HHS Public Access

Author manuscript

Psychol Sch. Author manuscript; available in PMC 2024 January 21.

Published in final edited form as:

Psychol Sch. 2023 January 21; 60(7): 2320–2341. doi:10.1002/pits.22856.

The Utility of the Behavior Assessment System for Children-2 Behavioral and Emotional Screening System and Strengths and Difficulties Questionnaire in Predicting Mental Disorders in the Project to Learn About Youth-Mental Health

Melissa L. Danielson[1], **Hannah D. Kassab**[2], **Mary Lee**[2], **Julie Sarno Owens**[2], **Steven W. Evans**[2], **Corey Lipton**[1,3,a], **Sana Charania**[1,3], **Hayley Elia Young**[1,4,b], **Lorraine F. Kubicek**[5], **Kate Flory**[6], **Steven P. Cuffe**[7]

[1]Centers for Disease Control and Prevention, National Center on Birth Defects and Developmental Disabilities, Atlanta GA

[2]Department of Psychology, Ohio University, Athens, OH

[3]Oak Ridge Institute for Science and Education, Oak Ridge, TN

[4]Rollins School of Public Health, Emory University, Atlanta, GA

[5]Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO

[6]Department of Psychology, University of South Carolina, Columbia, SC

[7]Department of Psychiatry, University of Florida College of Medicine, Jacksonville, FL

Abstract

We examined the predictive utility of the Behavior Assessment System for Children-2 Behavioral and Emotional Screening System (BASC-2-BESS) and Strengths and Difficulties Questionnaire (SDQ) in identifying students with a mental disorder. Data were collected in a two-stage study over 34 months with kindergarten-12th grade (K–12) students (aged 5–19 years) in four U.S. school districts. In Stage 1, teachers completed the BASC-2-BESS and the SDQ. In Stage 2, parents of 1,054 children completed a structured diagnostic interview to determine presence of a mental disorder. Results suggest that teacher versions of the BASC-2-BESS and SDQ have modest utility in identifying children meeting criteria for a mental disorder based on parent report. Area Under the Curve (AUC) statistics representing prediction of any externalizing disorder (.73 for both measures) were higher than the AUCs predicting any internalizing disorder (.58 for both

Corresponding Author: Melissa Danielson, MDanielson@cdc.gov, Tel: 1-404-498-3016.
[a]Present institution: Booz Allen Hamilton, Atlanta, GA
[b]Present institution: North Carolina Department of Health and Human Services, Raleigh, NC

Conflict of Interest: Dr. Owens has received funding from the Institute of Education Sciences. Dr. Evans has received funding from Institute of Education Sciences, National Institute of Mental Health, and Patient Center Outcomes Research Initiative. Dr. Kubicek has received funding from Substance Abuse and Mental Health Services Administration and Colorado Clinical and Translational Sciences Institute. All other authors declare they have no conflict of interest.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

measures). Findings can inform the use of teacher report in mental health screening, specifically the selection of measures when implementing screening procedures.

Keywords

mental health; screening; children; adolescents; psychometrics

Mental disorders including anxiety, depression, attention-deficit/hyperactivity disorder (ADHD), and conduct problems are common among youth in the United States (Bitsko et al., 2022; Costello et al., 2005; Danielson et al., 2021; Merikangas, He, Brody et al., 2010; Merikangas, He, Burstein et al., 2010). National prevalence estimates indicate that approximately 9–10% of children aged 3–17 years have been diagnosed with ADHD, 9% with an anxiety disorder, 3–4% with a mood disorder, and 8–9% with conduct problems (Bitsko et al., 2022; Lebrun-Harris et al., 2022). Given these estimates, every classroom likely contains one to five students who are experiencing a mental disorder. Additionally, these disorders can be associated with negative impacts for the affected students, their peers and teachers, the school district, and society at large (e.g., Kessler, 2012; Robb et al., 2011).

Elementary school students who struggle with mental health problems experience difficulties in peer, student-teacher, and family relationships, as well as academic achievement (Grover et al., 2007; Hoza, 2007; Langley et al., 2004; Masten et al., 2005; Moilanen et al., 2010; Suldo et al., 2011). As children enter adolescence, both internalizing and externalizing problems place students at risk for additional problems including failing grades, school dropout, substance use, and suicide attempts (Erskine et al., 2016; Esch et al., 2014; Jaycox et al., 2009; Kent et al., 2011). Given these impacts, it is important that these concerns are detected early in their trajectory so that evidence-based interventions that have the potential to shift this trajectory can be provided (Evans et al., 2018; Long et al., 1994; Rimestad et al., 2019). For some children, anxiety or depression do not emerge until adolescence; yet early detection at this time provides the opportunity for the use of evidence-based treatments (Higa-McMillan et al., 2016; Weersing et al., 2017).

Early detection can occur when universal screening procedures are used in schools (Essex et al., 2009; Hill et al., 2004). Indeed, screening for academic, social, emotional, and behavioral problems is part of Tier 1 universal supports in most school-based multi-tiered systems of supports (Stoiber & Gettinger, 2016). Screening measures are often designed to capture as many students at risk for problems as possible, so that no one in need is missed (i.e., false positives are acceptable at this stage of supports, but false negatives should be avoided). When screening for mental health concerns, a variety of informants (e.g., parent/guardian, teacher, self) can be used. In schools, teacher report can be especially useful because teachers see students interacting with a variety of individuals (e.g., friends, other peers, adults) and in a variety of contexts (e.g., completing academic work that may be challenging, less structured times like transitions, recess, lunch, and study halls), allowing them to provide information based on all of these factors.

Recent studies show that, even relative to parent-report, teachers provide unique information (e.g., Garcia-Rosales, et al., 2020; Tobia et al., 2019). Several studies document that teacher

report can be useful in identifying mental health problems in children (Headley et al., 2011; Owens et al., 2016; Splett et al., 2019). Specifically, for externalizing concerns (e.g., disruptive behavior, oppositional behavior), there is evidence that screening by teacher report can detect concerns in elementary (e.g., Owens et al., 2016; Eklund et al., 2009; Renshaw et al., 2009) middle (Chin et al., 2013; Stoep et al., 2005) and high school (Dowdy et al., 2016) that are predictive of later outcomes such as office disciplinary referrals, grades, and course failures (Eklund et al., 2017). Additionally, while parent/guardians are a useful source for identification of problems in younger children, it may be challenging for school personnel to obtain a high response rate from parents (Ellickson & Hawes, 1989; Schanding & Nowell, 2013). Further, although self-report may be optimal for detecting internalizing problems in adolescents compared to teacher and caregiver report (e.g., Aebi et al., 2017; Kuhn et al., 2017; Margherio et al., 2019), self-report is not optimal for detecting either internalizing or externalizing problems in young children or for detecting externalizing problems in adolescents, as young children tend to overestimate their functioning, as do youth with externalizing problems (Owens et al., 2007). Thus, the use of screening tools completed by teachers for students at various grade levels creates the opportunity to identify problems that warrant intervention.

Although universal screening practices are beneficial, selecting a screener can be challenging for educators who have to select from multiple options that vary in cost, length, and predictive validity (Harrison et al., 2013). Thus, a comparison between universal screening tools would provide useful information for decision makers in schools who must make choices about what measure would best fit their situation. The goal of this study is to examine and compare the predictive validity of two established and widely used mental health screeners: the Behavior Assessment System for Children-2 Behavioral and Emotional Screening System (BASC-2-BESS; Kamphaus & Reynolds, 2007) and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). Below, we review the evidence for the utility of the teacher-report versions of the BASC-2-BESS and the SDQ in predicting school outcomes and mental health problems, as well as the limitations of this literature.

Strengths and Difficulties Questionnaire

The SDQ is a widely used, short, no-cost measure with adequate psychometric properties developed (Goodman & Scott, 1999) to screen for social and behavioral concerns among students such as emotional problems, peer problems, conduct problems, and hyperactivity/inattention problems, as well as indicators of prosocial behavior. The teacher-report SDQ has been shown to have predictive validity for school-related outcomes. For example, in a sample of kindergarten children (N = 248), the teacher-report SDQ total score had moderate to high utility (area under the curve (AUCs) = .78 to .93) in predicting daily student behavior across the year, grades in kindergarten and 1st grade, and impairment (Owens et al., 2016), with positive predictive values ranging from .40 to .71. This study did not assess the predictive utility of the impact supplement scores (i.e., impairment items).

Additionally, the teacher-report total score on the SDQ has also been shown to have utility in predicting mental disorders. For example, in a British sample with teacher-reported data on over 13,000 children, Goodman and Goodman (2011) found that average teacher-report

SDQ total scores were predictive of population disorder prevalence in children aged 5 to 16 years who were also assessed by diagnostic interviews with parents. Similarly, in a sample of 195 children aged 3 to 16 years in foster care in Norway (Lehmann et al., 2014), teacher-report SDQ scores differentiated between children with and without a mental disorder, with AUCs ranging from .74 (emotional problems subscale predicting an internalizing disorder) to .86 (conduct problems subscale predicting an externalizing disorder). Teacher-report SDQ impact supplement scores were assessed in this study and had similar predictive utility (AUC = .75) as did total scores (AUC = .77) in the prediction of any disorder. This study did not examine differences in teacher prediction by age of the student. Considering the established difficulties with teachers identifying mental health concerns in adolescence, especially when risk for internalizing problems is high, understanding the change in prediction across development is important.

In summary, there is some evidence supporting the predictive utility of the teacher version of the SDQ in identifying mental disorder diagnoses. This utility has been replicated in two samples, but not in the United States and not within a school-based sample. Further, these studies did not examine predictive utility separately for elementary and secondary school students. Because elementary school teachers spend significantly more time with students and thus may have a more in-depth understanding of student problems compared to secondary school teachers, separate analyses are warranted. Lastly, only one study assessed the predictive utility of SDQ impact supplement scores, which provide additional information about the severity of the student's problems. Thus, including these items could improve predictive validity.

