UNIVERSITY OF CALIFORNIA
RIVERSIDE

An Evaluation of Direct Behavior Rating as a Screening Tool for Internalizing Problems

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Richard Joon Sung Kong

September 2017

Dissertation Committee:
Dr. Austin Johnson, Chairperson
Dr. H. Lee Swanson
Dr. Cathleen Geraghty Jenkinson

The Dissertation of Richard Joon Sung Kong is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

An Evaluation of Direct Behavior Rating as a Screening Tool for Internalizing Problems

by

Richard Joon Sung Kong

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, September 2017
Dr. Austin Johnson, Chairperson

Studies have shown that mental health issues in students often impact their educational performance. Internalizing problems, such as behaviors related to depression and anxiety, are less likely to be referred for school support. However, only 12-13% of schools screen for mental health risks, and current screening tools for mental health may be burdensome to school districts due to costs and lengthy administration times. The present study evaluated the criterion-related validity and screening accuracy of data derived from Direct Behavior Rating Single Item Scales (DBR-SIS) to evaluate its use as an efficient screener for depression and anxiety in schools. The study included 6 middle school teachers who completed DBR-SIS, and 114 seventh and eighth grade students who completed self-report rating scales for depression and anxiety. Results showed adequate coefficients in support of criterion validity of DBR-SIS. Additionally, results included preliminary support for the use of DBR-SIS in screening for depression and anxiety.

Table of Contents

LIST OF TABLES

Page

LIST OF FIGURES

Page

An Evaluation of Direct Behavior Rating as a Screening Tool for Internalizing Problems

Recent paradigm shifts in education have seen schools adopting a multi-tiered framework that focuses on prevention and data-based decision making. As compared to the previous model of "wait-to-fail," where students need to exhibit a profound deficit in order to be assessed for services, a multi-tiered system of support (MTSS) framework, such as Response to Intervention (RtI) and Positive Behavior and Intervention Supports (PBIS), is increasingly being implemented in order to provide preventative services and respond promptly to students in need (Stiffler & Dever, 2015). MTSS frameworks utilize research-based assessments to make decisions for all levels of student need. This includes applying technically-sound screening and progress monitoring assessments to determine interventions for students who are at-risk for academic, behavioral, social, and emotional issues. Research has indicated that MTSS for academic and behavior support can be highly effective in fostering several positive educational outcomes, such as fewer discipline referrals, improved performance on state tests, and better academic engagement time (e.g., Anderson et al., 2013; Gersten et al., 2009; Shapiro & Clemens, 2009; Vincent et al., 2011). Unfortunately, the support to implement MTSS in schools for mental health concerns has been limited, possibly due to limited school resources (Davis, Kruczek, & McIntosh, 2006). For instance, despite an increased awareness of the importance of school services for mental health needs, there continues to be a significant deficit of mental health professionals serving in the school setting (Jenkins, 1998). Additionally, there is limited research that supports school-based assessment for mental

health issues, therefore creating a barrier to implementing MTSS for mental health needs (McIntosh, Ty, & Miller, 2013).

Despite efforts to address mental health issues in youth, a significant gap exists between children with mental health needs and services received. For example, despite approximately a fifth of youth displaying mental issues (Costello et al., 2003), as few as one sixth of these youths may receive any mental health intervention (Weisz, 2004). Several organizations, such as the Council for Children with Behavioral Disorders (CCBD, 2012) have advocated for increased mental health services, which include training mental health practitioners in schools. One of the major recommendations included a systematic approach to screening for mental health risk to detect risks before symptoms escalate (CCBD, 2012; Stiffler & Dever, 2015). Considering the effectiveness of Multi-Tiered Systems of Support (MTSS) for academic and behavioral supports, schools can play a significant role in the early identification and prevention of mental health issues with school-aged children through a MTSS framework. Dowdy, Ritchey, and Kamphaus (2010) emphasized that mental health services in schools need to be prioritized in the MTSS service delivery system. As data-based decision making is a major component of MTSS, screening is a necessity in MTSS to inform prevention practices. Data-based decision making includes utilizing instruments that produce reliable and valid data and procedures to screen for at-risk students. Unfortunately, screening instruments for mental health issues at school are lacking in practice and research (Dowdy et al., 2015). This study proposes to use data derived from Direct Behavior Ratings (DBR) to screen for student risk for mental health issues in schools.

Mental Health in Schools

Evidence suggests that mental health directly influences the educational success of today's students (e.g., Atkins et al., 2003; Levitt, Saka, Romanelli, & Hoagwood, 2007). Prevention and treatment services in schools for mental health issues have been found to be quite effective (Durlak & Wells, 1997; Rones & Hoagwood, 2000). Despite this, many students experience problems related with mental health without receiving support (United States Public Health Service, 2000). According to prior studies, approximately one in five children and adolescents display symptoms that may meet standards for a psychiatric diagnosis (Costello et al., 2003). Unfortunately, Briggs-Gowan et al. (2000) found that only 16.8% of these youths are diagnosed as having mental health problems. From those identified with mental health problems, only about half of the children received some form of treatment in schools (Costello et al., 2003). Therefore, only a small percentage of students who are struggling with a mental health need are receiving the type of care that they require.

Presently, students exhibiting externalizing behavior problems, which are often related to behaviors present in children with AD/HD, conduct disorder, oppositional defiant disorder and bipolar disorder, are more readily identified and treated in schools (McIntosh et al., 2013). Achenbach and Edelbrock (1981) described that externalizing problems are characterized by behaviors such as verbally and physically fighting, deceiving, and being disobedient. Externalizing problems typically include behaviors that are often highly disruptive, and may include behaviors such as aggression that affect the well-being of peers (Whitcomb & Merrell, 2013). On the other hand, students with

internalizing behaviors, often characterized by symptoms related to anxiety and depression, are less likely to be referred to receive support (Whitcomb & Merrell, 2013). Internalizing disorders such as anxiety and depression are less likely to interrupt daily class instruction, thus making detection more difficult (Whitcomb & Merrell, 2013). Due to the nature of internalizing problems, specific related behaviors, such as isolation and withdrawal, are difficult to observe, as evidenced by previous research that resulted in only low to moderate agreement when rating internalizing problems (Cole et al., 1997). Nevertheless, internalizing disorders are a serious problem, especially considering that approximately 1 in 4 school-aged children will experience an anxiety disorder, and approximately 1 in 10 will experience a major depressive/dysthymic disorder (NIMH, 2010).

Comorbidity. Additionally, the high rates of comorbidity of anxiety and depression in youths have been recognized for several decades. In fact, some experts have proposed theories in which symptoms of anxiety and depression for youths should not be separated, and other experts have noted that anxiety and depression in youths occur in succession (e.g., Brady & Kendall, 1992; Cole, Peeke, Martin, Truglio, & Seroczynski, 1998). On the other hand, another theory proposes that there is a link between anxiety and depression in youths, but the constructs can be reasonably discriminated. Clark and Watson's (1991) tripartite model proposed that a common component can be identified in the overlap, or comorbidity, of the symptoms from depression and anxiety. Negative affect (NA) represents the degree to which a person feels negatively, which includes being upset, unpleasantly engaged, distressed, angry,

sad, or worried. Also, being calm and relaxed is a sign of a person with low NA. Clark and Watson also hypothesized positive affect (PA) and physiological hyperarousal (PH) as differentiating factors for depression and anxiety. PA is related to such feelings and behaviors as being active, delighted, pleased, or energetic. The absence of PA would be displayed in a person who is tired and sluggish. PH includes somatic issues such as being physically tense, short of breath, or dizzy. Clark and Watson hypothesized that combinations of these three factors (NA, PA, PH) may determine the separation of depression and anxiety. Although past research has demonstrated mixed support for Clark and Watson's tripartite model, current studies have seen more consistency in differentiating symptoms of depression and anxiety when the focus on symptoms is emphasized through PA and PH (Anderson & Hope, 2007).

Other studies have maintained that despite similarities, the distinction between anxiety and depression should be maintained for youth (Crowley & Emerson, 1996). Crowley and Emerson (1996) explored these constructs using confirmatory factor analysis (CFA). In this study, 273 fourth and fifth graders each completed two self-report rating scales, the Children's Depression Inventory (CDI; Kovacs & Beck, 1977) and Reynolds Child Depression Scale (RCDS; Reynolds, 1989) to measure depression, and two self-report rating scales, the Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1985) and State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973) for anxiety. The one-factor model was investigated as a negative affectivity (NA) factor, whereas the two-factor model investigated separate anxiety and depression factors. Based on the results of the analysis, the study found that the two-

factor model was a better fit than the one-factor model, based on the magnitude of the chi-square statistics, as well as the goodness-of-fit (GFI) and adjusted GFI indices. Therefore, although there were high correlations for self-reported anxiety and depression measures, the results of the CFA suggested that maintaining a distinction between anxiety and depression is supported.

School role. Schools are ideal for detecting mental health issues and delivering effective preventative supports and interventions (O'Connell et al., 2009). In addition, schools may be the most common arena for mental health service delivery (O'Connell et al., 2009; US Department of Health and Human Services, 1999). Yet, only 12-13% of schools screen for mental health risks, which is an improvement from 2% in 2005 (Bruhn et al., 2014). The increase in screening practices may be due to the emphasis on delivering school supports in a MTSS framework, where screening multiple times a year for student risk is highly emphasized. Nevertheless, with only 12-13% of schools screening for mental health risks, there continues to be a high percentage of students who experience mental health issues without receiving proper identification and treatment.

There is supporting evidence that mental health issues in students often impact their educational performance in schools. For instance, nearly 50% of students with mental disorders drop out of school, with less than half of the students who remain graduating high school with a diploma (US Public Health Services, 2000). Mental health issues are also related to the increase of several negative outcomes, such as drug use, crime, unsafe sex practices, and suicide attempts (e.g., Lewinson et al., 1994; Saluja et al., 2004). For instance, Swanson, Linskey, Quintero-Salinas, Pumariega, and Holzer III's

(1992) study surveyed 4157 adolescents on subjects related to depression, drug use, and suicidal ideation. The study found that over 48% of adolescents scored in the subthreshold for depression symptoms on the Center for Epidemiologic Studies' Depression Scale (CES-D). The subthreshold represents a score of at least 16 on the CES-D, which implies a significant presence of depressive symptoms. Moreover, Swanson and colleagues (1992) found that 26 to 43 percent of youths who scored significantly on the CES-D reported suicidal ideation. In another study, Galaif, Chou, Sussman, and Dent (1997) explored the relationship between depression, suicidal ideation, gateway drugs (e.g., cigarettes, alcohol, marijuana), and hard substances (e.g., cocaine, stimulants, inhalants). Results indicated numerous moderate to significant relationships between those who reported significant depression scores and higher rates of suicidal ideation, as well as with higher rates of use of gateway drugs and hard substances. The evidence suggesting the negative impact of mental health issues that go untreated in students is overwhelming. Fortunately, research has indicated that schools can play a significant part in delivering effective practice in the identification and treatment of mental health in students (Durlak & Wells, 1997). Despite the positive school influence on mental health in students, barriers exist that prevent schools from implementing prevention and intervention services.

School Barriers. Possible barriers to the identification and treatment of students who are experiencing or are at risk for mental health problems are the limited role and knowledge of school mental health personnel, and resources within schools. Traditional perceptions would argue that mental health diagnosis and treatment should occur in

clinics and hospitals (Hoagwood, Hibbs, Brent, & Jensen, 1995). This would therefore discourage schools from participating in mental health activities, particularly services related to identification and prevention (Levitt et al., 2007). These limitations impact the ability to conduct preventative practices, such as screening assessments for at-risk youths. Mental health specialists in schools (e.g., social worker, psychologists, and counselors) can only meet the needs of approximately 10% of children with mental health problems (Jenkins, 1998). This is compounded by the ratio of mental health professionals in schools, such as school psychologists, to the number of students served. For instance, the mean ratio of students to school psychologists, as reported in a 2009-2010 survey, is 1383 students to 1 school psychologist (Curtis et al., 2010). Curtis et al. (2010) found that the ratio grew even wider in urban and rural areas. In another study, Sullivan and Long (2010) found that approximately 15% of school psychologists were serving over 2000 students. This ratio is significantly beyond the 550 students to 1 psychologist recommendation from the National Association of School Psychologists (NASP, 2010). Therefore, only the students who are either most disruptive or present with severe mental health and behavioral symptoms are likely to be assessed and treated in schools (Graden, 2004; Jenkins, 1998).

Nevertheless, a survey by Stoiber and Vanderwood (2008) found that the importance of screening assessments is being recognized and valued in schools by mental health specialists such as school psychologists. The use of technically adequate screening assessment data is a critical component of identifying at-risk students and informing decisions on future interventions and progress monitoring assessments (Roberts et al.,

2005). Considering the resource deficit and the lack of availability of mental health professionals in schools, there is a significant need to develop and utilize widely available and efficient mental health screening tools for implementation in school settings. Therefore, this study proposes the evaluation of a mental health screening tool with minimal resource requirements.

Characteristics of Screening Measures

Screening serves the purpose of identifying students who are at-risk of a current or future problem (Glover & Albers, 2007; Johnson et al., 2009). By using screening assessments in identifying the levels of risk for certain characteristics of a student, researchers have been able to predict future problems of students. Screening assessments have also been used to identify students who have current issues. By using screening in schools, educators can have access to critical information for prevention-based strategies, as well as implement interventions for individual or groups of students.

The utilization of screening assessments is being encouraged by educational policy makers and organizations. For instance, the National Association of School Psychologists (NASP, 2010) has recommended using evidence-based practices such as universal screening for academic, behavioral, and emotional issues. Screening for the identification of at-risk youths for various academic needs have been researched thoroughly and utilized throughout the nation (e.g., Gersten et al., 2007; Jenkins, Hudson, & Johnson, 2007). Additionally, screening for behavioral needs is also on the rise with prevention based initiatives such as Positive Behavior Intervention and Supports (PBIS; Sugai et al., 2000). The U.S. Department of Education Office of Special Education and

Rehabilitative Services (2002) has urged schools to screen for emotional and behavioral difficulties so that at-risk children could be identified and more significant issues could be prevented through early classroom based preventative strategies and interventions. Screening practices are important in introducing preventative and intervention services in a timely manner (Fuchs et al., 2007; Jenkins, Hudson, & Johnson, 2007). Despite the emergence of screening for academics and behavior, screening procedures and instruments for mental health issues in schools are not as well developed and underutilized (Levitt et al., 2007). Part of the success of academic screening can be contributed to the ease of use, social validity (Messick, 1989), and the technical adequacy of the measures that are being implemented (Good, Gretchen, & Shinn, 1998). Therefore, it is important to address several measurement criteria of screening tools before evaluating their effectiveness.

Mental Health Screening Theory. From early on, prevention has been emphasized as a necessary step in effectively addressing health issues. Caplan (1964) applied the concept of a multi-tiered approach to prevention within public health, creating definitions that would differentiate between preventative services and treatment. Following Caplan, Cowen (1977) further defined prevention efforts as a way to reduce the issue before the onset of a disorder, and targeted towards a whole population or subgroup with known risk factors.

In 1983, Gordon proposed three types of prevention, which was later modified by the Institute of Medicine (IOM; 1994): (a) universal, (b) selected, and (c) indicated. Screening in universal prevention involves assessing the full population of interest, as

long as the benefits to screening clearly outweigh the negatives or risks. For example, babies in the United States go through several tests as newborns to screen for a variety of health and developmental risks. Medical health screenings are often utilized throughout a person's lifetime, and are delivered in various settings. Similarly, school-based universal screening typically involves testing all students in a set assessment, or series of assessments, for the purpose of identifying students who are at-risk for educationally related deficits. For instance, universal screenings for various academic areas are recommended on a range from yearly up to three times a year (Vellutino et al., 2007).

Whereas universal screenings target the entire population of interest, selected screening involves assessing people who present with elevated risks and characteristics. Selected screening is not offered to the whole population of interest, but instead usually requires a selection process, such as teacher nomination. Selected, as well as indicated, populations may have a greater likelihood of needing interventions since their risks are considered elevated (IOM, 1994). Selected screening in a multi-tiered approach can involve multiple gating, which includes a process by which multiple informants and/or procedures are implemented sequentially to increase accurate identification. For instance, the Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1992) involves three gates. First, teachers rank order their students for externalizing and internalizing problems. Next, a checklist for maladaptive behaviors is completed for each nominated student. Lastly, for the students whose scores on the checklist exceed the cut-off, a structured observation in multiple settings is completed.

Finally, indicated screening involves assessing a person individually who has already presented vulnerability to a disease or disorder. An example of this would be assessing all children in a special education class for emotional disturbance (ED) to see whether they would benefit from a specialized socio-emotional intervention.

In schools, all three types of screening may be present in different assessment procedures. Often, indicated screening may only be utilized within a special education setting, whereas selected screening can be integrated into both general and special education environments. In a MTSS framework, universal screening is considered a critical component of early identification and treatment. Universal screening, which utilizes brief instruments, is conducted on all students and can act as the first process for identifying at-risk students. Thus, it is important in school-based universal screening to use an instrument that accurately identifies at-risk students.

Standards for Screening Measures

The Standards for Psychological and Educational Testing ("Standards"; APA/AERA/NCME, 2014) noted that screening instruments should be appropriately standardized for the use of the population being measured, show consistency in its measurements, and display accuracy when identifying individuals who are at-risk. To address the above standards, there are several necessary components that a screening instrument must possess. Glover and Albers (2007) noted that technical adequacy (i.e., reliability and validity), intended use, and usability must be major factors in evaluating the effectiveness of a screening instrument.

Technical adequacy. Demonstrations of technical adequacy, specifically validity and reliability, are a critical component of evaluating the effectiveness of data derived from a screening measure. When evaluating a screener, the various types of validity (e.g., construct, criterion, concurrent, predictive, social and consequential) should be examined (Glover & Albers, 2007; Jenkins et al., 2007; Messick, 1998). Considering the purpose of screening is to accurately identify or predict risk factors in students, a common focus in screening studies is the concurrent and predictive validity of data derived from the measure. To establish the strength of the screening measure, it is important to examine the concurrent validity, commonly through correlating data between one or more well-established measures of a similar construct with the targeted screening measure (Jenkins et al., 2007).

Upon conducting correlational analyses, data from the screening measure may be evaluated for predictive validity. Receiver operating characteristic (ROC) curve analyses are beneficial in determining sensitivity and specificity according to differing thresholds, or cut scores. Cut scores can be created through various statistical analysis (e.g., logistic regression) or by taking the thresholds that are determined by a gold standard measure. Sensitivity is defined as the proportion of students who were correctly identified as being at-risk through the screening measurement, whereas specificity is the proportion of students who were correctly identified as not at-risk (Swets, 1988). Additionally, specificity and sensitivity indices are valuable in indicating whether the screening measure can accurately identify at-risk students at a later time (Glover & Albers, 2007). For instance, as sensitivity levels of the screener declines, more students who are truly at-

risk are missed. On the other hand, if specificity levels decline, more students who are not at-risk are being over identified as being at-risk. In addition, the positive predictive value (PPV), negative predictive value (NPV), and overall hit rate values could be calculated. The PPV estimates the proportion of correctly identified students from all students who were identified as being at-risk on the tested measure. The NPV estimates the proportion of correctly identified students from all students who were identified as being not at-risk on the tested measure. Low PPV estimates indicate that students are being over-identified, and therefore risk consuming more resources than necessary on students who are not at-risk. Generally, sensitivity and specificity values that fall between .75 and .80 have been considered acceptable in screenings for schools (Glover & Albers, 2007).

Another critical consideration is the area under the curve (AUC) of a ROC curve. As a ROC curve represents a plot of sensitivity and 1 - specificity for different possible cut-points of the measurement, the AUC measures the overall discrimination of the screening measure to classify those who are at-risk and those who are not. A screening measure that is no better than chance has an AUC of 0.5. On the other hand, a screening measure that perfectly classifies without any false positives and false negatives has an AUC of 1.0. The AUC point estimate represents the probability that a randomly chosen subject at-risk has a higher score than a subject who is not at-risk. An approximate guide for interpretation of AUC point estimates is that scores of fair value range from .70 to .80, good value from .80 to .90 and values above .90 are considered excellent value (Safari, Baratloo, Elfil & Negida, 2016).

Intended use and usability. In the school setting, intended use refers to how outcomes from the screening tool match the needs of the school it is being used in, and to what extent the measurement results are compatible with resources of the users (Glover & Albers, 2007). Glover and Albers (2007) emphasized the need for screening measures to have theoretical support in terms of how they are to be used, while also being empirically supported in the alignment of the constructs that the measure claims to be screening for.

Usability is an important factor in evaluating a screening tool, as it encompasses important dimensions of social validity such as acceptability, feasibility, and understanding (e.g., Kane, 2013; Messick, 1998; Wolf, 1978). Although a screening tool may be technically sound and appropriate for the setting, the perceived practicality to administer within a specific context needs to be considered (Glover & Albers, 2007). For example, the cost of the instrument may inform the acceptability and feasibility of the screening tool. Jenkins, Hudson, & Johnson (2007) note that screening should not only consider cost of material (fiscal resources), but also the time and personnel needed. Also, since universal screening is often administered by several school personnel, feasibility of administration, such as ease of instructions and scoring, need to be well thought out (Glovers & Albers, 2007). Finally, of significant importance is the usability of the data acquired from the screening tool. Specifically, if being used in an MTSS setting, the data should be helpful in providing information towards intervention or treatment decisions (Hayes, Nelson, & Jarrett, 1987; Merrell, 2000).

Utilization of Screening in Schools

In light of the emphasis on prevention, many schools have adopted screening at least once a year for several educational domains. Particularly for academics, and especially for reading, adequate screening measures exist based on their ability to identify students who are at-risk. The most common screening assessment tool for academic skills are based on curriculum based measurements (CBM), such as oral reading fluency (ORF). CBMs are quick and accurate measures of academic competency of the curriculum that students are expected to gain mastery over the course of the year (e.g., Deno, 1985; Fuchs et al., 2001; Reschly et al., 2009). For instance, several studies have found that measures of ORF excel at predicting future reading performance, as well as show strong technical adequacy across studies examining reliability and validity (Graney, Martinez, Missal, & Aricak, 2010; Goffreda, Diperna, & Pedersen, 2009).

The positive aspects of academic screening, especially in reading, have been well noted. The Institute of Education Sciences' What Works Clearinghouse evaluated screening as an integral part of delivering interventions in MTSS (Gersten et al., 2009). Gersten and colleagues (2009) synthesized existing literature and created a practice guide for the process of assisting primary grade students struggling with reading. In this document, screening all students for potential reading problems, and regularly monitoring the progress of students identified at risk for reading problems, was found to have sufficient evidence. Therefore, Gersten et al. (2009) found screening to be a critical step in identifying at-risk students to provide evidence based instruction at appropriate skill levels. Nevertheless, one area of concern for screening for reading was the specificity of

data from these measures, or correctly identifying students who were not at-risk (Gersten et al., 2009). The authors noted that although this was a concern, the over-identification could be addressed in further assessment methods (e.g., progress monitoring).

In addition to displaying accuracy with identifying students who are at-risk, there is flexibility in the use of CBM screening assessments. For example, ORF can be used as a progress monitoring tool due to the adequacy of its ability in measuring growth (Shapiro, 2011; Shinn, 2009). Established CBMs, such as ORF, have the benefit of being both accurate screeners and progress monitoring tools, thus fitting well with schools that utilize MTSS. The advantage of establishing sensitivity to growth gives the advantage of measuring the effectiveness of treatment, thus further informing treatment decisions. Therefore, having screening tools for behavior and mental health that emulate the technical abilities of academic CBMs (e.g., ORF) would be beneficial in not only screening for at-risk students, but also measuring the effectiveness of interventions provided.

Behavior. Screening measures for behavior have also been utilized and analyzed in classrooms. Although universal behavior screening is not as common in schools as academic screening (Eklund et al., 2009), a number of rating scales have been developed to identify at-risk students. There are several rating scales that are utilized for this purpose, such as the Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2015). The BESS is an adapted form of the Behavior Assessment System for Children, 3rd Edition (BASC-3), and is utilized for screening students who are at-risk for behavioral and emotional problems. The BESS is a behavior rating scale with 25 to 30

items that takes up to 10 minutes to complete per student. The BESS has teacher and parent report forms for students aged 2 through 18 years and 11 months, whereas the self-report forms are for students aged 8 to 18 years and 11 months. Kamphaus & Reynolds (2007) reported finding strong internal consistency coefficients (.92-.96) and 8-week test-retest reliability (.80-.91). Additionally, interrater reliability correlations were satisfactory, with .70 for teacher forms and .87 for parent forms. When the BESS teacher form was compared to the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001), correlations ranged from .29 for Internalizing, .69 for Externalizing, and .76 for Total Problems. The BESS parent form and the ASEBA parent form produced .64 for Internalizing, .66 for Externalizing, and .76 for Total Problems. Similarly, the BESS self-report and ASEBA youth report had correlations of .69 for Internalizing, .66 for Externalizing, and .77 for Total Problems. Overall, correlation coefficients showed strong relationships between the BESS and ASEBA. Unfortunately, the BESS teacher form and ASEBA resulted in a small correlation for Internalizing, thus indicating weak association.

In a study conducted by King, Reschly, and Appleton (2012), students in kindergarten through fifth grade (n = 496), regular education teachers (n = 25), and parents (n = 241) completed self-report, teacher, and parent forms of the BESS. The self-report forms were only distributed to students in third, fourth, and fifth grade. The study examined the properties of the BESS by analyzing convergent associations with other school behavioral risks and concurrent validity with school performance as the criterion. Convergent relationships between the BESS and academic (oral reading fluency) and

behavior (office disciple referrals, suspensions, attendance) outcomes were established. Low to moderate correlations (-.112 to .338) were found across each BESS form and academic and behavior criteria. Group differences were also examined between normal, elevated, and extremely elevated scores on the BESS. Significant group findings were found for various academic and behavior criteria, and all three BESS forms had significant group differences for office discipline referrals. Additionally, the BESS teacher and self-report forms noted significant group differences for academics as well. Considering the BESS' ability to discriminate between risk groups in both academics and behaviors, this study appears to provide supporting evidence for the BESS' screening utility for identifying students at-risk for school-related problems.

Another behavioral screening assessment used in schools and research is the Student Risk Screening Scale (SRSS; Drummond, 1994). The SRSS is a universal screening tool that uses a brief rating scale for teachers to rate the externalizing and internalizing behaviors of their classroom students. A strength of the SRSS is its usability, since the developers claim that a class of 25 students can be rated within ten to twenty minutes. An adapted SRSS version by Lane and colleagues (2012) expanded the SRSS to include internalizing symptoms, for a total of 14 items. A study by Lane et al. (2007) examined the SRSS with the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) with middle school students from a rural setting (n = 500) and an urban setting (n = 528). Internal consistency via Cronbach's alpha ranged from .78 to .85. Test-retest reliability, from 14 to 34 weeks, ranged from .56 to .80. Internal consistency and test-retest reliability appear to be within acceptable ranges. In regard to convergent

validity, Pearson correlation coefficients between the SRSS and the SDQ were moderate to high for Total (r = .66), Emotional Symptoms (r = .42), Conduct Problems (r = .51), Hyperactivity (r = .55) and Peer Problems (r = .41). With the SDQ as the criterion, the SRSS showed relationships that suggests similar constructs are being measured.

Comparable but slightly overall higher reliability and validity coefficients were found for elementary aged students in a study conducted by Ennis, Lane, and Oakes (2011). Pearson correlation coefficients between the SRSS and the SDQ were moderate to high with scores ranging from .71 to .80 for Total, .76 to .85 for Conduct Problems, .60 to .69 for Hyperactivity, and .41 to .58 for Peer Problems. However, the correlation coefficients for the SRSS and Emotional Symptoms were lower, ranging from .18 to .31. This study also examined the predictive validity of the SRSS with school disciplinary actions. Correlation coefficients suggested a significant relation between SRSS total scores and school suspensions for the same school year (r = .40) and the following school year (r = .32). Internal consistency coefficients exceeded .80 at all time points, and test-retest reliability coefficients ranged from .62 to .86. In accordance with the prior study by Lane et al. (2007), Ennis et al. (2011) found similar relationships between the SRSS and SDQ, and added to the literature supporting the SRSS as a behavior screener by showing significant relationships between the SRSS and school suspensions.

Behavior screeners such as the SRSS and the BESS have shown to be effective in screening for externalizing problems. On the other hand, current behavior screeners struggle to successfully screen for internalizing symptoms, such as depression and anxiety.

Mental health assessments

Mental health assessments can include one or more assessment methodologies such as rating scales, observations, and interviews. The purpose of mental health assessments is to measure different aspects of a person's psychological state of being to classify and to inform treatment (Levitt et al., 2007; Stiffler, 2015). Multiple-gating is a highly-recommended method of assessment as it requires using multiple sources and multiple methods of assessing for issues. As described by Whitcomb and Merrell (2014), the foundation of multiple gating involves a series, or gates, of assessments and decisions, in order to sequentially narrow down a larger population to a smaller population, who are more likely to demonstrate behaviors and symptoms across settings and time. Rating scales can be an important part of a multiple gating method of assessing for mental health. Additionally, rating scales are common assessment tools for informing a clinician's decision on classification or diagnosis (Whitcomb & Merrell, 2014).

Rating scales. Rating scales can be utilized as a broadband measure, which is a general measure of mental health, or a narrow band measure, which focuses on specific mental health areas (e.g. depression). Broadband rating scales typically include externalizing (e.g., conduct, aggression, hyperactivity) and internalizing behaviors (e.g., depression, anxiety). An example of a broadband rating scale is the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 2015, 2004, 1992), currently in its third edition. The BASC-3 includes teacher, parent, and self-report forms with primary scales for both externalizing and internalizing behaviors in one assessment (Reynolds & Kamphaus, 2015). Depending on the age of the child, forms range from 105 to 175 items.

The BASC-3 covers children and adolescents ages 2 to 21 years and 11 months and may take 10 to 30 minutes to complete the forms. The developers report strong Cronbach alpha coefficients for teacher (.86-.90), parent (.84-.89), and self-report (.81-.85).

A narrowband rating scale that has been utilized for over two decades in practice and research is the Children's Depression Inventory (CDI; Kovacs, 2010). Now in its second edition, the CDI is a narrow band measurement tool that looks specifically at depression in children. The CDI has a self-report, teacher, and parent forms. Additionally, a short version of the self-report form is available. The self-report form consists of 28 items, whereas the short version is made up of 12 items. The CDI forms a Total score, along with two scale scores, emotional problems and functional problems. Additionally, four subscales are also provided: negative mood, negative self-esteem, ineffectiveness, and interpersonal problems. The developers note that the administration time of the CDI is 5 to 15 minutes long. For all scores, T-scores are reported with 50 as the mean and standard deviation of 10.

Over the years, evidence supporting the reliability and validity of data derived from the CDI has accumulated (Carey et al., 1987). In 2004, Timbremont, Braet, and Dreessen examined the utility of the CDI for predicting a diagnosis of depressive disorder on the Structured Clinical Interview (SCID) in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). The study consisted of children ages 8 to 18 years of age ($N = 80$, $M = 12.21$, $SD = 2.72$) from inpatient and outpatient settings. The scores of the children from the SCID were obtained two weeks after the administration of the CDI. A ROC curve analysis determined the optimal cut points for this study. With an

optimal cut point of 16 on the CDI, indices for specificity (.84), sensitivity (.94), positive predictive value (.63), and negative predictive value (.98) were satisfactory for a screener. Additionally, a logistic regression analysis revealed that the CDI's overall classification accuracy was 84.8% with the SCID as criterion.

