediatrics, 15(1), 19-35.

US Department of Health and Human Services. (2015). Federal Interagency Forum on Child and Family Statistics. *America's Children: Key National Indicators of Well-Being*. http://www.childstats.gov/pdf/ac2015/ac_15.pdf

Van Roy B, Veenstra M, Clench-Aas J (2008) Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *J Child Psychol Psychiatry*, **49**, 1304-12.

Violence Institute of New Jersey at UMDNJ. (2007). Searchable inventory of instruments assessing violent behavior and related constructs in children and adolescents: Strengths and difficulties questionnaire. Retrieved from http://vinst.umdnj.edu/VAID/TestReport.asp?Code5SEDQS

Visser, S. N., Danielson, M. L., Bitsko, R. H., Holbrook, J. R., Kogan, M. D., Ghandour, R. M., ... & Blumberg, S. J. (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(1), 34-46.

Visser, S. N., Danielson, M. L., Cuffe, S. P., McKeown, R. E., Bitsko, R. H., Claussen, A. H., Bard, D. E., and Wolraich, M. L. (2011). Health risk behaviors over time among youth with ADHD. Poster presented at the American Academy of Child and Adolescent Psychiatry Annual Meeting, Toronto, Canada.

Walker, H. M., Nishioka, V., Zeller, R., Severson, H., & Feil, E. (2000). Causal factors and potential solutions for the persistent under-identification of students having emotional or behavioral disorders in the context of schooling. Assessment for Effective Intervention, 26, 29–40.

Weist, M. D., Rubin, M., Moore, E., Adelsheim, S., & Wrobel, G. (2007). Mental health screening in schools. Journal of School Health, 77(2), 53-58.

Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgalarrondo, P., et al. (2004b). The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. European Child & Adolescent Psychiatry, 13(Suppl 2), II/47–54

Wolraich, M. L., McKeown, R. E., Visser, S. N., Bard, D., Cuffe, S., Neas, B., ... & Beck, L. (2012). The prevalence of ADHD: its diagnosis and treatment in four school districts across two states. *Journal of attention disorders*, 1087054712453169.

Youth in Mind. (n.d.). SDQ: Information for researchers and professionals about the strengths & difficulties questionnaires. Retrieved from http://www.sdqinfo.org/