Behavior Assessment System for Children-2 Behavioral and Emotional Screening System (BASC-2-BESS)

The BASC-2-BESS was developed with the goal of screening for emotional and behavior problems (Kamphaus & Reynolds, 2007). Several studies evaluating the teacher-report BASC-2-BESS have focused on predictive utility of student outcomes. In a study with 3rd and 4th grade students (*N*=48), Renshaw et al. (2009) found evidence that total scores correctly identified students with (88%) and without (67%) problems in academic achievement, academic engagement, and behavioral performance according to report card data. Similarly, in a larger elementary school sample (*N* = 496), total scores had a sensitivity of 68.4% and specificity of 77.1% in identifying students referred to the office one or more times by midyear (King et al., 2012). A study examining predictive validity with kindergarten students (*N* = 248) found that total scores had moderate to high utility (AUCs = .84 to .90) in predicting daily student behavior across the year, grades in kindergarten and 1st grade, and impairment (Owens et al., 2016), with positive predictive values ranging from .32 to .71 across outcomes. This study directly compared the SDQ to the BASC-2-BESS and found them both to be significant predictors of all outcomes, and both demonstrated incremental predictive validity beyond the prediction of kindergarten-entry literacy scores. However, the SDQ was a stronger predictor of daily behavior and peer relationships, whereas the BASC-2-BESS was a stronger predictor of student grade point average (GPA). Taken together, these studies show that the teacher-report BASC-2-BESS scores suggest

promise for predicting school-related student outcomes, but that the SDQ and BASC-2-BESS may have different capacities to predict outcomes among school-aged children and adolescents.

Compared to the SDQ, the utility of BASC-2-BESS teacher-report scores in predicting mental disorders has been less studied. In fact, we are aware of only one study assessing the association between teacher BASC-2-BESS scores and mental health problems. With a diverse student sample (N = 359 K – 5th graders), Kamphaus and colleagues (2007) examined correlations between the teacher-report BASC-2-BESS screener completed by teachers in Year 1 of the study with subscales of the Behavioral Assessment System for Children (BASC) in Year 2 of the study. Correlations between the teacher-report BASC-2-BESS total scores and BASC ratings of conduct problems, psychoticism, and social skills ranged from .48 to .50; correlations with depression (r = .37), withdrawal (r = .35), and anxiety (r = .20) subscales were lower. Thus, we may expect teacher ratings of the BASC-2-BESS to better predict externalizing disorders than internalizing disorders. One limitation of this study is that the BASC-2-BESS items were derived from the BASC; thus, there may be some inflation in these statistics as a function of shared item variance.

In summary, additional studies assessing the utility of the teacher-report BASC-2-BESS scores in identifying mental disorders among elementary and secondary students would advance the literature in this area. There is some evidence supporting the predictive utility of the teacher BASC-2-BESS in identifying mental health problems (dimensionally), but this study lacked independence in measurement (Kamphaus et al., 2007). There is more evidence of the utility of the teacher BASC-2-BESS in identifying school problems. However, most studies of the teacher version included elementary school students only. Lastly, there is one study that directly compares the utility of the SDQ and the BASC-2-BESS; however, this sample was restricted to kindergarten students (Owens et al., 2016). Given the high prevalence of mental health problems for students in kindergarten–12th grade (K–12) and the importance of early intervention in modifying risk trajectories, additional research that informs the screening decisions of school personnel is warranted.

Current Study

We examined the predictive utility of teacher-report scores from the BASC-2-BESS and the SDQ (subscales and impact supplement) in identifying students who meet criteria for a mental disorder. This study advances the literature by: (a) examining the predictive utility of two screening tools in the same sample, (b) examining the utility of teacher ratings for elementary and secondary school students separately, and (c) assessing multiple disorders (internalizing and externalizing disorders broadly, as well as specific disorders). Based on the literature, we hypothesized that both screening tools would have significant predictive utility in identifying those who meet criteria for a mental disorder at both the elementary and secondary levels, but that the prediction at the elementary level may be stronger given the relative depth of student-teacher relationships in younger grades compared to older grades. We also hypothesized that both measures may more strongly predict externalizing disorders relative to internalizing disorders. Ultimately, this comparative study of universal screeners

provides insight regarding the choice of screening measures to facilitate early detection and referral of youth who could likely benefit from mental health services.

Method

Participants

Data were collected for the Project to Learn About Youth-Mental Health (PLAY-MH), a multi-site, community-based, cross-sectional study designed to produce population-based estimates of mental disorder prevalence, disorder co-occurrence, and receipt of associated mental health treatment in K–12 students (aged 5–19 years). Details on the study methods and demographic characteristics of the samples have been published elsewhere (Danielson, et al., 2021; Wanga, et al., 2022). Data were collected using a two-stage study design. In Stage 1, teachers completed the BASC-2-BESS and the SDQ on 18,107 students. Teachers also completed a screener for the presence of tics[1] that was incorporated into the sample selection (see next paragraph), but this screener was not evaluated for predictive utility in this analysis and no data from the tic screener were included in any of the analyses presented in this manuscript. Of the students who were screened by both the BASC-2-BESS and SDQ, parents of 1,054 children completed a structured diagnostic interview to determine presence of a mental disorder in Stage 2.

Data were collected between 2014 and 2018 in four school districts in four states (Colorado, Florida, Ohio, South Carolina) that collectively included urban, suburban, and rural areas (Figure 1). The Stage 1 screening results were used to stratify students into high risk and low risk groups by student sex (male, female; additional sex options not available in the original data collection) and school level (elementary school, middle/high school) and an approximately equal number of students in each strata were randomly sampled for participation in the second stage of data collection. In Stage 2, parents of the sampled students completed a structured diagnostic interview that was used to identify students who met diagnostic criteria for a set of externalizing and internalizing disorders. The South Carolina site was funded for two complete iterations of data collection (PLAY-MH and Re-PLAY-MH in 2014 through 2017; Wanga, et al., 2022), while the other three sites collected data in one round. Only the data from the second iteration of SC data collection (Re-PLAY-MH) are included in this analysis because of the larger sample size in Stage 2 of the second data collection and limited availability of BASC-2-BESS data in Stage 1 of the first data collection. The overall sample size for each site's data collection for Stage 1 ranged from 4,198 (Colorado) to 6,937 (South Carolina) and for Stage 2 ranged from 160 (Ohio) to 572 (South Carolina). Stage 1 completion rates ranged from 71.9% (South Carolina) to 78.1% (Ohio), and Stage 2 response rates ranged from 3.5% (Ohio) to 22.6% (Colorado) among the families sampled and contacted for participation in Stage 2. Data were collected over 21 months (Ohio) to 34 months (Florida), and the median time between the Stage 1 screener and the Stage 2 interview ranged from 8 months in Ohio to 21 months in Florida. All study procedures including informed consent were reviewed and approved

[1]One of the primary objectives of the PLAY-MH study was to estimate the prevalence of tic disorders in a school-based sample. To facilitate this objective, a tic screener was included in the Stage 1 data collection and teacher report of tics was included as a criterion for a student to be categorized into the high risk group.

by the respective Institutional Review Boards associated with each site's project team and the administrators in each school district. All authors certify responsibility of the results reported herein.

Measures and Case Definitions

Behavior Assessment System for Children-2 Behavioral and Emotional Screening System (BASC-2-BESS)—The BASC-2-BESS was developed in the United States with the goal of being a screener for emotional and behavior problems (Kamphaus & Reynolds, 2007). Versions are available for children in preschool through grade 3 by parent (30 items) and teacher (27 items) report as well as for youth in grades 3 through 12 by parent, teacher, and self-report. Respondents can choose between four options based on frequency of symptom occurrence (1=Never, 2=Sometimes, 3=Often, 4=Almost always). Scoring instructions are available in the BASC-2-BESS user manual (Kamphaus & Reynolds, 2007). Typically, only the BASC-2-BESS total score is reported and studied, however, there is some support for a four-factor structure (internalizing, hyperactivity–inattention, school problems, personal adjustment) for students in grades 3 to 5, with adequate internal consistencies (α's ranging from .83 to .94; Dever et al., 2012) and in high school students (Harrell-Williams et al., 2015), albeit with slightly lower internal consistencies (α's ranging from .72 to .83). Total scores are invariant across males and females (Kamphaus & Reynolds, 2007), and students of different ethnicities (Harrell-Williams et al., 2015). Only the BASC-2-BESS total score was used in the current study.

Strengths and Difficulties Questionnaire (SDQ)—The SDQ is a widely-used, short, no-cost measure with versions for different informants: for youth aged 3 to 11 years, parent and teacher versions are available, and for youth aged 11 to 16 years, parent, teacher, and self-report versions are available. All versions have 25 items and five subscales: emotional problems, peer problems, conduct problems, hyperactivity/inattention problems, and prosocial behavior. Respondents can select between three options (i.e., Not True, Somewhat True, Certainly True) based on how applicable a statement is for a child. In addition, all versions include impact supplement items assessing the extent to which the child's problems create distress, burden, and/or interfere with functioning (e.g., in relationships, classroom, leisure activities). Forms and scoring instructions are available at sdqinfo.org. There is evidence for both a five-factor structure (Bosik et al., 2021; Goodman & Scott, 1999; Hall et al., 2019) and a two-factor structure wherein the conduct problems and hyperactivity/inattention problems subscales are combined to form an externalizing dimension and the emotional problems and peer problems subscales are combined to represent an internalizing dimension (Goodman et al., 2010). The factor structure of the SDQ is invariant across gender and ethnicity, at least in a US sample (He et al., 2013). In the current study, the SDQ total score (i.e., sum across all subscales except the prosocial subscale) was used to determine risk status, as suggested by Goodman (1997).