Another narrowband rating scale that is widely used is the Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1985, 2008), currently in the second edition. The RCMAS consists of 49 items and also consists of a short version of 10 items. The child responds to each item by circling Yes or No on the protocol. This self-report measurement tool is normed for children from 6 to 19 years old. According to the developers, the RCMAS requires 10 to 15 minutes to complete. The RCMAS is designed to yield scores for three anxiety related subscales: Physiological Anxiety, Worry, and Social Anxiety. In addition, a Total Anxiety score is reported. Two additional scales, Inconsistent Responding Index and Defensiveness, are calculated to measure aspects of invalid or biased responses. Subscale scales are reported as T-scores with a mean of 50 and standard deviation of 10. The Total Anxiety score is reported as a standard score with a mean of 100 and standard deviation of 15. Construct validity of the RCMAS has been examined by the relationship with other accepted anxiety measures. Correlations between the RCMAS and the State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973) ranged from moderate (.65) to strong (.85) over several studies (Reynolds, 1980; Reynolds, 1982). The RCMAS has also been used in research as the criterion to validate other anxiety measures, presenting the RCMAS as a widely-

recognized standard for anxiety measurement (e.g., Inderbitzen-Nolan & Walters, 2000; Kamps, Roberts, & Varela, 2005; Muris et al., 1998).

Interviews. Interviews are a common form of assessment in schools, clinics, and hospitals to gain information about a client's background and concerns. In fact, Watkins, Campbell, Nieberding, and Hallmark (1995) found that interviewing is the most commonly utilized form of assessment for diagnosing mental health disorders. Interviews can include interviewing the subject, depending on the age and maturity, as well as interviewing other impactful people such as parents, caregivers, and teachers. Interviews are conducted in different forms and methods, ranging from unstructured and informal, to highly structured interview schedules. Not only are interviews used to extract information, but they can also be a procedure in which rapport is built with the client. Additionally, diagnostic interviews, such as the Diagnostic Interview for Children and Adolescents (DICA; Reich, Welner, & Herjanic, 1997) or Diagnostic Interview Schedule (DIS; Robins, Helzer, Croughan, & Ratcliss, 1981) provide questions related to diagnostic criteria from the Diagnostic Statistical Manual of Mental Disorders, currently in its fifth edition (DSM-5).

Observations. Direct observations are a valuable but potentially time-consuming method of assessment. Direct observation refers to a procedure in which target behaviors are operationally defined and the behaviors are systematically observed and recorded (Whitcomb & Merrell, 2013). Observations can provide valuable information through narrative descriptions, as well as with coding behavioral data (e.g., partial interval, time sampling, duration recording). On the other hand, best practices for conducting

observations includes observing across multiple times and settings, which is time consuming (Whitcomb & Merrell, 2013). Additionally, proper training is imperative in producing reliable observational data. When multiple observers are involved, Kazdin (1981) recommends the use of observer reliability checks (e.g., interobserver agreement) to ensure consistency.

Although there are well-established mental health assessments, the utilization of screening measures for mental health issues is lacking. Existing school-based screening measures for mental health often are encompassed in broader behavior rating scales such as the BESS, in which technical adequacy is lacking for the Internalizing subscale. In addition, well-established mental health assessments, as well as interviews and observations, can be time consuming, especially since rating scales such as the RCMAS-2 and CDI-2 can take up to 15 minutes per student. Lengthy administration times may significantly reduce the likelihood of implementation in a school (Simon, Dirksen, & Bögels, 2013). Instead, a screening tool that echoes the efficiency of academic screeners (e.g., CBM) would be beneficial.

Direct Behavior Rating

Alternative methods of screening in schools have also been utilized successfully in research and in practice. The need to develop alternative methods of behavior screening have been noted considering the continued lack of behavior screening in schools, possibly due to the difficulty of implementing time-consuming rating scales such as the BESS and the SSBD. One standardized method of screening that has gained research support is Direct Behavior Rating (DBR), which is a behavior assessment

method that combines concepts of systematic direct observation (SDO) and behavior rating scales (Chafouleas, Riley-Tillman, & Sugai, 2007). By using the feasibility of rating scales and the flexibility and repeatability of systematic direct observation, DBR attempts to combine the strengths of these common forms of assessment. In establishing DBR as a unique screening and progress monitoring tool, Chafouleas, Riley-Tillman, and Christ (2009) suggested these core features: (a) rating occurs immediately after observation, (b) rater has firsthand experience with target being rated, and (c) minimal inference is needed in discerning the behaviors.

Procedures for how the standard DBR single item scales (DBR-SIS) are used in school settings require observation and ratings of each student of interest in specific behaviors (Chafouleas et al., 2009). Most DBR-SIS research has focused on three behaviors: academic engagement (AE), disruptive behavior (DB), and respectful behavior (RB). Raters are encouraged to complete DBR-SIS ratings over multiple consecutive days so as to increase the reliability of aggregated estimates derived from these ratings (Chris, Riley-Tillman, Chafouleas, & Boice, 2010)

DBR as a screening measure. In the first study to examine DBR-SIS as a screener, Chafouleas, Kilgus, and Hernandez (2009) investigated the use of DBR-SIS with kindergarten students for the Fall (n = 20) and Spring (n = 18). Target behaviors on DBR-SIS included academic engagement (AE) and disruptive behavior (DB). Student scores on DBR-SIS were derived from the mean scores across 10 ratings. To establish DBR-SIS as a screening tool, Chafouleas and colleagues compared DBR-SIS target behaviors with a criterion measure, Social Skills Rating System (SSRS; Gresham &

Elliott, 1990). Correlations between DBR-SIS and the SSRS indicated moderate to strong associations (.40-.88) in screening for behavioral issues.

Kilgus, Chafouleas, Riley-Tillman, and Welsh (2012) investigated DBR-SIS as a screener with a larger sample of second grade students (n = 120). The targets included DB, AE, and compliance (CO). The mean of DBR-SIS ratings for each target for each student was utilized for analyses. In addition to examining DBR-SIS relative to the BESS (Kamphaus & Reynolds, 2007) and the Social Skills Improvement System (SSIS; Elliott & Gresham, 2007), ROC curve analysis was performed on DBR-SIS targets with the BESS as the criterion. Results from the study noted that DBR-SIS single targets offered either a cut score associated with high sensitivity and lower than desired specificity, or conversely, lower than desired sensitivity with high specificity. Kilgus and colleagues also included combinations of targets, which they termed "multiple gating" procedures, in the ROC curve analysis. Multiple gating procedures with all three target behaviors resulted in the best sensitivity (.91) and specificity (.83) values. The results from this study showed promising support for DBR-SIS as a possible option as a behavior screening tool in schools.

Whereas prior studies of DBR-SIS as a screener involved single grade levels, Chafouleas, Kilgus, Jaffery, Riley-Tillman, Welsh, and Christ (2013) evaluated DBR as a school-based behavior screener with a large sample across several primary grade levels. The study included a sample of 831 students, grades kindergarten through eight. The teachers observed and rated students on DBR-SIS, BESS (Kamphaus & Reynolds, 2007), and Student Risk Screening Scale (Drummond, 1994). DBR-SIS consisted of the

standard three behavior targets of DB, AE and RB. The mean of DBR-SIS ratings for each target for each student was utilized for analyses. Using cut scores from the BESS, DBR-SIS targets were evaluated on their ability to accurately classify students based on single targets, as well as in combination of targets. The study hypothesized that the results would indicate that DBR would perform well in comparison to other assessments in screening for school-based behavior problems. Results showed DBR-SIS had good accuracy, ranging from .51-90 for sensitivity, .47-.83 for specificity, .94-.98 for negative predictive power, and .14-.41 for positive predictive power, with either the DB, AE, or a combination of the two targets on various cut scores depending on grade levels. Thus, DB and AE appeared to generally have classification accuracy indices within the moderate to strong ranges. On the other hand, RS did not perform well when evaluated as a single target or combined with DB or AE, with indices in the weak to moderate ranges for specificity, sensitivity, positive predictive power, and negative predictive power. Overall, DBR-SIS was found to be relatively more accurate for elementary grades than middle school grades.

Whereas Chafouleas et al.'s (2013) study compared DBR-SIS with other standardized rating scales, Riley-Tillman and colleagues (2008) analyzed DBR-SIS against systematic direct observation (SDO). In this study, external observers utilized SDO to observe on-task and disruptive behavior, while classroom teachers used DBR-SIS to observe and rate the same target behaviors. Participants included 15 teachers, with each teacher observing one student who has been identified as exhibiting disruptive behavior. The teachers taught either the elementary (n = 10) or middle school level (n =

5). The two target behaviors were rated on DBR-SIS on a scale from 0 to 5, with 5 indicating the highest occurrence of the target behavior. SDO data was collected on a 20 second interval over 15 minutes. On-task behavior was observed and coded on a momentary time sampling procedure, and disruptive behavior was recorded on a frequency count. Each external observer and teacher completed the observations over three occasions. Correlations between SDO and DBR-SIS data for both target behaviors were promising. The correlation coefficients of the mean of the three observations for SDO and DBR were significant, with on-task behavior at .81, and .87 for disruptive behavior.

In a recent study by Miller et al. (2015), DBR-SIS was compared with other screening measures for social, emotional, and behavioral risk. Data were collected during the 2011-2012 school year across twenty schools representing rural, suburban and urban settings. Teachers of grades 1, 2, 4, 5, 7, and 8 were included in the study. All students in participating classes were eligible for inclusion with parental permission, and up to ten students were randomly selected for inclusion in the screening procedure. This study utilized five screening measures: (a) Direct Behavior Rating – Single Item Scales, (b) Social Skills Improvement System (SSIS; Elliott & Gresham, 2007), (c) Behavior and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), (d) office disciplinary referrals (ODR) and (e) school nomination through a student record review form (SRRF). Screening occurred three times in the school year of fall, winter, and spring. For each screening period, DBR-SIS was administered twice daily across five days, and up to five students were rated concurrently. After completion of the first five

students, teachers completed a second set of five students for five days. For each student, a mean score was calculated, which was used as the value for subsequent analyses. DBR-SIS had moderate to strong correlations with the BESS and SSIS, but had a weak correlation with ODR's. The classification accuracy indices for DBR-SIS, SSIS and ODR with the BESS as the criterion resulted in mixed results. DBR-SIS and the SSIS consistently had strong sensitivity indices above .80 and specificity above .70. On the other hand, ODR had low sensitivity indices ranging from .21 to .36 and high specificity above .90.

Similar to Miller et al.'s (2015) study, Johnson et al. (2016) evaluated the classification accuracy of DBR-SIS across three time points, using the BESS as the true class measure. Data was also collected in the 2011-2012 school year from teachers of grades 1, 2, 4, 5, 7, and 8. Standard DBR-SIS target behaviors of AE, DB, and RS were used. Additionally, a composite score of the three target behaviors was reported. Data derived from DBR-SIS included a mean score for each student, which was used for the analyses. Results indicated that AUC point estimates were moderately accurate for each target with AE ranging from .84 to .88, DB from .79 to .85, RS from .76 to .81, and DBR-SIS composite from .84 to .88. Additionally, sensitivity indices across grades and time points generally exceeded .70 for individual targets and .75 for DBR-SIS composite. Specificity indices were slightly lower with individual targets exceeding .68 and the composite exceeding .70.

DBR for mental health. Although the literature on the use of DBR-SIS looks promising, it is apparent that current applications of DBR-SIS are utilized for problems

related to externalizing behaviors. In regard to utilizing DBR-SIS as a form of assessment for mental health, especially internalizing behaviors, the literature is scarce. A recent study by von der Embse, Scott, and Kilgus (2014) investigated the sensitivity to change and the concurrent validity of DBR-SIS for measuring academic anxiety. The purpose of the study was to investigate the use of DBR-SIS as a brief anxiety assessment for progress monitoring purposes. Participants included 115 undergraduate college students from a larger intervention study, selected from introductory psychology classes. Students were divided into three groups: (a) students wearing a heart rate variability monitor (HRV) and completed self-monitoring (SM) with DBR-SIS (N =57), (b) SM only (N =55), and (c) HRV only (N =55). Students were given the opportunity to earn a substantial gift card for overall test performance on a modified GRE exam. Students were assessed with a pre- and post-Test Anxiety Inventory (TAI; Spielberger, 1980), DBR-SIS, and modified GRE exam. The single items for DBR included cognitive type anxiety (DBR-C), social type anxiety (DBR-P), and physiological type anxiety (DBR-A). Pearson correlation coefficients in evaluating the concurrent validity of DBR-SIS suggested small to medium ranges for pretest TAI and DBR-SIS (.22 to .47), and small to large ranges for posttest (.23 to .59). Overall, posttest TAI and mean DBR-SIS variables had higher mean correlations (.47) than pretest correlations (.31). Regarding sensitivity to change, the results did not correspond between DBR-SIS and the TAI. Results were predominantly not significant. Upon further evaluation, DBR-SIS demonstrated greater change than the TAI across the testing situation. Thus, the study by von der Embse et al. (2014) did not demonstrate support for DBR-SIS as a viable tool in measuring anxiety.

However, there were several limitations to the study, which include the absence of school-aged students, the narrow focus on test anxiety, and the use of the TAI as an assessment for sensitivity to change.

A dissertation study by Rohrer (2015) examined DBR-SIS for screening and progress monitoring for mental health issues in middle school students. The study included 26 teachers and 437 students from two middle schools. Rohrer utilized the standard DBR-SIS behaviors of AE, DB, and RB, but also added another single item to encompass internalizing behaviors. Labeled as "unhappy," the target behavior contained behavioral descriptors present in negative affectivity and low positive affect. The study included DBR-SIS ratings from teachers and student self-ratings. Criterion measures for this study included the Teacher Observation of Classroom Adaptation-Checklist (TOCA-C; Koth, Bradshaw, & Leaf, 2009) internalizing subscale, the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001), and the BESS (Kamphaus & Reynolds, 2007). Correlation coefficients between the TOCA-C and DBR-SIS for internalizing problems were strong for both the fall (.53) and spring (.53). The BESS was only administered in the fall, but DBR-SIS for internalizing and the BESS were moderately correlated (.32). On the other hand, the PHQ-8 and DBR-SIS showed small correlation coefficients in the fall (.11) and spring (.26). For the purpose of investigating screening utility of DBR-SIS for internalizing problems, ROC curves were computed with various cutoff scores from the PHQ-8, which included scores for mild, moderate, moderately severe, and severe depression. From fall ratings, only moderate depression was statistically significant, although *AUC* indicated poor accuracy (.590). Spring ratings

featured several statistically significant findings with good accuracy for severe depression (*AUC* = .899) and moderately severe depression (*AUC* = .811) and poor accuracy for moderate depression (*AUC* = .640). Student-rated DBR-SIS for internalizing problems showed moderate accuracy for mild depression (*AUC* = .752), and good accuracy for moderate depression (*AUC* = .853), moderately severe depression (*AUC* = .901), and severe depression (*AUC* = .905). The results of this study show that DBR-SIS is capable of accurately discriminating different levels of internalizing problems.

Unfortunately, there are limitations to Rohrer's (2015) study that may impact generalization. A major limitation acknowledged in the study was that the TOCA-C, which had the strongest correlations with DBR-SIS internalizing problems, has unknown psychometric properties. Thus, having a gold-standard criterion would have established properties of data derived from DBR-SIS related to internalizing problems with substantially more defensibility.

Purpose of Study

Despite evidence supporting the use of DBR-SIS for screening and progress monitoring for externalizing issues, there are only two known studies addressing DBR-SIS for internalizing issues, with only Rohrer's (2015) dissertation focusing on school-aged children. For the purpose of addressing gaps in mental health screening assessments, procedures, and utilization, this study focused on the criterion validity and screening adequacy of DBR-SIS with target definitions for Anxiety, Depression, and Disruptive Behavior. Limitations of prior studies included the use of assessments with unknown technical adequacy. This study differentiates from the limited research in utilizing DBR-

SIS to screen for internalizing issues by proposing to apply multiple cut points on gold standard measures. Additionally, this study will examine internalizing issues on the DBR-SIS by disaggregating into separate single items for anxiety and depression rather than adding a single item for internalizing as in Rohrer's (2015) study. By evaluating DBR-SIS for the use of mental health screening accuracy, schools would have the option of a low-cost and time efficient screener in DBR-SIS.

The purpose of this study is to examine the relationship between the researcher-proposed screening measure (DBR-SIS) with that of gold standard mental health rating forms. In addition, DBR-SIS will be evaluated based on classification accuracy, as well as usability. The following research questions will be addressed through this study:

(1a) What is the strength of the relationship between scores derived from DBR-SIS Anxiety, DBR-SIS Depression, DBR-SIS Disruptive Behavior and DBR-SIS Total score with an anxiety self-rating scale, specifically the RCMAS-2?

(1b) What is the strength of the relationship between scores derived from DBR-SIS Anxiety, DBR-SIS Depression, DBR-SIS Disruptive Behavior and DBR-SIS Total score with a depression self-rating scale, specifically the CDI-2?

(2a) To what extent does scores derived from DBR-SIS Anxiety accurately identify students at-risk for anxiety, and secondarily for depression?

(2b) To what extent does scores derived from DBR-SIS Depression accurately identify students at-risk for depression, and secondarily for anxiety?

(2c) To what extent does scores derived from DBR Disruptive Behavior accurately identify students at-risk for depression and anxiety?

(2d) To what extent does scores derived from DBR Total score accurately identify students at-risk for depression and anxiety?

Method

Participants

Students. Overall, 114 students (41 male) from a middle school within a suburban southern California region participated in the study. The student sample consisted of 16 sixth grade students, 48 seventh grade students, and 50 eighth grade students. From the total sample, 111 completed all components of the study. The students were selected from general education classrooms in which teachers agreed to participate; thus, the sample was one of convenience. All students that signed assent forms and returned parental consents were included in the study. The school's racial/ethnic demographics were as follows: 70% Hispanic, 21% Caucasian, 4% African American, 3% Asian, and 1% Other. Approximately 75% of students in the school qualify for the Free and Reduced Lunch Program.

Teachers. Teachers for this study were selected through convenience sampling. Approval for research was granted by the superintendent and the principal of a middle school for the district. Eligible teachers included all full-time teachers in the middle school, regardless of subject taught. The principal recommended several teachers as potential participants. A total of six teachers agreed to participate in the study. The teachers participated voluntarily and signed consents to participate in the research study. All teachers possess a credential for teaching and are full-time general education teachers. This study excluded substitute teachers, whether short- or long-term, as well as part-time teachers. All study procedures were approved by the University of California, Riverside Institutional Review Board.

Measures

Direct Behavior Rating – Single Item Scale. The single item targets for this study are an adaptation of the standard Direct Behavior Rating Single Item Scale (DBR-SIS; Chafouleas et al., 2009). DBR-SIS is a combination of an observation and a rating form in which a teacher rates target behaviors that are observed over a set amount of time. The ratings are based on the percent of the duration of observed time that the student exhibits the target behaviors. The form includes a horizontal line for each of the three target behaviors (i.e., Anxiety, Depression, Disruptive Behavior), with equally spaced marks to represent gradients. The behaviors are operationally defined and rated on a scale from 0 to 10 along a line in which 0 represents the absence of the behavior, or 0% of the behavior occurring, and 10 represents the most frequent possible observation, as represented by the behavior occurring during 100% of the observation period. The ratings are also represented by descriptive anchors, where "never" is aligned with 0, "sometimes" is aligned with 5, and "always" is aligned with 10.

Anxiety and Depression target behaviors were operationally defined for this study. Definitions were derived from the descriptions of symptoms from gold standard measures (e.g., CDI-2, RCMAS-2). The Anxiety target behavior is operationally defined as expresses being worried or nervous, and/or complaining about physical pains or illnesses (e.g., acting scared or nervous about common situations, restless, asking repeatedly for directions or clarification, complaining of stomachaches/headaches or other physical symptoms). The Depression target behavior is operationally defined as low or restricted activity levels and/or withdrawing from or avoiding social interactions (e.g.,

not participating in games and group activities, not talking to peers, preferring to be alone, persistently tired or tearful). In addition, disruptive behavior in children have been found to be an indicator of current or future mental health issues (e.g., Campbell, 2002; Durlak & Wells, 1997; Williford & Shelton, 2008). Having been explored extensively in DBR-SIS literature, the Disruptive Behavior target provides an established definition in this measure. The standard definition of Disruptive Behavior in DBR-SIS, which is included in this study's form, is a student action that interrupts regular school or classroom activity (e.g., out of seat, fidgeting, playing with objects, acting aggressively, talking/yelling about things that are unrelated to classroom instruction).

Revised Children's Manifest Anxiety Scale – Second Edition (RCMAS-2). The RCMAS-2, which is the second edition of the RCMAS (Reynolds & Richmond, 1984; 2008), is a self-report rating scale that has been used in both clinical and research settings for the measurement of anxiety symptoms in youth ages six through 19 years. According to Reynolds and Richmond (2008), the statements on RCMAS-2 was written at an elementary reading level. The 49 yes/no items are scored to create a Total Anxiety score and three subscales: Physiological anxiety, Worry, and Social anxiety. Additionally, a Defensiveness scale is reported. The RCMAS-2 reliability estimates are well-established. Internal consistency estimates are based on the full standardization sample (N = 3,086) ages 6 through 19. Cronbach's alpha estimate for the Total Anxiety score is .92, which is considered strong. Test-retest coefficients, based on 100 school children, across a 1-week interval, resulted in .76 for the Total Anxiety score, which is considered adequate. Acceptable coefficients were observed for the subscales, as well.

Cronbach's alpha estimate for the Physiological Anxiety, Worry, and Social Anxiety scales ranges from .75 to 86, and test-retest ranges from .64 to .73.

Evidence of validity can be seen in several studies. Initially, a factor analytic study by Reynolds and Richmond (1979) was conducted on the test development sample (N = 329) and a three-factor model was recommended: Physiological, worry/ oversensitivity, and social concerns/concentration. Furthermore, in 1981, Reynolds and Page conducted a factor analysis of the RCMAS on the standardization sample (N = 4,972). Reynolds and Page found a five-factor solution to be the clearest to interpret, which included the original three factors, as well as two Lie factors. More recently, the RCMAS-2 standardization sample was utilized and had a similar five-factor solution (Reynolds & Richmond, 2008). The social concerns/concentration scale was renamed as Social Anxiety to better fit the items, and the Lie factors were split to Defensiveness and Inconsistent Responding. Additionally, as mentioned earlier in the paper, the RCMAS showed moderate to strong correlations (.65-.85) with the STAIC over multiple studies (Reynolds, 1980; Reynolds, 1982) in support for concurrent validity.

Children's Depression Inventory – Second Edition (CDI-2). The CDI-2, which is the second edition of the CDI (Kovacs, 1991, 2010), is a rating scale that offers a self-report form to measure the degree of depression symptoms in youth ages 7-17 years. The statements on the CDI-2 was written at a grade 2 reading level (Kovacs, 2010). The self-report form consists of 28 items that provides a Total score and two scale scores: Emotional Problems and Functional Problems. Each scale score is comprised of two subscales. Emotional Problems is comprised of negative mood/physical symptoms

subscale and negative self-esteem subscale. Functional Problems is comprised of ineffectiveness subscale and interpersonal problems subscale. The CDI, and the current CDI-2's reliability estimates are well-established. Internal consistency estimates are based on the full standardization sample (N = 1266) ages 7 through 17. Cronbach's alpha estimates range from .83 to .89 and test-retest coefficients range from .74 to .77. Additionally, the CDI has the advantage of possibly being the most widely researched children's self-report rating scale for the purpose of measuring depression in children (Whitcomb & Merrell, 2013).

Evidence of validity can be observed through a factor analysis study of the standardization sample by the test developers. The first edition of the CDI best loaded onto a 5-factor model, in which Anhedonia was an additional subscale. Due to addition of new items and deletion of prior items based on clinical changes (Bae, 2012), the Anhedonia subscale is absent from the most current edition, CDI-2. Evidence also supports the concurrent validity of data derived from the CDI-2. For instance, studies found that the data from the CDI-2 demonstrate moderate to strong correlations with the CES-D (r = .58), Children's Depression Rating Scale (r = .51 to .78), Depression Symptom Complex of the Child Assessment Schedule (r = .53), Child Behavior Checklist (r = .49 to .79), and Reynolds Adolescent Depression Scale (r = .68 to .94) (Doerfler, Felner, Rowlison, Raley, & Evans, 1988; Eason, Finch, Brasted, & Saylor, 1985; Hodges, McKnew, Cytryn, Stern, & Kline, 1982; Kazdin et al., 1983; Shain, Naylor, & Alessi, 1990).

For this study, the CDI-2 Total score and the two scale scores (Emotional Problems and Functional Problems) were utilized for analysis. Both scale scores encompass subscales (i.e., Negative mood/Physical Symptoms, Negative Self-Esteem, Ineffectiveness, Interpersonal Problems), but these subscales were not used for analysis purposes. According to the publishers, the Total score and scale scores have the strongest reliability and validity in relation to depressive symptoms (Kovacs, 2010).

Procedures

Middle school teachers and student participants were obtained through convenience sampling from a middle school in southern California. As an incentive for participation, teachers were provided with a $50 gift card at the completion of data collection. Teachers also signed a consent to participate in this study prior to receiving training.

Consent to participate as a subject of observation for DBR-SIS, as well as participation to complete self-report rating forms were sent home to all students in two classrooms of each participating teachers. From the students who returned signed consents, 10 students from each class were randomly selected for teacher rated DBR-SIS observations. The 10 selected students from each class were randomly assigned to one of two groups. Following procedures from Chafouleas et al. (2013) and Johnson et al. (2016), teachers observed and rated students using DBR-SIS in group one twice daily for five consecutive school days, and the second group of students were observed and rated twice daily for the next five consecutive school days. The first rating occurred for behaviors observed from the first half of the period, which was approximately 25

minutes, and the second rating occurred for behaviors observed from the second half of the period. Thus, a total of 10 data points for each DBR-SIS target was collected for each student. As noted in previous studies using generalizability theory for determining reliable decisions (Briesch et al., 2010; Chafouleas et al., 2010), as well as with studies investigating DBR-SIS as a screening measurement (e.g., Chafouleas et al., 2013; Kilgus et al., 2012), a student must have at least five data points for each target to be included in the study. Five data points was identified to be the minimal criteria for reliable decisions.

Prior to conducting DBR-SIS, teachers were trained by a graduate researcher over a course of two sessions. The two sessions combined took approximately 60 to 70 minutes to complete. In the first session, which took place in a conference room in the school building, participating teachers received information on the rationale of the current study, which included information on the effects of mental health issues in students at school, and the need for a screener that accurately identifies internalizing symptoms. Additionally, teachers received information of DBR-SIS, including a description of the standard DBR-SIS, as well as procedures on how to complete DBR-SIS. In this training, teachers were also introduced to the DBR-SIS with the target behaviors exclusive for this study (Depression, Anxiety). Teachers were provided with folders that included forms for the first and second group of students along with printed instructions. Instructions included the timeframes for completing the DBR-SIS forms, which was at the half way mark of the period and at the end of the period. All six of the participating teachers, as well as the school's principal and administrative assistant, were present for the first session of the training.

The second session involved the teachers independently completing a 20 to 30-minute online training module related to the use of DBR-SIS (www.directbehaviorratings.org). Teachers received the web link for the training module in an email, and all six teachers responded to the email when they completed the training. In addition, teachers were provided with a shared folder online that consisted of the principles of the research and DBR-SIS, data collection schedule, and contact information for the researcher if any questions needed to be addressed. No student identification information was included in online materials. All training and DBR-SIS observations were conducted in the late Spring semester.

The week after DBR-SIS data were collected by the teachers, students were pulled into an unused classroom in small groups (six to 10 students) and the researcher administered the CDI-2 and RCMAS-2 self-report rating scales. Students were provided with partitions on their desks to protect their privacy when completing the scales. Instructions were provided in writing, as well as read verbally by the researcher. To reduce pretest sensitization as a threat to validity, the self-report rating forms were administered after the completion of DBR-SIS ratings. This was to reduce the possibility that the questions on the self-report forms (CDI-2, RCMAS-2) may influence the behaviors of the students while being observed and rated by the teachers. Additionally, the presentation of self-report rating scales was counterbalanced with the purpose of controlling for order effects. Prior to pulling the groups for administration of the self-report rating scales, each group was randomly assigned to the order of CDI-2 followed by RCMAS-2, or a RCMAS-2 followed by the CDI-2.

To estimate the score reliability of target behaviors on DBR-SIS, intraclass correlation (ICC) coefficients were estimated. The study utilized Shrout and Fleiss' (1979) suggested formula for calculating data reliability with ICC coefficients. The calculation, based on a one-way analysis of variance, examines the variance between students and within observations.

ICC (1,*k*) = [(Between Mean Squares) – (Within Mean Squares)] / [(Between Mean Squares) + (*k* -1)(Within Mean Squares)]

In this study, ICC (1,*k*) values are the proportion of total variance attributable to students for mean DBR-SIS ratings of randomly selected students and observation periods (Johnson et al., 2016). DBR-SIS Depression had an ICC coefficient of .74, .76 for Anxiety, .91 for Disruptive Behavior, and .83 for Total score. Mean DBR-SIS rating scores appeared to have acceptable levels of reliability as variance of scores can be attributed to the student. DBR-SIS Disruptive Behavior, which is an established target behavior on the standard DBR-SIS form, had a comparable ICC coefficient as prior studies (e.g., Johnson et al., 2016).

Lastly, with the completion of the ten consecutive school days of data collection, teachers were asked to complete a questionnaire that provides information on the social validity, or perceived usability, of DBR-SIS. The Usage Rating Profile – Assessment (URP-A; Chafouleas, Miller, Briesch, Neugebauer, & Riley-Tillman, 2012) was distributed for teachers to complete. This instrument consists of 28 statements regarding the usage of DBR-SIS in which teachers rate each statement on a Likert-type scale. The Likert scale consists of six possible ratings: (1) *strongly disagree*, (2) *disagree*, (3)

slightly disagree, (4) *slightly agree*, (5) *agree*, and (6) *strongly agree*. The URP-A is designed to measure perceptions related to six factors: Acceptability, Understanding, Home School Collaboration, Feasibility, System Climate, and System Support. Five of the six teachers completed and returned the form.

Design and analyses. All analyses were computed using IBM SPSS Statistics Version 24. Scores for CDI-2 and RCMAS-2 were calculated with the respective developers' scoring software. Norm band for T-scores obtained for this study were gender and age specific. Data was reviewed for missingness and error, such as values entered that were outside the range for each respective assessment.

Risk status for each student was determined by student self-ratings from the CDI-2 and RCMAS-2 T-score thresholds for at-risk, as recommended by the testing manuals. The CDI-2 identifies a T-score of 65 and above as an at-risk, or elevated status. Additionally, a T-score of 70 and above on the CDI-2 is considered to be in the very elevated status. T-Scores of 1.5 and 2 standard deviations above the mean are often given additional consideration in terms of classification in behavioral rating scales. These classifications are associated with the level of symptoms that a student may show, thus requiring further assessment and intervention. On the other hand, the RCMAS-2 manual identified a T-score of above 60 to be in the moderately problematic range and a T-score of above 70 to be in the extremely problematic range.