Categorization of Risk Status in Stage 1—Using guidelines from the BASC-2-BESS user manual (Kamphaus & Reynolds, 2007) and published thresholds (Goodman, 1997), students were categorized in the high-risk strata for sampling for Stage 2 if they had a

t-score on the BASC-2-BESS greater than 60, an SDQ total score greater than 11, or had any teacher-reported tics on the tics screening measure.

Diagnostic Interview Schedule for Children-IV (DISC-IV)—In Stage 2, parents completed the Diagnostic Interview Schedule for Children-IV (DISC-IV; Shaffer, et al., 2000) to determine if their child met the Diagnostic and Statistical Manual of Mental Disorders version IV (DSM-IV; American Psychiatric Association, 1994) criteria for selected mental disorders. The disorder modules included in the Stage 2 interview for all four sites were separation anxiety disorder, generalized anxiety disorder, panic disorder, agoraphobia, obsessive-compulsive disorder (OCD), post-traumatic stress disorder, major depressive disorder, dysthymic disorder, mania, hypomania, attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder, and conduct disorder; the social phobia module was also completed in each site except Ohio. For most modules, a child was considered to have the disorder if they met symptom criteria and had at least two moderate or one severe rating of impairment for the given disorder in the past year. For OCD and panic disorder, only symptom criteria in the past year needed to be met, consistent with disorder criteria set forth in the DSM-IV. For ADHD, a child was considered to have the disorder if they met symptom criteria in the past year, had at least two moderate or one severe rating of impairment, and had at least two ADHD symptoms reported by their teacher on the BASC-2-BESS or the SDQ.

Statistical Analysis

The data from all four sites were combined and treated as a convenience sample for this study's psychometric analysis. Students who were missing the BASC-2-BESS, SDQ, or DISC-IV (for Stage 2 analyses only) were excluded from the analytic sample. The percentage of students who were grouped into the high-risk category based on their teacher screener was calculated for the BASC-2-BESS (t-score greater than 60) and SDQ (total score greater than 11). Among those who were selected for and participated in Stage 2, the percentage who met criteria for any disorder, any internalizing disorder, any externalizing disorder, and each individual disorder were calculated by high-risk status on the two screeners. Sensitivity, specificity, positive predictive value, and negative predictive value were also calculated by high-risk status for both screeners for any disorder, any internalizing disorder, any externalizing disorder, and each individual disorder. Receiver operating characteristic (ROC) curves were plotted and the area under the curve (AUC) was calculated to predict the presence of any disorder, any internalizing disorder, and any externalizing disorder by the BASC-2-BESS t-score and the SDQ total and subscale (emotion, conduct, hyperactivity, peer, prosocial, and impact) scores.

Results

The study sample's demographic characteristics are displayed in Table 1. Among the 18,107 students who were screened in Stage 1 with both the BASC-2-BESS and SDQ, 15.9% screened as high-risk on both measures, 4.1% screened high on the BASC-2-BESS only, and 6.6% screened high on the SDQ only (Table 2). A higher percentage of males screened high on both screeners than females at both the elementary and middle/high school levels.

A higher percentage of students at the Colorado and Florida sites screened high on both screeners than at the Ohio and South Carolina sites.

Of those with both BASC-2-BESS and SDQ screening data and who participated in Stage 2 (n=1,054), 267 students (25.3%) met criteria for at least one externalizing or internalizing disorder. Of these students, 59.2% screened as high-risk on either the BASC-2-BESS or SDQ (Table 3). Among those who met criteria for an internalizing disorder, 46.0% screened high on either the BASC-2-BESS or SDQ. Among those who met criteria for an externalizing disorder, 68.0% screened high on either screener. Among those meeting criteria for one or more of the 12 disorders included in this study, an equal or higher percentage of students screened high on the SDQ than on the BASC-2-BESS. The disorders with the highest percentage of students who screened high on either measure were mania/hypomania (n = 5; 100.0%), ADHD (n = 88; 87.1%) and conduct disorder (n = 18; 72.0%), and the disorders with the lowest percentage of students who were in the high-risk group were generalized anxiety (n = 5; 26.3%), panic disorder (n = 3; 33.3%), and social phobia (n = 13; 36.1%). Of those who screened high on the BASC-2-BESS and/or SDQ, 40.1% met criteria for at least one disorder (compared to 16.5% of those who screened low on both; data not shown). Of those who screened high on the BASC-2-BESS regardless of screening status on the SDQ, 40.1% met criteria for at least one disorder (compared to 19.9% of low screens). Of those who screened high on the SDQ regardless of screening status on the BASC-2-BESS, 41.1% met criteria for at least one disorder (compared to 17.2% of low screens).

The sensitivity, specificity, positive predictive value, and negative predictive value for the BASC-2-BESS and the SDQ in predicting the presence of any disorder, any internalizing disorder, any externalizing disorder, and each of the individual disorders are presented in Table 4. The sensitivity was the same or higher for the SDQ than the BASC-2-BESS for all disorders and disorder groupings, while the specificity was higher for the BASC-2-BESS for all disorders and disorder groupings. Positive predictive values for both measures ranged from 0.6% to 41.1%, and negative predictive values were above 80% for all disorders and disorder groupings by both measures. The AUC for the prediction of meeting criteria for any disorder among all students was .69 for both the BASC-2-BESS and the SDQ total score (Table 5). The AUC for the SDQ subscales ranged from .57 for the emotion subscale to .70 for the hyperactivity subscale. The AUC for the total BASC-2-BESS in predicting any disorder was higher for elementary school students (.72) than for middle (.64) and high school students (.65). The AUC for the total SDQ predicting any disorder was higher for elementary (.72) and middle school (.67) students than for high school students (.60). The AUCs were higher for the prediction of any externalizing disorder among elementary school and middle school students than for the prediction of any internalizing disorder at any level, whereas the AUCs for internalizing and externalizing disorders were similar for high school students. The AUCs were higher for males than females for the prediction of any disorder and any internalizing disorder for both the BASC-2-BESS and the SDQ total score, whereas the AUCs were relatively similar for males and females for the prediction of any externalizing disorder (Table 6). Figure 2 presents the receiver operating characteristic (ROC) curves for the BASC-2-BESS and SDQ total scores prediction of the presence of any disorder, any internalizing disorder, and any externalizing disorder. The ROC curves show

the larger area under the curve for the prediction of externalizing disorders than internalizing disorders for both measures.

Discussion

Using a large school-based sample of students in grades K–12, we examined the predictive utility of the teacher-report BASC-2-BESS and the SDQ (including subscales and impact supplement items) in identifying students meeting criteria for a mental disorder, and with specific types of mental health problems (internalizing, externalizing). Consistent with other studies, the screening data suggest that approximately a quarter of students may be struggling with social, emotional, or behavioral problems and that teachers report higher problems in males than females, in both elementary and secondary samples (Costello et al. 2005, Merikangas, He, Brody et al., 2010; Merikangas, He, Burstein, et al., 2010). Broadly speaking, our findings suggest that teacher versions of the BASC-2-BESS and SDQ have modest utility in identifying children meeting criteria for a mental disorder based on parent report. As hypothesized, the AUCs representing the prediction of any externalizing disorder (.73 for both measures) were higher than the AUCs predicting any internalizing disorder (.58 for both measures). The strongest true positive rates were found for those meeting criteria for mania and ADHD, where 100.0% and 85.2% of those meeting criteria for these disorders, respectively, were detected by a high screen on the SDQ total score. Otherwise, true positive rates were below 65% for all other disorders besides mania and ADHD by both the SDQ and the BASC-2-BESS. It should also be noted that there were few cases of mania (n=5), which may limit the generalizability of 100% true positive rate. The AUCs suggested a slight advantage in detecting any disorder or internalizing disorders in males relative to females, whereas the AUCs were similar between males and females for externalizing disorders. The high specificity and negative predictive values suggest that low scores on either measure offers confidence in a low likelihood of a child meeting criteria for a parent-reported mental disorder. Conversely, the low to moderate sensitivity values indicate that a considerable percentage of students with a mental disorder will not be identified by either of the two measures.

On one hand, the overall modest predictive utility may not be surprising given the cross-informant methodology, as there is less shared variance between teacher screening scores and parent report of mental health problems, as compared to when teacher screening scores are used to predict other teacher-based measures (e.g., Kamphaus & Reynolds, 2007; Owens et al., 2016). Further, the procedures used in this study are akin to Tier 1 universal screenings in a school-based community sample, where false positives are to be expected (McCormick et al., 2009). On the other hand, the modest utility suggests caution may be warranted when using the teacher version of BASC-2-BESS or SDQ to indicate the likelihood of a child meeting criteria for a mental disorder. Because obtaining teacher report is aligned with evidence-based assessment procedures for most childhood problems (Youngstrom et al., 2020), additional research on the predictive utility of teacher-based screening tools could offer further evidence for school-based screening efforts. Below, we interpret the findings and discuss implications for research and practice.

Elementary School Level

As hypothesized, the pattern for both the BASC-2-BESS and SDQ total scores suggests that the prediction of the presence of any disorders (AUCs = .72 for both measures) and any externalizing disorders (AUCs = .75 for both measures) was strongest at the elementary school level. This likely represents the relatively greater amount of time teachers spend with students at the elementary school level relative to the middle and high school levels. The AUCs detected in this study are slightly lower than those found in the Owens et al. (2016) study and those found in the Goodman and Goodman (2011) study. However, Owens and colleagues were predicting school outcomes (e.g., GPA, daily behavior notes) determined and/or observed by the same teacher who completed the screening tool, which may have shared variance. Findings may differ from Goodman and Goodman because they used different SDQ cut-points to define risk status. Although the AUCs are modest (.75 for both measures), these data suggest that the BASC-2-BESS and SDQ total scores may be adequate screeners to detect which elementary school students are likely to meet criteria for an externalizing disorder. In contrast, the AUCs for the BASC-2-BESS and SDQ total scores detecting the presence of an internalizing disorder at elementary school were low and not much better than a chance prediction (AUCs = .59 for both measures), suggesting caution in using scores from either measure if detection of internalizing disorders is the goal.

When predicting any disorder at the elementary level, there were no benefits to using the SDQ subscale scores relative to the SDQ total score, though the impact score was typically more predictive than most of the other subscale scores. Given that completion of the impact scores take less time to complete than the full SDQ (25 items), additional research on the psychometric properties of the impact scores could inform screening implementation practices. Teachers may be better able to detect functional impairment than any given set of symptoms. Only a few studies have examined the predictive utility of teacher-rated SDQ impact scores (e.g., Lehmann et al., 2014), yet this may be a fruitful area for future research and for screening purposes, as it may increase feasibility to have teachers rate impairment.

Secondary School Level

For middle and high school students, the pattern of results showed some variability. Namely, for externalizing disorders, both the BASC-2-BESS and the SDQ totals had slightly larger AUCs for middle school students (AUCs = .71 and .74, respectively) than for high school students (AUCs = .67 and .59, respectively), with the SDQ hyperactive subscale (.75) and impact scores (.74) having the highest AUCs for middle school students. For internalizing scores, the pattern was the opposite. We found higher AUCs for high school students (.64 and .62, respectively) than middle school students (.53 and .56, respectively), although the AUCs were low for both groups of students. This relative difference between internalizing and externalizing problems may be associated with the relative increase in the manifestation of internalizing problems as adolescents age from pre-teens to teenagers (Merikangas, He, Burstein, et al., 2010), as well as the change in the presentation of some externalizing problems, during adolescence, like for ADHD (e.g., hyperactivity manifests more as internal restlessness than external fidgeting (e.g., Turgay et al., 2012)). Despite this relative pattern, the AUCs predicting internalizing problems in secondary students are quite modest. However, similar to the utility for elementary school students, the SDQ may be

a useful screener for middle school students when interested in risk for an externalizing disorder. As recommended by Margherio and colleagues (2019) and others (Aebi et al., 2017), adolescent self-report may be more helpful than teacher report for detection of internalizing problems in middle and high school students.

Implications

This is the first study to examine utility of two teacher-based screening tools in the same sample to identify mental health problems in elementary and secondary school students in the United States. First, understanding the comparative utility of the BASC-2-BESS and SDQ is important, as screening is a critical part of Tier 1 services within multi-tier frameworks (i.e., part of the process of identifying and delivering services to students in need). Screening may be particularly important to consider at significant timepoints throughout development. For instance, universal screening tools may help address the needs of students with disruptive behavior problems in elementary school, as a number of evidence-based practices (such as classroom behavioral interventions) are available for students in this age range (Evans et al., 2018). Similarly, screening for internalizing problems in late elementary, middle, and high school allows school personnel to determine who may benefit most from Tier 2 (targeted services) for adjustment problems, anxiety, and/or depression. Catching at-risk students at these critical junctures provides rationale for research into which screeners are most effective at identifying students in need of services.

Second, when using the BASC-2-BESS with elementary school students or the SDQ as a screener with elementary or middle school students, the scores have moderate utility in detecting those with externalizing manifestations (ADHD and conduct disorder) and in ruling out the likelihood of any disorders. Thus, scores from the BASC-2-BESS or SDQ may serve as a reasonable starting point to consider students who may need additional assessment. Indeed, gated screening procedures (i.e., universal screening then targeted follow-up of those deemed high risk by initial screening results) may be warranted to enhance diagnostic certainty (Kilgus et al., 2016). Another approach to enhancing understanding of a diagnostic likelihood is to use a probability nomogram method. A probability nomogram allows the user to leverage the properties of the measure (e.g., sensitivity and specificity) and local base rates within a setting of interest to enhance understanding of diagnostic probability (see Van Meter, 2020 for discussion). Nomograms can be utilized to predict individual risk level and can also be leveraged to integrate multiple sources of information, which can be useful for assessment purposes since cross-informant reports are often relied upon for identifying concerns in children and adolescents. Tools such as nomograms that take into account more data (i.e., multiple informant reports) and context-specific information (i.e., setting-specific base rates) to inform diagnostics have the potential to contribute in a meaningful way to our understanding of how to more accurately identify at-risk students.

Third, in comparing these two screeners, it is important to note that the SDQ is free and publicly available and the BASC-2-BESS is not. Given that this study did not identify any incremental utility of the BASC-2-BESS over the SDQ, this price differential may be a consideration for those choosing which screener to use. Because cost may be a

barrier to using screening tools for some schools, research on the comparative utility of the SDQ with other publicly available measures could further inform school decision making. A developing list of cost-free measures can be found at https://en.wikiversity.org/wiki/Evidence-based_assessment/Which_Questionnaire_Should_I_Use%3f. In addition, since the data collection for this study, several additional social, emotional, and behavioral screening tools with adequate psychometric properties have been developed, including the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS; Kilgus et al., 2016) and the Social Skills Improvement System Social and Emotional Learning Brief Scales–Student Form (SSIS SEL*b*-S; Anthony et al., 2021). A newer version of the BASC-BESS was also published in 2015 (midway through the completion of this study). Items on the BASC-3-BESS are similar to those on the BASC-2-BESS, and there is currently no published evidence to suggest that the two versions are not comparable in terms of prediction (Kettler, et al., 2017). However, testing the comparative utility of each of these newer measures in the prediction of mental disorders could offer additional information on the relative validity of available screeners. Similarly, various cut-off scores and algorithms that include SDQ symptom and impact scores from teacher and/or parent report for identifying risk status (e.g., Goodman et al., 2004; Nielsen et al., 2019), could be examined to determine if these methods improve the predictive utility of the SDQ scores and/or help school personnel make decisions about identification and referral rates.

Additionally, when selecting a mental health screening measure, it is important to choose a culturally-relevant tool that has been validated and normed on a sample similar to the district's student population. The SDQ is available in over 75 languages and is one of the mostly widely-used mental health measures in the world with many published works originating from countries outside the U.S. (https://sdqinfo.org). Thus, especially when working with diverse student populations, this may be an important consideration for which screener to use.

Lastly, we acknowledge that obtaining teacher report is aligned with evidence-based assessment procedures for most childhood behavioral and emotional problems (Youngstrom et al., 2020). Thus, it may be beneficial for school personnel, primary care physicians, and mental health care providers to seek out teacher-based ratings, as they can offer critical information about child symptoms and functioning. If educators or providers have access to BASC-2-BESS or SDQ scores for elementary or middle school students, this study suggests that the scores may offer information about the child's risk for externalizing problems that could add information to parent report, though it is important to remember that rating scale data from any informant simply represents one piece of the diagnostic picture (Youngstrom et al., 2020).

Limitations

This study was subject to a number of limitations. First, although the DSM-5 was published in 2013, this study relied on DSM-IV-based case definitions and may not reflect epidemiological patterns for the more recent diagnostic criteria. However, the criteria for disorders considered for this study were comparable in the latest version of the DSM-5 (APA, 2013). Second, the case definitions for each disorder were based primarily on

parent responses to a structured interview and may not match diagnoses given by clinical judgement nor have they been compared against medical records. The study case definition for ADHD also included criteria based on data collected in the screener, potentially affecting the independence of the screener to predict disorder status. Relatedly, the diagnostic interview may not identify children with a mental disorder who are adequately treated. Third, because time elapsed between when the teacher screener was administered and when the parent diagnostic interview was completed, the student's symptoms and impairment may have changed over this interval (e.g., if the student received school- or community-based mental health treatment during this time). Further, changes in symptoms and impairment over time (with or without treatment) might have occurred more frequently for internalizing disorders than externalizing disorders given the episodic nature of some internalizing disorders (e.g., depression) versus the chronicity of externalizing disorders (e.g., ADHD; American Psychiatric Association, 2013). This might account, in part, for differences in prediction of the disorder categories. Fourth, the response rates for Stage 2 were relatively low, which may have resulted in some participation bias. Fifth, we acknowledge that the primary purpose of a universal screener is to identify students at risk who may benefit from services (rather than the prediction of a mental health disorder). Nonetheless, it is important to understand the utility of these measures in predicting disorder presence, as it is important to understand the utility of these measures to predict a variety of key youth outcomes. Finally, the data in this analysis were treated as a convenience sample and the results may not be generalizable to other populations with different demographic characteristics. Relatedly, this study focused on the psychometrics of the BASC-2-BESS and SDQ for the overall study population and by child sex and school level. It will be important for future work to delve into other characteristics that may exhibit differential levels of predictive utility, such as by race and ethnicity.

Conclusions

In this study we examined comparative utility of the teacher versions of the BASC-2-BESS and SDQ screening tools in identifying parent-reported mental health problems in elementary and secondary school students in the United States. Results suggest that teacher versions of the BASC-2-BESS and SDQ have modest utility in identifying children meeting criteria for a mental disorder based on parent report. AUCs representing the prediction of any externalizing disorder were higher than the AUCs predicting any internalizing disorder and higher for elementary school students than secondary school students. Low scores on either measure have utility in ruling out the likelihood of any disorder. Findings have implications for the utility of effective universal screening tools. If interpreted cautiously and used as a part of a larger process to identify children with mental disorders, teacher report can have benefits for maximizing the potential of improving student outcomes (Eklund et al., 2009). Studies like this one can help improve the identification of at-risk students in need of school-based services by informing the selection of measures used for implementation of school-based mental health screening.

Acknowledgements:

The authors would like to acknowledge the work of Rebecca Bitsko, Angelika Claussen, and Joseph Holbrook to inform the conceptualization of this study and provide feedback on the manuscript, and Maria Gonzalez for analytic

support in finalizing the study results. This study was supported by the Disability Research and Dissemination Center (DRDC) cooperative agreement U01DD001007, which was funded by the Centers for Disease Control and Prevention (CDC).