To identify a score for screening purposes, the mean of each DBR-SIS target was calculated for each student. Student mean scores of five to ten observations were calculated for each DBR-SIS target. Additionally, a mean total score, which is

represented by the sum of DBR-SIS targets (anxiety, depression, disruptive behavior), was also calculated for each student, representing a composite score. All observations of students met the minimum criteria of at least five observations. The percentage of missing scores were expected to be low for this study, as similar studies have reported missing data on 2.8% to 5.2% (Johnson et al., 2016). Students who were observed and rated in DBR-SIS, but were absent for the self-rating scales were also excluded from this study. Only 2.6%, or 3 students, of missing data was reported from students who did not complete the CDI-2 and RCMAS-2. The resulting means for the individual target items range from 0 (no presence of behavior) to 10 (high presence of behavior), and the total score ranges from 0 to 30.

Research Question 1. Research question 1a (i.e., strength of relationship of DBR-SIS and RCMAS-2) was addressed through Pearson's correlations to examine the relationship between teacher rated DBR-SIS targets (Depression, Anxiety, Disruptive Behavior) and Total score, and student rated RCMAS-2 for each subscale (Physiological, Worry, Social, Defensive) and Total score.

The same method was utilized to address research question 1b (i.e., strength of relationship of DBR-SIS and CDI-2) in examining the relationship of DBR-SIS targets with that of CDI-2 Total and scale scores (Emotional, Functional).

Research Question 2. A total of 44 Receiver Operating Characteristics (ROC) curves were computed to analyze the diagnostic accuracy of the targets and total score of DBR-SIS in identifying students who are at-risk for anxiety and depression with the RCMAS-2 and CDI-2 as the gold standard tests. DBR-SIS targets (Anxiety, Depression,

Disruptive Behavior) and total scores were evaluated based on elevated (T-score > 64) and very elevated (T-score > 69) threshold levels on the CDI-2. Each DBR-SIS target (Anxiety, Depression, Disruptive Behavior) and Total score was evaluated based on moderately problematic (T-score > 60) and extremely problematic (T-score > 70) threshold levels on the RCMAS-2. Classification indices were examined for AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) with 95% confidence intervals. The alpha level was set at .05 for all significance tests.

Bootstrapping procedures were implemented to address the low base rate of students who may display significantly elevated depression and anxiety scores. AUC indices were bootstrapped to generate 95% confidence intervals.

Regarding screening instruments, there is a preference for inclusion (sensitivity) as it is important to ensure that minimal students who are at-risk are left undetected (Glover & Albers, 2007; Levitt et al., 2007). Several authors and studies have recommended minimum sensitivity and specificity values of .75 or .80 (e.g., Carran & Scott, 1992; Carter, Briggs-Gowan, & Davis, 2004; Glover & Albers, 2007; Gredler, 2000; Meisels, 1989). In considering preference towards values of classification accuracy indices, this study used a systematic approach for identifying optimal cut-scores. Primarily, a preference for scores with a minimum of .80 for sensitivity and specificity values were used in determining the optimal cut-score on a ROC curve. If multiple scores on a ROC curve had minimum specificity and sensitivity values of .80, the score with the highest values, with preference given to sensitivity values, determined the optimal cut

score. When no cut-scores satisfied a minimum of .80 for both sensitivity and specificity, the cut-score with a sensitivity of at least .80 with the highest possible specificity value was considered. If no cut-scores met the above criteria, the cut-score with the most balanced, or equally weighted, sensitivity and specificity values was selected (Johnson et al., 2016).

Research question 2 examined the classification accuracy of specific targets on DBR-SIS by utilizing ROC curve analyses, examining values around sensitivity and specificity. Research question 2a examined to what extent DBR-SIS Anxiety target can accurately identify students at-risk for anxiety, and secondarily for depression. Research question 2b examined to what extent DBR-SIS Depression target can accurately identify students at-risk for depression, and secondarily for anxiety. Lastly, Research question 2c and 2d examined to what extent DBR-SIS Disruptive Behavior target and Total score can accurately identify students at-risk for anxiety and depression.

Results

Description of Data

Signed parent consent forms from 114 middle school students were collected for this study. All 114 students were initially included in the study, which included the teacher rated DBR-SIS and the student self-rating forms. A total of 111 were considered as demonstrating complete data, which included a minimum of five DBR-SIS observations and the completion of both student self-rating forms. Six participating full-time teachers from general education classes were included in this study. All participating students attended classes from these teachers. These teachers participated voluntarily and were referred by the principal of the school. Students were randomly assigned to one of two groups for observation purposes for the teacher rated DBR-SIS. Each teacher completed DBR-SIS for 10 selected students from two periods (n = 20) from those who returned consents from their classes. Only students with complete data were included in the study (i.e., present for a minimum of five data collections; completion of the CDI-2 and RCMAS-2). All students completed the two self-rating forms (CDI-2, RCMAS-2) over the course of a single day. Students who were absent completed the self-rating forms the following day. There were eleven students absent for the initial testing day, and eight of the students were present on the make-up day. The presentation order of the CDI-2 and RCMAS-2 were counterbalanced with the purpose of controlling for order effects. Additionally, the self-rating forms were administered after the completion of teacher rated DBR-SIS to address pretest sensitization as a threat to validity.

Descriptive Statistics

Table 1 displays descriptive statistics for the following measures organized by grade level: DBR-SIS Total, DBR-SIS Anxiety, DBR-SIS Depression, DBR-SIS Disruptive Behavior, CDI-2 Total, CDI-2 Emotional, CDI-2 Functional, RCMAS-2 Total, RCMAS-2 Physiological, RCMAS-2 Worry, RCMAS-2 Social, and RCMAS-2 Defensiveness. DBR-SIS Total means ranged from .93 to 2.01 (*SD* = 1.23-2.01) across grade levels, with a total combined mean of 1.67 (*SD* = 1.68). DBR-SIS Anxiety, Depression, and Disruptive Behaviors means ranged from .18 to 1.14 (*SD* = .44-1.50) across grade levels, with a combined mean ranging from .45 to .76 (*SD* = .59-1.20). Means for CDI-2 Total and scale scores ranged from 56.54 to 59.92 (*SD* = 10.72-14.09), and means for RCMAS-2 Total and subscales ranged from 45.50 to 56.29 (*SD* = 7.72-11.94). See Figure 1 for CDI-2 Total distribution and Figure 2 for RCMAS-2 Total distribution.

Proportion At-risk. Table 2 presents the proportion of students who were found to be at-risk on the CDI-2 across two thresholds: T-score above 64 (*elevated*) and T-score above 69 (*very elevated*). For CDI-2 Total scores, approximately 29.7% of students had scores that were at or above the *elevated* threshold, and 18.9% of students had scores that were at or above the *very elevated* threshold. Similarly, 29.7% of students scored at or above the *elevated* scores threshold and 16.2% of students scored at or above the *very elevated* threshold on CDI-2 Emotional. For CDI-2 Functional scores, 28.8% of students met the threshold for *elevated* scores, and 14.4% of student scores were at or above the *very elevated* threshold.

Table 3 presents the proportion of students who were found to be at-risk on the RCMAS-2 across two thresholds: T-score above 60 (*moderately problematic*) and T-score above 70 (*extremely problematic*). For RCMAS-2 Total scores, approximately 23.4% of student had scores at or above the *moderately problematic* threshold, and 1% of students had scores at or above the *extremely problematic* threshold. A total of 17.1% of students met the *moderately problematic* threshold and 1.8% met the *extremely problematic* threshold for RCMAS-2 Physiological scores. RCMAS-2 Worry presented with 27.9% of students scoring at or above the *moderately problematic* threshold and 3.6% at or above the *extremely problematic* threshold. In RCMAS-2 Social, 20.7% of students scored at or above *moderately problematic* and 3.6% students scored at or above the *extremely problematic* threshold. Lastly, 15.3% of students scored within the mod*erately problematic* threshold in the RCMAS-2 Defensiveness score. No scores for RCMAS-2 Defensiveness were within the *extremely problematic* range.

Self-report Rating Scales. Correlation coefficients were produced to examine the relationship between the two rating scales, CDI-2 and RCMAS-2. All correlation coefficients between CDI-2 and RCMAS-2 scores were significant with magnitudes ranging from $r = -.44$ to $r = .65$ ($p < .01$). The total scores for RCMAS-2 and CDI-2 displayed the strongest relationship with a coefficient of $r = .65$ ($p < .01$). RCMAS-2 Total was highly correlated with the CDI-2 Emotional ($r = .63$, $p < .01$) and CDI-2 Functional ($r = .57$, $p < .01$). CDI-2 Total was highly correlated with the RCMAS-2 Physiological ($r = .63$, $p < .01$), RCMAS-2 Worry ($r = .63$, $p < .01$), RCMAS-2 Social ($r = .60$, $p < .01$), and RCMAS-2 Defensiveness ($r = -.50$, $p < .01$). The RCMAS-2

Defensiveness subscale was negatively correlated with CDI-2 scores, ranging from r = -.440 to r = -.50, p < .01). These correlations produced similar coefficients as prior research on anxiety and depression rating scales, specifically with Crowley and Emerson's (1996) comparison of CDI and RCMAS.

Research Question 1a

In addressing the first part of Research question 1, Pearson's correlations were conducted to examine the relationship between DBR-SIS total and target scores, as measured by the subscales and total score of the RCMAS-2. The results of the correlational analyses are reported in Table 4. Coefficients between all measures of DBR-SIS and RCMAS-2 ranged from r = .19 to r = .51. The correlation for DBR-SIS Anxiety score and the RCMAS-2 Total score had the strongest correlation at .51. DBR-SIS Disruptive Behavior resulted in the lowest correlations with RCMAS-2 scores, ranging from .09 to .24. Additionally, DBR-SIS Depression and RCMAS-2 had low to moderate correlations, ranging from r = .31 to r = .43.

Research Question 1b

In addressing the second part of Research question 1, Pearson's correlations were used to examine the relationship between DBR-SIS total and target scores, as measured by the scale and total score of the CDI-2. The results of the correlational analyses are reported in Table 4. Coefficients between all measures of DBR-SIS and CDI-2 ranged from r = .19 to r = .67. The DBR-SIS Depression score and the CDI-2 scores produced the strongest correlations, ranging from r = .61 to r = .67. The weakest relationship was between DBR-SIS Disruptive Behavior and CDI-2 scores, ranging from r = .19 to r = .25.

Research Question 2

ROC curve analyses were conducted to address Research question 2. Figures 3 – 46 provide graphs of ROC curves for each analysis. Classification accuracy indices are calculated for DBR-SIS screening measure, relative to established standardized measures for mental health (RCMAS-2, CDI-2). The indices are calculated with the following equations:

Sensitivity = valid positives / (valid positives + false negatives)

Specificity = valid negatives / (valid negatives + false positives)

PPV = valid positives / (valid positives + false positives)

NPV = valid negatives / (valid negatives + false negatives)

Classification indices were evaluated based on optimal cut scores, for each selected threshold on the CDI-2 and RCMAS-2, based on developers' recommendations. CDI-2 identifies T-scores above 64 (*elevated*) and above 69 (*very elevated*) as thresholds, and the RCMAS-2 identifies T-scores above 60 (*moderately problematic*) and above 70 (*extremely problematic*) as thresholds. In total, 44 ROC curves were generated to calculate classification indices for this study. ROC curves were not generated for RCMAS-2 *extremely problematic* scores due to the low base rate, possibly due to sample size limitations. In examining the proportion of students who fell within the *extremely problematic* scores, only 1 to 3 students were identified as having T-scores above 70 for each subscale and Total score. Low base rates lead to issues in analysis related to

significant inaccuracies in estimated parameters associated with ROC curve analyses (Hanczar et al., 2010). Tables 4-6 represent the summary of area under the curve (AUC) point estimates for CDI-2 Total and scale scores. Tables 7-11 represent the summary for AUC point estimates for RCMAS-2 Total and subscale scores. Tables 13-14 summarize the values on DBR-SIS and classification indices (i.e., sensitivity, specificity, PPV, NPV) of ROC curve for *elevated* and *very elevated* cut-scores on the CDI-2. Table 15 summarizes the values on DBR-SIS and classification indices of ROC curve for *moderately problematic* cut-scores on the RCMAS-2.

AUC statistics are valuable in evaluating the diagnostic accuracy of DBR-SIS as a mental health screening tool. AUC point estimates range from .50, which corresponds to chance, and 1.00, which indicates perfect precision. Each AUC point estimate was above .50, which indicate each DBR-SIS target performed better than chance at identifying students at-risk for depression and anxiety, respective of CDI-2 and RCMAS-2. Most confidence intervals were also within the range of above .50, with the exception of DBR-SIS Disruptive Behavior in CDI-2 Emotional, CDI-2 Functional, and RCMAS-2 Physiological. AUC point estimates for DBR-SIS Depression in both CDI-2 Total thresholds (*elevated, very elevated*) were above .80, as were DBR-SIS Anxiety for RCMAS-2 Total (*moderately problematic*). AUC point estimates for DBR-SIS Total was above .80 in CDI-2 Total for the *very elevated* threshold and for the RCMAS-2 Total for the *moderately problematic* threshold.

AUC for Depression. AUC point estimates for DBR-SIS Depression ranged from .80 to .92 across both thresholds of the CDI-2. DBR-SIS Depression had AUC point

estimates above .90 for CDI-2 scale and total scores within the *very elevated* threshold. AUC point estimates for DBR-SIS Anxiety ranged from .64 to .73 in CDI-2 across both thresholds. Fair or moderate point estimates were found only in DBR-SIS Anxiety for CDI-2 Total (*very elevated*) and CDI-2 Emotional (*elevated*). No AUC point estimates met or exceeded .70 for DBR-SIS Disruptive Behaviors. AUC point estimates for DBR-SIS Disruptive Behaviors ranged from .61 to .66 in CDI-2 across both thresholds. DBR-SIS Total scores resulted in all AUC point estimates exceeding .70 for CDI-2 across both thresholds. The AUC estimates ranged from .76 to .86. DBR-SIS Total scores were especially accurate, with AUC point estimates above .80, on all subscales and Total scores across both thresholds, apart from CDI-2 Total score *elevated* threshold (*AUC* = .79) and CDI-2 Functional score *elevated* threshold (*AUC* = .76).

AUC for Anxiety. The *extremely problematic* threshold for RCMAS-2 was not included due to a low base rate. All reported analyses for RCMAS-2 is based on the threshold of a T-score above 60, *moderately problematic*. AUC point estimates for DBR-SIS Anxiety ranged from .71 to .87 across RCMAS-2 subscale and total scores. DBR-SIS Anxiety was especially accurate with an AUC point estimate of .87 on the RCMAS-2 Total score. DBR-SIS Depression also exceeded .70 in AUC point estimates across RCMAS-2 subscale and total scores. AUC point estimates ranged from .71 to .76. No AUC point estimates met or exceed .70 for DBR-SIS Disruptive Behaviors in the RCMAS-2, other than the Defensiveness subscale (*AUC* = .75). AUC point estimates for DBR-SIS Disruptive Behaviors ranged from .53 to .75. DBR-SIS Total scores resulted in all AUC point estimates exceeding .70 for RCMAS-2 across all subscales and total score.

The AUC estimates ranged from .72 to .81. DBR-SIS Total was especially accurate on RCMAS-2 Worry subscale and Total score with AUC point estimates above .80.

The Total scores for the CDI-2 and RCMAS-2 displayed the strongest point estimates for all DBR-SIS targets and total score. Therefore, sensitivity, specificity, PPV, and NPV indices were derived from Total scores of CDI-2 and RCMAS-2.

Cut-scores, Sensitivity, and Specificity. For the CDI-2 Total *elevated* scores threshold, cut scores for each DBR-SIS target behavior and total score are as follows: .24 for Depression, .23 for Anxiety, .23 for Disruptive Behavior, and 1.70 for Total. Optimal cut scores for each target and total with sensitivity (SN) and specificity (SP) values of .80 and above was not achieved in any DBR-SIS targets. Specificity values ranged from .58 to .73, and sensitivity values ranged from .61 to .82. Sensitivity and specificity values were highest for DBR-SIS Depression (.82, .73), followed by DBR-SIS Total score (.73, .72), DBR-SIS Disruptive Behavior (.67, .64), and DBR-SIS Anxiety (.61, .58).

For the CDI-2 Total *very elevated* scores threshold, cut scores for each DBR-SIS target behavior and total score are as follows: .65 for Depression, .05 for Anxiety, .47 for Disruptive Behavior, and 2.37 for Total. Optimal cut scores for each target and total with sensitivity (SN) and specificity (SP) values of .80 and above were achieved by DBR-SIS Depression and Total scores. Specificity values ranged from .52 to .85, and sensitivity values ranged from .67 to .86. DBR-SIS Depression produced the highest values with sensitivity and specificity values of .86 and .85, respectively, followed by DBR-SIS Total, which produced sensitivity and specificity values of .81 and .80. Although DBR-SIS Anxiety produced a good sensitivity value (.81), specificity was of weak value (.52).

Lastly, DBR-SIS Disruptive Behavior produced sensitivity and specificity values of .67 and .64.

For the RCMAS-2 Total *moderately problematic* threshold, cut-scores for each DBR-SIS target behavior and total score are as follows: .44 for Anxiety, .41 for Depression, .18 for Disruptive Behavior, and 1.76 for Total. Optimal cut-scores for each target and total with sensitivity (SN) and specificity (SP) values of .80 and above was only achieved in DBR-SIS Anxiety target (SN = .89, SP = .82). Specificity values ranged from .60 to .82, and sensitivity values ranged from .72 to .89. DBR-SIS Total produced sensitivity and specificity values of .81 and .74, respectively. DBR-SIS Disruptive Behavior produced sensitivity and specificity values of .81 and .60. Lastly, DBR-SIS Depression produced sensitivity and specificity values of .72 and .71.

PPV and NPV. Results from past studies related to DBR-SIS indicate that PPV point estimates were expected to be lower than NPV point estimates (e.g., Johnson et al., 2016). Each target behavior and total score is expected to have low PPV estimates and high NPV values. This can be expected as the lower the PPV, there is a greater chance the screener is over-identifying students as at-risk. DBR-SIS target and total scores ranged in PPV point estimates from .38 to .56 for CDI-2 *elevated* threshold. For CDI-2 *very elevated*, PPV point estimates ranged from .28 to .58. Similarly, PPV point estimates ranged from .38 to .61 for RCMAS-2 *moderately problematic*. NPV point estimates ranged from .78 to.90 across all DBR-SIS targets and CDI-2 and RCMAS-2 thresholds.

Usage Rating Profile – Assessment. The URP-A was utilized in evaluating the usability of the DBR-SIS for screening for depression and anxiety in middle school

students. Of the six teachers who participated in this study, five completed the URP-A. The factors on URP-A are the following: Acceptability, Understanding, Home School Collaboration, Feasibility, System Climate, and System Support (Miller et al., 2014). Acceptability, which is related to the teacher's perception on the appropriateness and interest of the DBR-SIS, had a mean rating of 4.47. The Understanding factor (M = 5.07) determines the teacher's knowledge of content and procedures of the assessment tool. Items for Home School Collaboration determines the teacher's perception on the necessity of collaboration in using DBR-SIS. In the current study, Home School Collaboration factor had a mean score of 3.80. Teachers' expressed their ease of use in their ratings for items on the Feasibility factor with a mean score of 4.20. The System Climate factor evaluated how well DBR-SIS fit in with the current school climate (M = 4.80). System Support factor had a mean score of 2.80. This factor determines to what extent teachers felt they would benefit from additional support to implement DBR-SIS. A lower score on the System Support factor may indicate that the teachers perceived that they had a greater ability to implement DBR-SIS independently (Miller et al., 2014).

Discussion

This study focused on the evaluation of data derived from Direct Behavior Rating Single-Item Scales to screen for internalizing problems, specifically depression and anxiety, in schools with middle grade students. Based on current literature, DBR-SIS has been shown to be an efficient screener and progress monitoring tool for externalizing behaviors (e.g., Chafouleas et al., 2009; Fabiano et al., 2009, Johnson et al., 2016). Only a few recent studies have examined DBR-SIS for internalizing problems, as a combined measure of various characteristics of internalizing problems, or for a single item described for a broad internalizing problem (Rohrer, 2016; von der Embse et al., 2015). The present study differed from past studies in that single items on the DBR-SIS were specific to and labeled as depression and anxiety. Prior studies either studied different characteristics of one internalizing issues such as anxiety (von der Embse et al., 2015) or created a single target with a broad definition related to internalizing problems (Rohrer, 2015). Additionally, past studies of DBR-SIS targeting internalizing problems utilized measures that lacked technical adequacy in assessing for internalizing issues. The current study aimed at utilizing gold standard measures, specifically the CDI-2 and RCMAS-2, in evaluating DBR-SIS cut scores and overall diagnostic accuracy for depression and anxiety screening, as well as evaluating general concurrent and divergent validity. Lastly, with using the CDI-2 and RCMAS-2, different thresholds were utilized based on recommendations from the publishers of each respective rating form.

In examining the descriptive statistics of this study, CDI-2 mean scores, ranging from 56.54 to 59.92 (*SD* = 10.72 to 14.09), were higher than expected across all grade

levels. Therefore, a high proportion of students (29-30%) were found to be in the *elevated* range (T-score > 64) across CDI-2 scale and total scores. The proportion of students (14-19%) within the *very elevated* range (T-score > 69) appeared to be more aligned with studies of depression in youth (Costello et al., 2003; NIMH, 2010; SAMHSA, 2015). For instance, SAMHSA (2015) stated that up to 16% of adolescents live with major depression symptoms. On the other hand, mean scores of the RCMAS-2 were as expected, ranging from 45.50 to 56.29 (*SD* = 7.72 to 11.94). The RCMAS-2 Total score was especially uniform with mean scores from 48.50 to 52.10 (*SD* = 8.15 to 11.29). Percentages of students at-risk on the RCMAS-2 for the lower threshold *(moderately problematic*) ranged approximately from 15% to 28%. This is in line with recent studies that estimates approximately 15% to 25% of school-aged children experience an anxiety disorder (Beesdo, Knape, & Pine, 2011; Costello et al., 2003; NIMH, 2010). Conversely, in the higher threshold (*extremely problematic*) recommended by the RCMAS-2 publisher's manual, only 0% to 3.6% of students fell in this category. As previously mentioned, due to the exceptionally low base rate on the *extremely probl*ematic threshold for RCMAS-2, this threshold was excluded from further investigation in this study.

Two main research questions were investigated in this study. First, criterion validity was examined by the strength of relationship between DBR-SIS target behaviors and total score, which was evaluated against gold standard measures, CDI-2 and RCMAS-2. Second, DBR-SIS target behavior and total score classification accuracy was investigated based on different thresholds on the CDI-2 and RCMAS-2. Specifically, the primary concern was whether DBR-SIS Anxiety was accurate in identifying students at-

risk for anxiety based on the RCMAS-2, and whether DBR-SIS Depression was accurate based on CDI-2. Being an established target behavior in DBR-SIS research, Disruptive Behavior was evaluated against both gold standard measures. Additionally, representing the composite score, DBR-SIS Total was evaluated against both gold standard measures.

The present study examined the concurrent validity of DBR-SIS targets and found most relationships to be statistically significant with moderate strength, especially when examining similar constructs, most importantly the CDI-2 and DBR-SIS Depression, as well as the RCMAS-2 and DBR-SIS Anxiety. The importance of this finding is that the proposed targets of Depression and Anxiety on DBR-SIS appear to measure the construct related to its definitions, respectively. Data from the target behaviors on DBR-SIS also supported evidence of discriminant validity. When DBR-SIS target behaviors were compared to unrelated constructs (e.g., Disruptive Behavior to CDI-2 and RCMAS-2), relationships were not significantly correlated or were of weak strength. Measures between Anxiety and Depression (e.g., DBR-SIS Depression to RCMAS-2) were statistically significant and were of weak to moderate strength. This appears to be aligned with studies that support the theory of depression and anxiety sharing underlying constructs, but also demonstrating unique constructs respective of depression and anxiety (Clark & Watson, 1991). Additionally, ICC estimates showed sufficient evidence that variance in mean DBR-SIS ratings was attributed to student differences.

Classification accuracy indices provide preliminary support for the use of DBR-SIS as a screening tool for internalizing problems. When analyzing the use of DBR-SIS for depression, several findings were of importance. First, AUC point estimates were

calculated for DBR-SIS target and total scores against CDI-2 across all subscales and total score. For depression, DBR-SIS Depression and CDI-2 Total score from the *elevated* (*AUC* = .84, *95% CI* [.74 - .93]) and *very elevated* (*AUC* = .92, *95% CI* [.87 - .97]) thresholds resulted in good accuracy. This suggests that DBR-SIS Depression has good accuracy in identifying students on both thresholds from student self-ratings on the CDI-2 Total. However, similarly high AUC point estimates were noted for CDI-2 Emotional and Functional scales, and therefore there was no clear deviation from CDI-2 Total score. Considering standardized delivery of the CDI-2 requires the completion of the whole form, the Total score was utilized for analyzing sensitivity, specificity, PPV, and NPV indices. In identifying depression through DBR-SIS, as expected the Depression target resulted in larger sensitivity and specificity values than other targets on DBR-SIS. Specifically, the optimal cut-score on DBR-SIS Depression was associated with a sensitivity of .86 and specificity of .85 for the *very elevated* threshold, and a sensitivity of .82 and specificity of .73 for the *elevated* threshold on the CDI-2 Total score. Considering the intended use of a mental health screener, higher sensitivity marks were favored over specificity when evaluating cut-scores. Although DBR-SIS Depression on the lower threshold of the CDI-2 Total score produced weaker indices, both sensitivity and specificity were near or within recommended minimal estimates (.75 to .80; Glover & Albers, 2007).

Significant findings were noted in the evaluation of DBR-SIS for anxiety. AUC point estimates were calculated for DBR-SIS target and total scores for the *moderately problematic* threshold on RCMAS-2 across all subscales and total score. As expected,

DBR-SIS Anxiety target and RCMAS-2 Total score resulted in good accuracy (AUC = .87, 95% CI [.79 - .96]). AUC point estimates for DBR-SIS Anxiety were considerably weaker on RCMAS-2 subscales, with point estimates ranging from .71 to .80. Similar to the strength of the total score on CDI-2, the RCMAS-2 Total score was utilized for analyzing sensitivity, specificity, PPV, and NPV indices. DBR-SIS Anxiety resulted in the strongest classification accuracy indices for identifying students with *moderately problematic* anxiety scores. The optimal cut-score on DBR-SIS Anxiety was associated with a sensitivity of .89 and specificity of .82.

An encouraging finding is the classification accuracy similarities of DBR-SIS Depression and Anxiety targets with other mental health screeners (e.g., Kovacs, 2010; Levitt et al., 2007; Rischel et al., 2005). This is of importance as administering and scoring DBR-SIS may be relatively more efficient in time and costs than established gold standard assessments. This may suggest DBR-SIS as a screening tool that may be used more widely in schools than other mental health screeners which are typically reserved for special education purposes.

With CDI-2 and RCMAS-2 as the established measures, PPV and NPV indices were calculated. As previously stated, NPV indices were expected to be higher than PPV indices. Generally, cut-scores on DBR-SIS targets over-identified students at risk. For instance, only 56% of students identified at-risk by the DBR-SIS Depression optimal cut-score on the *elevated* threshold were at-risk on the CDI-2. Similarly, 58% of students identified at-risk with the optimal cut-score on DBR-SIS Depression were considered at-risk on the *very elevated* threshold for CDI-2. On the other hand, NPV indicated that 90%

and 96% were accurately identified as not at-risk on *elevated* and *very elevated* thresholds, respectively. Likewise, DBR-SIS Anxiety resulted with 61% of students being accurately identified at-risk and 96% as being accurately identified not at-risk on the *moderately problematic* threshold for RCMAS-2 Total score. The relatively low PPV would suggest that DBR-SIS would not be appropriate as the sole instrument in making decisions for intervention consideration. However, results from this study show DBR-SIS may be a valuable tool as part of a multiple gating screening procedure where a gold standard measure such as a CDI-2 or RCMAS-2 is followed by the initial identification by DBR-SIS screening.

It is worth noting that DBR-SIS Disruptive Behavior, and to a certain extent the Total score, did not contribute indices that were more accurate than DBR-SIS Anxiety and Depression. For instance, the Disruptive Behavior target resulted in AUC point estimates that ranged from .53 to .69 for RCMAS-2, and .60 to .66 for CDI-2. These point estimates are considered weak and near chance. Additionally, the optimal cut-score on the Disruptive Behavior target resulted in weak to moderate indices for specificity and sensitivity across CDI-2 and RCMAS-2 thresholds. Although DBR-SIS Disruptive Behavior optimal cut-score had decent classification indices of .81 for sensitivity and .60 for specificity on RCMAS-2, the Anxiety target and Total score were conclusively more accurate.

One of the barriers of screening for depression and anxiety is that internalizing symptoms are difficult to detect, especially in comparison to externalizing symptoms. This study found that teachers, without extensive training, through observations and

ratings with DBR-SIS Anxiety and Depression targets, can be precise in detecting at-risk students for internalizing issues. Although PPV indices showed DBR-SIS Depression and Anxiety over-identified students, there is also a possibility that teachers could detect aspects of student depression and anxiety that was not explained by student self-rating scales.

A goal of this study was to evaluate a school-based mental health screening tool that would be favorably perceived by teachers. The URP-A was administered to evaluate the social validity of DBR-SIS for use in schools by the teachers involved in this study. Five of the six teachers completed the URP-A, and ratings were generally positive. Teachers were asked to complete 28 questions on Likert scales which measured six factors related to usability. The Acceptability, Understanding, Feasibility, System Climate, and System Support all garnered mean scores that indicated a positive perception from the teachers. The teachers especially thought that their knowledge of the content and procedure of the DBR-SIS was strong, as the Understanding factor was the most highly rated. Not surprisingly, Home School Collaboration obtained the lowest scores as administration and scoring of DBR-SIS is independent of collaboration with students' home environment.

Limitations

This study aimed to strengthen the literature on mental health screening, particularly with internalizing problems, through examining the classification properties of data derived from DBR-SIS. Researcher-developed target definitions for DBR-SIS Anxiety and Depression were validated against gold standard measures, the CDI-2 and

RCMAS-2. DBR-SIS achieved acceptable classification accuracy indices to screen for depression and anxiety across various thresholds. Additionally, the study design was carefully examined to address threats to internal validity. For instance, self-rating forms were presented after the completion of DBR-SIS ratings to reduce pretest sensitization. Other procedures included random assignment of students to groups, random observation order, and counterbalancing presentation of the self-rating scales. Nevertheless, the following limitations are recognized.

One limitation of the study was that the classification criteria was derived strictly from the recommended thresholds from the publisher's manual of CDI-2 and RCMAS-2. Other studies have developed alternate thresholds for the CDI-2 (Bang, Park, & Kim, 2015), which may have impacted the extent to which students express thoughts and feelings of depression. The current study utilized T-score based thresholds (e.g., T-score > 60), whereas other studies utilized raw score based thresholds (Comer & Kendall, 2005). This raw score threshold was mentioned as an option in the first edition of the Children's Depression Inventory (Kovacs, 1992), although several studies have found differing raw score thresholds than the publisher's manual (Bang et al., 2015; Cowell et al., 2005). Although prevalence rates on the CDI-2 *elevated threshold* (T-score > 69) and RCMAS-2 *moderately problematic* threshold (T-score > 60) were similar to rates from prior studies (Costello et al., 2003; NIMH, 2010; SAMHSA, 2015), there is a possibility that using updated raw score thresholds may have provided stronger classification indices.

For this study, grade levels were combined into one group for analyses. Considering that there were three grade levels present (Grades 6, 7, 8), there is a possibility that optimal cut-scores may have differed for some grade levels. Although the current study did not have a sample size large enough to consider parsing grade levels, this may be an area that requires caution in interpretation. Additionally, only middle school students were part of this study, therefore adopting the optimal cut-scores for other grade levels (e.g., kindergarten through Grade 5 and high school) would not be advised. Consideration for DBR-SIS Anxiety and Depression targets in grade levels outside of middle school would require establishing optimal cut-scores for those grade groups.