References

Aebi M, Kuhn C, Banaschewski T, Grimmer Y, Poustka L, Steinhausen H-C, & Goodman R (2017). The contribution of parent and youth information to identify mental health disorders or problems in adolescents. Child & Adolescent Psychiatry & Mental Health, 11, 1–12. Doi: 10.1186/s13034-017-0160-9 [PubMed: 28077965]

American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders, 4th edn. American Psychiatric Association, Washington, DC.

American Psychiatric Association. (2013). DSM-5 fact sheets. Psychiatry.org – DSM-5 Fact Sheets. Retrieved May 4, 2022, from https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/dsm-5-fact-sheets

Anthony CJ, Brann KL, Elliott SN, & Garis EJ (2021). Examining the structural validity of the SSIS SEL brief scales—Teacher and student forms. Psychology in the Schools, 59(2), 260–280. Doi: 10.1002/pits.22607

Bitsko RH, Claussen AH, Lichstein J, Black LI, Jones SE, Danielson ML, Hoenig JM, … Ghandour RM (2022). Mental health surveillance among children — United States, 2013–2019. MMWR Supplement, 71(Suppl-2), 1–42. Doi: 10.15585/mmwr.su7102a1

Bosik SW, Harris B, Alibrahim N, & Donovan C (2021). Construct validity of the teacher version of the Strengths and Difficulties Questionnaire in a sample of elementary-aged children in the United States. Contemporary School Psychology, 1–17. Doi: 10.1007/s40688-021-00371-8 [PubMed: 33589870]

Chin JK, Dowdy E, & Quirk MP (2013). Universal screening in middle school: Examining the Behavioral and Emotional Screening System. Journal of Psychoeducational Assessment, 31(1), 53–60. Doi: 10.1177/0734282912448137

Costello EJ, Egger H, & Angold A (2005). 10-year research update review: The epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. Journal of the American Academy of Child and Adolescent Psychiatry, 44(10), 972–986. Doi: 10.1097/01.chi.0000172552.41596.6f. [PubMed: 16175102]

Danielson ML, Bitsko RH, Holbrook JR, Charania SN, Claussen AH, McKeown RE, … Flory K (2021). Community-based prevalence of externalizing and internalizing disorders among school-aged children and adolescents in four geographically dispersed school districts in the United States. Child Psychiatry and Human Development, 52(3), 500–514. Doi: 10.1007/s10578-020-01027-z [PubMed: 32734339]

Dever BV, Mays KL, Kamphaus RW, & Dowdy E (2012). The factor structure of the BASC-2 Behavioral and Emotional Screening System teacher form, child/adolescent. Journal of Psychoeducational Assessment, 30(5), 488–495. Doi: 10.1037/pas0000458

Dowdy E, Harrell-Williams L, Dever BV, Furlong MJ, Moore S, Raines T, & Kamphaus RW (2016). Predictive validity of a student self-report screener of behavioral and emotional risk in an urban high school. School Psychology Review, 45(4), 458–476. Doi: 10.17105/SPR45-4.458-476

Eklund K, Renshaw TL, Dowdy E, Jimerson SR, Hart SR, Jones CN, & Earhart J (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. California School Psychologist, 14, 89–95. Doi: 10.1007/BF03340954

Eklund K, Kilgus S, von der Embse N, Beardmore M, & Tanner N (2017). Use of universal screening scores to predict distal academic and behavioral outcomes: A multilevel approach. Psychological Assessment, 29(5), 486–499. Doi: 10.1037/pas0000355 [PubMed: 27504906]

Ellickson PL, & Hawes JA (1989). An assessment of active versus passive methods for obtaining parental consent. Evaluation Review, 13(1), 45–55. Doi: 10.1177/0193841X8901300104 [PubMed: 11659371]

Erskine HE, Norman RE, Ferrari AJ, Chan GCK, Copeland WE, Whiteford HA, & Scott JG (2016). Long-term outcomes of attention-deficit/hyperactivity disorder and conduct disorder: A systematic

review and meta-analysis. Journal of the American Academy of Child and Adolescent Psychiatry, 55(10), 841–850. Doi: 10.1016/j.jaac.2016.06.016 [PubMed: 27663939]

Esch P, Bocquet V, Pull C, Couffignal S, Lehnert T, Graas M, … Ansseau M (2014). The downward spiral of mental disorders and educational attainment: a systematic review on early school leaving. BMC Psychiatry, 14(1), 237. Doi: 10.1186/s12888-014-0237-4 [PubMed: 25159271]

Essex MJ, Kraemer HC, Slattery MJ, Burk LR, Boyce WT, Woodward HR, & Kupfer DJ (2009). Screening for childhood mental health problems: Outcomes and early identification. Journal of Child Psychology and Psychiatry, 50(5), 562–570. Doi: 10.1111/j.1469-7610.2008.02015.x [PubMed: 19432682]

Evans SW, Owens JS, Wymbs BT, & Ray AR (2018). Evidence-based psychosocial treatments for children and adolescents with attention deficit/hyperactivity disorder. Journal of Clinical Child and Adolescent Psychology, 47(2), 157–198. Doi: 10.1080/15374416.2017.1390757 [PubMed: 29257898]

Garcia-Rosales A, Vitoratou S, Faraone SV, Rudaizky D, Banaschewski T, Asherson P … Chen W. (2021). Differential utility of teacher and parent–teacher combined information in the assessment of attention deficit/hyperactivity disorder symptoms. European Child & Adolescent Psychiatry, 30(1), 143–153. Doi: 10.1007/s00787-020-01509-4 [PubMed: 32246275]

Goodman A, & Goodman R (2011). Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. Journal of Child Psychology & Psychiatry, 52(1), 100–108. Doi: 10.1111/j.1469-7610.2010.02278.x [PubMed: 20722879]

Goodman A, Lamping DL, & Ploubidis GB (2010). When to use broader internalizing and externalizing subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. Journal of Abnormal Child Psychology, 38(8), 1179–1191. Doi: 10.1007/s10802-010-9434-x [PubMed: 20623175]

Goodman R (2001). Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child and Adolescent Psychiatry, 40(11), 1337–1345. Doi: 10.1097/00004583-200111000-00015 [PubMed: 11699809]

Goodman R (1997). The strengths and difficulties questionnaire: A research note. Journal of Child Psychology and Psychiatry, 38(5), 581–586. Doi: 10.1111/j.1469-7610.1997.tb01545.x [PubMed: 9255702]

Goodman R, Ford T, Corbin T, & Meltzer H (2004). Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. European Child & Adolescent Psychiatry, 13(Suppl 2), ii25–ii31. doi: 10.1007/s00787-004-2005-3 [PubMed: 15243783]

Goodman R, & Scott S (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? Journal of Abnormal Child Psychology, 27(1), 17–24. Doi: 10.1023/a:1022658222914 [PubMed: 10197403]

Grover RL, Ginsburg GS, & Ialongo N (2007). Psychosocial outcomes of anxious first graders: A seven-year follow-up. Depression and Anxiety, 24(6), 410–420. Doi: 10.1002/da.20241 [PubMed: 17041929]

Hall CL, Guo B, Valentine AZ, Groom MJ, Daley D, Sayal K, & Hollis C (2019). The validity of the Strengths and Difficulties Questionnaire (SDQ) for children with ADHD symptoms. PloS one, 14(6). Doi: 10.1371/journal.pone.0218518

Harrell-Williams LM, Raines TC, Kamphaus RW, & Dever BV (2015). Psychometric analysis of the BASC–2 Behavioral Emotional Screening System (BESS) Student Form: Results from high school student samples. Psychological Assessment 27, 738–743. Doi: 10.1037/pas0000079 [PubMed: 25642926]

Harrison JR, Vannest KJ, & Reynolds CR (2013). Social acceptability of five screening instruments for social, emotional, and behavioral challenges. Behavioral Disorders, 38(3), 171–189. Doi: 10.1177/019874291303800305

He J-P, Burstein M, Schmitz A, & Merikangas KR (2013). The strengths and difficulties questionnaire (SDQ): The factor structure and scale validation in U.S. adolescents. Journal of Abnormal Child Psychology, 41(4), 583–595. Doi: 10.1007/s10802-012-9696-6 [PubMed: 23183936]

Headley C, & Campbell M (2011). Teachers' recognition and referral of anxiety disorders in primary school children. Australian Journal of Educational and Developmental Psychology, 11, 78–90.

Higa-McMillan CK, Francis SE, Rith-Najarian L, & Chorpita BF (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. Journal of Clinical Child & Adolescent Psychology, 45(2), 91–113. Doi: 10.1080/15374416.2015.1046177 [PubMed: 26087438]

Hill LG, Lochman JE, Coie JD, & Greenberg MT (2004). Effectiveness of early screening for externalizing problems: Issues of screening accuracy and utility. Journal of Consulting and Clinical Psychology, (5), 809–820. Doi: 10.1037/0022-006X.72.5.809 [PubMed: 15482039]

Hoza B (2007). Peer functioning in children with ADHD. Journal of Pediatric Psychology, (6), 655–663. Doi: 10.1093/jpepsy/jsm024 [PubMed: 17556400]

Jaycox LH, Stein BD, Paddock S, Miles JNV, Chandra A, Meredith LS, … Burnam MA (2009). Impact of teen depression on academic, social, and physical functioning. Pediatrics, 124(4), e596–e605. Doi: 10.1542/peds.2008-3348 [PubMed: 19736259]

Kamphaus RW, Thorpe JS, Winsor AP, Kroncke AP, Dowdy ET, & Vandeventer MC (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. Educational and Psychological Measurement, 67(2), 342–356. Doi: 10.1177/0013164407067002 1001

Kamphaus RW & Reynolds CR (2007). BASC-2 Behavioral and Emotional Screening System Manual. Bloomington, MN: Pearson Clinical Assessment.