Additionally, even though DBR-SIS Anxiety and Depression optimal cut-scores produced adequate sensitivity and specificity indices, the scale may have been too large (zero through 10), subsequently causing nominal cut-scores. For example, despite DBR-SIS Anxiety having sensitivity and specificity indices over .80, the optimal cut-score was .44. In other words, a cut-score of .44 refers to a mean DBR-SIS score of .44 when aggregated across multiple ratings and days. Since the DBR-SIS scale ranges from zero to 10, with zero indicating no behavior during the observation period and 10 indicating behavior occurred 100% of the time, .44 indicates that behavior occurred for an average of 4.4% of the observation period. This small number may be difficult to detect within a 10-point scale and may have poor face value by school based users.

DBR-SIS Disruptive Behavior resulted in the weakest correlation coefficients with other measures and poorest classification indices in identifying depression and anxiety. This likely also affected the DBR-SIS Total score. Considering that the DBR-

SIS used in this study was tested with three behaviors being observed for a composite DBR-SIS Total score, subtracting one target would possibly have different results. So, even though DBR-SIS Total score resulted in sufficient classification indices for both anxiety and depression, there is a possibility that there could have been more robust conclusions based on DBR-SIS Total scores if the Disruptive Behavior target did not influence the composite scores. This study suggests that future research exclude the Disruptive Behavior target on the DBR-SIS form and focus on the Anxiety and Depression targets and a Total score comprised of the two targets.

Future Directions

Findings from this study provide hopeful groundwork for the utility of DBR-SIS as a screening tool for internalizing symptoms. Therefore, future exploration of DBR-SIS and mental health could build on this study in numerous ways. For instance, examining DBR-SIS Anxiety and Depression targets across multiple time periods would provide important direction of the various uses of DBR-SIS. When implemented in schools, typical behavior and academic screening occur through multiple time periods through the school year (e.g., fall, winter, spring). Considering the single time period that the current study utilized, it may be beneficial to mimic the three times a year that most screening occurs. This would provide information on classification accuracy for fall, winter, and spring, as well as potentially explore sensitivity to change. Prior studies on DBR-SIS with standard behaviors (AE, DB, RB) have shown to be accurate throughout different time points, and have also been successful in detecting change as a progress monitoring tool (Chafouleas, 2011). Establishing classification accuracy across differing time periods

throughout the school year, as well as evaluating the progress monitoring utility of DBR-SIS Anxiety and Depression targets, would support use throughout all tiers within a multi-tiered system of support.

Considering the low classification accuracy of DBR-SIS Disruptive Behavior, exclusion of this target behavior may be warranted from future DBR-SIS internalizing problems research. As suggested earlier, one consideration may include utilizing a DBR-SIS form with only Anxiety and Depression targets. Another procedure may be to substitute Disruptive Behavior target with another well-established target behavior, Academic Engagement, from the standard DBR-SIS form. Academic Engagement, or a lack thereof, may be a good fit especially considering internalizing problems are associated with behaviors (e.g., avoidance, withdrawal) that lead to difficulties with engaging with school tasks.

As noted earlier, lower PPV indices suggest that DBR-SIS Anxiety and Depression targets tend to over-identify students at-risk on the RCMAS-2 and CDI-2, respectively. As a standalone screener, the over-identification of students may lead to significant impact in schools by the overuse of mental health resources and negatively affecting the ability to focus on students who are truly in need. However, considering the promising classification accuracy of DBR-SIS, being part of a multiple gating procedure is likely to alleviate issues of over-identification. Multiple gating procedures are considered best practice in assessing for student mental health (McConaughy & Ritter, 1995). DBR-SIS has been proposed as part of a multiple gating procedure, involving combinations of DBR-SIS targets or combining DBR-SIS with other assessments (e.g.,

Kilgus et al., 2012, Miller et al., 2015). Some studies have suggested that DBR-SIS could be the first screener of a two-part screening procedure (Johnson et al., 2016; Rorher, 2015). To utilize DBR-SIS as part of a two-part screening procedure, future studies would benefit from using a multi-informant method, with teacher completed DBR-SIS as the initial screener, and then implementing a student self-rating scale. This process would follow the foundation of multiple gating by helping to sequentially narrow down a larger group, leading to the identification of a smaller group that is more likely to demonstrate the symptoms of concern (Whitcomb & Merrell, 2014).

Implications of utilizing DBR-SIS as a screening tool within a MTSS framework appears promising. As previously mentioned, mental health services, such as screening and interventions, are infrequent in schools (Briggs-Gowan et al., 2000). Therefore, identifying students for mental health needs with DBR-SIS may lead to identifying students without having the resources to provide interventional services. Another process of multiple gating that could be considered would be to use DBR-SIS as the first gate, and then implementing a class-based intervention prior to conducting a second gate of assessments. Considering the positive effects of MTSS in delivering services, providing a meaningful intervention prior to conducting the second gate of assessments may be beneficial in identifying at-risk students by evaluating response to the class-based intervention.

Summary

Screening for mental health in schools is burdensome to implement due to costs and time consumption, thus decreasing its likelihood of being implemented. Furthermore,

internalizing problems can be difficult to detect and therefore students with symptoms related to depression and anxiety may go without interventions. Developing a low cost and time efficient screener for internalizing problems may help increase the chances of students with internalizing issues to receive proper intervention as soon as possible. To address the lack of a favorable screener for internalizing problems, this study developed specific target definitions, Depression and Anxiety, for DBR-SIS and evaluated the screening abilities of the resulting data. In evaluating the use of DBR-SIS for screening for internalizing problems, multiple thresholds on gold standard measures of depression (CDI-2) and anxiety (RCMAS-2) were utilized.

The first promising finding from this study was the evidence of criterion validity of the Anxiety and Depression targets of DBR-SIS. Using the CDI-2 and RCMAS-2 as the gold standard measurements, DBR-SIS targets showed discriminant validity as well. Additionally, ICC coefficients were promising in the variance of scores being attributable to student differences, with DBR-SIS Disruptive Behavior target having similar ICC coefficients to prior studies.

ROC curve analyses showed promising screening utility with DBR-SIS. Specifically, the Depression target on DBR-SIS showed good sensitivity and specificity when evaluated against the total score of the CDI-2 elevated and very elevated thresholds. Similarly, DBR-SIS Anxiety target showed good sensitivity and specificity against the total score of RCMAS-2 on the moderately problematic threshold. Although prior research has shown that disruptive behavior is associated with mental health issues (Gould et al., 1998; Zahn-Waxler, Dougan, & Slattery, 2000), this study did not find this

to be the case when evaluating the classification accuracy of DBR-SIS Disruptive Behavior target.

In conclusion, the results of this study provide a clearer understanding of the abilities of DBR-SIS to screen for internalizing problems, and provides preliminary guidelines to cut-scores for various thresholds. The main implications of this study were that schools may benefit from utilizing DBR-SIS for internalizing problems due to promising technical adequacy in screening, as well as the possibility of low cost and time-efficient implementation. The high false positives when screening with DBR-SIS was anticipated, especially considering prior findings from research on standard targets of DBR-SIS. Thus, it may be advisable to implement a two-part multiple gating procedure to confirm at-risk status of students identified with DBR-SIS. Lastly, teachers participating in this study provided significant support for the usability of DBR-SIS as a screener for internalizing problems. Although only six teachers were part of the study, the favorable results from the URP-A regarding the use of DBR-SIS are reflective of past research findings. Although more research is necessary to fully support DBR-SIS as a screener for internalizing symptoms, the results of this study build upon the growing research focused on mental health screening.

References

Achenbach, T. M., & Edelbrock, C. S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development*, 1-82.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (2014). *Standards for educational and psychological testing*. Amer Educational Research Assn.

Anderson, E. R., & Hope, D. A. (2008). A review of the tripartite model for understanding the link between anxiety and depression in youth. *Clinical Psychology Review*, *28*(2), 275-287.

Anderson, C. M., Horner, R. H., Rodriguez, B. J., & Stiller, B. (2013). Building systems for successful implementation of function-based support in schools. *International Journal of School & Educational Psychology*, *1*(3), 141-153.

Bae, Y.H. (2012). Test review of Children's Depression Inventory 2. *Journal of Psychoeducational Assessment, 30*(3), 304-308.

Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V. *Psychiatric Clinics of North America*, *32*(3), 483-524.

Brady, E. U., & Kendall, P. C. (1992). Comorbidity of anxiety and depression in children and adolescents. *Psychological Bulletin*, *111*(2), 244-255.

Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2010). Generalizability and dependability of behavior assessment methods to estimate academic engagement: A comparison of systematic direct observation and direct behavior rating. *School Psychology Review*, *39*(3), 408-421.

Briggs-Gowan, M. J., Horwitz, S. M., Schwab-Stone, M. E., Leventhal, J. M., & Leaf, P. J. (2000). Mental health in pediatric settings: Distribution of disorders and factors related to service use. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*(7), 841-849.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry*, *36*(1), 113-149.

Caplan, G. (1964). *Principles of preventive psychology*. New York, NY: Basic Books.

Carran, D.T., & Scott, K.G. (1992). Risk assessment in preschool children: Research implications for the early detection of educational handicaps. *Topics in Early Childhood Special Education, 12*, 196-211.

Carter, A.S., Briggs-Gowan, M., & Davis, N.O. (2004). Assessment of young children's social emotional development and psychopathology: Recent advances and recommendations for practice. *Journal of Child Psychology and Psychiatry, 45*, 109-134.

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., Welsh, M., & Christ, T. J. (2013). Direct behavior rating as a school-based behavior screener for elementary and middle grades. *Journal of School Psychology*,*51*(3), 367-385.

Chafouleas, S. M., Kilgus, S. P., & Hernandez, P. (2009). Using Direct Behavior Rating (DBR) to screen for school social risk: A preliminary comparison of methods in a kindergarten sample. *Assessment for Effective Intervention*, *34*(4), 214-223.

Chafouleas, S. M., Miller, F. G., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2012). Usage Rating Profile-Assessment. *Storrs, CT: University of Connecticut*.

Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct Behavior Rating (DBR): An emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention*, *34*(3), 195-200.

Chafouleas, S., Riley-Tillman, T. C., & Sugai, G. M. (2007). *School-based behavioral assessment: Informing intervention and instruction*. New York, NY: Guilford Press.

Christ, T. J., Riley-Tillman, T. C., Chafouleas, S. M., & Boice, C. H. (2010). Direct behavior rating (DBR): Generalizability and dependability across raters and observations. *Educational and Psychological Measurement*, *70*(5), 825-843.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*(3), 316-336.

Cole, D. A., Truglio, R., & Peeke, L. (1997). Relation between symptoms of anxiety and depression in children: A multitrait-multimethod-multigroup assessment. *Journal of Consulting and Clinical Psychology*, *65*(1), 110-119.

Cole, D. A., Peeke, L. G., Martin, J. M., Truglio, R., & Seroczynski, A. D. (1998). A longitudinal look at the relation between depression and anxiety in children and adolescents. *Journal of Consulting and Clinical Psychology*,*66*(3), 451-460.

Comer, J. S., & Kendall, P. C. (2005). High-end specificity of the children's depression inventory in a sample of anxiety-disordered youth. *Depression and anxiety*, *22*(1), 11-19.

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, *60*(8), 837-844.

Council for Children with Behavioral Disorders., Cheney, D., & Council for Exceptional Children. S*tudents with emotional or behavioral disorders: Current approaches for positive outcomes*. Arlington, VA: Published by the Council for Children with Behavioral Disorders.

Cowell, J. M., Gross, D., McNaughton, D., Ailey, S., & Fogg, L. (2005). Depression and suicidal ideation among Mexican American school-aged children. *Research and theory for nursing practice*, *19*(1), 77-94.

Cowen, E. L. (1977). Psychologists and primary prevention: Blowing the cover story. *American Journal of Community Psychology*, *5*(4), 481-489.

Crowley, S. L., & Emerson, E. N. (1996). Discriminant validity of self-reported anxiety and depression in children: Negative affectivity or independent constructs? *Journal of Clinical Child Psychology*, *25*(2), 139-146.

Curtis, M.J., Castillo, J.M. & Gelley, C. (2010) School psychology 2010: Demographics, employment, and the context for professional practices. *NASP Communique*, *40*(7).

Davis, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research, 5*, 194-197.

Davis, A. S., Kruczek, T., & McIntosh, D. E. (2006). Understanding and treating psychopathology in schools: Introduction to the special issue. *Psychology in the Schools*, *43*(4), 413-417.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*(3), 219-232.

Doerfler, L. A., Felner, R. D., Rowlison, R. T., Raley, P. A., & Evans, E. (1988). Depression in children and adolescents: a comparative analysis of the utility and construct validity of two assessment measures. *Journal of Consulting and Clinical psychology*, *56*(5), 769-772.

Dowdy, E., Furlong, M., Raines, T. C., Bovery, B., Kauffman, B., Kamphaus, R. W., … & Murdock, J. (2015). Enhancing school-based mental health services with a preventive and promotive approach to universal screening for complete mental health. *Journal of Educational and Psychological Consultation*, *25*(2-3), 178-197.

Durlak, J. A., & Wells, A. M. (1997). Primary prevention mental health programs for children and adolescents: A meta-analytic review. *American Journal of Community Psychology*, *25*(2), 115-152.

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.

Eason, L.J., Finch, A.J., Brasted, W., & Saylor, C.F. (1985). The assessment of depression and anxiety in hospitalized pediatric patients. *Child Psychiatry and Human Development, 16,* 57-64.

Ennis, R. P., Lane, K. L., & Oakes, W. P. (2011). Score reliability and validity of the Student Risk Screening Scale: A psychometrically sound, feasible tool for use in urban elementary schools. *Journal of Emotional and Behavioral Disorders*, *20*(4), 241-259.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239-256.

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children*, *73*(3), 311-330.

Fuchs, D., & Fuchs, L. S. (2005). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Teaching Exceptional Children*, *38*(1), 57-61.

Galaif, E. R., Chou, C. P., Sussman, S., & Dent, C. W. (1998). Depression, suicidal ideation, and substance use among continuation high school students. *Journal of Youth and Adolescence*, *27*(3), 275-299.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting Students Struggling with Reading: Response to Intervention (RtI) for Elementary and Middle Schools. NCEE 2009-4060. *What Works Clearinghouse*.

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

Goffreda, C. T., Diperna, J. C., & Pedersen, J. A. (2009). Preventive screening for early readers: Predictive validity of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). *Psychology in the Schools*, *46*(6), 539-552.

Good III, R.H., Jefferson, Gretchen, & Shinn, M.R. (1998). *Advanced applications of curriculum-based measurement*. New York, NY: Guilford Press.

Gould, M. S., King, R., Greenwald, S., Fisher, P., Schwab-Stone, M., Kramer, R., ... & Shaffer, D. (1998). Psychopathology associated with suicidal ideation and attempts among children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*(9), 915-923.

Graden, J. L. (2004). Arguments for change to consultation, prevention, and intervention: Will school psychology ever achieve this promise? *Journal of Educational and Psychological Consultation, 15*, 345–359.

Graney, S. B., Martínez, R. S., Missall, K. N., & Aricak, O. T. (2010). Universal screening of reading in late elementary school: R-CBM versus CBM Maze. *Remedial and Special Education, 31,* 368–377.

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics*, *26*(6), 822-830.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238-247.

Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*(5), 683-687.

Hodges, K., McKnew, D., Cytryn, L., Stern, L., & Kline, J. (1982). The Child Assessment Schedule (CAS) diagnostic interview: a report on reliability and validity. *Journal of the American Academy of Child Psychiatry*, *21*(5), 468-473.

Inderbitzen-Nolan, H. M., & Walters, K. S. (2000). Social Anxiety Scale for Adolescents: Normative data and further evidence of construct validity. *Journal of Clinical Child Psychology*, *29*(3), 360-371.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review*,*36*(4), 582-600.

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice*, *24*(4), 174-185.

Johnson, A. H., Miller, F. G., Chafouleas, S. M., Welsh, M. E., Riley-Tillman, T. C., & Fabiano, G. A. (2016). Evaluating the technical adequacy of DBR-SIS in tri-annual behavioral screening: A multisite investigation. *Journal of School Psychology*, *54*, 39-57.

Kamps, J. L., Roberts, M. C., & Varela, R. E. (2005). Development of a new fear of hypoglycemia scale: preliminary results. *Journal of Pediatric Psychology*, *30*(3), 287-291.

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, *42*(4), 448-457.

Kazdin, A.E., French, N.H., Unis, A.S. (1983). Child, mother, and father evaluations of depression in psychiatric inpatient children. *Journal of Abnormal Child Psychology, 11,* 167-180.

Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & Welsh, M. E. (2012). Direct behavior rating scales as screeners: a preliminary investigation of diagnostic accuracy in elementary school. *School Psychology Quarterly*, *27*(1), 41-50.

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.

Kovacs, M. (2004). Children's depression inventory (CDI). *Toronto: Multi-Health Systems Inc*.

Lane, K. L., Parks, R. J., Kalberg, J. R., & Carter, E. W. (2007). Systematic screening at the middle school level: Score reliability and validity of the Student Risk Screening Scale. *Journal of Emotional and Behavioral Disorders*, *15*(4), 209-222.

Lane, K. L., Oakes, W. P., Ennis, R. P., Cox, M. L., Schatschneider, C., & Lambert, W. (2011). Additional evidence for the reliability and validity of the Student Risk Screening Scale at the high school level: A replication and extension. *Journal of Emotional and Behavioral Disorders*, 1063426611407339.

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.

Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1994). Psychosocial risk factors for future adolescent suicide attempts. *Journal of Consulting and Clinical Psychology*, *62*(2), 297-305.

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382-386.

McIntosh, K., Ty, S.V., & Miller, L.D. (2013). Effects of school-wide positive behavior support on internalizing problems: Current evidence and future directions. *Journal of Positive Behavior Interventions*.

McKenzie, J. F., Wood, M. L., Kotecki, J. E., Clark, J. K., & Brey, R. A. (1999). Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health Behavior*, *23*(4), 311-318.

Meisels, S.J. (1989). Can developmental screening tests identify children who are developmentally at risk? *Pediatrics, 83*, 578-585.

Merrell, K.W. (2000). Informant report: Rating scale measures. In E.S. Shapiro & T.R. Kratochwill (Eds.), *Conducting school-based assessment of child and adolescent behaviors* (pp. 203-234). New York, NY: Guilford.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11.

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), 184-196.

Muris, P., Merckelbach, H., Mayer, B., van Brakel, A., Thissen, S., Moulaert, V., & Gadet, B. (1998). The screen for child anxiety related emotional disorders (SCARED) and traditional childhood anxiety measures. *Journal of Behavior Therapy and Experimental Psychiatry*, *29*(4), 327-339.

National Association of School Psychologists (2010). National association of school psychologists model for comprehensive and integrated school psychological services. *School Psychology Review, 39*(2), 320-333.

O'Connell, M. E., Boat, T., & Warner, K. E. (Eds.). (2009). *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Washington D.C.: National Academies Press.

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

Reynolds, C.R., & Kamphaus, R.W. (2007). *Behavior assessment system for children* (2nd Ed.) Bloomington, MN: Pearson.

Reynolds, C. R., & Richmond, B. O. (1985). *Revised children's manifest anxiety scale*. Los Angeles: Western Psychological Services.

Reynolds, C.R., & Kamphaus, R.W. (2015). *Behavior assessment system for children* (3rd Ed.) Bloomington, MN: Pearson.

Riley-Tillman, T. C., Chafouleas, S. M., Sassu, K. A., Chanese, J. A., & Glazer, A. D. (2008). Examining the agreement of direct behavior ratings and systematic direct observation data for on-task and disruptive behavior. *Journal of Positive Behavior Interventions*, *10*(2), 136-143.

Roberts, A. R., & Ottens, A. J. (2005). The seven-stage crisis intervention model: A road map to goal attainment, problem solving, and crisis resolution. *Brief Treatment and Crisis Intervention*, *5*(4), 329-335.

Rohrer, D. (2015). *Developing an integrated screening and progress monitoring tool for internalizing problems in middle school students* (Unpublished doctoral dissertation). University of Missouri-Columbia, Missouri.

Rones, M., & Hoagwood, K. (2000). School-based mental health services: A research review. *Clinical Child and Family Psychology Review*, *3*(4), 223-241.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94-104.

Safari, S., Baratloo, A., Elfil, M., & Negida, A. (2016). Evidence based emergency medicine; part 4: pre-test and post-test probabilities and Fagan's nomogram. *Emergency*, *4*(1), 48-51.

Saluja, G., Iachan, R., Scheidt, P. C., Overpeck, M. D., Sun, W., & Giedd, J. N. (2004). Prevalence of and risk factors for depressive symptoms among young adolescents. *Archives of Pediatrics & Adolescent Medicine*, *158*(8), 760-765.

SAMHSA. (2015). Results from the 2014 National Survey on Drug Use and Health: Summary of National Findings. pp 1-6. Retrieved from: http://www.samhsa.gov/data/NSDUH.

Shain, B. N., Naylor, M., & Alessi, N. (1990). Comparison of self-rated and clinician-rated measures of depression in adolescents. *The American Journal of Psychiatry*, *147*(6), 793-795.

Shapiro, E. S., & Clemens, N. H. (2009). A conceptual model for evaluating system effects of response to intervention. *Assessment for Effective Intervention*, *35*(1), 3-16.

Shapiro, E. S. (2011). *Academic skills problems: Direct assessment and intervention*. New York, NY: Guilford Press.

Simon, E., Dirksen, C. D., & Bögels, S. M. (2013). An explorative cost-effectiveness analysis of school-based screening for child anxiety using a decision analytic model. *European Child & Adolescent Psychiatry*, *22*(10), 619-630.

Stiffler, M.C., & Dever, B.V. (2015). *Mental health screening at school*. Switzerland: Springer International Publishing.

Stoiber, K. C., & Vanderwood, M. L. (2008). Traditional Assessment, Consultation, and Intervention Practices: Urban School Psychologists' Use, Importance, and Competence Ratings. *Journal of Educational & Psychological Consultation, 18*, 264-292.

Streiner, D.L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristic curves. T*he Canadian Journal of Psychiatry, 52*, 121-128.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., ... & Turnbull, H. R. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions*, *2*(3), 131-143.

Sullivan, A. L., & Long, L. (2010). Examining the changing landscape of school psychology practice: A survey of school-based practitioners regarding response to intervention. *Psychology in the Schools, 47*(10), 1059-1070.

Swanson, J. W., Linskey, A. O., Quintero-Salinas, R., Pumariega, A. J., & Holzer, C. E. (1992). A binational school survey of depressive symptoms, drug use, and suicidal ideation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *31*(4), 669-678.

Timbremont, B., Braet, C., & Dreessen, L. (2004). Assessing depression in youth: relation between the Children's Depression Inventory and a structured interview. *Journal of Clinical Child and Adolescent Psychology*, *33*(1), 149-157.

Vincent, C., Spaulding, S., & Tobin, T. J. (2009). A reexamination of the psychometric properties of the school-wide evaluation tool (SET). *Journal of Positive Behavior Interventions, 12*(3), 161-179.

von der Embse, N. P., Scott, E. C., & Kilgus, S. P. (2015). Sensitivity to change and concurrent validity of direct behavior ratings for academic anxiety. *School Psychology Quarterly*, *30*(2), 244-259.

Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmark, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice*, *26*(1), 54-60.

Weisz, J. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*. New York, NY: Cambridge University Press.

Whitcomb, S., & Merrell, K. W. (2013). *Behavioral, social, and emotional assessment of children and adolescents*. New York, NY: Routledge.

Williford, A. P., & Shelton, T. L. (2008). Using mental health consultation to decrease disruptive behaviors in preschoolers: adapting an empirically-supported intervention. *Journal of Child Psychology and Psychiatry*, *49*(2), 191-200.

Zahn–Waxler, C., Klimes–Dougan, B., & Slattery, M. J. (2000). Internalizing problems of childhood and adolescence: Prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Development and psychopathology*, *12*(3), 443-466.

Table 1

Descriptive Statistics for DBR-SIS, CDI-2, and RCMAS-2

	6th Grade (n = 14)		7th Grade (n = 48)		8th grade (n = 49)	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
DBR Total	.93	1.23	1.53	1.41	2.01	1.96
DBR Anxiety	.18	.44	.55	.65	.45	.55
DBR Depression	.24	.45	.55	.74	.42	.50
DBR Disruptive	.51	.85	.43	.75	1.14	1.50
CDI-2 Total	58.00	12.25	57.81	12.21	59.92	13.22
CDI-2 Emotional	56.93	10.72	56.54	14.09	58.55	13.50
CDI-2 Functional	57.57	13.21	56.81	11.89	59.16	12.40
RCMAS-2 Total	48.50	8.15	52.10	11.29	51.63	10.78
RCMAS-2 Physiological	45.50	9.72	49.94	11.31	48.51	9.43
RCMAS-2 Worry	50.79	8.80	53.10	11.94	53.49	9.33
RCMAS-2 Social	48.86	9.03	50.67	11.38	51.49	10.87
RCMAS-2 Defensive	56.29	7.72	50.56	9.95	50.86	9.02

*CDI-2 and RCMAS-2 reported in T-scores

Table 2

Percentage of students at risk CDI-2 T-scores

	>64 (elevated)	*>69 (very elevated)*
Total	29.7%	18.9%
Emotional	29.7%	16.2%
Functional	28.8%	14.4%

Table 3

Percentage of students at risk RCMAS-2 T-scores

	>60 (moderately problematic)	*>70 (extremely problematic)*
Total	23.4%	.90%
Physiological	17.1%	1.8%
Worry	27.9%	3.6%
Social	20.7%	3.6%
Defensiveness	15.3%	0%

Table 4

Correlations

	1	2	3	4	5	6	7	8	9	10	11	12
1 DBR-Anx	--											
2 DBR-Dep	.395**	--										
3 DBR-DB	.141	.138	--									
4 DBR-Tot	.587**	.597**	.812**	--								
5 CDI Tot	.330**	.674**	.240*	.530**	--							
6 CDI Emo	.355**	.624**	.249**	.527**	.904**	--						
7 CDI Fun	.263**	.607**	.185	.442**	.927**	.714**	--					
8 RCMAS Tot	.509**	.432**	.186	.462**	.652**	.633**	.572**	--				
9 RCMAS Phy	.491**	.413**	.087	.374**	.630**	.618**	.533**	.764**	--			
10 RCMAS Wor	.454**	.329**	.236*	.442**	.501**	.480**	.441**	.924**	.548**	--		
11 RCMAS Soc	.382**	.386**	.154	.382**	.599**	.586**	.550**	.861**	.510**	.774**	--	
12 RCMAS Def	-.098	-.311**	-.199*	-.281**	-.500**	-.486**	-.440**	-.402**	0.517**	-.275**	-.298**	--

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

Table 5

Area Under the Curve for CDI-2 Total

Test	AUC (95% C.I>) Elevated T-score >64	AUC (95% C.I.) Very Elevated T-score >69
DBR-Depression	.835 (.743 - .926)**	.921 (.870 - .973)**
DBR-Anxiety	.659 (.543 - .776)**	.703 (.580 - .827)**
DBR-Disruptive	.646 (.533 – 759)*	.639 (.508 - .771)*
DBR-Total	.794 (.705 - .883)**	.848 (.771 - .924)**

** Asymptotic significance* < .05. *** Asymptotic significance* < .01

Table 6

Area Under ROC Curves for CDI-2 Emotional

Test	AUC (95% C.I>) Elevated T-score >64	AUC (95% C.I.) Very Elevated T-score >69
DBR-Depression	.870 (.788 - .953)**	.913 (.854 - .972)**
DBR-Anxiety	.726 (.620 - .832)**	.673 (.529 - .817)*
DBR-Disruptive	.604 (.487 - .721)	.657 (.514 - .799)*
DBR-Total	.817 (.734 - .900)**	.855 (.768 - .942)**

* *Asymptotic significance* < .05. ** *Asymptotic significance* < .01

Table 7

Area Under ROC Curves for CDI-2 Functional

Test	AUC (95% C.I>) Elevated T-score >64	AUC (95% C.I.) Very Elevated T-score >69
DBR-Depression	.800 (.695 - .905)**	.905 (.846 - .964)**
DBR-Anxiety	.670 (.555 - .785)**	.644 (.497 - .791)
DBR-Disruptive	.611 (.494 - .727)	.661 (.516 - .806)*
DBR-Total	.761 (.660 - .862)**	.829 (.738 - .920)**

* *Asymptotic significance* < .05. ** *Asymptotic significance* < .01

Table 8

Area Under ROC Curves for RCMAS-2 Total

Test	AUC (95% C.I.) Moderately Problematic T-score > 60
DBR-Depression	.708 (.584 - .833)**
DBR-Anxiety	.874 (.791 - .958)**
DBR-Disruptive	.688 (.578 - .799)**
DBR-Total	.813 (.723 - .903)**

* *Asymptotic significance* < .05. ** *Asymptotic significance* < .01

Table 9

Area Under ROC Curves for RCMAS-2 Physiological

Test	AUC (95% C.I.) Moderately Problematic T-score > 60
DBR-Depression	.763 (.626 - .900)**
DBR-Anxiety	.706 (.563 - .850)**
DBR-Disruptive	.534 (.394 - .675)
DBR-Total	.722 (.597 - .847)**

* *Asymptotic significance* < .05. ** *Asymptotic significance* < .01

Table 10

Area Under ROC Curves for RCMAS-2 Worry

Test	AUC (95% C.I.) Moderately Problematic T-score > 60
DBR-Depression	.728 (.613 - .842)**
DBR-Anxiety	.799 (.695 - .904)**
DBR-Disruptive	.688 (.576 - .800)**
DBR-Total	.811 (.718 - .905)**

** Asymptotic significance* < .05. *** Asymptotic significance* < .01

Table 11

Area Under ROC Curves for RCMAS-2 Social

Test	AUC (95% C.I.) Moderately Problematic T-score > 60
DBR-Depression	.738 (.618 - .858)**
DBR-Anxiety	.759 (.644 - .874)**
DBR-Disruptive	.643 (.526 - .760)*
DBR-Total	.736 (.623 - .849)**

** Asymptotic significance* < .05. *** Asymptotic significance* < .01

Table 12

Area Under ROC Curves for RCMAS-2 Defensiveness

Test	AUC (95% C.I.) Moderately Problematic T-score > 60
DBR-Depression	.657 (.535 - .780)*
DBR-Anxiety	.570 (.420 - .720)
DBR-Disruptive	.746 (.638 - .853)**
DBR-Total	.737 (.624 - .851)**

** Asymptotic significance* < .05. *** Asymptotic significance* < .01

Table 13

DBR classification indices and cut scores for CDI-2 Total (elevated)

	Cut score	SN	SP	PPV	NPV
DBR-Depression	.24	.82	.73	.56	.90
DBR-Anxiety	.23	.61	.58	.38	.78
DBR-Disruptive	.23	.67	.64	.43	.81
DBR- Total	1.7	.73	.72	.52	.85

Table 14

DBR classification indices and cut scores for CDI-2 Total (very elevated)

	Cut score	SN	SP	PPV	NPV
DBR-Depression	.65	.86	.85	.58	.96
DBR-Anxiety	.05	.81	.52	.28	.92
DBR-Disruptive	.47	.67	.64	.30	.89
DBR- Total	2.37	.81	.80	.49	.95

Table 15

DBR classification indices and cut scores for RCMAS-2 Total (T-score >60)