Kent K, Pelham W, Molina B, Sibley M, Waschbusch D, Yu J, … Karch KM (2011). The academic experience of male high school students with ADHD. Journal of Abnormal Child Psychology, 39(3), 451–462. Doi: 10.1007/s10802-010-9472-4 [PubMed: 21103923]

Kessler RC (2012). The costs of depression. Psychiatric Clinics, 35(1), 1–14. Doi: 10.1016/j.psc.2011.11.005 [PubMed: 22370487]

Kettler RJ, Feeney-Kettler KA, & Dembitzer L (2017). Social, emotional, and behavioral screening: A comparison of two measures and two methods across informants. Journal of School Psychology, 64, 93–108. doi: 10.1016/j.jsp.2017.05.002 [PubMed: 28735610]

Kilgus SP, Eklund K, Nathaniel P, Taylor CN, & Sims WA (2016). Psychometric defensibility of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) Teacher Rating Scale and multiple gating procedure within elementary and middle school samples. Journal of school psychology, 58, 21–39. doi: 10.1016/j.jsp.2016.07.001 [PubMed: 27586068]

King K, Reschly AL, & Appleton JJ (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. doi: 10.1177/0734282912440673

Kuhn C, Aebi M, Jakobsen H, Banaschewski T, Poustka L, Grimmer Y, … Steinhausen H-C (2017). Effective mental health screening in adolescents: Should we collect data from youth, parents or both? Child Psychiatry & Human Development, 48(3), 385–392. doi: 10.1007/s10578-016-0665-0 [PubMed: 27363421]

Langley AK, Bergman RL, McCracken J, & Piacentini JC (2004). Impairment in childhood anxiety disorders: Preliminary examination of the child anxiety impact scale–parent version. Journal of Child and Adolescent Psychopharmacology, 14(1), 105–114. doi: 10.1089/104454604773840544 [PubMed: 15142397]

Lebrun-Harris LA, Ghandour R, Kogan MD & Warren MD (2022). Five-year trends in US children's health and well-being, 2016–2020. JAMA Pediatrics, 176(7). doi: 10.1001/jamapediatrics.2022.0056

Lehmann S, Heiervang ER, Havik T, & Havik OE (2014). Screening foster children for mental disorders: properties of the Strengths and Difficulties Questionnaire. PLoS ONE, 9(7), 1–11. doi: 10.1371/journal.pone.0102134

Long P, Forehand R, Wierson M, & Morgan A (1994). Does parent training with young noncompliant children have long-term effects? Behaviour Research and Therapy, 32(1), 101–107. doi: 10.1016/0005-7967(94)90088-4 [PubMed: 8135705]

Margherio SM, Evans SW, & Owens JS (2019). Universal screening in middle and high schools: Who falls through the cracks. School Psychology Quarterly, 34(6), 591–602. Doi: 10.1037/spq0000337

Masten AS, Glenn IR, ., Jeffrey D., L., Keith B, B., Jelena O, Jennifer R,R., … Auke T. (2005). Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. Developmental Psychology, (5), 733–746. Doi: 10.1037/0012-1649.41.5.733. [PubMed: 16173871]

McCormick E, Thompson K, Vander Stoep A, & McCauley E (2009). The case for school-based depression screening: Evidence from established programs. Report on Emotional & Behavioral Disorders in Youth, 9(4), 91–96. [PubMed: 26451134]

Merikangas KR, He JP, Brody D, Fisher PW, Bourdon K, & Koretz DS (2010). Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. Pediatrics, 125(1), 75–81. Doi: 10.1542/peds.2008-2598 [PubMed: 20008426]

Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, … Swendsen J (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). Journal of the American Academy of Child and Adolescent Psychiatry, 49(10), 980–989. Doi: 10.1016/j.jaac.2010.05.017 [PubMed: 20855043]

Moilanen KL, Shaw DS, & Maxwell KL (2010). Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. Development & Psychopathology, 22(3), 635–653. Doi: 10.1017/S0954579410000337 [PubMed: 20576184]

Nielsen LG, Rimvall MK, Clemmensen L, Munkholm A., Elberling H., Olsen EM. …Jeppesen P. (2019). The predictive validity of the Strengths and Difficulties Questionnaire in preschool age to identify mental disorders in preadolescence. PLOS One, 4(6). doi: 10.1371/journal.pone.0217707

Owens JS, Goldfine ME, Evangelista NM, Hoza B, & Kaiser NM (2007). A critical review of self-perceptions and the positive illusory bias in children with ADHD. Clinical Child and Family Psychology Review, 10(4), 335–351. Doi: 10.1007/s10567-007-0027-3 [PubMed: 17902055]

Owens JS, Holdaway AS, Serrano VJ, Himawan LK, Watabe Y, Storer J, … Andrews N (2016). Screening for social, emotional, and behavioral problems at kindergarten entry: The utility of two teacher rating scales. School Mental Health, 8(3), 319–331. Doi: 10.1007/s12310-016-9176-1

Renshaw TL, Eklund K, Dowdy E, Jimerson SR, Hart SR, Earhart J Jr., & Jones CN (2009). Examining the relationship between scores on the “Behavioral and Emotional Screening System” and student academic, behavioral, and engagement outcomes: An investigation of concurrent validity in elementary school. California School Psychologist, 14, 81–88. Doi: 10.1007/BF03340953

Rimestad ML, Lambek R, Zacher Christiansen H, & Hougaard E (2019). Short-and long-term effects of parent training for preschool children with or at risk of ADHD: a systematic review and meta-analysis. Journal of Attention Disorders, 23(5), 423–434. doi: 10.1177/1087054716648775 [PubMed: 27179355]

Robb JA, Sibley MH, Pelham WE, Foster EM, Molina BS, Gnagy EM, & Kuriyan AB (2011). The estimated annual cost of ADHD to the US education system. School Mental Health, 3(3), 169–177. doi: 10.1007/s12310-011-9057-6 [PubMed: 25110528]

Schanding GT Jr., . & Nowell KP. (2013). Universal screening for emotional and behavioral problems: Fitting a population-based model. Journal of Applied School Psychology, 29(1), 104–119. doi: 10.1080/15377903.2013.751479

Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME (2000). NIMH diagnostic interview schedule for children version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry, 39(1), 28–38. doi: 10.1097/00004583-200001000-00014 [PubMed: 10638065]

Splett JW, Garzona M, Gibson N, Wojtalewicz D, Raborn A, & Reinke WM (2019). Teacher recognition, concern, and referral of children’s internalizing and externalizing behavior problems. School Mental Health, 11(2), 228–239. doi: 10.1016/s0031-3955(01)00014-1

Stoiber KC, & Gettinger M (2016). Multi-tiered systems of support and evidence-based practices. In Jimerson SR, Burns MK, & VanDerHeyden AM(Eds.), Handbook of Response to Intervention: The Science and Practice of Multi-Tiered Systems of Support (pp. 121–141). New York, NY: Springer.

Stoep AV, McCauley E, Thompson KA, Herting JR, Kuo ES, Stewart DG, … Kushner S (2005). Universal emotional health screening at the middle school transition. Journal of Emotional

and Behavioral Disorders, 13(4), 213–223. doi: 10.1177/10634266050130040301 [PubMed: 21430789]

Suldo S, Thalji A, & Ferron J (2011). Longitudinal academic outcomes predicted by early adolescents' subjective well-being, psychopathology, and mental health status yielded from a dual factor model. The Journal of Positive Psychology, 6(1), 17–30. doi: 10.1080/17439760.2010.536774

Tobia V, Greco A, Steca P, & Marzocchi GM (2019). Children's wellbeing at school: A multi-dimensional and multi-informant approach. Journal of Happiness Studies, 20(3), 841–861. doi: 10.1007/s10902-018-9974-2

Turgay A, Goodman DW, Asherson P, Lasser RA, Babcock TF, Pucci ML, & Barkley R (2012). Lifespan persistence of ADHD: The life transition model and its application. Journal of Clinical Psychiatry, 73(2), 192–201. doi: 10.4088/JCP.10m06628 [PubMed: 22313720]

Van Meter A (2020). The prediction phase of evidence-based assessment. In Youngstrom EA, Princestein MJ, Mash EJ, and Barkley RA (Eds). Assessment of disorders in childhood and adolescents, 5th ed. (pp.30–48). New York, NY: The Guilford Press.th

Wanga V, Danielson ML, Bitsko RH, Holbrook JR, Lipton C, Claussen AH, … Flory K (2022). Stability of mental disorder prevalence estimates among school-aged children and adolescents: Findings from the community-based Project to Learn About Youth-Mental Health (PLAY-MH) and Replication-PLAY-MH (Re-PLAY-MH), 2014–2017. Annals of Epidemiology, 72, 82–90. doi: 10.1016/j.annepidem.2022.05.007 [PubMed: 35661706]

Weersing VR, Jeffreys M, Do MCT, Schwartz KT, & Bolano C (2017). Evidence base update of psychosocial treatments for child and adolescent depression. Journal of Clinical Child & Adolescent Psychology, 46(1), 11–43. doi: 10.1080/15374416.2016.1220310 [PubMed: 27870579]

Youngstrom EA., Prinstein MJ., Mash EJ, and Barkley RA (Eds.). (2020) Assessment of Disorders in Childhood and Adolescence, Fifth Edition. New York, NY: The Guilford Press.