	Cut score	SN	SP	PPV	NPV
DBR-Depression	.41	.72	.71	.43	.89
DBR-Anxiety	.44	.89	.82	.61	.96
DBR-Disruptive	.18	.81	.60	.38	.91
DBR- Total	1.76	.81	.74	.49	.93

Figure 1

CDI-2 Total Distribution

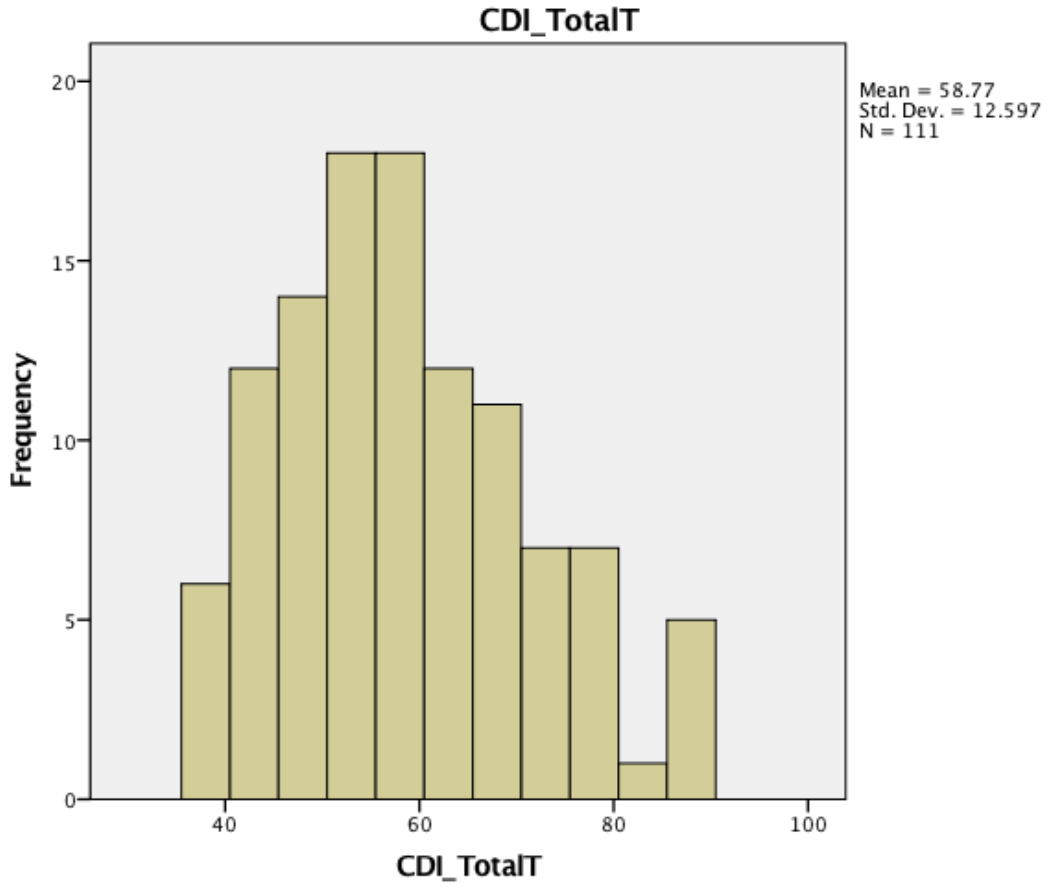


Figure 2

RCMAS-2 Distribution

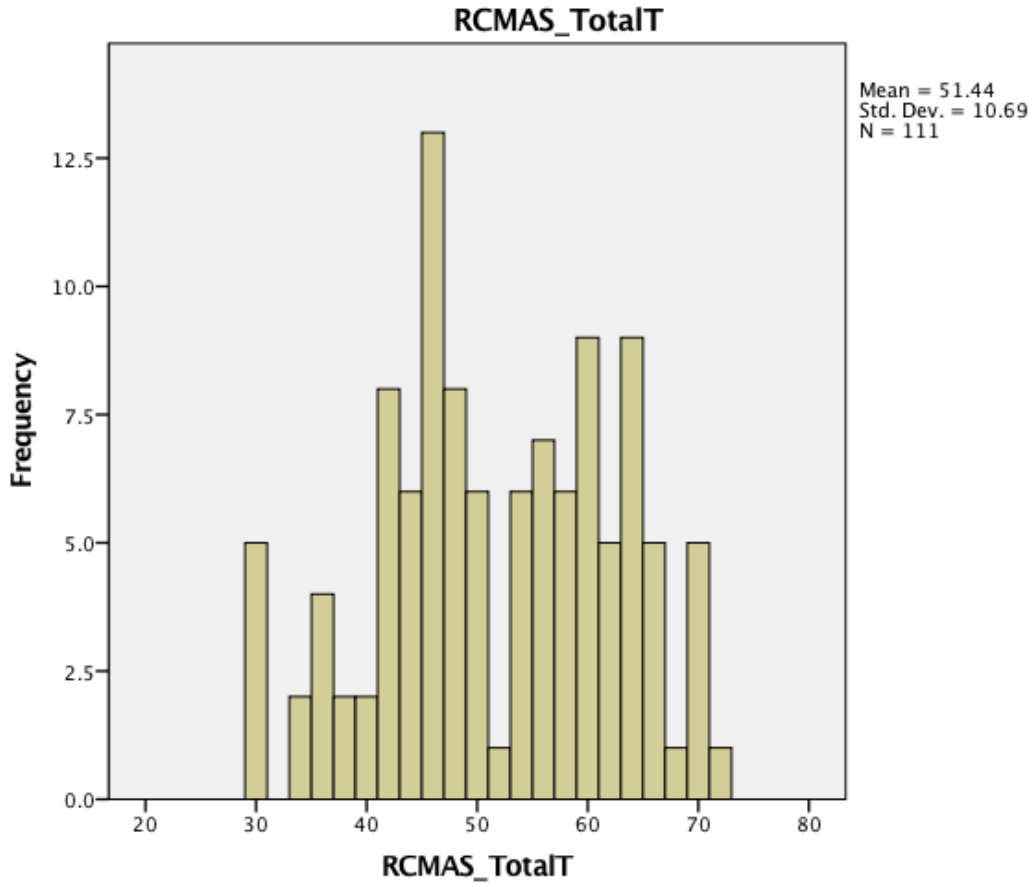
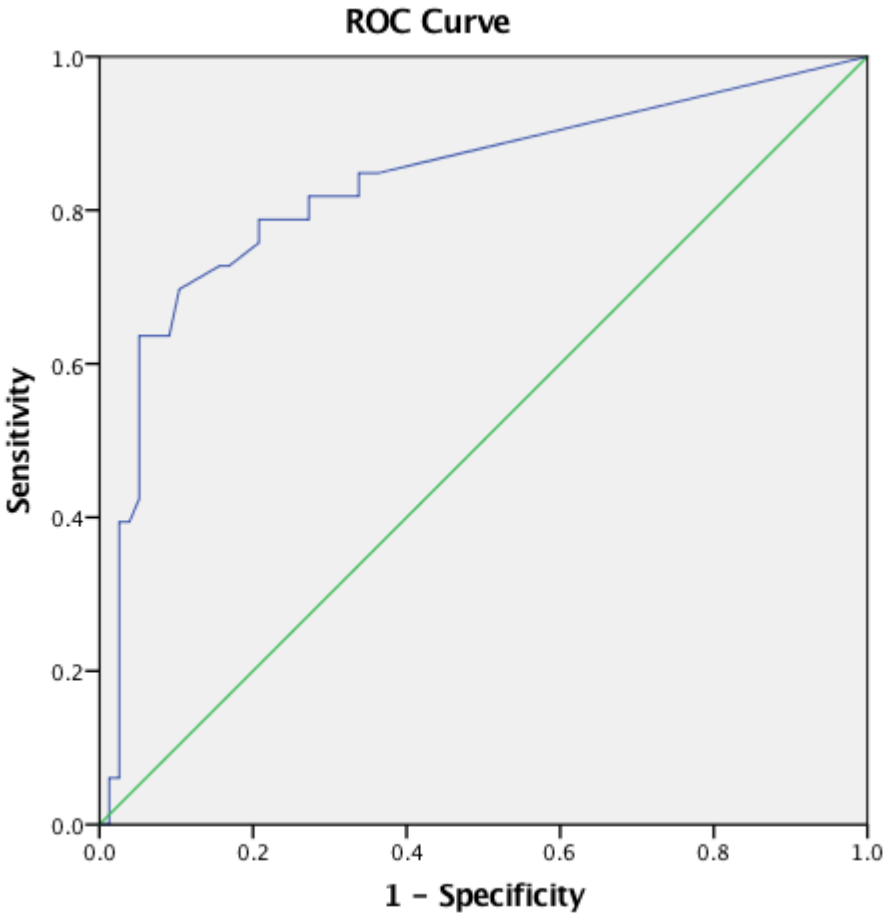