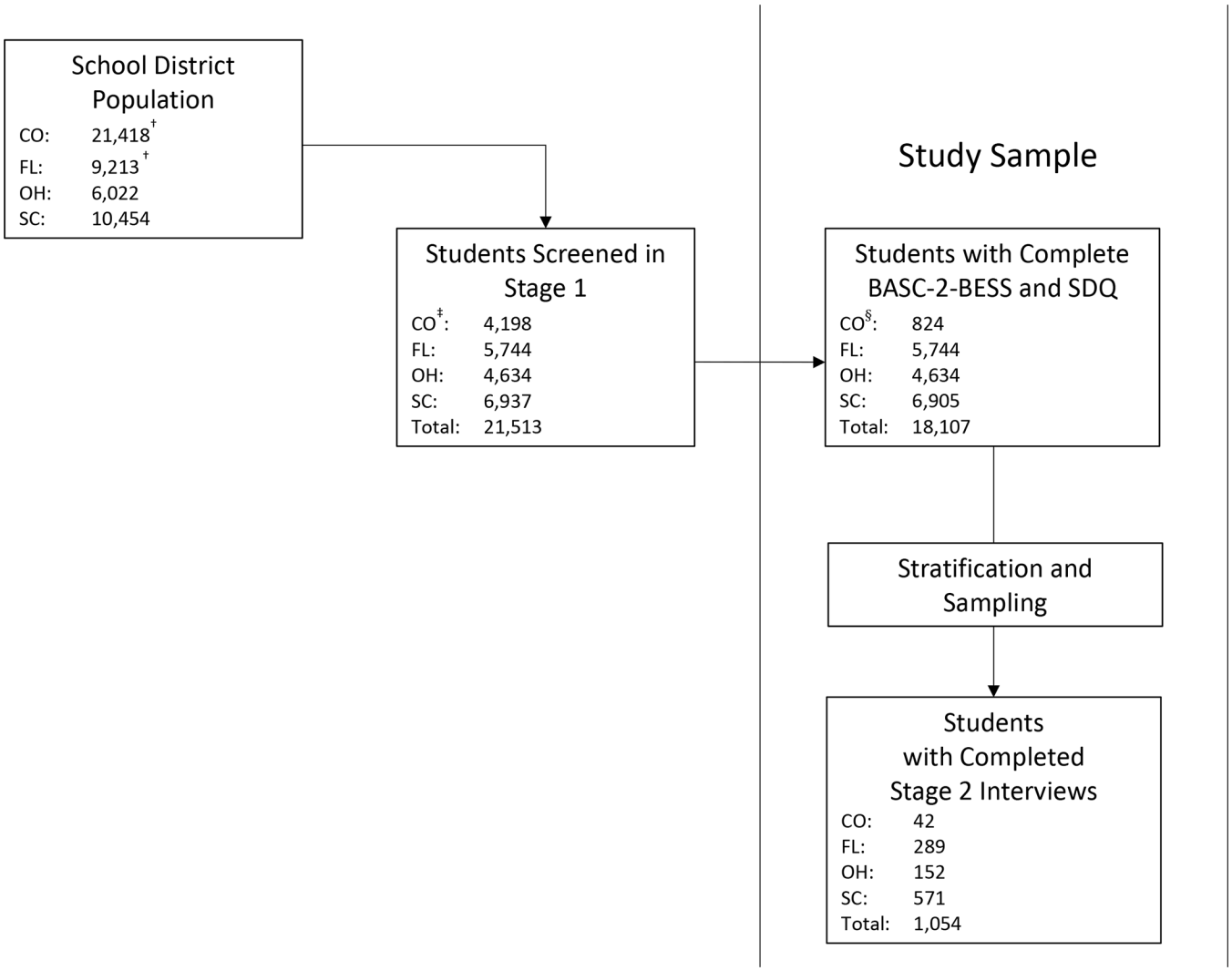


Figure 1:

Flow chart showing how the study population was derived by site

CO: Colorado; FL: Florida; OH: Ohio; SC: South Carolina

† Population in participating schools; not all schools in the CO or FL districts participated in the PLAY-MH study

‡ The Colorado site sampled students in participating schools for Stage 1 screening; all students in participating schools in other sites were generally eligible for Stage 1 (see Danielson, et al., 2021 for details).

§ Colorado site collected the SDQ on only a subset of students in Stage 1; students with only BASC-2-BESS data were excluded from these analyses.

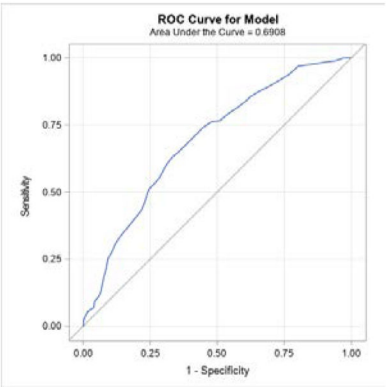
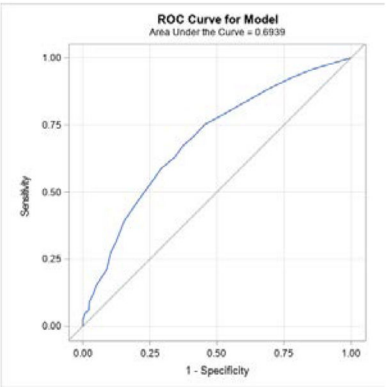
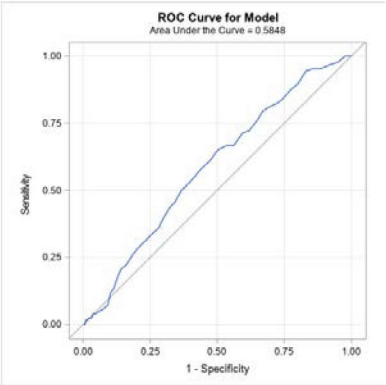
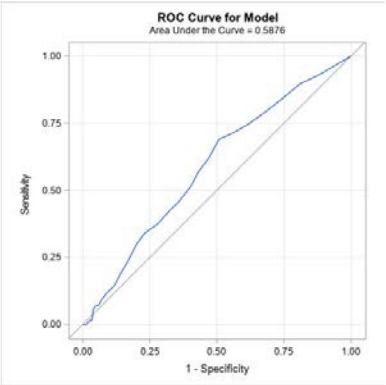
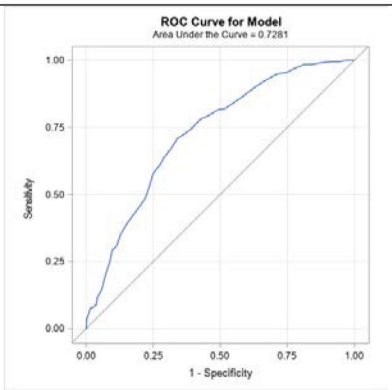
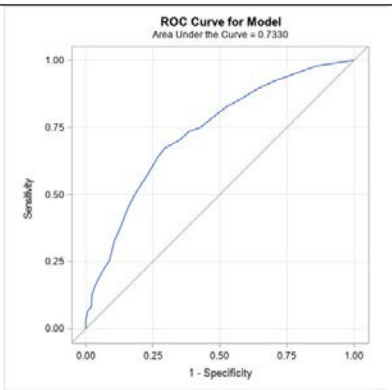
Any disorder	BASC-2-BESS	SDQ
		
Any internalizing disorder		
Any externalizing disorder		

Figure 2.

ROC curves showing BASC-2-BESS and SDQ total scores predicting any disorder, any internalizing disorder, and any externalizing disorder on the DISC, Project to Learn About Youth-Mental Health.

ROC=Receiver Operating Characteristic; BASC-2-BESS = Behavior Assessment System for Children-2 Behavioral and Emotional Screening System; SDQ = Strengths and Difficulties Questionnaire

Psychol Sch. Author manuscript; available in PMC 2024 January 21.

Table 1

Demographic characteristics of study participants with Stage 1 BASC-2-BESS and SDQ screening data and Stage 2 DISC-IV interview data, Project to Learn about Youth-Mental Health

	Stage 1 n (%)	**Stage 2 n (%)**
Sex		
Male	9417 (52.0%)	554 (52.6%)
Female	8690 (48.0%)	500 (47.4%)
Grade level at Stage 1[†]		
Elementary school (K–5th grade)	7367 (40.7%)	651 (61.8%)
Middle school (6th–8th grade)	4664 (25.8%)	216 (20.5%)
High school (9th–12th grade)	6076 (33.6%)	187 (17.7%)
Ethnicity/race		
Non-Hispanic White	11419 (63.1%)	576 (54.7%)
Non-Hispanic Black	4365 (24.1%)	323 (30.7%)
Hispanic	1318 (7.3%)	72 (6.8%)
Any other race/ethnicity	1005 (5.6%)	83 (7.9%)
Free/reduced price lunch		
Yes	9379 (51.8%)	614 (58.3%)
No	8728 (48.2%)	440 (41.8%)
Site		
Colorado[≠]	824 (4.6%)	42 (4.0%)
Florida	5744 (31.7%)	289 (27.4%)
Ohio	4634 (25.6%)	152 (14.4%)
South Carolina	6905 (38.1%)	571 (54.2%)

BASC-2-BESS= Behavior Assessment System for Children-2 Behavioral and Emotional Screening System; SDQ=Strengths and Difficulties Questionnaire; DISC-IV=Diagnostic Interview Schedule for Children-IV

[†]Child age was not collected at Stage 1, but elementary school (K–5th grade) typically includes students aged 5–13 years, middle school (6th–8th grade) typically includes students aged 11–15 years, and high school (9th–12th grade) typically includes students aged 14–19 years.

[≠]Teachers in the Colorado site completed the BASC-2-BESS for all screened students but the SDQ for only a subset of screened students; the data presented in these analyses include only Colorado students with both a completed BASC-2-BESS and a completed SDQ.