Figure 3

DBR-SIS Depression ROC curve for CDI-2 Total Elevated

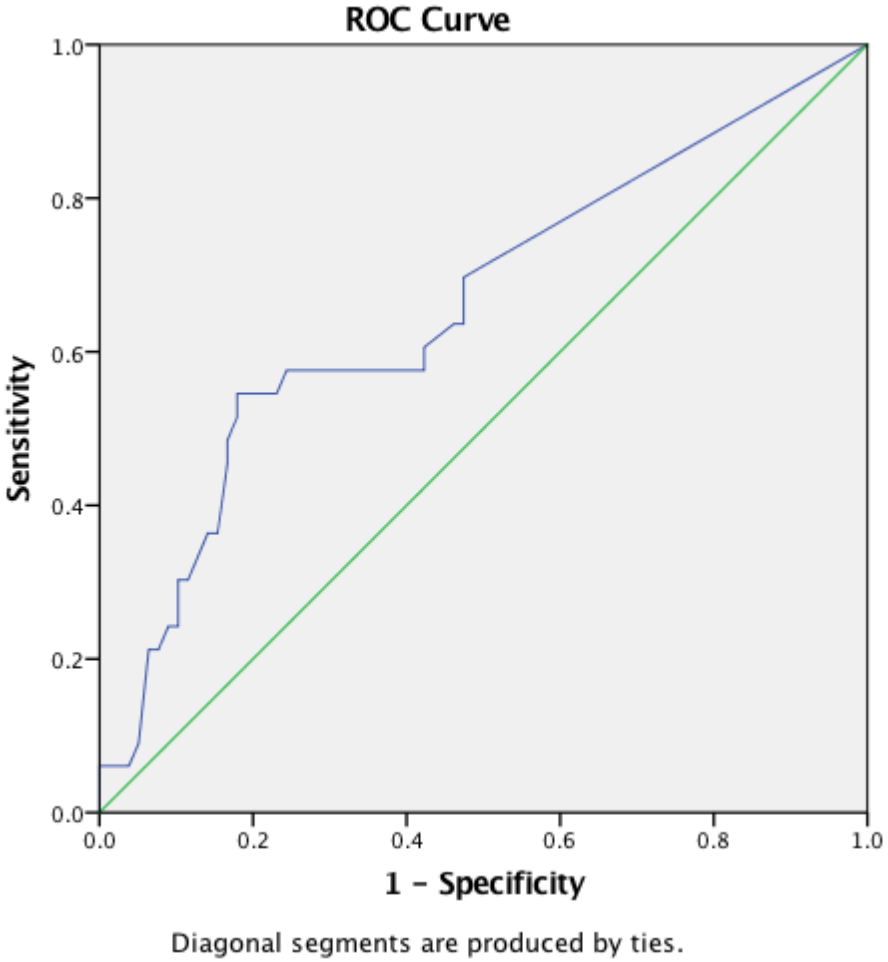


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.835	.047	.000	.743	.926

Figure 4

DBR-SIS Anxiety ROC curve for CDI-2 Total Elevated

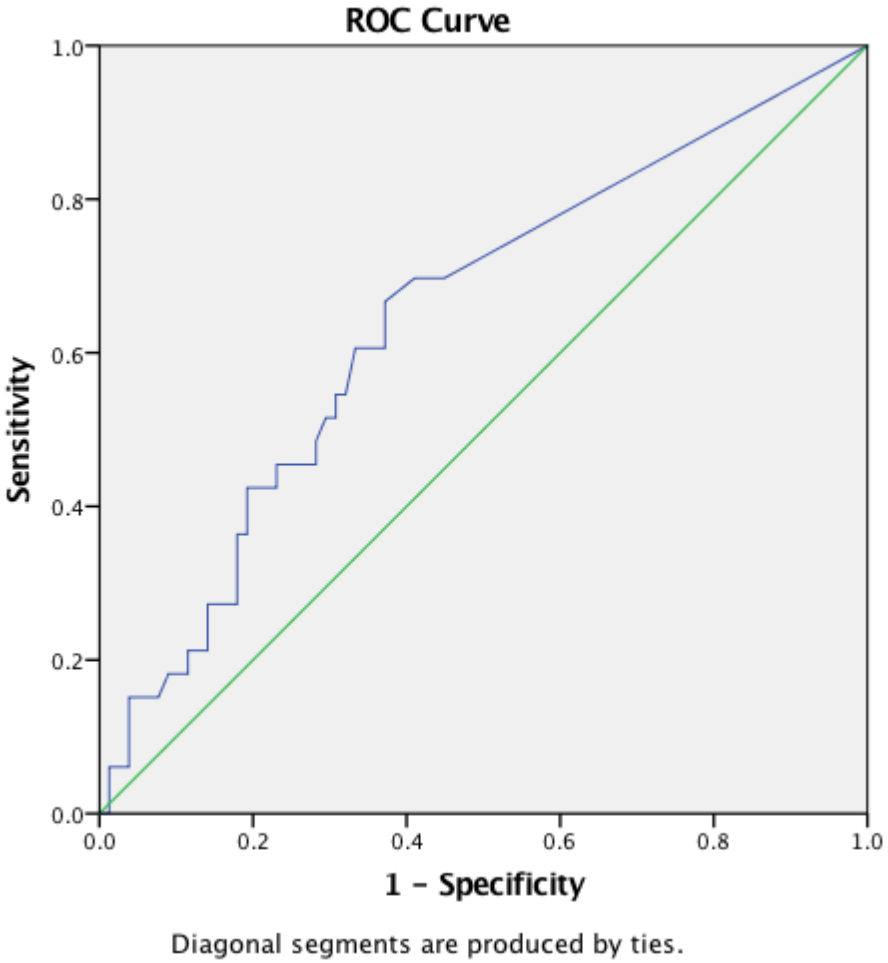


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.659	.059	.008	.543	.776

Figure 5

DBR-SIS Disruptive Behavior ROC curve for CDI-2 Total Elevated

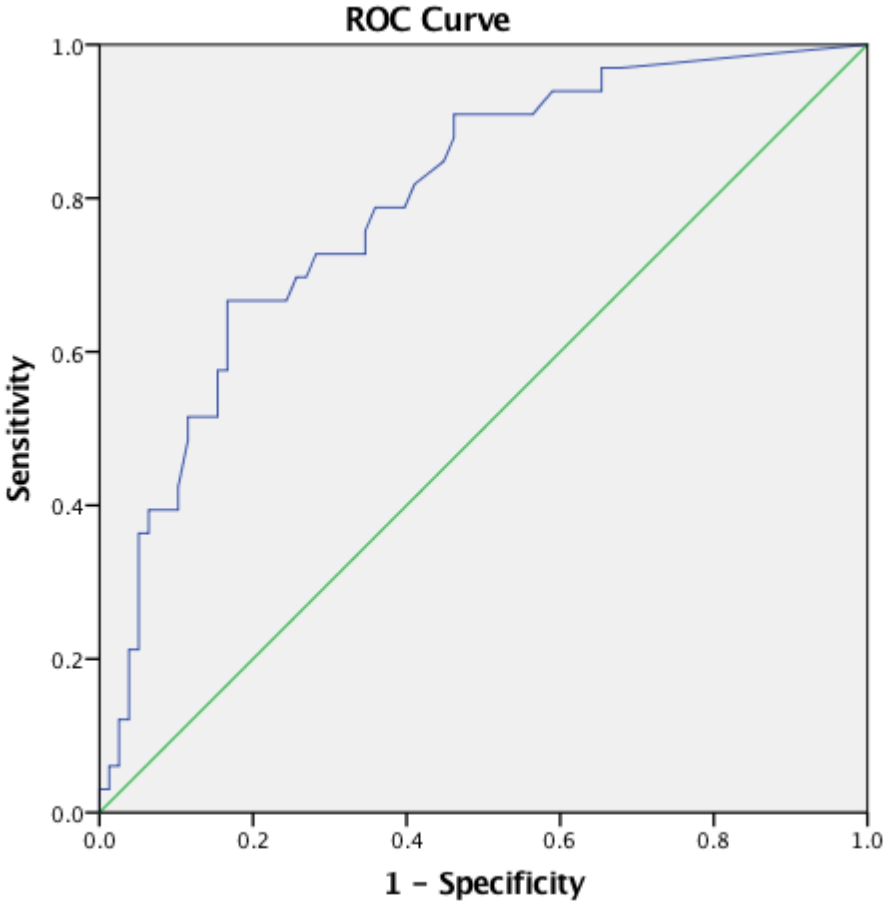


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.646	.058	.015	.533	.759

Figure 6

DBR-SIS Total ROC curve for CDI-2 Total Elevated

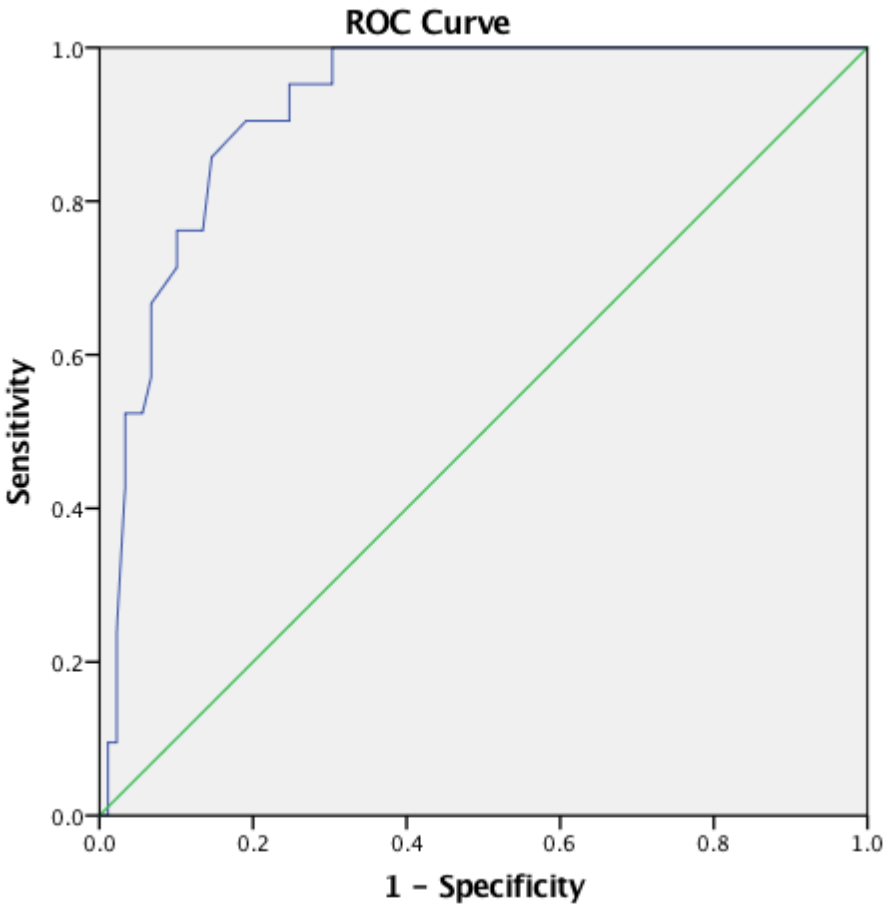


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.794	.045	.000	.705	.883

Figure 7

DBR-SIS Depression ROC curve for CDI-2 Total Very Elevated

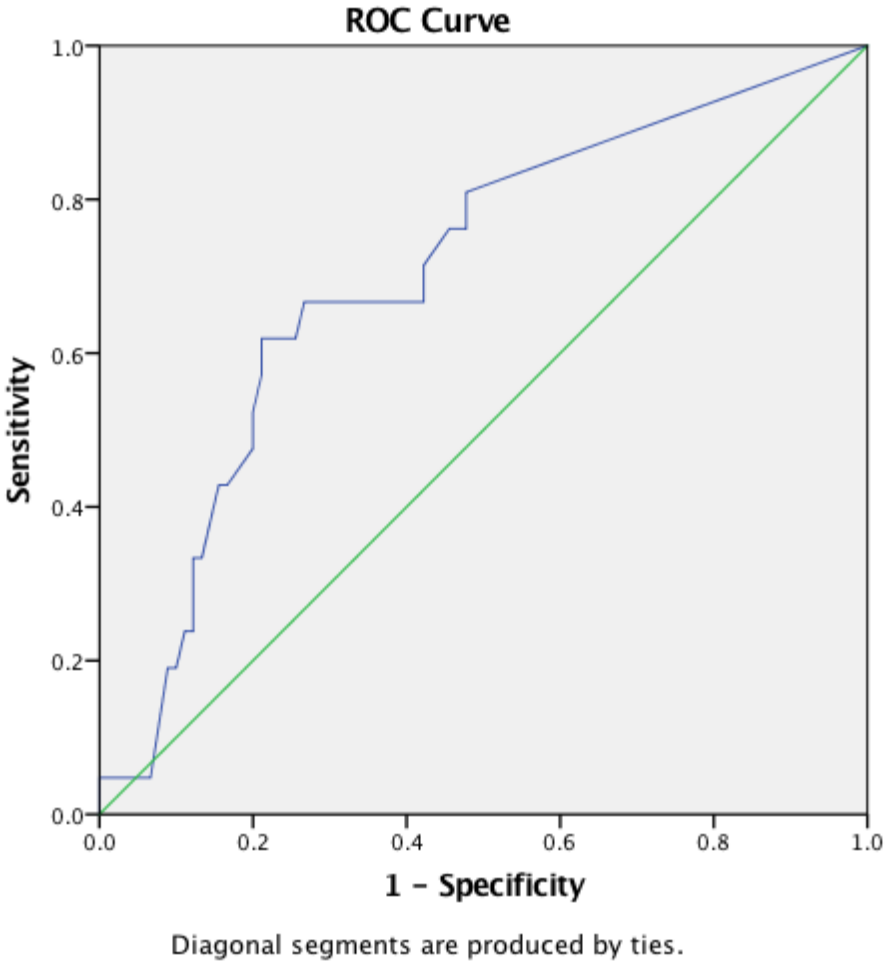


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.921	.026	.000	.870	.973

Figure 8

DBR-SIS Anxiety ROC curve for CDI-2 Total Very Elevated

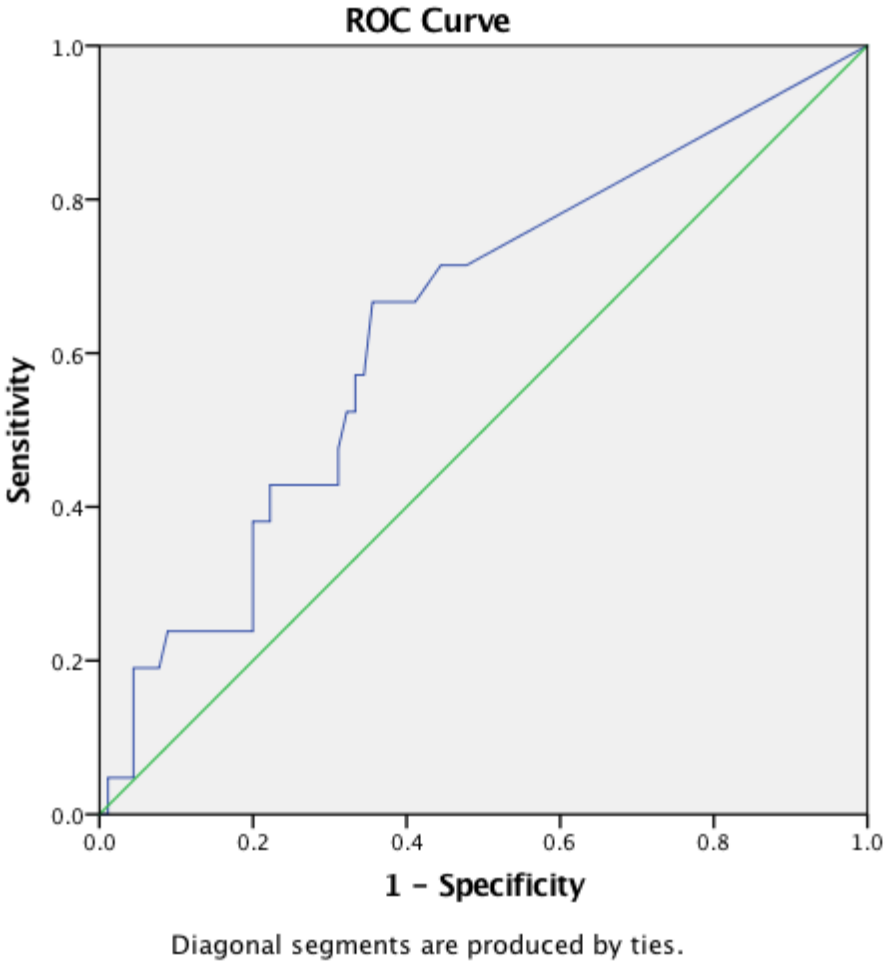


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.703	.063	.004	.580	.827

Figure 9

DBR-SIS Disruptive Behavior ROC curve for CDI-2 Total Very Elevated

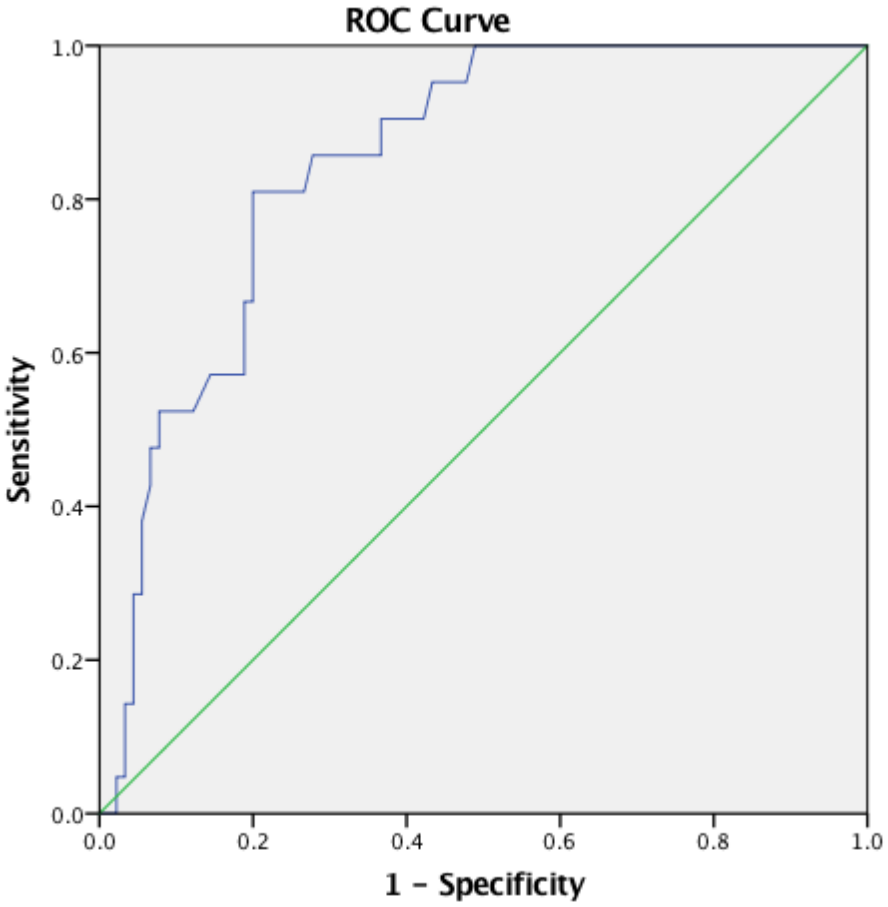


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.639	.067	.047	.508	.771

Figure 10

DBR-SIS Total ROC curve for CDI-2 Total Very Elevated

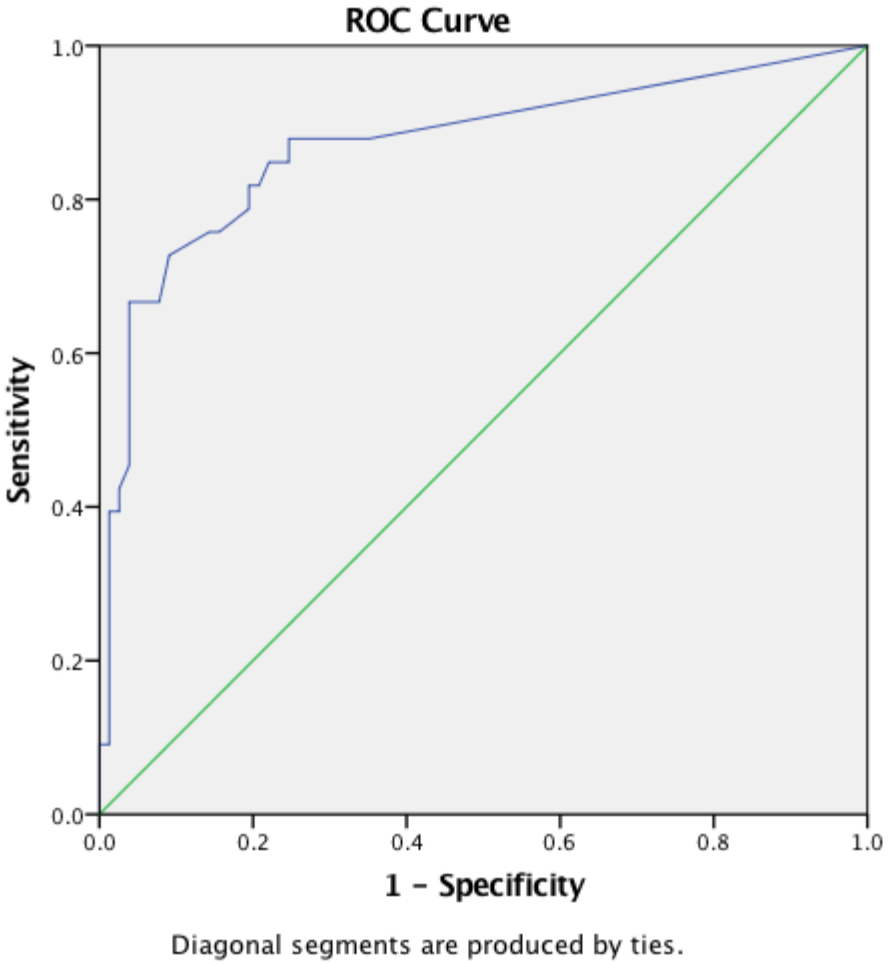


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.848	.039	.000	.771	.924

Figure 11

DBR-SIS Depression ROC curve for CDI-2 Emotional Elevated

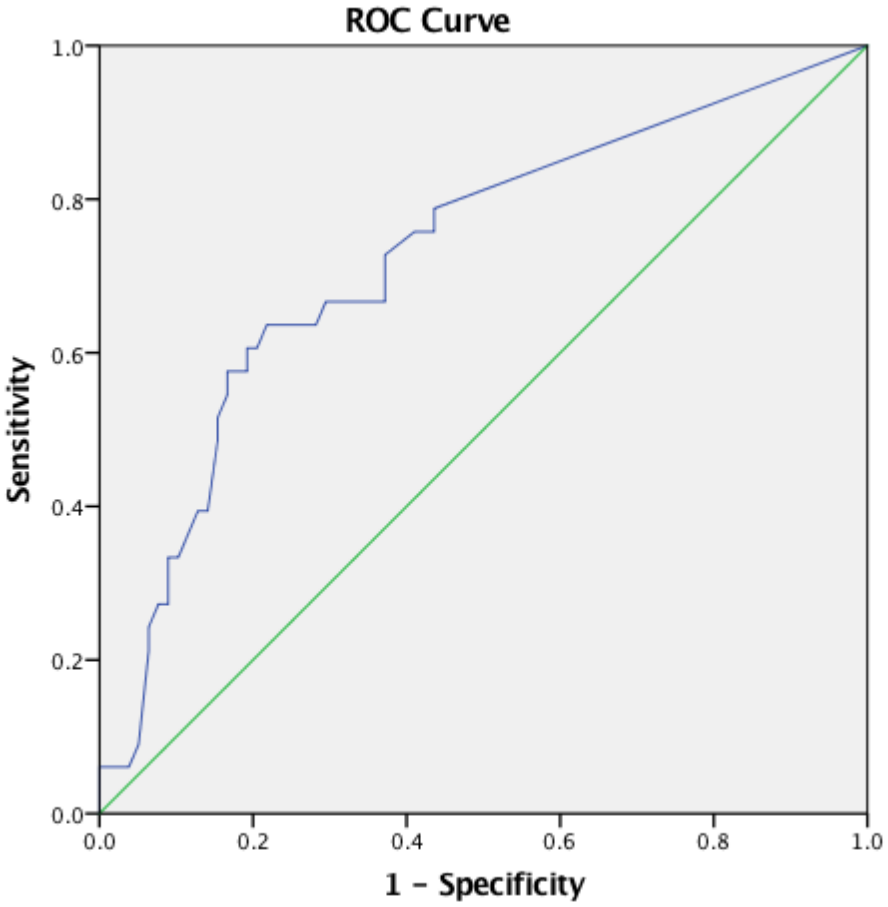


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.870	.042	.000	.788	.953

Figure 12

DBR-SIS Anxiety ROC curve for CDI-2 Emotional Elevated

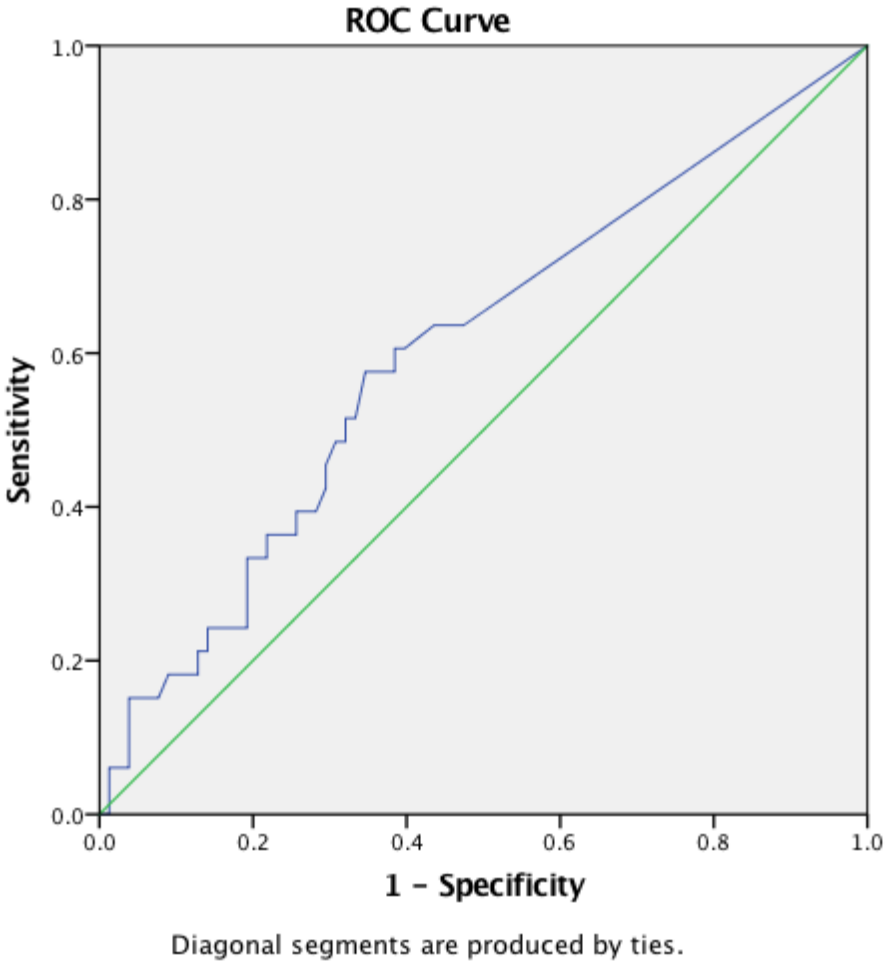


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.726	.054	.000	.620	.832

Figure 13

DBR-SIS Disruptive Behavior ROC curve for CDI-2 Emotional Elevated

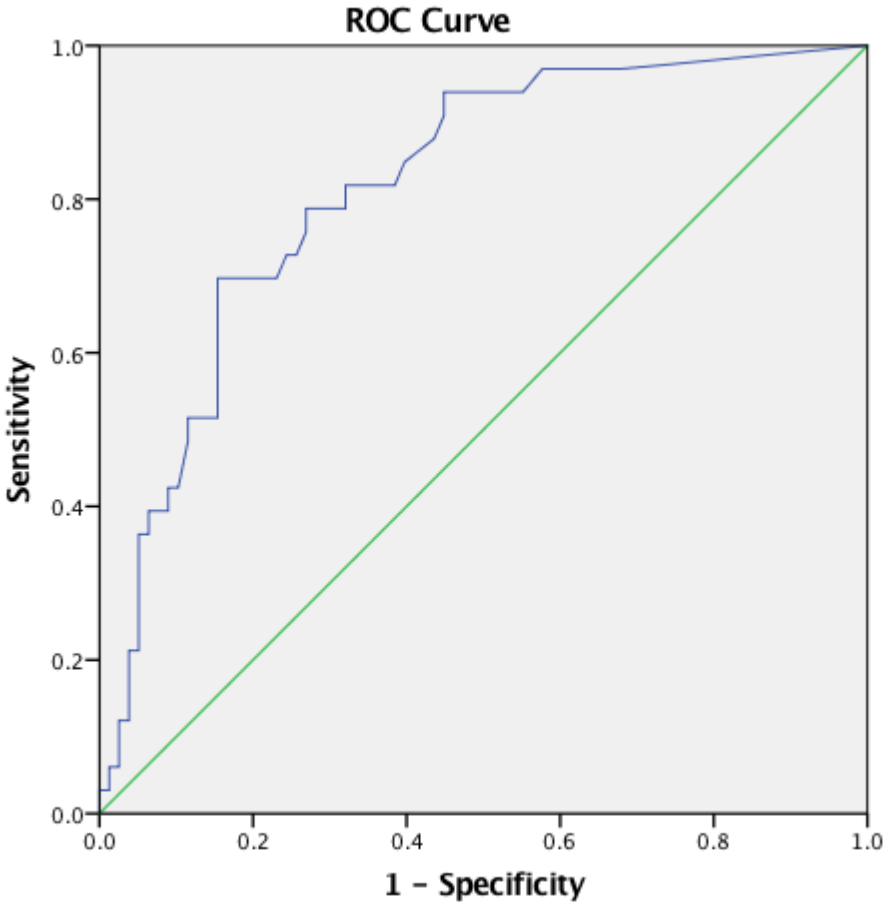


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.604	.060	.085	.487	.721

Figure 14

DBR-SIS Total ROC curve for CDI-2 Emotional Elevated

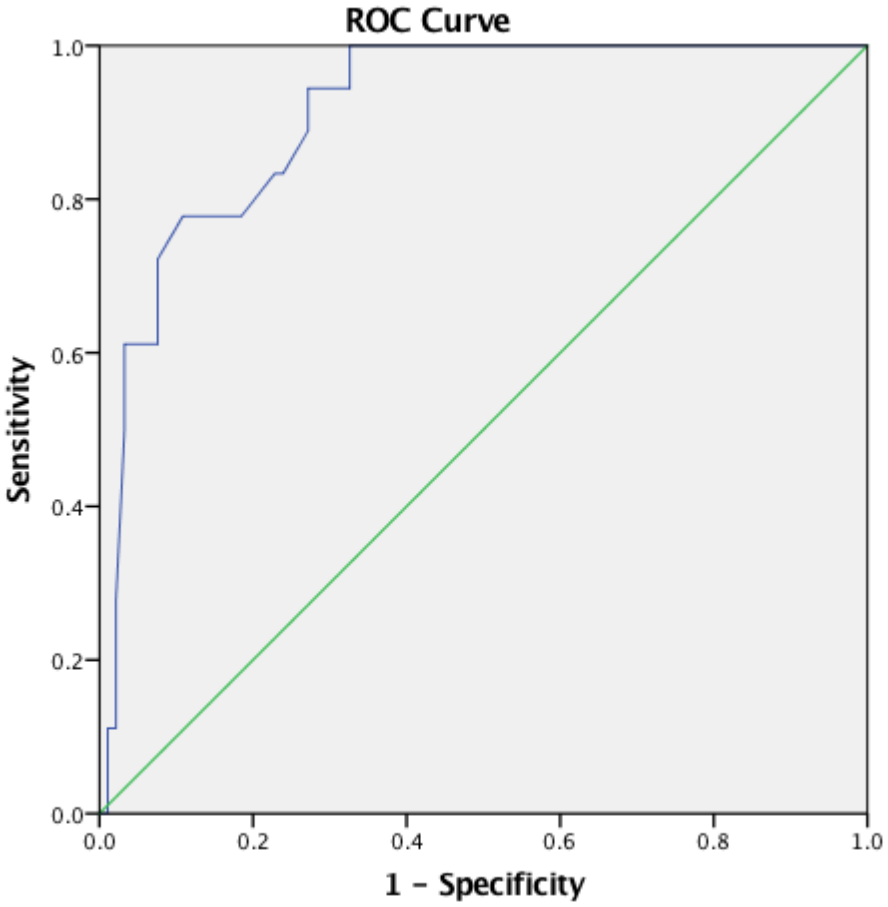


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.817	.042	.000	.734	.900

Figure 15

DBR-SIS Depression ROC curve for CDI-2 Emotional Very Elevated

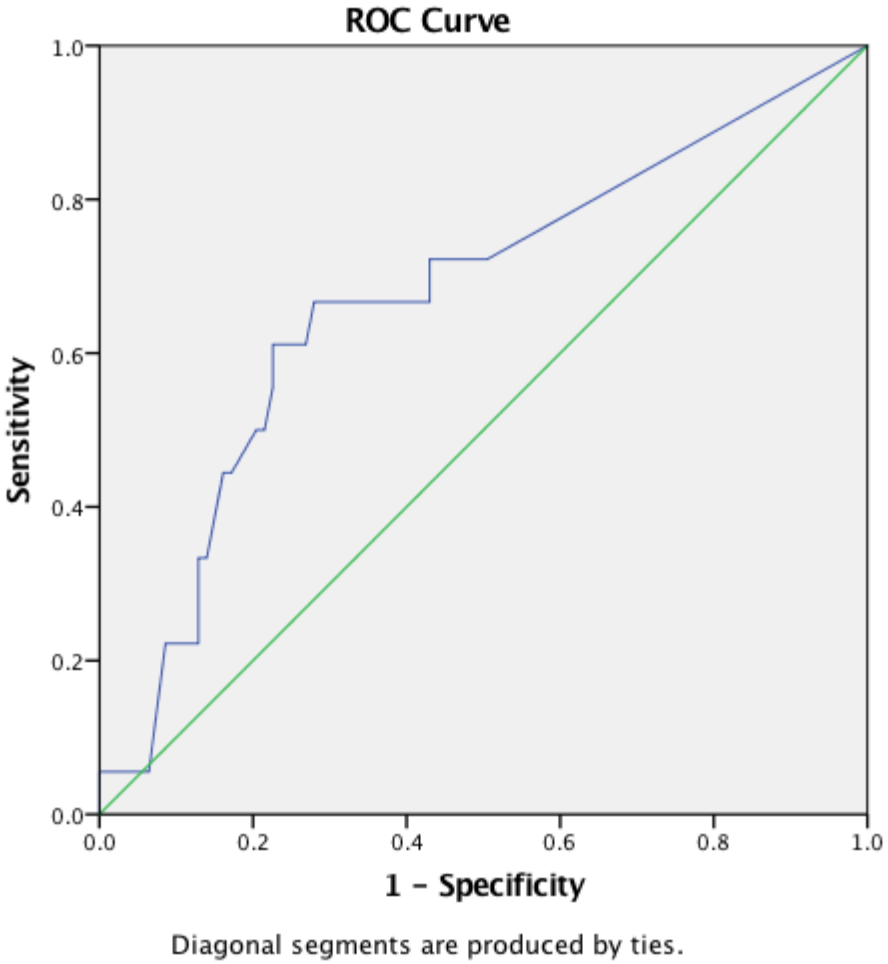


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.913	.030	.000	.854	.972

Figure 16

DBR-SIS Anxiety ROC curve for CDI-2 Emotional Very Elevated

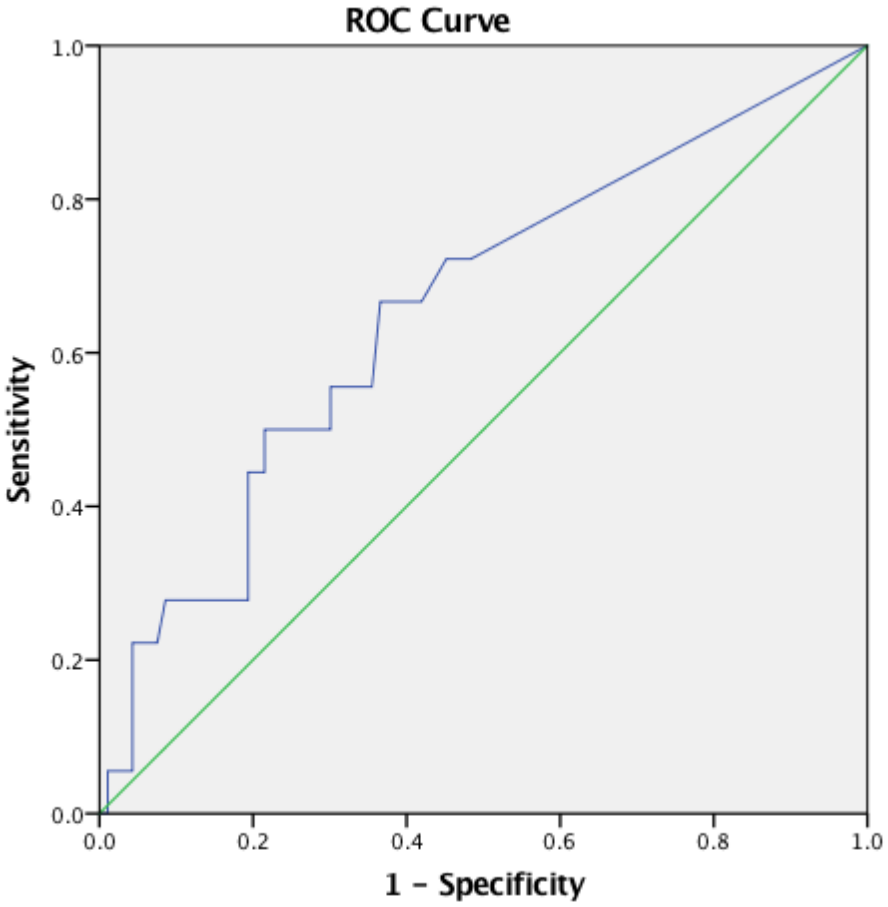


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.673	.073	.021	.529	.817

Figure 17

DBR-SIS Disruptive Behavior ROC curve for CDI-2 Emotional Very Elevated

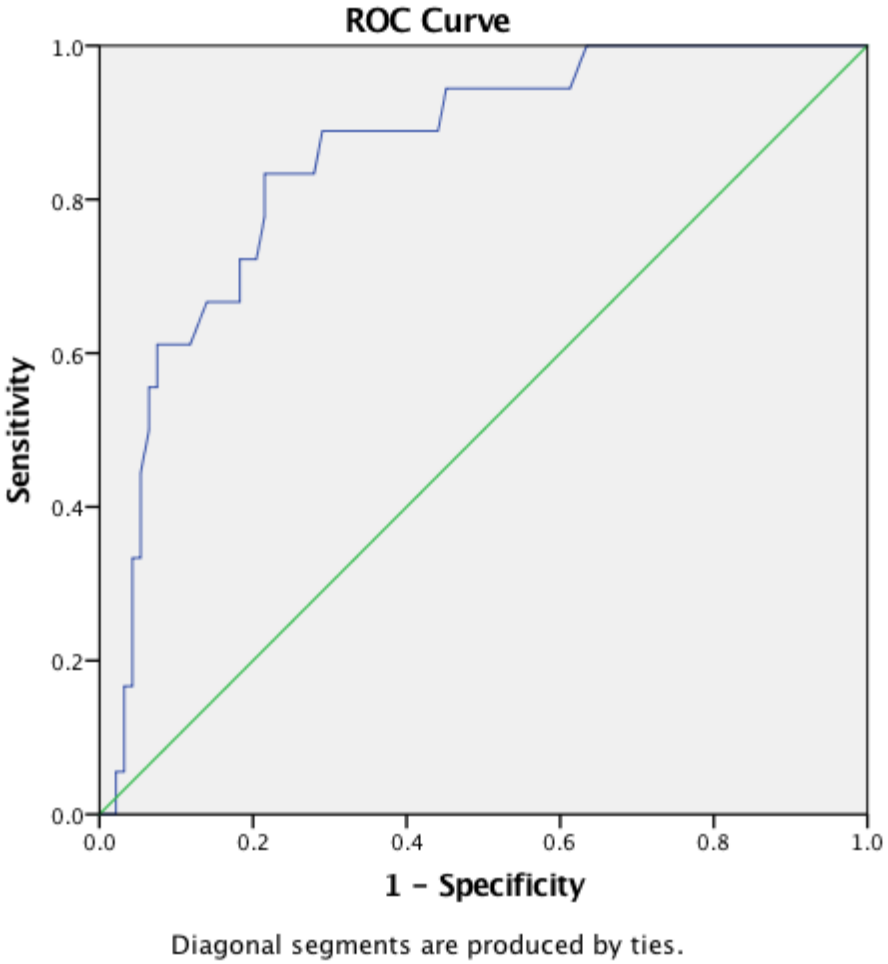


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.657	.073	.036	.514	.799

Figure 18

DBR-SIS Total ROC curve for CDI-2 Emotional Very Elevated

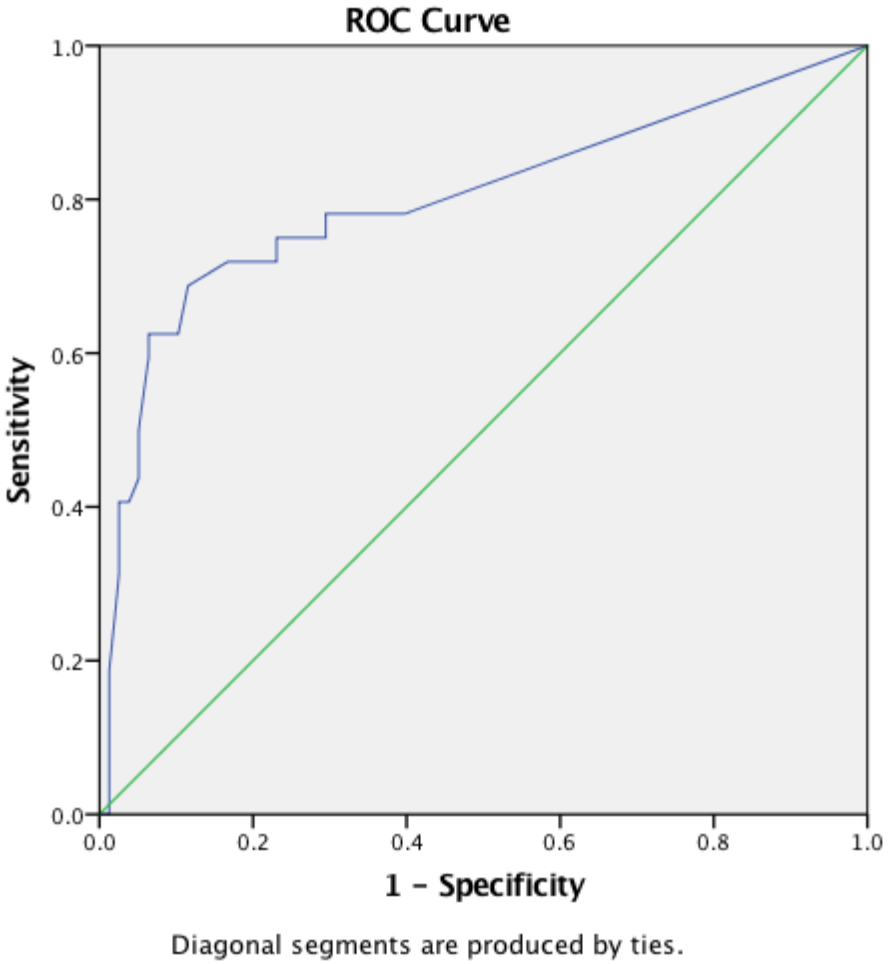


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.855	.044	.000	.768	.942

Figure 19

DBR-SIS Depression ROC curve for CDI-2 Functional Elevated

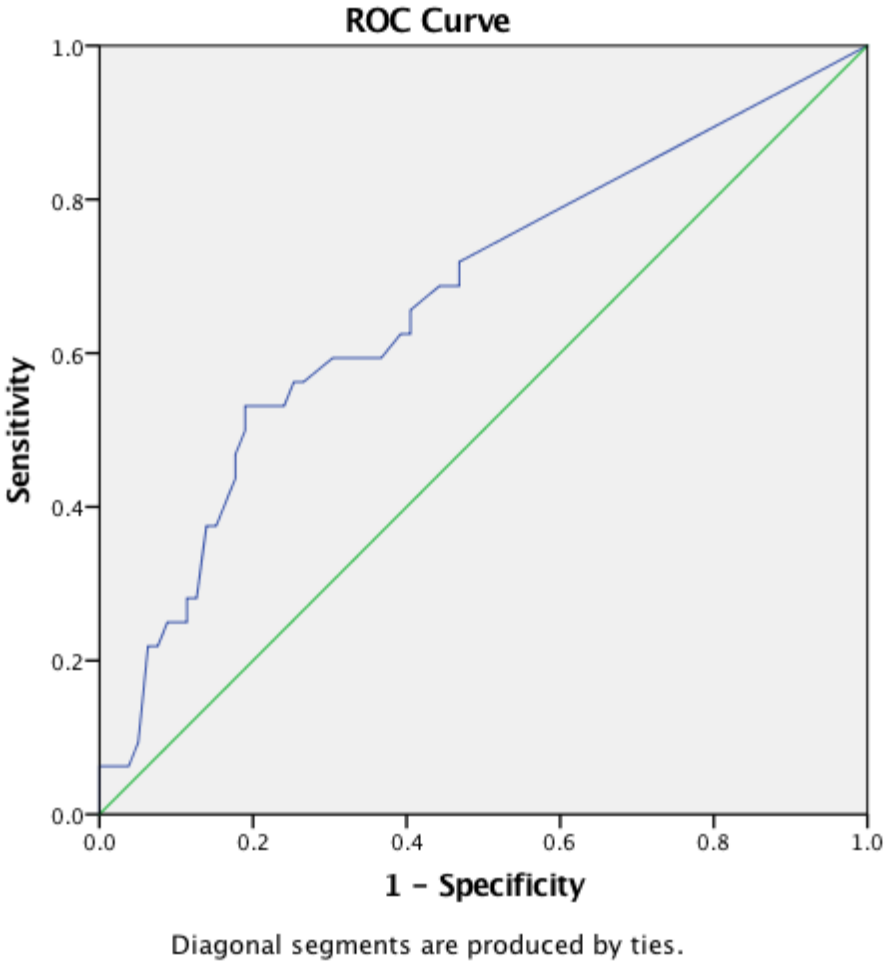


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.800	.054	.000	.695	.905

Figure 20

DBR-SIS Anxiety ROC curve for CDI-2 Functional Elevated

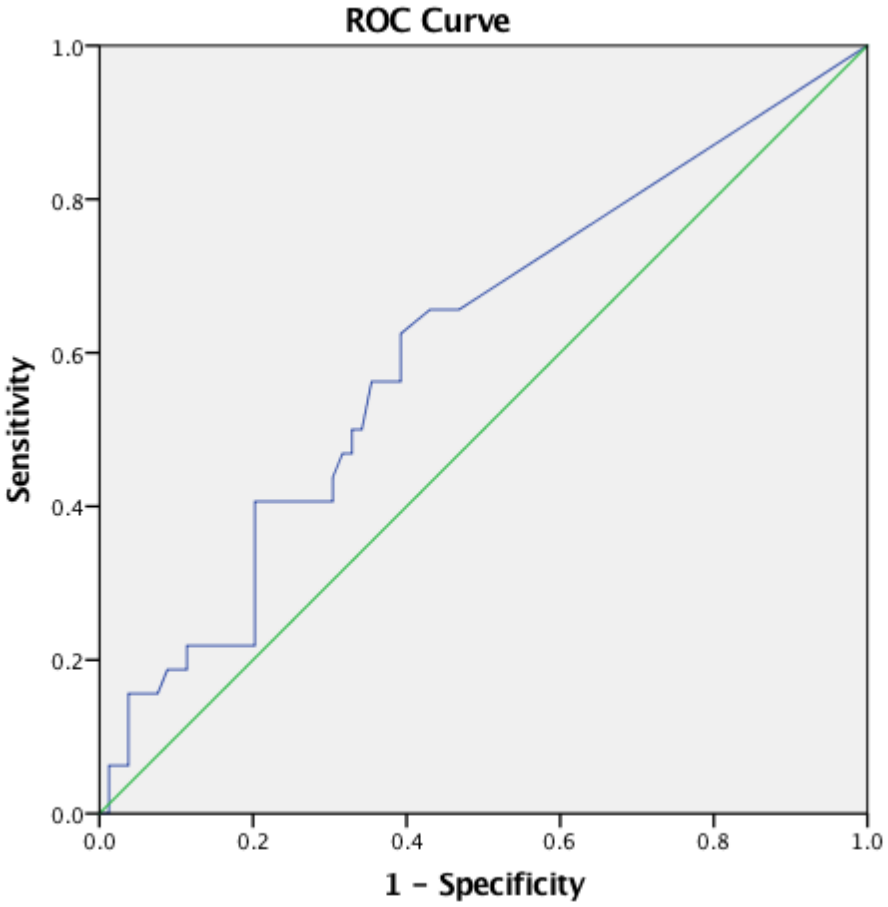


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.670	.059	.005	.555	.785

Figure 21

DBR-SIS Disruptive Behavior ROC curve for CDI-2 Functional Elevated

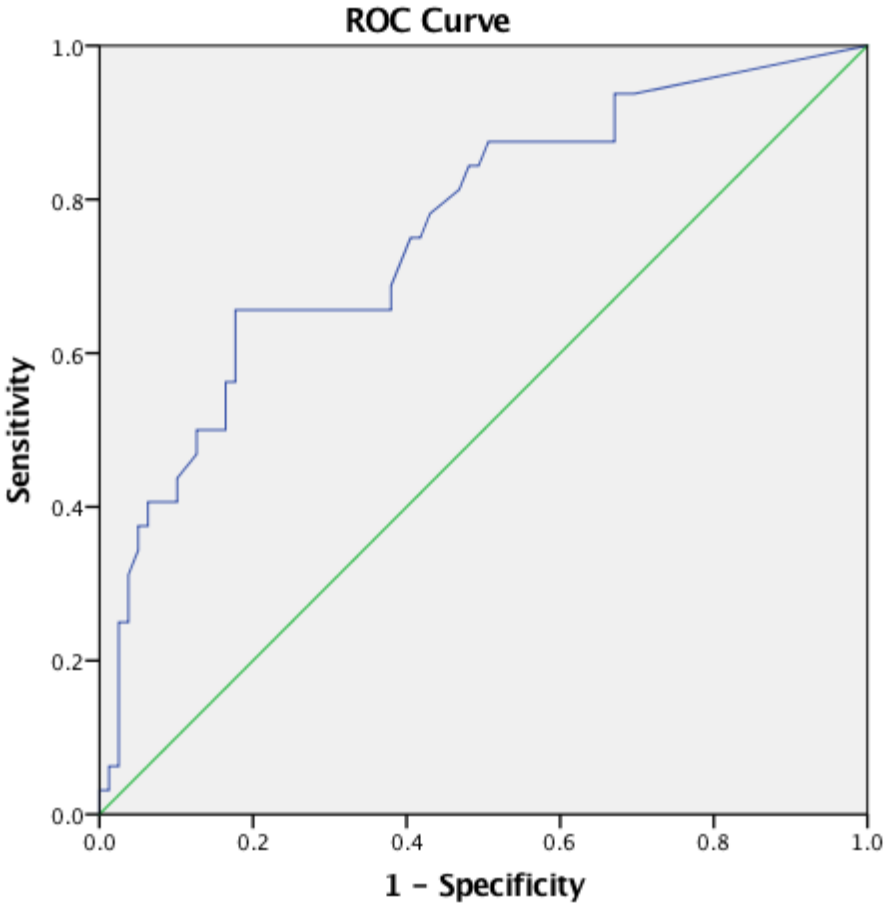


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.611	.060	.069	.494	.727

Figure 22

DBR-SIS Total ROC curve for CDI-2 Functional Elevated

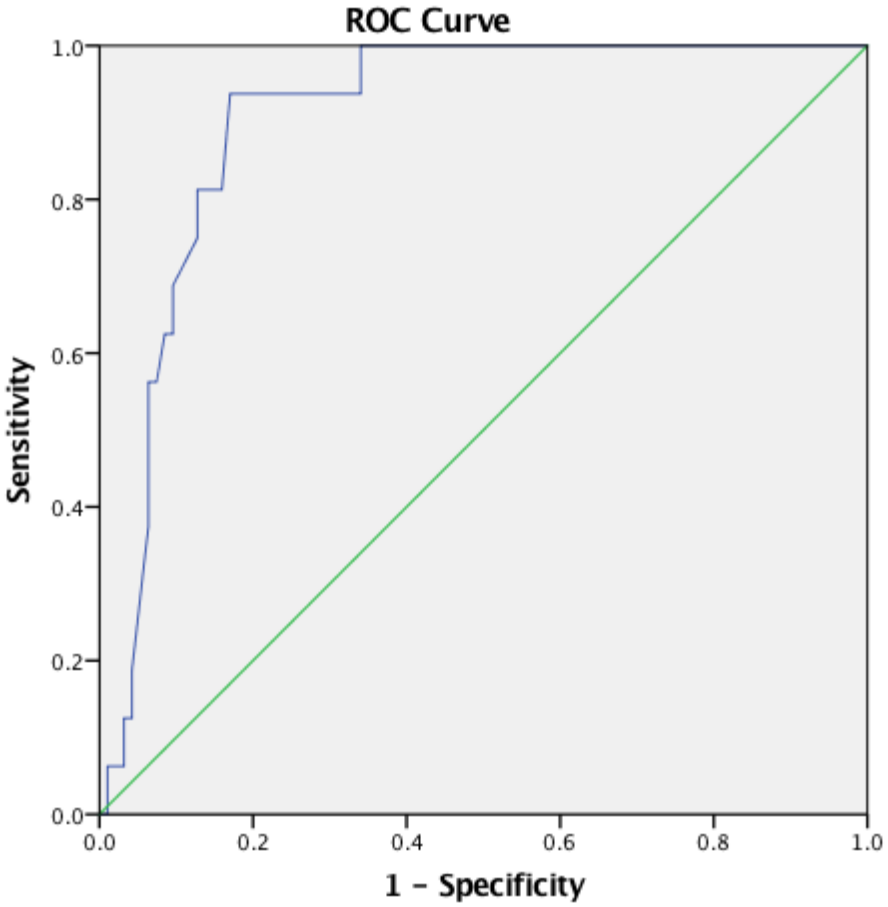


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.761	.051	.000	.660	.862

Figure 23

DBR-SIS Depression ROC curve for CDI-2 Functional Very Elevated

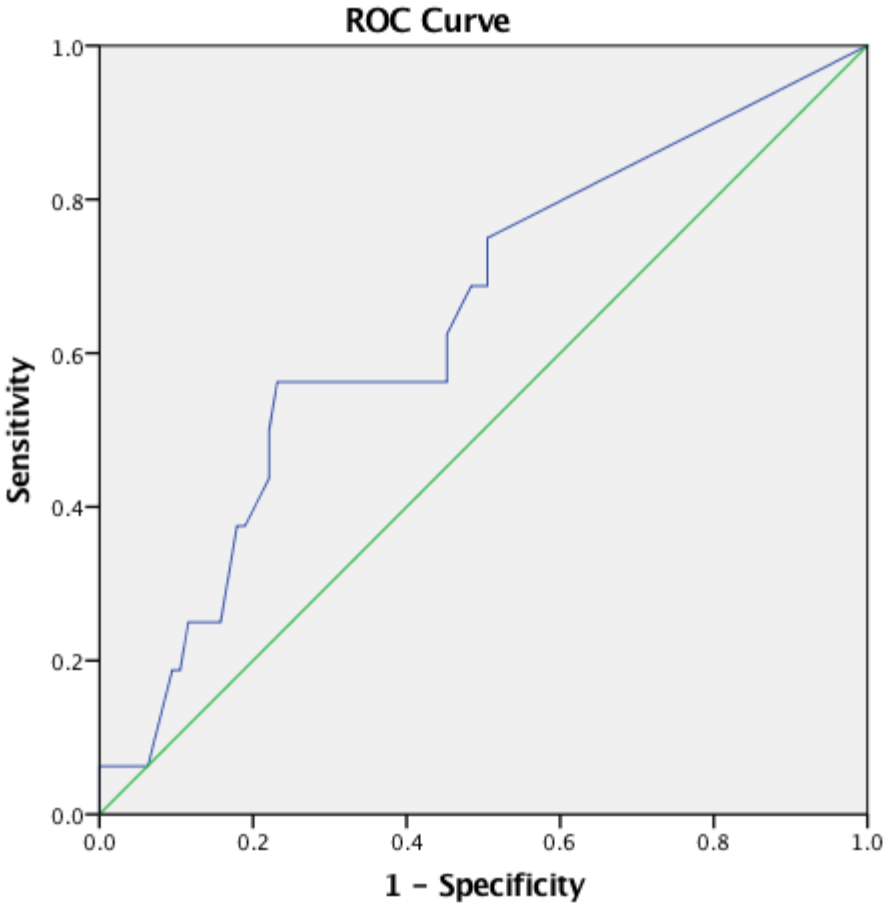


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.905	.030	.000	.846	.964

Figure 24

DBR-SIS Anxiety ROC curve for CDI-2 Functional Very Elevated

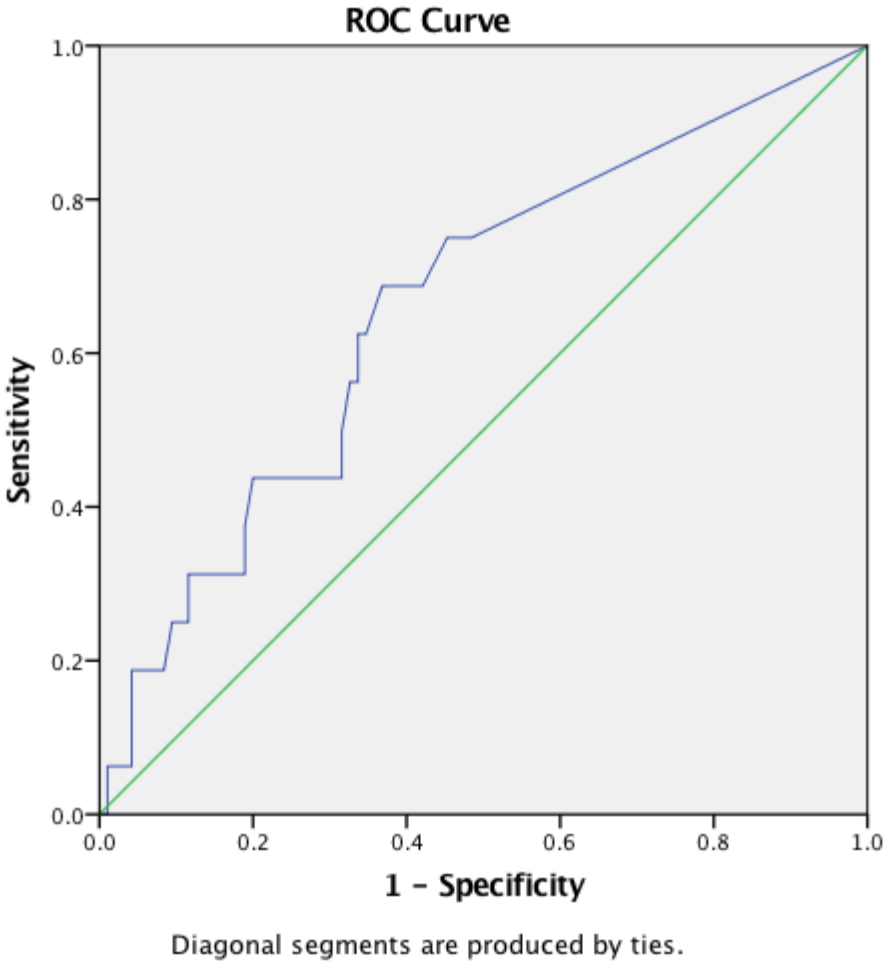


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.644	.075	.066	.497	.791

Figure 25

DBR-SIS Disruptive Behavior ROC curve for CDI-2 Functional Very Elevated

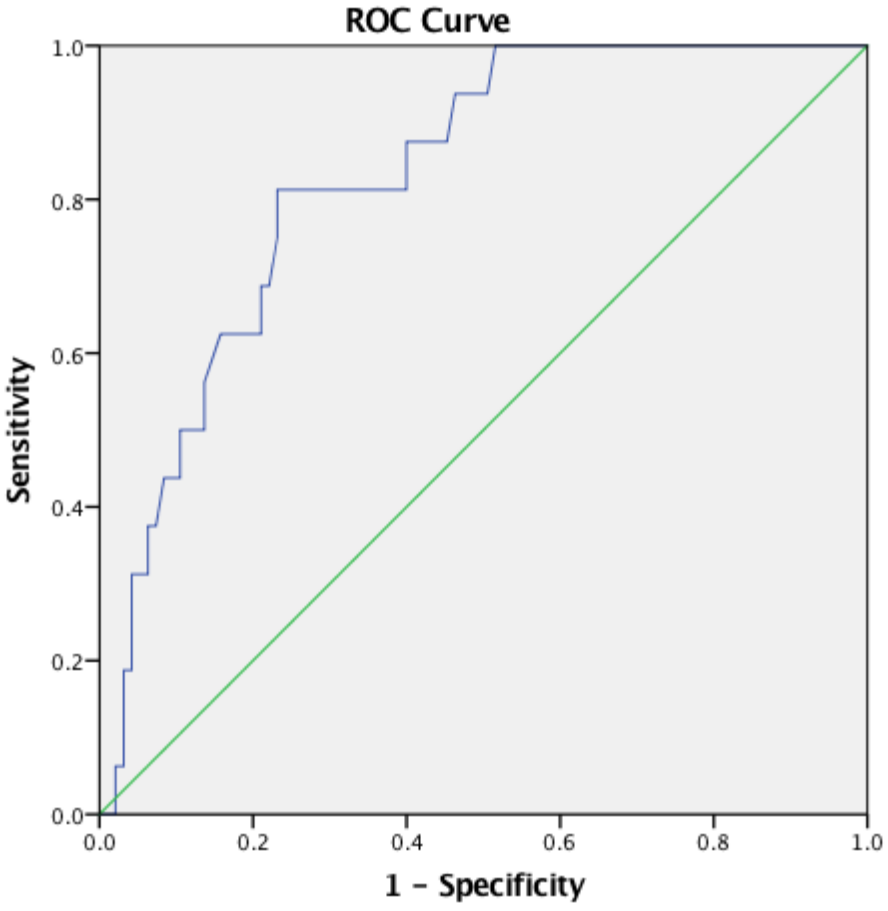


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.661	.074	.040	.516	.806

Figure 26

DBR-SIS Total ROC curve for CDI-2 Functional Very Elevated

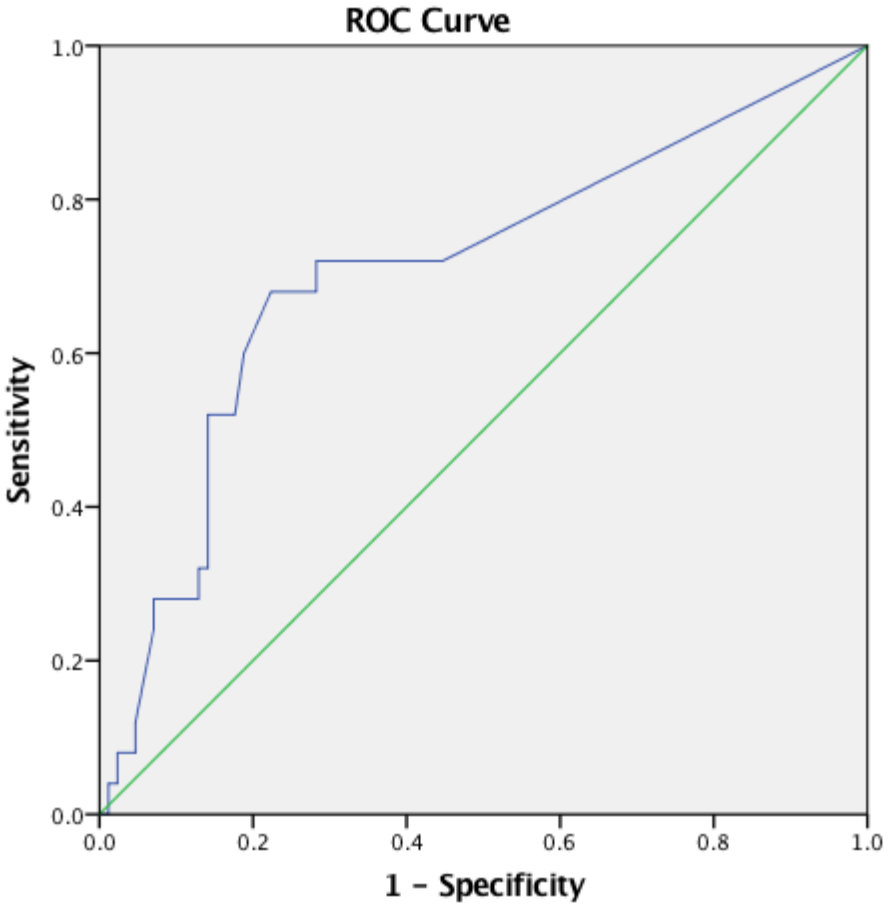


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.829	.046	.000	.738	.920

Figure 27

DBR-SIS Depression ROC curve for RCMAS-2 Total Moderately Problematic

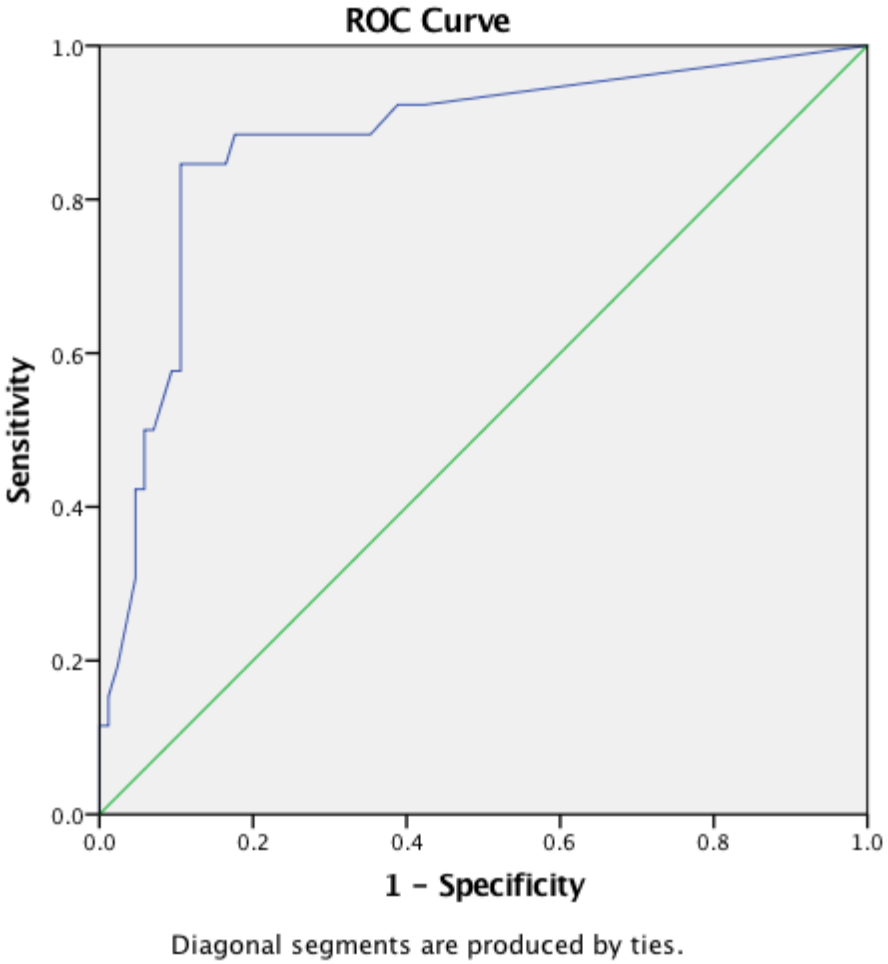


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.708	.064	.002	.584	.833

Figure 28

DBR-SIS Anxiety ROC curve for RCMAS-2 Total Moderately Problematic

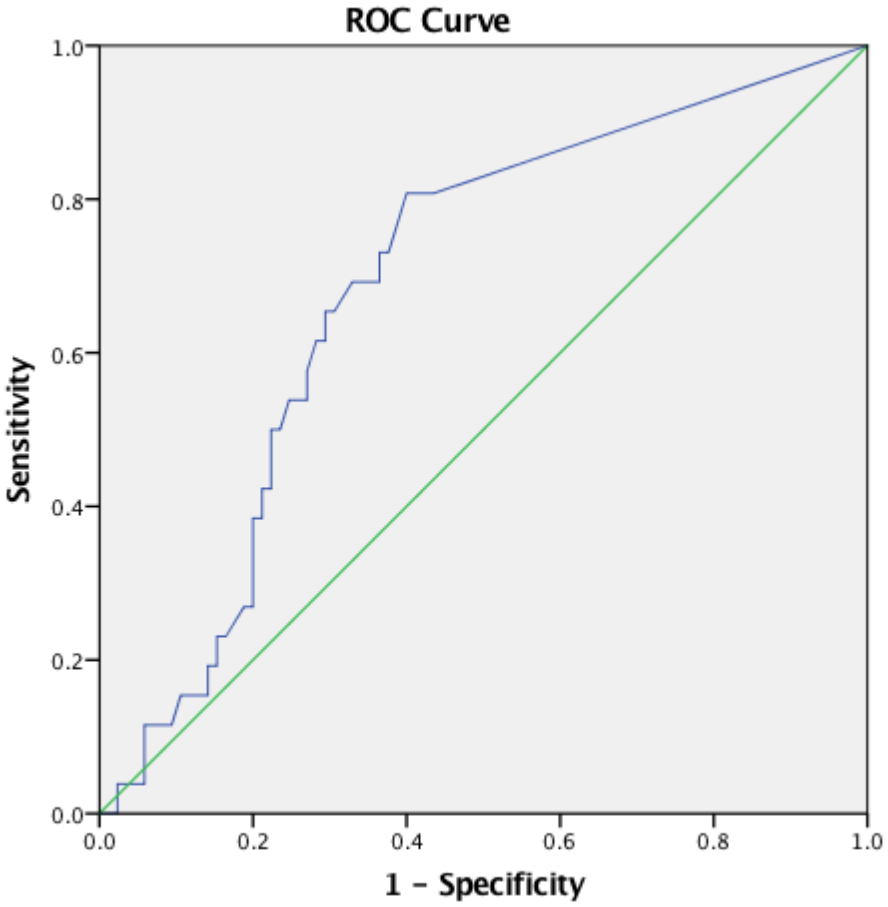


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.874	.042	.000	.791	.958

Figure 29

DBR-SIS Disruptive Behavior ROC curve for RCMAS-2 Total Moderately Problematic

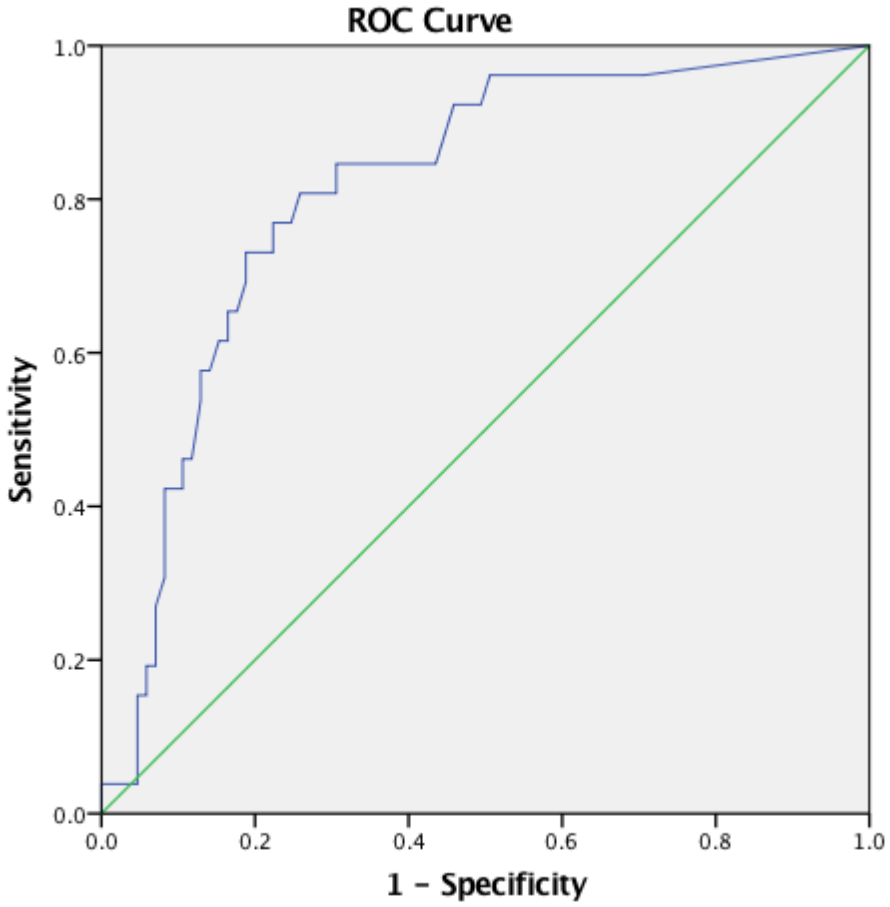


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.688	.056	.004	.578	.799

Figure 30

DBR-SIS Total ROC curve for RCMAS-2 Total Moderately Problematic

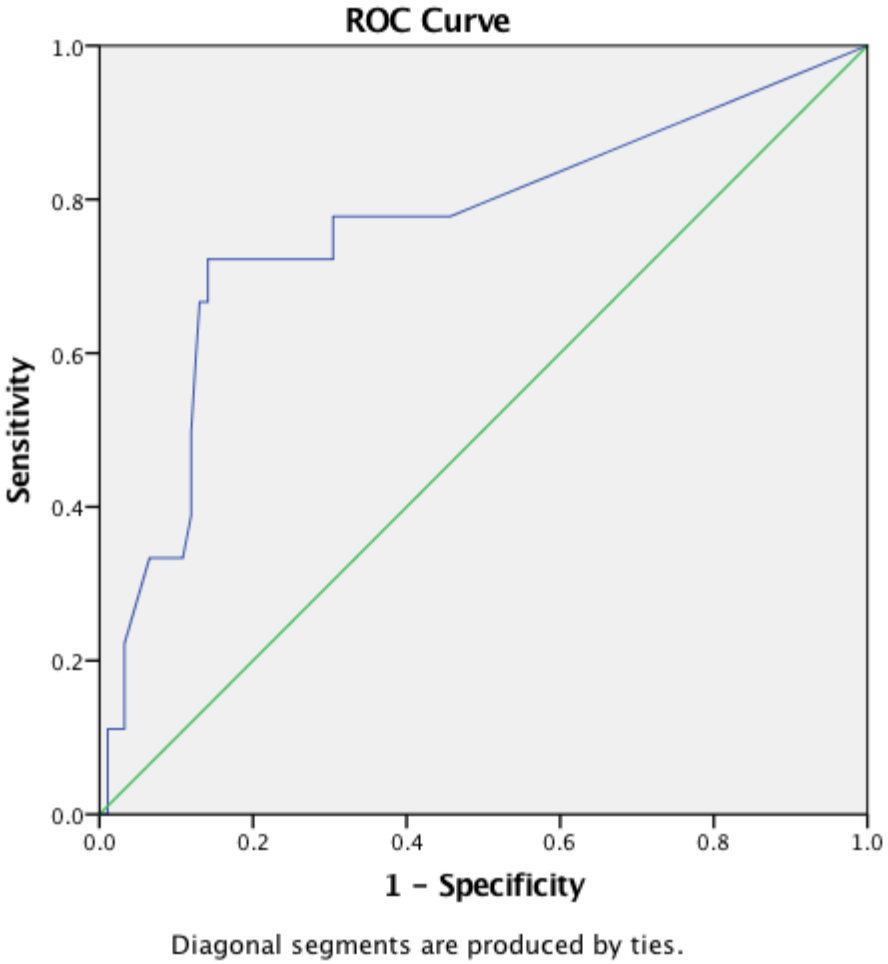


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.813	.046	.000	.723	.903

Figure 31

DBR-SIS Depression ROC curve for RCMAS-2 Physiological Moderately Problematic

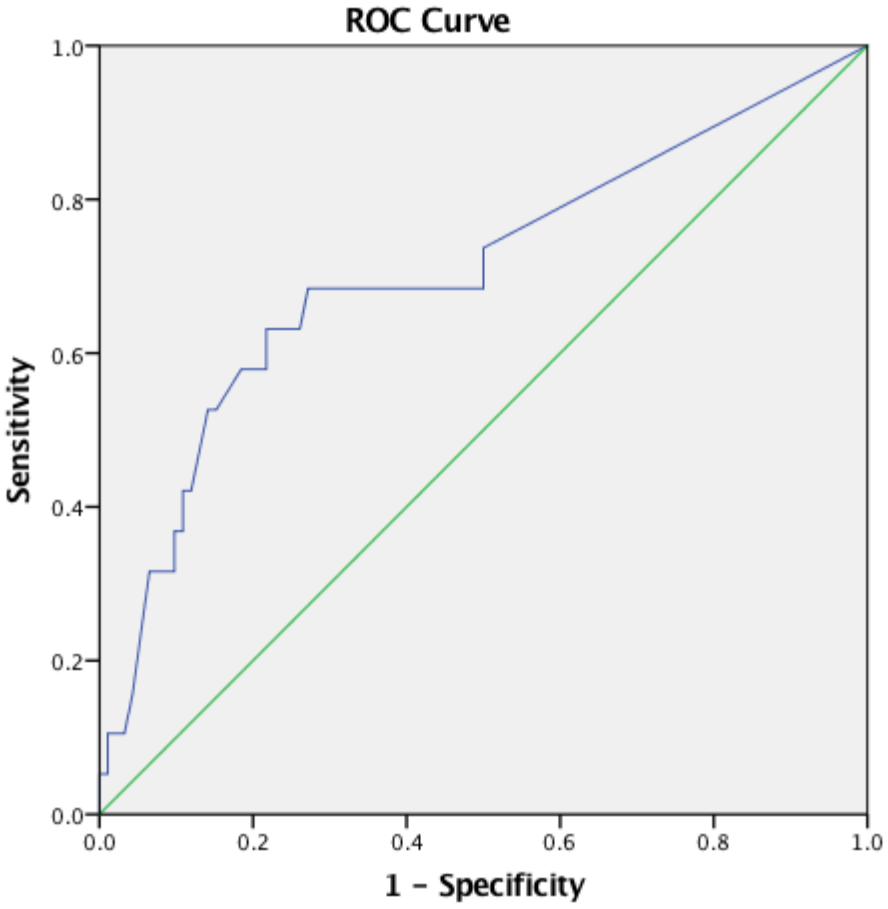


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.763	.070	.000	.626	.900

Figure 32

DBR-SIS Anxiety ROC curve for RCMAS-2 Physiological Moderately Problematic

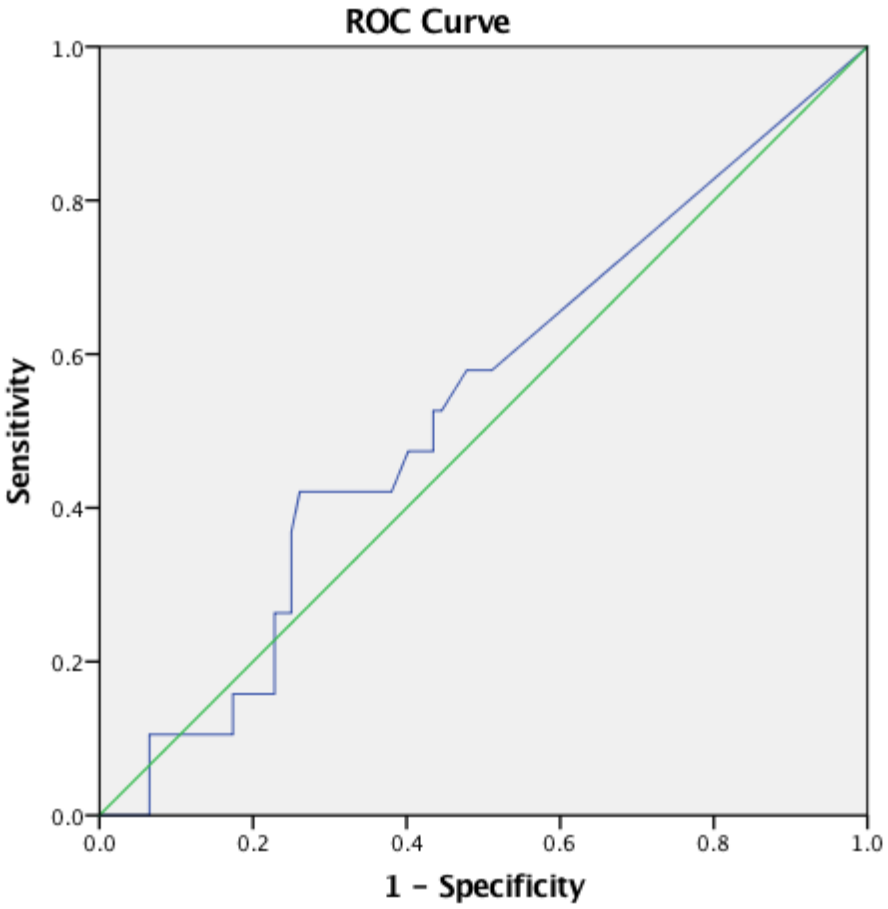


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.706	.073	.005	.563	.850

Figure 33

DBR-SIS Disruptive Behavior ROC curve for RCMAS-2 Physiological Moderately Problematic