Table 2

Screening status on the SDQ and BASC-2-BESS by sex, school level, and site, Project to Learn About Youth-Mental Health

		Screened high on both BASC-2-BESS and SDQ	Screened high on BASC-2-BESS only	Screened high on SDQ only	Low screen on BASC-2-BESS and SDQ	
	n	%	%	%	%	Chi-square p-value
All	18,107	2,881 (15.9%)	733 (4.1%)	1,199 (6.6%)	13,294 (73.4%)	
Males	9,417	1,692 (18.0%)	253 (2.7%)	883 (9.4%)	6,589 (70.0%)	
Elementary school males	3,839	696 (18.1%)	61 (1.6%)	428 (11.2%)	2,654 (69.1%)	<0.0001†
Middle/high school males	5,578	996 (17.9%)	192 (3.4%)	455 (8.2%)	3,935 (70.5%)	
Females	8,690	1,189 (13.7%)	480 (5.5%)	316 (3.6%)	6,705 (77.2%)	
Elementary school females	3,528	439 (12.4%)	138 (3.9 %)	133 (3.8%)	2,818 (79.9%)	<0.0001†
Middle/high school females	5,162	750 (14.5%)	342 (6.6%)	183 (3.6%)	3,887 (75.3%)	
Site						
Colorado	824	191 (23.2%)	26 (3.2%)	67 (8.1%)	540 (65.5%)	
Florida	5,744	1,151 (20.0%)	238 (4.1%)	472 (8.2%)	3,883 (67.6%)	<0.0001
Ohio	4,634	700 (15.1%)	252 (5.4%)	233 (5.0%)	3,449 (74.4%)	
South Carolina	6,905	839 (12.2%)	217 (3.1%)	427 (6.2%)	5,422 (78.5%)	

BASC-2-BESS = Behavior Assessment System for Children-2 Behavioral and Emotional Screening System; SDQ = Strengths and Difficulties Questionnaire

†Chi-square test p-values<0.0001 were significant for all male vs. female comparison (i.e., elementary school males vs. middle/high school males, elementary school females vs. middle/high school females, and males vs. females at each developmental level).

Table 3

Percentage of students who met criteria for mental disorders in Stage 2 who screened as high-risk on the BASC-2-BESS and SDQ in Stage 1, Project to Learn About Youth-Mental Health

	Disorder N	High-risk by BASC-2-BESS or SDQ n (%)	High-risk by BASC-2-BESS n (%)	High-risk by SDQ n (%)	Low Screen on BASC-2 BESS and SDQ n (%)
Any disorder	267	158 (59.2%)	114 (42.7%)	147 (55.1%)	109 (40.8%)
Any internalizing	126	58 (46.0%)	43 (34.1%)	54 (42.9%)	68 (54.0%)
Generalized anxiety	19	5 (26.3%)	3 (15.8%)	5 (26.3%)	14 (73.7%)
Social phobia†	36	13 (36.1%)	9 (25.0%)	12 (33.3%)	23 (63.9%)
Separation anxiety	44	21 (47.7%)	12 (27.3%)	21 (47.7%)	23 (52.3%)
Panic disorder	9	3 (33.3%)	3 (33.3%)	3 (33.3%)	6 (66.7%)
Agoraphobia	17	9 (52.9%)	8 (47.1%)	8 (47.1%)	8 (47.1%)
Obsessive compulsive disorder	26	15 (57.7%)	12 (46.2%)	14 (53.9%)	11 (42.3%)
Post-traumatic stress disorder	4	2 (50.0%)	2 (50.0%)	2 (50.0%)	2 (50.0%)
Any depressive disorder‡	31	18 (58.1%)	13 (41.9%)	17 (54.8%)	13 (41.9%)
Mania/hypomania	5	5 (100.0%)	4 (80.0%)	5 (100.0 %)	0 (0.0%)
Any externalizing	197	134 (68.0%)	95 (48.2%)	126 (64.0%)	63 (32.0%)
Attention-deficit/hyperactivity disorder	101	88 (87.1%)	67 (66.3%)	86 (85.2%)	13 (12.9%)
Oppositional defiant disorder	117	69 (59.0%)	49 (41.9%)	64 (54.7%)	48 (41.0%)
Conduct disorder	25	18 (72.0%)	11 (44.0%)	16 (64.0%)	7 (28.0%)

BASC-2-BESS = Behavior Assessment System for Children-2 Behavioral and Emotional Screening System; SDQ = Strengths and Difficulties Questionnaire

† The social phobia module was not administered in the Ohio site.

‡ This group includes major depressive disorder and dysthymic disorder.

Table 4

Sensitivity, specificity, positive predictive value, and negative predictive value for BASC-2-BESS and SDQ total scores, Project to Learn About Youth-Mental Health.

	BASC-2-BESS				SDQ			
	Sensitivity	**Specificity**	**Positive Predictive Value**	**Negative Predictive Value**	**Sensitivity**	**Specificity**	**Positive Predictive Value**	**Negative Predictive Value**
Any disorder	42.7%	78.4%	40.1%	80.1%	55.1%	73.2%	41.1%	82.8%
Any internalizing	34.1%	74.0%	15.1%	89.2%	42.9%	67.2%	15.1%	89.7%
Generalized anxiety	15.8%	72.9%	1.1%	97.9%	26.3%	65.9%	1.4%	98.0%
Social phobia†	25.0%	73.0%	3.2%	96.5%	33.3%	66.0%	3.4%	96.6%
Separation anxiety	27.3%	73.1%	4.2%	95.8%	47.7%	66.5%	5.9%	96.7%
Panic disorder	33.3%	72.8%	1.1%	99.2%	33.3%	65.8%	0.9%	99.1%
Agoraphobia	47.1%	73.4%	2.9%	98.8%	47.1%	66.3%	2.3%	98.7%
Obsessive compulsive disorder	46.2%	73.5%	4.3%	98.2%	53.9%	66.4%	3.9%	98.3%
Post-traumatic stress disorder	50.0%	73.2%	0.7%	99.7%	50.0%	66.1%	0.6%	99.7%
Any depressive disorder‡	41.9%	73.5%	4.6%	97.7%	54.8%	66.7%	4.8%	98.0%
Mania/hypomania	80.0%	73.3%	1.4%	99.9%	100.0%	66.3%	1.4%	100.0%
Any externalizing	48.2%	77.9%	33.5%	86.7%	64.0%	72.9%	35.2%	89.8%
ADHD	66.3%	77.4%	23.9%	95.5%	85.2%	71.6%	24.4%	97.8%
Oppositional defiant disorder	41.9%	75.1%	17.4%	91.1%	54.7%	68.8%	18.1%	92.4%
Conduct disorder	44.0%	74.0%	4.1%	98.1%	64.0%	67.5%	4.8%	98.7%

BASC-2-BESS = Behavior Assessment System for Children-2 Behavioral and Emotional Screening System; SDQ = Strengths and Difficulties Questionnaire; ADHD = Attention-Deficit/Hyperactivity Disorder

† The social phobia module was not administered in the Ohio site.

‡ This group includes major depressive disorder and dysthymic disorder.

Table 5

Area under the curve statistics for BASC-2-BESS and SDQ scores by DISC results and school level, Project to Learn About Youth-Mental Health.

	School Level	BASC-2-BESS Total	SDQ Total	SDQ Emotion	SDQ Conduct	SDQ Hyper	SDQ Peer	SDQ Prosocial	SDQ Impact†
Any disorder	All students	0.691	0.694	0.574	0.664	0.695	0.614	0.616	0.671
	Elementary	0.716	0.718	0.581	0.673	0.719	0.640	0.626	0.681
	Middle	0.635	0.672	0.609	0.628	0.671	0.582	0.620	0.669
	High	0.653	0.603	0.533	0.662	0.600	0.560	0.608	0.617
Any internalizing disorder	All students	0.585	0.588	0.571	0.575	0.547	0.581	0.568	0.606
	Elementary	0.586	0.587	0.574	0.571	0.551	0.568	0.548	0.608
	Middle	0.532	0.564	0.588	0.455	0.522	0.554	0.581	0.615
	High	0.642	0.617	0.537	0.634	0.566	0.638	0.600	0.624
Any externalizing disorder	All students	0.728	0.733	0.575	0.697	0.749	0.629	0.626	0.701
	Elementary	0.748	0.752	0.578	0.701	0.758	0.661	0.639	0.694
	Middle	0.706	0.738	0.637	0.684	0.751	0.628	0.645	0.742
	High	0.670	0.591	0.535	0.671	0.637	0.513	0.625	0.627

BASC-2-BESS = Behavior Assessment System for Children-2 Behavioral and Emotional Screening System; SDQ = Strengths and Difficulties Questionnaire

†The Florida site did not include the SDQ impact questions in the screening phase.

Table 6

Area under the curve statistics for BASC-2-BESS and SDQ Scores by DISC results and child sex, Project to Learn About Youth-Mental Health

	Child Sex	BASC-2-BESS Total	SDQ Total	SDQ Emotion	SDQ Conduct	SDQ Hyper	SDQ Peer	SDQ Prosocial	SDQ Impact†
Any disorder	All students	0.691	0.694	0.574	0.664	0.695	0.614	0.616	0.671
	Males	0.712	0.711	0.590	0.690	0.700	0.608	0.631	0.674
	Females	0.665	0.661	0.559	0.624	0.666	0.618	0.584	0.650
Any internalizing disorder	All students	0.585	0.588	0.571	0.575	0.547	0.581	0.568	0.606
	Males	0.611	0.607	0.557	0.607	0.576	0.580	0.582	0.615
	Females	0.556	0.565	0.589	0.536	0.492	0.579	0.545	0.584
Any externalizing disorder	All students	0.728	0.733	0.575	0.697	0.749	0.629	0.626	0.701
	Males	0.724	0.728	0.592	0.701	0.723	0.617	0.636	0.691
	Females	0.738	0.725	0.558	0.683	0.757	0.643	0.592	0.697

BASC-2-BESS = Behavior Assessment System for Children-2 Behavioral and Emotional Screening System; SDQ = Strengths and Difficulties Questionnaire

†The Florida site did not include the SDQ impact questions in the screening phase.