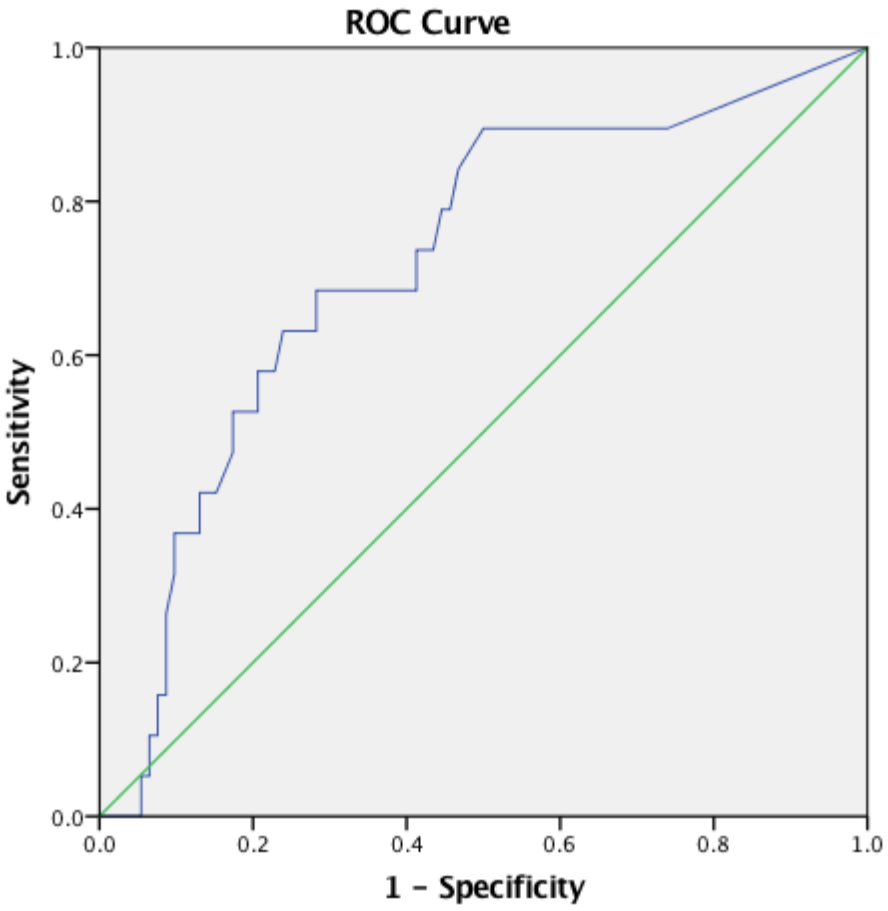


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.534	.072	.639	.394	.675

Figure 34

DBR-SIS Total ROC curve for RCMAS-2 Physiological Moderately Problematic

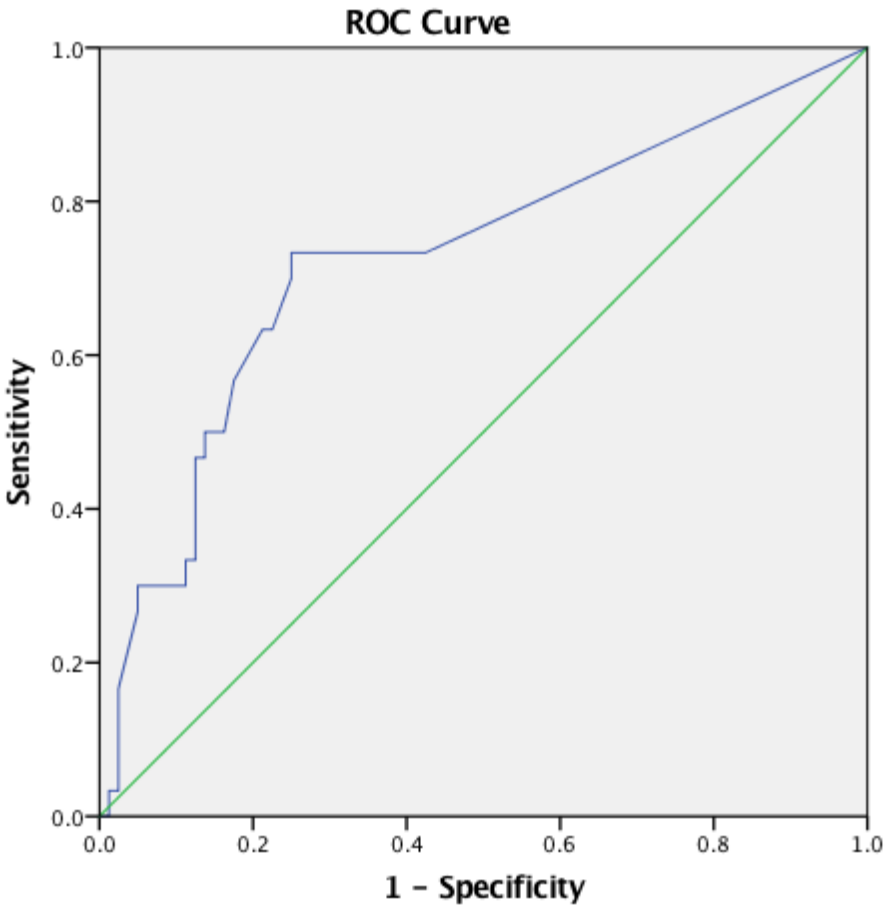


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.722	.064	.002	.597	.847

Figure 35

DBR-SIS Depression ROC curve for RCMAS-2 Worry Moderately Problematic

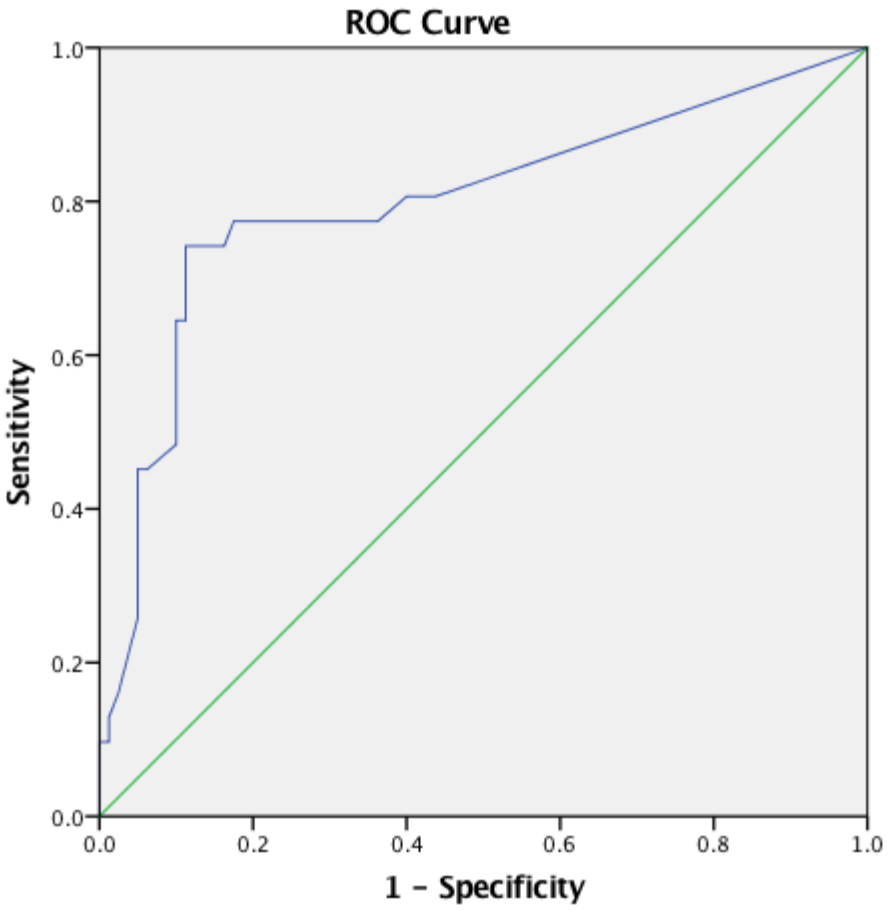


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.728	.058	.000	.613	.842

Figure 36

DBR-SIS Anxiety ROC curve for RCMAS-2 Worry Moderately Problematic

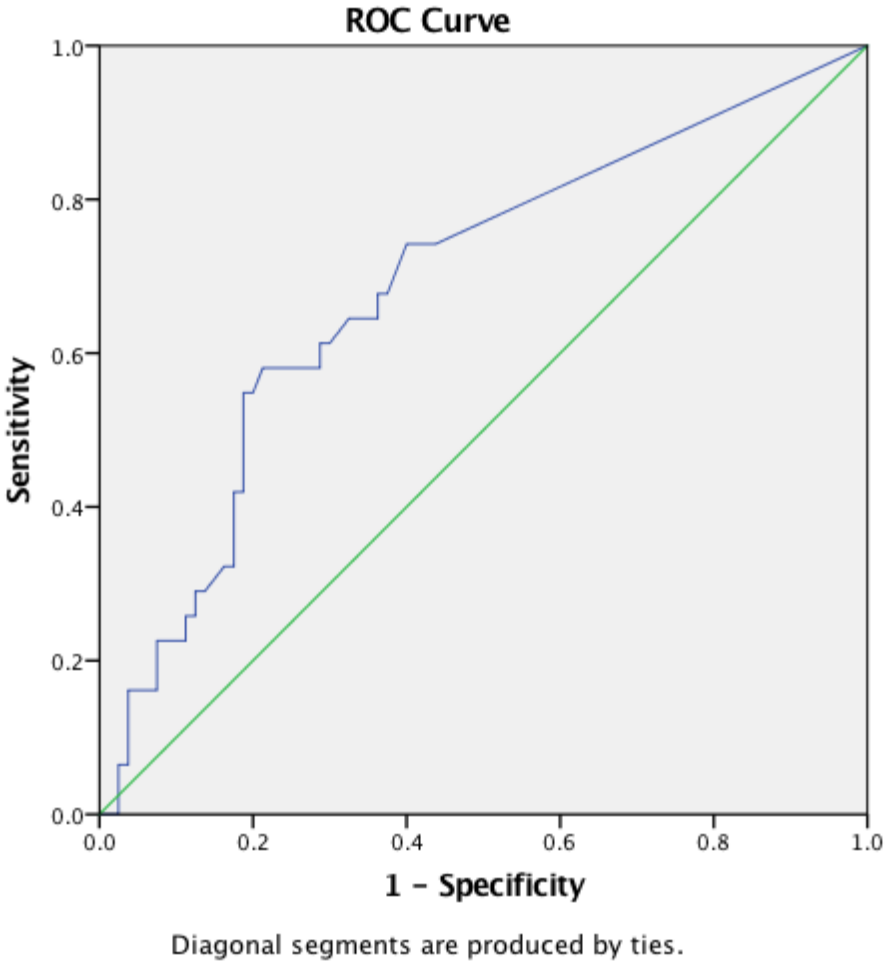


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.799	.053	.000	.695	.904

Figure 37

DBR-SIS Disruptive Behavior ROC curve for RCMAS-2 Worry Moderately Problematic

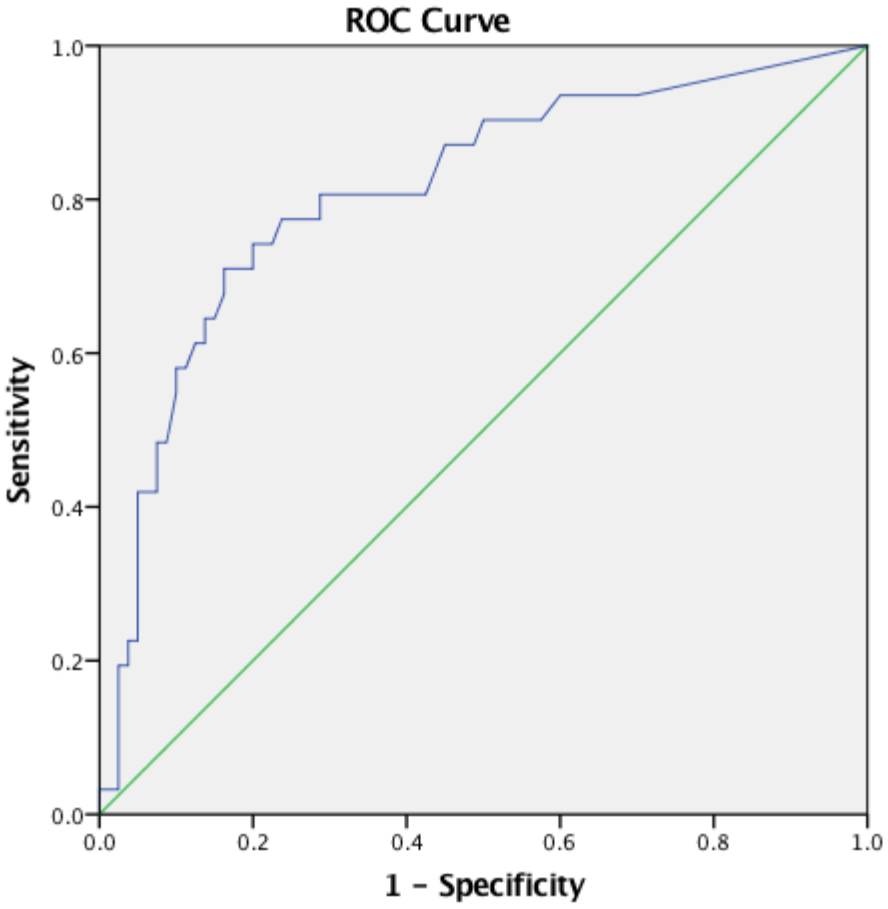


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.688	.057	.002	.576	.800

Figure 38

DBR-SIS Total ROC curve for RCMAS-2 Worry Moderately Problematic

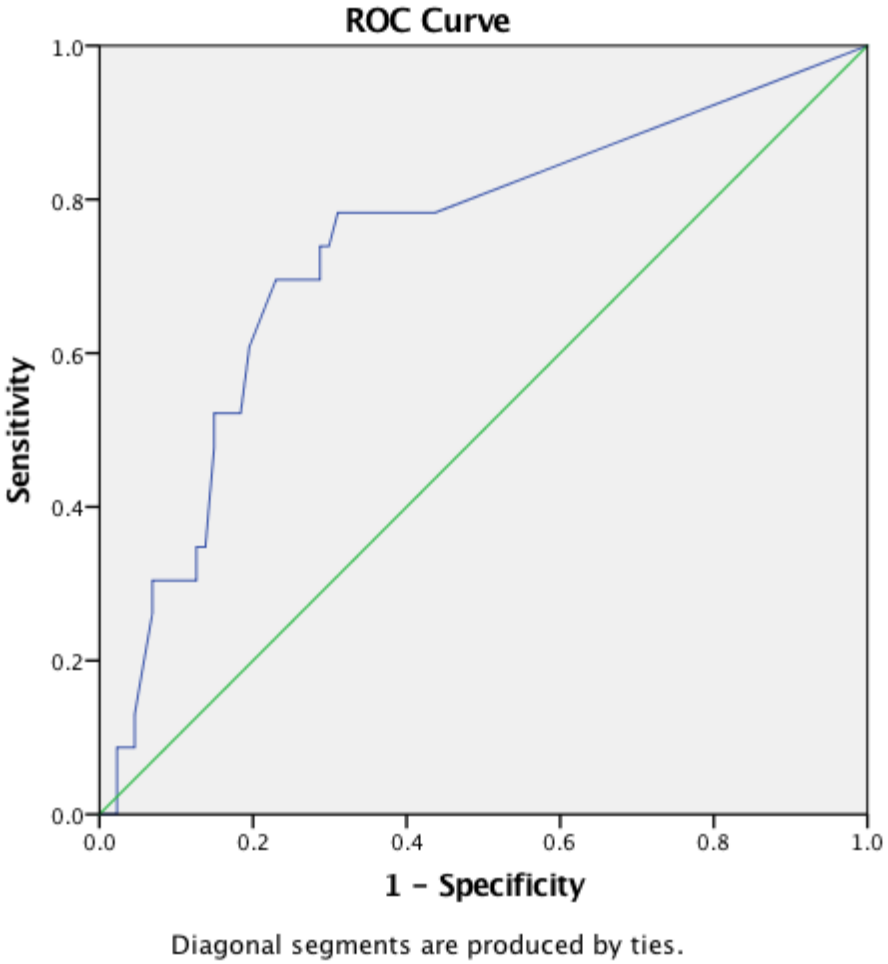


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.811	.048	.000	.718	.905

Figure 39

DBR-SIS Depression ROC curve for RCMAS-2 Social Moderately Problematic

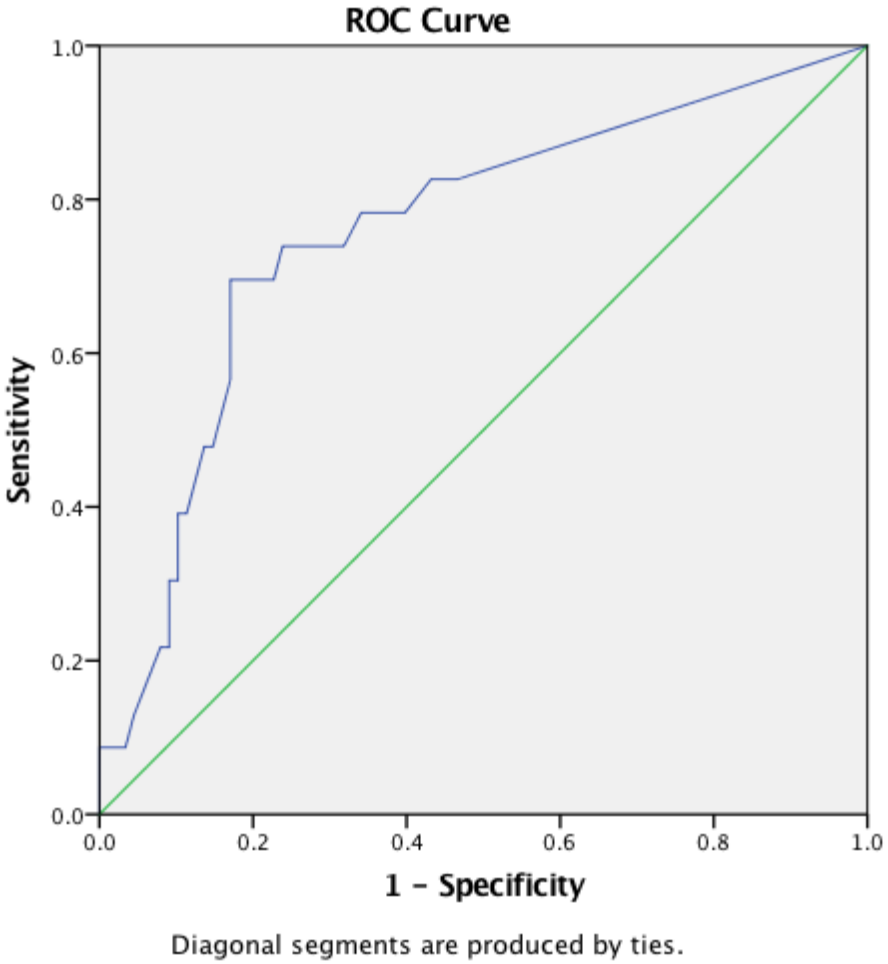


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.738	.061	.000	.618	.858

Figure 40

DBR-SIS Anxiety ROC curve for RCMAS-2 Social Moderately Problematic

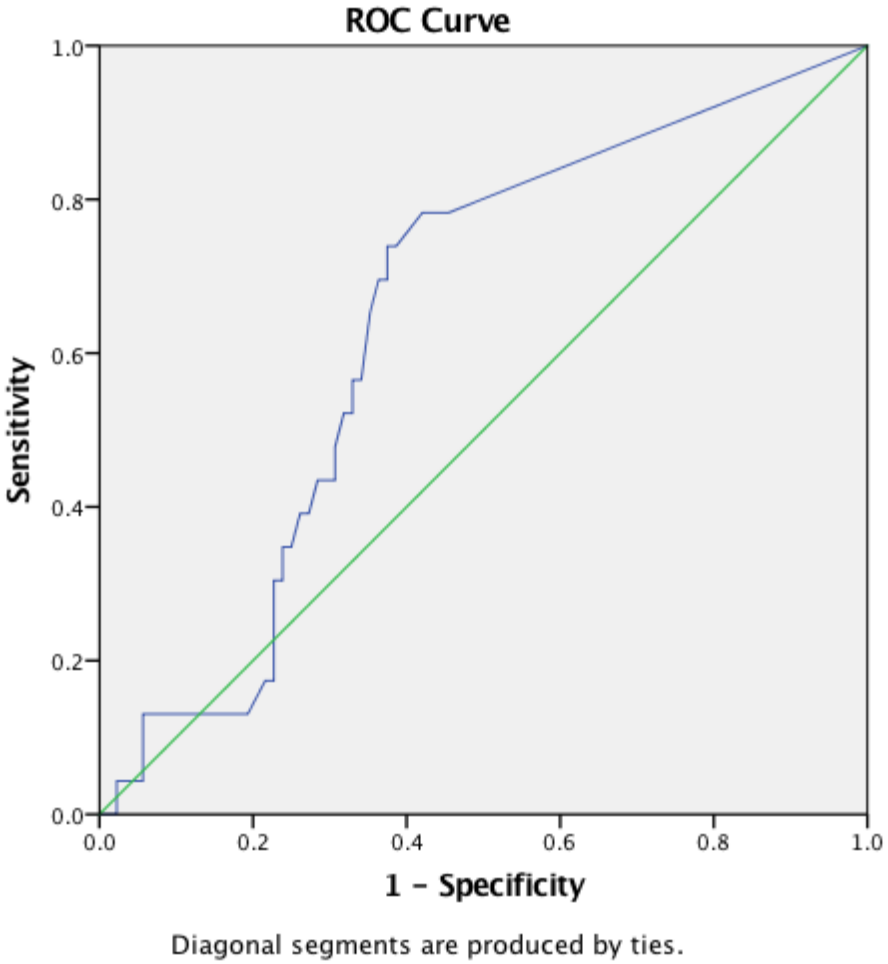


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.759	.059	.000	.644	.874

Figure 41

DBR-SIS Disruptive Behavior ROC curve for RCMAS-2 Social Moderately Problematic

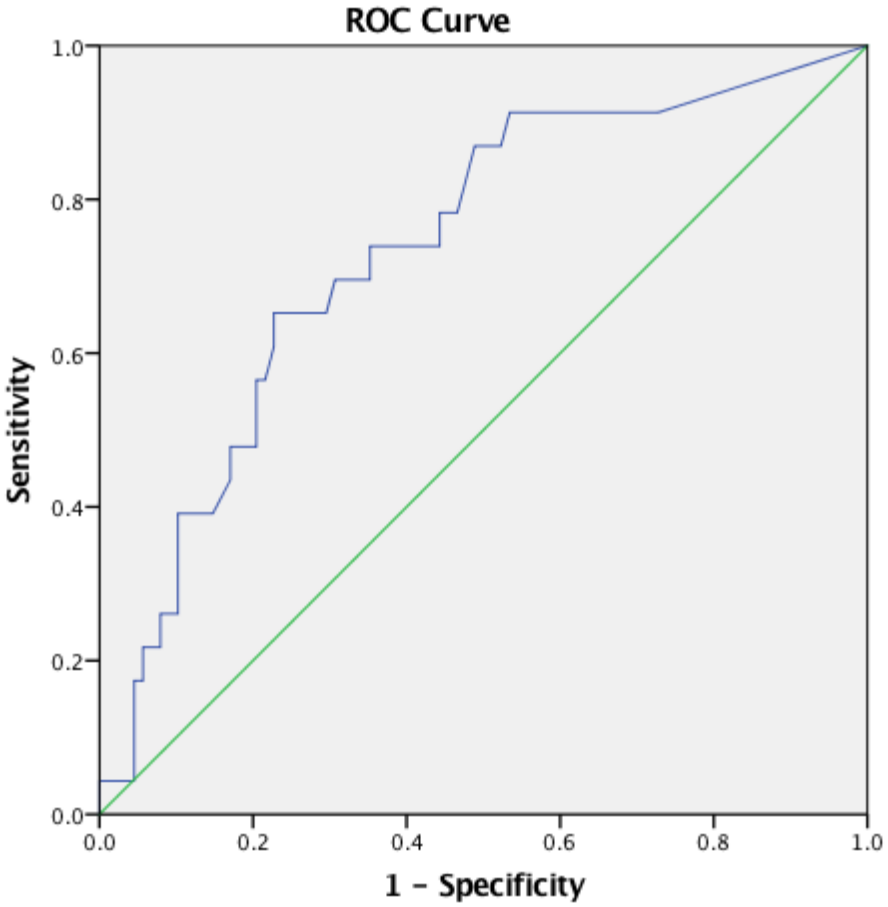


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.643	.060	.035	.526	.760

Figure 42

DBR-SIS Disruptive Behavior ROC curve for RCMAS-2 Social Moderately Problematic

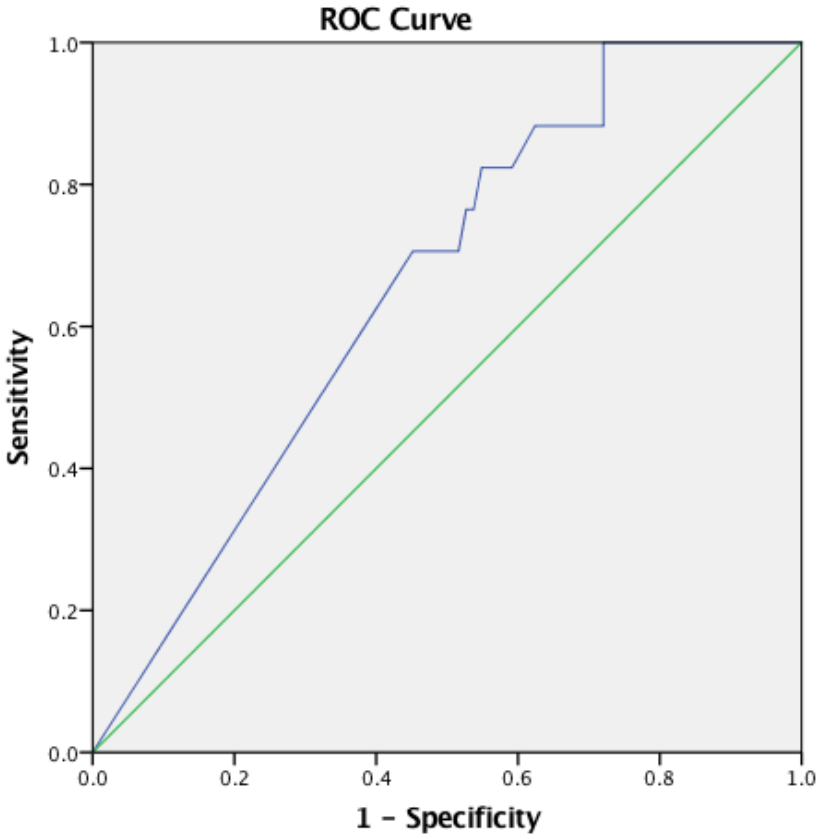


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.736	.058	.001	.623	.849

Figure 43

DBR-SIS Depression ROC curve for RCMAS-2 Defensiveness Moderately Problematic

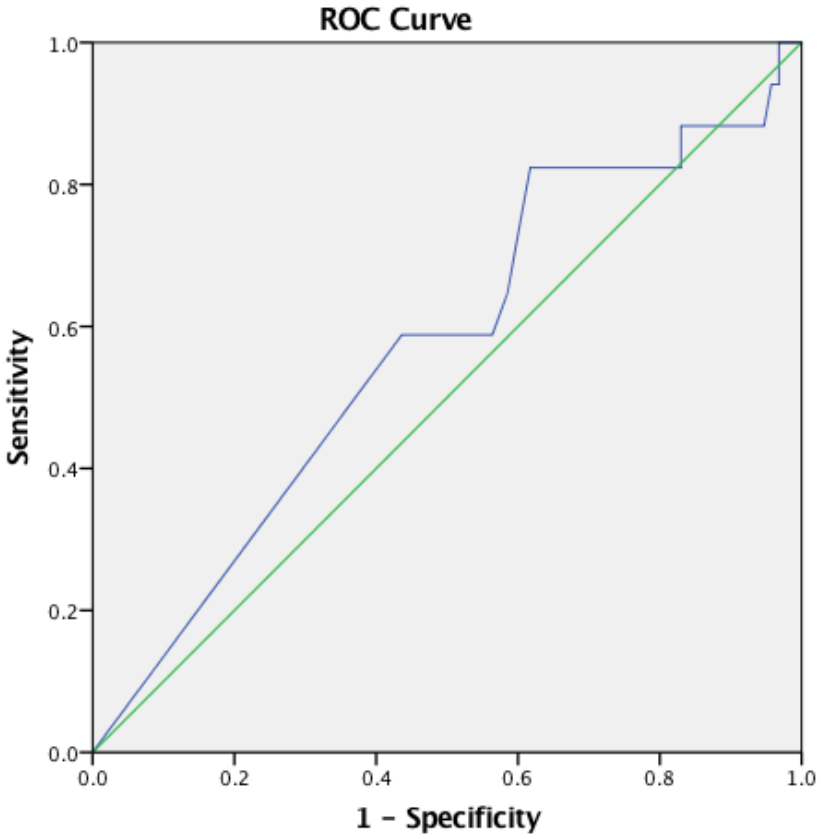


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.657	.062	.039	.535	.780

Figure 44

DBR-SIS Anxiety ROC curve for RCMAS-2 Defensiveness Moderately Problematic

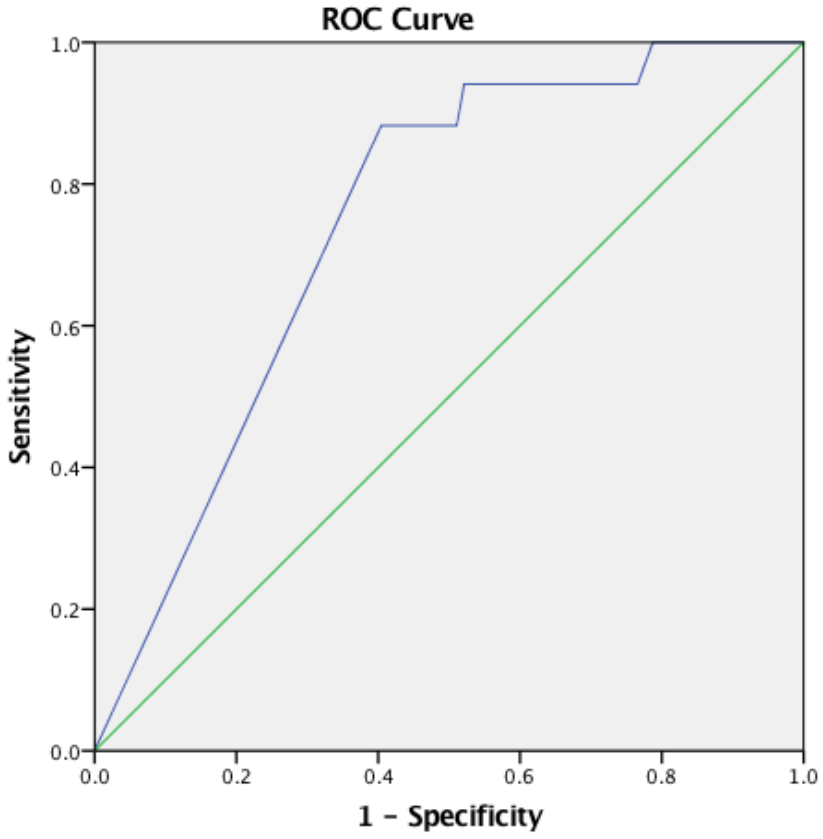


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.570	.077	.359	.420	.720

Figure 45

DBR-SIS Disruptive Behavior ROC curve for RCMAS-2 Defensiveness Moderately Problematic

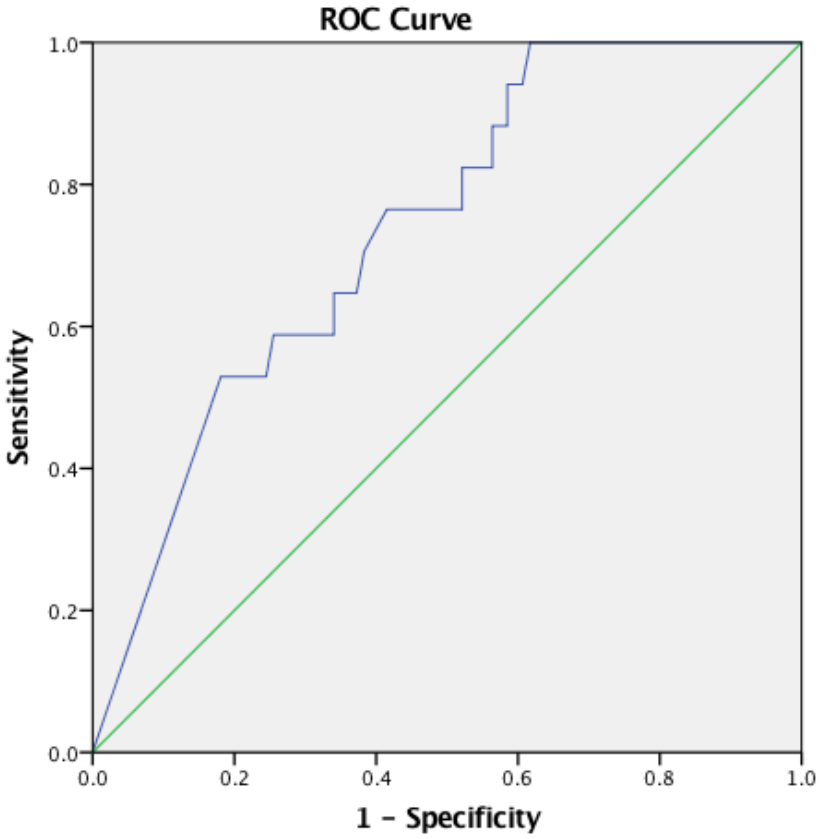


Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.746	.055	.001	.638	.853

Figure 46

DBR-SIS Total ROC curve for RCMAS-2 Defensiveness Moderately Problematic



Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.737	.058	.002	.624	.851